

Probleme und Perspektiven der psychometrischen Traitforschung

Zur Anzahl und Generalität von Persönlichkeitsdimensionen

Dr. phil. André Beauducel

Habilitationsschrift – eingereicht an der Fakultät
Mathematik und Naturwissenschaften
der Technischen Universität Dresden

Dresden, November 2001

Inhalt

Vorwort und Danksagung.....	IV
1. Einleitung.....	1
2. Forschungsstrategische Einordnung und Ziele der Arbeit.....	4
2.1 Zur Überwindung der Defizite psychometrischer Traitforschung durch experimentelle und biopsychologische Ansätze.....	5
2.2 Ziele der Arbeit.....	15
3. Zur empirisch-methodischen Unverbindlichkeit der Traitforschung.....	21
3.1 Bestandsaufnahme: Zum Einsatz multivariater Verfahren.....	21
3.2 Vorschläge: Verbindlichere Traitforschung durch anspruchsvolleren Einsatz multivariater Verfahren.....	25
3.2.1 Zur konfirmatorischen Faktorenanalyse.....	26
3.2.2 Zur exploratorischen Faktorenanalyse.....	30
3.2.2.1 Extraktionskriterien: Bestandsaufnahme und Auswahl.....	31
3.2.2.2 Extraktionsmethoden: Bestandsaufnahme und Auswahl.....	49
3.2.2.3 Rotationsmethoden: Bestandsaufnahme und Auswahl.....	64
3.2.2.4 Probanden- und Variablenanzahl in exploratorischen Faktorenanalysen.....	81
3.2.2.5 Faktorenanalyse binärer Daten.....	82
3.2.2.6 Liste der begründeten Spezifikationen für die exploratorischen Faktorenanalysen.....	92
4. Zur theoretisch-interpretativen Unverbindlichkeit der Traitforschung.....	93
4.1 Zur Repräsentativität von Traitmodellen.....	93
4.2 Zum Problem der Generalität von Traits.....	96

4.3	Vorschläge zur vergleichenden Generalitätsbewertung von Traits.....	108
4.3.1	Inklusions-Indizes für Skalen.....	109
4.3.2	Inklusions-Indizes für Faktorwerte.....	119
5. Auswahl der Lösungsvorschläge und Traitmodelle für die empirische Untersuchung.....		130
5.1	Auswahl der Lösungsvorschläge.....	130
5.2	Auswahl der Traitmodelle.....	132
6. Empirische Anwendung der erarbeiteten Vorschläge auf konkurrierende Traitmodelle		134
6.1	Beschreibung der Untersuchung.....	134
6.1.1	Probandengruppe	134
6.1.2	Inventare.....	138
6.1.3	Durchführung.....	139
6.1.4	Statistische Analyse.....	141
6.1.5	Interne Konsistenzen der analysierten Fragebogenskalen.....	141
6.2	Modellspezifische, getrennte Faktorenanalysen für die Fragebogen.....	144
6.2.1	Eysenck: EPQ-R.....	144
6.2.2	Cattell: 16 PF-R.....	152
6.2.3	Costa & McCrae: NEO-PI-R.....	160
6.2.4	Zuckerman: SSS-V und ZKPQ.....	170
6.2.5	Zusammenfassende Diskussion der Ergebnisse zu den modellspezifischen Analysen.....	184
6.3	Modellübergreifende, gemeinsame Faktorenanalysen für die Fragebogen.....	187
6.3.1	Modellübergreifende Analysen für die Modelle von Cattell, Eysenck sowie Costa und McCrae.....	187
6.3.1.1	Faktorenanalysen erster Ordnung.....	187
6.3.1.2	Faktorenanalysen zweiter Ordnung.....	200

6.3.2	Modellübergreifende Analysen für die Modelle von Zuckerman, Eysenck und Costa und McCrae.....	206
6.3.2.1	Faktorenanalysen erster Ordnung.....	206
6.3.2.2	Faktorenanalysen zweiter Ordnung.....	215
6.3.3	Diskussion der Ergebnisse der modellübergreifenden Analysen.....	221
6.4	Anwendung der Inklusions-Indizes zur Ermittlung der Generalität.....	226
6.4.1	Zur Validität der Generalitäts-Indizes anhand des Vergleichs zwischen Faktoren erster und zweiter Ordnung.....	226
6.4.2	Zur Generalität der Neurotizismus-Faktoren des EPQ-R, NEO-PI-R und ZKPQ.....	231
6.4.3	Zur Generalität der Extraversions-Faktoren des EPQ-R und NEO-PI-R sowie der Soziabilitäts- und Aktivitäts-Faktoren im ZKPQ.....	234
6.4.4	Zum Generalitätsvergleich modellspezifischer Faktoren mit den in Kapitel 6.3 ermittelten Extraversions- und Neurotizismus-Faktoren Faktoren zweiter Ordnung.....	237
7.	Zusammenfassende Diskussion und Ausblick.....	240
7.1	Ergebnisse der theoretischen und methodischen Überlegungen.....	241
7.2	Faktorenanalytische Ergebnisse der empirischen Untersuchungen.....	242
7.3	Ergebnisse des empirischen Generalitätsvergleichs.....	250
7.4	Grenzen dieser Arbeit, verbleibendes Arbeitsprogramm und Perspektiven.....	254
	Zusammenfassung.....	258
	Literatur.....	260
	Anhang (extra gebunden).....	274

Vorwort und Danksagung

Viele Personen haben zum Gelingen der hier beschriebenen Untersuchungen beigetragen. Zuerst sei den anonymen Teilnehmern¹, die die Fragebogen ausfüllten, um eine individuelle Rückmeldung in Form eines Persönlichkeitsprofils zu erhalten, herzlich gedankt. Durch ihr Interesse an der Untersuchung trugen sie mit dazu bei, psychologische Erkenntnisse in den gesellschaftlichen Alltag zu transportieren. Besonderen Dank möchte ich auch Frau Brigitte Drubig aussprechen, die auf ruhige, konzentrierte Art, sowohl den überwiegenden Teil der Telefonate mit den Probanden, als auch die Dateneingabe hervorragend durchführte. Ihre durchgängig zuverlässige Arbeit stellte für mich eine große Entlastung dar.

Umfassende wissenschaftliche Arbeiten spiegeln stets auch die fachlichen und menschlichen Impulse wider, denen die Verfasser der Arbeiten ausgesetzt waren und sind. Durch die Nennung derer, die mich und mein Fachverständnis in diesem Sinne besonders beeinflusst und geprägt haben, möchte ich im folgenden meinen Dank aussprechen.

Ein früher, indirekter Impuls zu dieser Arbeit ging vom Betreuer meiner Dissertation, Prof. Dr. A.O. Jäger aus, der mir die Bedeutsamkeit breit angelegten, nicht zu sehr am Alltagsgeschäft einzelner Zeitschriften-Publikationen orientierten, Arbeitens vermittelte. Meine Entscheidung, eine Monographie zu schreiben, beruhte mitunter auf der Einsicht in den wissenschaftlichen Nutzen eines „langen Atems“ in der Forschung, wie ich ihn bei Herrn Prof. Jäger erleben konnte. Es bedurfte eines solchen Impulses, um, trotz der aktuellen Veränderungen der Qualifikationsanforderungen, die Motivation zu einer Monographie zu erzeugen.

Der nächste wichtige Impuls zu dieser Arbeit stammt von Prof. Dr. B. Brocke, der zuerst mein Interesse für die Untersuchung von Persönlichkeitseigenschaften (Traits) außerhalb des Leistungsbereiches weckte. Neben diesem frühen Impuls, beeinflusste mich Prof. Brocke durch seine forschungsstrategische, multimodale Perspektive und vor allem durch seine Kritik an der konventionellen, psychometrischen Traitforschung. In vielen Diskussionen tauschten wir, teilweise kontrovers, jedoch überwiegend in Übereinstimmung, unsere forschungsstrategischen und inhaltlichen Positionen aus. Gerade diese offenen, undogmatisch geführten, kontroversen Diskussionen mit Prof. Brocke waren stimulierend für die Erstellung dieser Arbeit.

Aufgrund seiner hohen transsituationalen und transtemporalen Konsistenz stellt der Dank, den ich Dr. Martin Kersting schulde, beinahe eine stabile Disposition, einen Trait dar. Auch wenn über die Heritabilität dieses Traits Zweifel bestehen mögen, möchte ich nicht versäumen, auf die massive kausale Determination dieses „Dank-Traits“ hinzuweisen: Dr. Kerstings gesunde,

¹ Personen und Funktionsbezeichnungen gelten jeweils in weiblicher und männlicher Form.

kritische, aber wohlwollende Distanz zur Forschungspraxis in der Psychologie, unsere zahlreichen Diskussionen zu den Inhalten dieser Arbeit und natürlich Diskussionen zur Realität der psychologischen Diagnostik, schärften mir das Fachverständnis und erfreuten das Gemüt.

Beim Gemüt angekommen, möchte ich auf das angenehme, kameradschaftliche Verhältnis zu meinen Kollegen Dr. Stefan Debener, Christine Kern, Dipl.-Psych. Sirko Rabe, Doreen Schade, Dipl.-Psych. Anja Schmidt, Dipl.-Psych. Doreen Schulze, Dipl.-Psych. Alexander Strobel und Kathrin Wohlfahrt hinweisen. Das kollegiale und humorvolle Klima und die vielen, mitunter über berufliche Fragen weit hinausgehenden Gespräche, waren eine große Hilfe bei der langwierigen Arbeit. Darüber hinaus möchte ich Dr. Stefan Debener und Dipl.-Psych. Alexander Strobel herzlich dafür danken, dass sie neben ihren intensiven beruflichen Verpflichtungen die Zeit fanden, Teile meiner Arbeit durchzusehen. Dipl.-Psych. Sirko Rabe und Dipl.-Psych. Anja Schmidt danke ich für ihre Bereitschaft zur stetigen, fachlichen und über das fachliche hinausgehenden Auseinandersetzung mit meinen Vorstellungen.

Obgleich die Struktur der vorliegenden Arbeit in der Einleitung genauer erläutert wird, möchte ich vorab auf eine technische Besonderheit, den umfangreichen, getrennt gebundenen Anhang, eingehen. Da in dieser Arbeit eine mitunter unzulängliche Praxis der Durchführung und Darstellung von Faktorenanalysen kritisiert wird, wurde hier maximale Transparenz der Ergebnisdarstellungen hergestellt. Daher wurden auch die teilweise extrem umfangreichen Ladungsmatrizen im Anhang gedruckt, damit der interessierte Leser die Möglichkeit hat, die in dieser Arbeit vorgenommenen Interpretationen nachzuvollziehen und zu relativieren. Es handelt sich also weniger um ein für das Verständnis dieser Arbeit unbedingt erforderliches Material, sondern eher um Material, das eine weiterführende Auseinandersetzung mit den hier vorgenommenen Interpretationen ermöglicht.

Die Vorbemerkungen abschließend möchte ich hier noch die Hoffnung äußern, dass – auch bei geänderter Gesetzeslage – noch viele Kollegen Gelegenheit zu integrativen, monographischen Arbeiten finden mögen.

Dresden, im November 2001

1. Einleitung

Die psychometrische Traitforschung, also die primär korrelative Forschung zu relativ stabilen Persönlichkeitsdimensionen, zählt trotz einiger Kritik (z.B. Mischel, 1968; Mischel & Peake, 1983) nach wie vor zu den dominierenden Bereichen der Differentiellen und Persönlichkeitspsychologie. Diese Forschung reicht bis in die 30er Jahre zurück (z.B. Allport & Odbert, 1936) und weist eine große Zahl konkurrierender Modelle auf (z.B. Cattell, 1972, 1994a, b; Cattell & Dreger, 1977; Costa & McCrae 1992b; Costa & McCrae, 1995; Eysenck & Eysenck, 1985; Guilford, Zimmermann & Guilford, 1976; Hofstee, De Raad & Goldberg, 1992). Dennoch erscheinen die Debatten, die in der Traitforschung geführt werden, oft wenig weiterführend: So wird seit den frühen Kontroversen zwischen Cattell, Guilford und Eysenck lediglich über den optimalen Auflösungsgrad und die Anzahl der zur angemessenen Beschreibung von Persönlichkeitstraits erforderlichen Dimensionen sowie um das Generalitätsniveau der verschiedenen Traits diskutiert (z.B. Cattell, 1972, 1994; Eysenck, 1991; Guilford, 1975). Dabei wird das Generalitätsniveau meist im Sinne der transsituationalen Konsistenz (vgl. Mischel, 1968) verstanden, d.h. eine hohe Generalität bedeutet, dass ein Trait in vielen Situationen relevant sein soll. Der Terminus Generalität wird auch im Rahmen dieser Arbeit im Sinne der Generalisierbarkeit über Situationen verwendet. Dies ist im Einklang mit dem Vorschlag von Michel und Conrad (1982), die die Invarianz gegenüber Bedingungen als Generalität, die Invarianz gegenüber Stichproben als Universalität und die Invarianz gegenüber Zeitpunkten als Stabilität bezeichnen.

Auch die Diskussion neuerer Traitssysteme fokussiert häufig auf Fragen der Faktorenanzahl und Generalität (z.B. Andresen, 1995; Barbaranelli & Caprara, 1996; Becker, 1999; Benet-Martínez & Waller, 1997; Costa & McCrae, 1995; Di Blas & Forzi, 1998). Insgesamt kann festgestellt werden, dass seit der Interaktionismus-Debatte nur selten grundsätzlichere Probleme der psychometrischen Traitforschung diskutiert werden (s. aber Block, 1995; Brocke, 2000; Smedslund, 1987; Wittmann, 1988). Das Feststellen der Unverbindlichkeiten und Defizite der psychometrischen Traitforschung hat jedenfalls zu einem hohen Maß an Ernüchterung bezüglich der Perspektiven der psychometrischen Traitforschung geführt (z.B. Block, 1995).

In der vorliegenden Monographie werden gerade die Bereiche, in denen die Traitforschung zur Perpetuierung von Diskussionen neigt, genauer untersucht. Somit fokussiert auch die vorliegende Arbeit auf die immer wieder diskutierten Fragen der Faktorenanzahl und Generalität. Im Unterschied zu den Diskussionen in Barbaranelli und Caprara (1996), Cattell (1994a), Costa und McCrae (1992b, 1995), Eysenck (1991) und Goldberg (1990), bei denen in der Regel allein die Vorzüge eines Modells gegenüber allen anderen Modellen betont werden, steht hier kein spezielles Persönlichkeitsmodell im Vordergrund. Vielmehr soll die Untersuchung

Gründe für die teilweise wenig fruchtbaren Diskussionen aufdecken. Auf dieser Grundlage sollen dann Vorschläge zur Überwindung der Unverbindlichkeiten und damit auch der sich perpetuierenden Diskussionen unterbreitet werden.

Natürlich wird nicht erwartet, dass eine einzelne Arbeit die Probleme einer ganzen Forschungstradition bewältigen kann. Dies umso weniger, als die Durchsetzung wissenschaftlicher Positionen und Standards zu einem großen Teil ein sozialer Vorgang ist (siehe z.B. Westmeyer, 1995). Dennoch sollen hier einige Vorschläge zur Bewältigung zentraler Probleme der psychometrischen Traitforschung unterbreitet werden.

Gerade in Zeiten, in denen Zeitschriftenaufsätze zu aktuellen Forschungsfragen die Wissenschaft zunehmend bestimmen, werden die Vorzüge einer Monographie besonders deutlich: In einer Monographie können – mit einigem Abstand vom Forschungsalltag – verschiedene Ansätze theoretischen und methodischen Vorgehens umfassend reflektiert und besonders problematische Bereiche identifiziert werden. Zugleich können für die identifizierten Problembereiche Lösungsvorschläge unterbreitet werden und anhand empirischer Untersuchungen evaluiert werden. Die enge Verzahnung von Problemidentifikation, Lösungsvorschlägen und empirischer Überprüfung kann in einer Serie von Zeitschriftenbeiträgen nur schwer vorgenommen werden. Wegen der breiteren Integrationsmöglichkeiten und der größeren Unabhängigkeit vom Forschungsalltag erscheinen Monographien auch weiterhin als notwendiges Komplement zu Zeitschriftenbeiträgen.

Im weiteren Verlauf der Arbeit wird dargelegt, inwiefern die Probleme der psychometrischen Traitforschung auf einer Kombination bestimmter methodischer und theoretischer Freiräume beruhen, die dazu führen, dass die Qualitätsunterschiede zwischen relevanten Trait-Modellen kaum evaluiert werden können. Es wird dargelegt, wie gerade das Wechselspiel zwischen Freiheitsgraden bei der Bestimmung der Dimensionszahl und Freiheitsgraden bei der Generalitätszuweisung, dazu führt, dass sich die Unverbindlichkeiten potenzieren. Daher wird in der vorliegenden Arbeit der Versuch unternommen, beide Aspekte der Unverbindlichkeit zugleich anzugehen, was die bereits benannte enge Verzahnung verschiedenster Aspekte der psychometrischen Traitforschung mit sich bringt.

Zum Aufbau der Arbeit

In Kapitel 2 erfolgt die genaue Begründung und Spezifikation der Ziele der Arbeit sowie die forschungsstrategische Einordnung. Als Ergebnis der Analysen in Kapitel 2 wird einerseits eine primär *empirisch-methodische Unverbindlichkeit*, insbesondere bei der Bestimmung der Dimensionsanzahl, und andererseits eine primär *theoretisch-interpretative Unverbindlichkeit*, insbesondere bei der Zuschreibung von Generalität zu Konstrukten, festgestellt. Möglichkeiten zur Überwindung der *empirisch-methodischen Unverbindlichkeit* werden in Kapitel 3, Möglichkeiten zur Überwindung der *theoretischen Unverbindlichkeit* in Kapitel 4 dargestellt. Die begründete Auswahl der in einer empirischen Untersuchung geprüften Lösungsvorschläge sowie der dazu einbezogenen Trait-Modelle erfolgt in Kapitel 5. Die empirische Untersuchung und die Anwendung der erarbeiteten Lösungsvorschläge auf die Daten wird in Kapitel 6 beschrieben. In Kapitel 7 erfolgt schließlich eine zusammenfassende Diskussion der Vorschläge und ihrer Evaluation.

2. Forschungsstrategische Einordnung und Ziele der Arbeit

Die seit langem bekannten Probleme der psychometrischen Traitforschung haben – wenn auch nicht durchgängig – Überlegungen zur angemessenen Forschungsstrategie in der Persönlichkeitspsychologie ausgelöst. Beispielsweise beklagte Cronbach (1957, 1975) das historisch gewachsene Schisma zwischen der experimentellen und korrelativen Forschungstradition. In der Folge wurden in der Persönlichkeitspsychologie wiederholt Vorschläge für eine Forschungsstrategie unterbreitet, die neben der psychometrischen Tradition verstärkt experimentelle und biopsychologische Daten einbezieht (z.B. Brocke, 2000; Eysenck, 1997; Fahrenberg, 1987; Gray, 1991; Zuckerman, 1994).

Die Integration verschiedener Datenebenen umgreift zwei verschiedene Aspekte: Zum einen können verschiedene Datenebenen im Sinne verschiedener Modalitäten unterschieden werden. So unterscheidet Zuckerman (1992) die psychometrische Ebene, soziales Verhalten, Konditionierung, Physiologie, Biochemie, Neurologie und Genetik. Sicher wären auch andere Strukturierungen denkbar, aber die große Zahl verschiedener Datenebenen wird in Zuckermans Unterscheidung sehr deutlich. Andererseits können auf die Daten der verschiedenen Ebenen auch verschiedene Korrelationstechniken angewendet werden, wie sie beispielsweise im Datenquader von Cattell (1957, 1988) dargestellt werden. So können zu einem Zeitpunkt Variablen über Personen (R-Technik) oder Personen über Variablen (Q-Technik) korreliert werden. Es können auch für eine Person Variablen über Messzeitpunkte korreliert werden (P-Technik). In dieser Weise können zunächst sechs verschiedene Korrelationstechniken unterschieden werden. Wittmann (1988) beschreibt vier dreidimensionale Datenboxen, je eine für Prädiktoren, Kriterien, kontrollierte Treatments und nicht kontrollierte Treatments. Der Vorteil der letztgenannten Einteilung ist, dass sie die in der Psychologie häufig auftretenden Untersuchungsarten verdeutlicht. So werden viele Korrelationsstudien zwischen potentiellen Prädiktoren ohne Treatmentvariablen durchgeführt, so dass lediglich die interindividuellen Unterschiede beschrieben werden. Viele Studien untersuchen Treatment-Variablen und erfassen so intraindividuelle Varianzen. Die von Zuckerman (1992) genannten Datenarten können also mit den Korrelationstechniken kombiniert werden, so dass sich ein enormer Möglichkeitsraum für die psychologische Forschung ergibt.

Im folgenden können die vielen Möglichkeiten, die eine Ausschöpfung der verschiedenen Datenarten und der verschiedenen Datenboxen bietet, nicht betrachtet werden. Es sei lediglich darauf hingewiesen, dass in der Psychologie bisher eher kleinere Ausschnitte der insgesamt

denkbaren Möglichkeiten extrem intensiv untersucht wurden, während andere, potentiell interessante Möglichkeiten bisher kaum beachtet wurden.

Im Rahmen dieser Arbeit soll jedoch auf die Bereiche fokussiert werden, in denen seit langem intensive persönlichkeitspsychologische Forschung stattfindet. Dies sind die Bereiche der experimentellen, der biopsychologischen und der psychometrischen Traitforschung. Der Fokus auf experimentelle und biopsychologische Ansätze impliziert nicht, dass die Möglichkeiten die psychometrische Traitforschung durch eine Integration sozialpsychologischer, klinischer und anderer Ansätze zu bereichern, als gering eingeschätzt werden. Für die experimentellen und biopsychologischen Ansätze erscheinen die Möglichkeiten der Integration mit der psychometrischen Traitforschung jedoch besonders vielversprechend (z.B. Eysenck, 1981, 1997; Gray, 1983, 1991; Zuckerman, 1994), vielleicht auch weil in diesem Bereich seit längerem versucht wird integrativ zu arbeiten. Daher wird im folgenden eine intensivere Behandlung der Möglichkeiten und Probleme dieser Forschungsansätze vorgenommen.

2.1 Zur Überwindung der Defizite psychometrischer Traitforschung durch experimentelle und biopsychologische Ansätze

Es wird – wohl zu recht – davon ausgegangen, dass die Erweiterung korrelativer, psychometrischer Traittheorien zu multimodal-kausalen Traittheorien (s. Brocke, 2000) in vielen Bereichen eine wesentliche Weiterentwicklung ermöglicht (Cronbach, 1957; Eysenck, 1997; Stelmack, 1997; Zuckerman, 1994; Zuckerman, Kuhlman, Joireman, Teta & Kraft, 1993). Diese Sicht sollte jedoch nicht dazu führen, Optimierungen allein aus der Hinzunahme weiterer Datenebenen bzw. Forschungsansätze zu erhoffen. Hier wird vielmehr die Ansicht vertreten, dass ohne eine Reduzierung der Unverbindlichkeiten, die der psychometrischen Traitforschung *immanent* sind, auch die notwendige Integration weiterer Forschungsansätze ihr innovatives Potential in der Traitforschung bei weitem nicht voll entfalten kann. Diese Annahme gründet sich auf den Umstand, dass die Hinzunahme weiterer Messebenen, sowie experimenteller Befunde, zunächst ein neues Problem aufwirft: Das Problem der Integration verschiedener Mess- und Theorieebenen. Mit welchem Gewicht sollen die Befunde verschiedener Messebenen bei der Theoriebildung berücksichtigt werden? Auf die teilweise geringen Erfolge einer multivariat-multimodalen Forschungsstrategie wird beispielsweise in Wittmann (1987) hingewiesen. Andererseits kann die Integration korrelativer und experimenteller Befunde für verschiedene Persönlichkeitskonstrukte bei verschiedenen Paradigmen mit Gewinn erfolgen (z.B. Bartussek, Becker, Diedrich, Naumann & Maier, 1996; Brocke, Tasche & Beauducel,

1996, 1997; Debener, 2001; Netter, Hennig & Rohrman 1999; Stelmack & Pivik, 1996; Zuckerman, 1994), wobei mit Gewinn hier – im Einklang mit Wittmann (1987) – zunächst der Nachweis replizierbarer empirischer Phänomene gemeint ist. Beispielsweise konnten Beauducel, Debener, Brocke und Kayser (2000) für das Augmenting/Reducing-Paradigma zeigen, dass die Bedingungen erfolgreicher Replikation von Zusammenhängen zwischen biopsychologisch-experimentellen Maßen und psychometrischen Traits häufig nicht genügend Beachtung finden, da dieselben empirischen Phänomene mit nicht systematischen, großen experimentellen Variationen untersucht werden.

Ein weiteres Problem ist, dass in vielen Studien dieser Art sowie in den Studien zur „Aptitude Treatment Interaction“ (z.B. Cronbach, 1975) Befunde für „Aptitudes“ (nach Cronbach können auch Persönlichkeitstrait „aptitudes“ darstellen) berichtet werden, ohne dass sich aus der Integration experimenteller Befunde Hinweise für eine Restrukturierung der psychometrisch gewonnenen „Aptitudes“ ergeben. Vielmehr werden für gegebene psychometrische Traits oder „Aptitudes“ experimentelle Befunde aufgezeigt, die dann ein besseres theoretisches Verständnis der jeweiligen Merkmale ermöglichen sollen. Anstatt die psychometrischen Traits als gegeben anzusehen und ihnen biopsychologische und experimentelle Befunde beizustellen, kann die Perspektive auch gewechselt werden, indem nicht der psychometrische Trait als Ausgangspunkt angesehen wird, sondern als gleichwertiger Bestandteil einer den Messebenen übergeordneten Traitdimension. Ein derartiger Ansatz ergibt sich beispielsweise aus der kriterienorientierten Faktorenrotation (Eysenck, 1950) oder aus der Verwendung der Kommunalitätenanalyse (z.B. Cooley & Lohnes, 1976), um Traits im Zentrum mehrerer Datenebenen (z.B. psychometrisch, experimentell und biopsychologisch) zu lokalisieren (s. Brocke, Beauducel & Tasche, 1999). Diese simultane, aber auch gleichwertige Betrachtung mehrerer Datenebenen ist in der Traitforschung bisher kaum zum Tragen gekommen. Es bleibt abzuwarten, welches Innovationspotential mit diesem Ansatz verbunden ist und inwieweit er zur Reduzierung der Unverbindlichkeiten der psychometrischen Traitforschung beitragen kann.

Die Integration experimenteller und biopsychologischer Befunde in die psychometrische Traitforschung erfolgte in der Regel nicht über die simultane Verrechnung entsprechender Varianzen, sondern über eine vermeintliche theoretische Integration entsprechender Befunde. Beispiele für diese Forschungsstrategie werden im folgenden anhand der biopsychologisch-experimentellen Persönlichkeitsmodelle von Eysenck, Gray, Zuckerman und Cloninger skizziert, wobei hier keine vollständige Darstellung der Modelle erfolgt:

Eysenck

Eysenck (1967, 1981) integrierte in seiner Persönlichkeitstheorie psychometrische Befunde mit experimentellen und biopsychologischen Befunden. Dadurch entstand ein Theoriennetz, das die Ableitung vielfältiger Hypothesen erlaubt (s. Brocke & Battmann, 1992). Das Problem der unterschiedlichen Gewichtung von Befunden aus verschiedenen Messebenen wird dabei in Eysencks Persönlichkeitstheorie besonders deutlich. So konnten Amelang und Ullwer (1990, 1991) nur wenige der experimentellen Befunde zu Eysencks Theorie replizieren. Zur Erklärung der Diskrepanzen zu Eysencks Befunden vermuten Amelang und Ullwer (1991, S. 44) neben interkulturellen Unterschieden „untypische Effekte einer idiosynkratischen Kombination von Autoren, Merkmalsträgern, Versuchsanordnungen, Auswertungen und Publikationsstrategien.“ Inzwischen ist die Theorieentwicklung ausgehend von Eysencks Ansatz weiter fortgeschritten, aber auch die erfolgversprechenden Weiterentwicklungen sind recht heterogen (z.B. Stelmack, 1997; Brocke et al., 1997), d.h. auch hier werden verschiedene Daten (z.B. Leistungsdaten, Ereigniskorrelierte Potenziale) mit sehr unterschiedlichem Gewicht integriert.

Gray

Besonders interessant ist in diesem Kontext, dass Gray (1983, 1991) auf der Basis experimenteller und biopsychologischer Befunde eine Restrukturierung von Eysencks Persönlichkeitsdimensionen Extraversion und Neurotizismus vorgeschlagen hat. Aufgrund seiner Befunde schlug Gray (1973) vor, eine 45°-Rotation der Faktorenachsen Extraversion und Neurotizismus vorzunehmen, um so die nach seiner Auffassung fundamentalen Dimensionen Impulsivität und Ängstlichkeit zu erhalten. Dies kann als Beispiel für den Einfluss experimenteller Forschung auf die psychometrische Restrukturierung von Daten gelten. Allerdings konnte die experimentelle und biopsychologische Stringenz von Grays Theorie nur mit geringem Erfolg in eine Entwicklung psychometrischer Inventare umgesetzt werden (Wilson, Gray & Barrett, 1990). Eine Optimierung psychometrischer Skalen zur Erfassung von Impulsivität und Ängstlichkeit als Alternativen zu Extraversion und Neurotizismus konnte bisher nur durch eine partielle Loslösung von Grays ursprünglichen verhaltenstheoretischen Konzepten erfolgen (Carver & White, 1994; Strobel, Beauducel, Debener & Brocke, in Druck). Die psychometrische Seite von Eysencks Theorie (Eysenck & Eysenck, 1975; Eysenck & Eysenck, 1991) konnte durch Grays experimentell und biopsychologisch motivierte Umstrukturierungen (z.B. Wilson et al., 1990) nicht ersetzt werden.

Zuckerman

Auch die auf Faktorenanalysen von Fragebogen beruhenden Teile von Zuckermans Theorie (Zuckerman et al., 1993; Zuckerman, 1994) sind zwar interessant und vielversprechend, konvergieren jedoch nur begrenzt mit Eysencks Persönlichkeitsmodell (s. Eysenck, 1991) und noch weniger mit den psychometrischen Teilen von Grays Theorie. Im übrigen zeigen sich teilweise auch Probleme bei der psychometrischen Erfassung von Sensation Seeking (z.B. Beauducel, Brocke, Strobel & Strobel, 1999; Haynes, Miles & Clements, 2000). Auch die psychophysiologischen Befunde zu Zuckermans Theorie sind keinesfalls einheitlich (z.B. Zuckerman, 1990; Brocke, Beauducel, John, Debener & Heilemann, 2000). Beispielsweise wird das Augmenting/Reducing Paradigma in den verschiedensten Varianten durchgeführt (s. Beauducel et al., 2000) und mit Sensation Seeking korreliert. Die Korrelationen treten über die Studien hinweg an wechselnden Ableitpositionen, für unterschiedliche Parameter des evozierten Potentials und unterschiedliche Subskalen von Sensation Seeking auf (s. Zuckerman, 1990).

Cloninger

Das Modell von Cloninger (1987; Cloninger, Przybeck & Svrakic, 1991) mit den drei Konstrukten Novelty Seeking, Harm Avoidance and Reward Dependence weist Parallelen zu den Modellen von Gray, Zuckerman und Eysenck auf. Der Grad der Überlappungen mit anderen Modellen ist jedoch unterschiedlich und insgesamt schwer einzuschätzen (s. aber Weyers, Krebs & Janke, 1995). Der Ausgangspunkt von Cloningers Modellentwicklung war die Annahme, dass sich zugrundeliegende bio-genetische Strukturen nicht direkt in der phänotypischen Struktur von Selbstbeschreibungen manifestieren müssen (Cloninger et al., 1991). Daher führte Cloninger eine zunächst deduktiv orientierte Skalenkonstruktion durch und optimierte die Skalen erst in einem zweiten Schritt an der Empirie. Auf diese Weise wollte Cloninger zu Dimensionen gelangen, die interindividuelle Unterschiede in Neurotransmittersystemen abbilden. Der postulierte Zusammenhang von Cloningers Dimensionen mit Transmittersystemen hat auch die Untersuchung des Zusammenhangs der Dimensionen mit spezifischen, für die Transmittersysteme charakteristischen molekulargenetischen Polymorphismen angeregt (z.B. Ebstein et al. 1996; Strobel, Wehr, Michel & Brocke, 1999). Aufgrund der besonderen Orientierung Cloningers bei der Skalenkonstruktion würde eine extrem hohe Überlappung der in Cloningers Modell postulierten Persönlichkeitsdimensionen mit anderen, induktiv anhand von Selbstbeschreibungen gewonnenen Persönlichkeitsdimensionen, Cloningers ursprüngliche Intention bei der Modellentwicklung in Frage stellen. Somit verdeutlicht der Vergleich von Cloningers Modellbildung mit der von Eysenck, Gray oder Zuckerman, wie bei der Kombination psychometrischer Forschung mit experimenteller und

biopsychologischer Forschung, unterschiedliche Strategien und Optimierungskriterien verwendet werden können.

Probleme der Integration verschiedener Datenebenen

Offenbar führt die Einbeziehung experimenteller und biopsychologischer Befunde nicht automatisch zu einer Erhöhung der Konvergenz zwischen verschiedenen Traitsystemen, was mit den Freiheitsgraden bei der Integration vieler verschiedener Datenarten in Verbindung gebracht werden kann. Es bedarf eines minimalen Konsens bei der Bewertung der Relevanz unterschiedlicher experimenteller und biopsychologischer Befunde und bei der Integration der komplexen Befunde in psychometrische Traittheorien. Natürlich kann über viele methodologische Kriterien schnell Konsens gewonnen werden. Es versteht sich von selbst, dass Replizierbarkeit und Effektstärken eine Rolle spielen müssen. Aber für inhaltlich motivierte Gewichtungen erscheint ein Konsens nur schwer erreichbar. Beispielsweise führt Eysenck (1997) deutlich andere experimentelle und biopsychologische Befunde als Grundlage von Extraversion an als Stelmack (1997).

Auch ist darauf hinzuweisen, dass bei der einfachen Betrachtung von Datenebenen wie der experimentellen, psychophysiologischen, neurochemischen, molekulargenetischen und psychometrischen Datenebene leicht übersehen werden kann, dass auch noch innerhalb dieser Datenebenen die unterschiedlichsten Datenarten auftreten. Allein auf der psychophysiologischen Messebene erscheint es oft schwierig Kovariationen zwischen unterschiedlichen psychophysiologischen Reaktionsarten nachzuweisen (z.B. Fahrenberg, 1987).

Auch ist bei der Integration experimenteller und biopsychologischer Befunde häufig ein pragmatischer Ekklektizismus vorzufinden (z.B. Zuckerman, 1994). Vermutlich ist dieser Ekklektizismus mangels genauerer Kenntnis der Zusammenhänge zwischen der neurochemischen, psychophysiologischen und psychometrischen Ebene auch zur Zeit – zumindest bei post-hoc Integrationen unterschiedlicher Befunde – nur schwer zu vermeiden. Daher erscheint eine kontinuierliche Prüfung der Stringenz der theoretischen Ableitungen derartiger Integrationsversuche notwendig. An dieser Stelle können wissenschaftstheoretische Rekonstruktionen der komplexen integrativen Systeme hilfreich sein (s. Brocke & Battmann, 1985, 1992). Auch das strukturalistische Forschungsprogramm (z.B. Westmeyer, 1992) könnte durch eine Erhöhung der Stringenz der theoretischen Aussagen einen Beitrag zur Bewältigung der Komplexität mehrmodaler Theorienbildung leisten. Natürlich ergeben sich auch methodische Probleme und Grenzen bei der Integration verschiedener Datenarten, die nicht durch eine Erhöhung der Stringenz der theoretischen Integrationen überwunden werden können. Bei-

spielsweise illustriert Asendorpf (1995) die Unabhängigkeit und ggf. sogar Gegenläufigkeit differentiell- und allgemein-psychologischer Varianzen.

Durch die obigen Ausführungen sollte nicht der Eindruck entstehen, dass experimentelle und biopsychologische Forschungen für die psychometrische Traitforschung von geringer Relevanz seien. Die Einseitigkeiten und Grenzen ausschließlich korrelativ angelegter psychometrischer Traitforschung werden vielmehr als bekannt vorausgesetzt. Ebenso wird vorausgesetzt, dass diese Einseitigkeiten langfristig überwunden werden müssen (s. Brocke, 2000; Cronbach, 1957; Underwood, 1975). Durch die obige Darstellung sollte jedoch verdeutlicht werden, dass die Probleme und Schwierigkeiten, die in den einzelnen Bereichen der Traitforschung auftreten, nicht ohne weiteres auf der Makroebene durch eine Erhöhung der Vielfalt der betrachteten Datenebenen bewältigt werden können. Vielmehr setzt eine gelungene Integration experimenteller, biopsychologischer, und ggf. sozialpsychologischer, klinischer mit psychometrischen Befunden in multimodalen oder auch kausalen Traittheorien jeweils eine optimale methodische Stringenz und Replizierbarkeit der Befunde und Teilmodelle in den jeweiligen Teildisziplinen voraus.

Das der vorliegenden Arbeit zugrundeliegende forschungsstrategische Argument lautet somit wie folgt: Es erscheint zunächst erfolgversprechender, innerhalb der jeweiligen Teilbereiche bzw. Datenebenen zu einheitlichen Bewertungsmaßstäben für Befunde zu gelangen als über verschiedene Teilbereiche hinweg. Zumindest sollte parallel zur Arbeit an Integrationsversuchen an der Erhöhung der Stringenz der methodischen Standards innerhalb der Teilbereiche gearbeitet werden. Beispielsweise können bezüglich der Handhabung multivariater ordnungsanalytischer Verfahren, insbesondere der Faktorenanalyse allerhand methodische Standards diskutiert und festgelegt werden (z.B. Gorsuch, 1983; Velicer, Eaton & Fava, 2000; Zwick & Velicer, 1986). Aber für die Gewichtung experimenteller und anderer Befunde bei der Integration in psychometrische Traittheorien kann bisher – neben der selbstverständlich erforderlichen Replizierbarkeit – kaum ein allgemeinverbindlicher Maßstab festgelegt werden, da die Qualität der Befunde und die Interpretationsmöglichkeiten zu heterogen sind. Zur Verdeutlichung zwei Fragen, die sich in diesem Zusammenhang stellen: Sind Befunde aus Zweifach-Wechselwirkungen höher zu gewichten als die Befunde aus Dreifach-Wechselwirkungen? Sind PET-Daten bedeutender als EEG- oder Reaktionszeit-Daten? Natürlich kann die Gewichtung verschiedener Befunde wiederum von den zugrundeliegenden Theorien und von deren Gegenstandsbereichen abhängig sein. Dies führt zu einer Zirkularität: Die theoretischen Vorannahmen bestimmen die Auswahl und Gewichtung relevanter Befunde, die die Grundlage der Theorie sein sollen. Durch eine Betonung der theoretischen Integrations-Perspektiven dürfen

die methodischen Aspekte der Gewichtung der Befunde jedoch nicht vernachlässigt werden. Die Gefahr besteht unter anderem deswegen, weil bisher keine methodischen Standards für die Integration der Befunde verschiedener Datenarten bestehen.

Eysenck (1991), der den Bedarf nach übergeordneten Bewertungskriterien explizit formuliert, unterbreitet einige Vorschläge für Bewertungskriterien in der Persönlichkeitsforschung. Dabei unterscheidet er zwischen sogenannten taxonomischen Kriterien, die für den psychometrischen Bereich relevant sind und sogenannten kausalen Kriterien, die die biopsychologische und experimentelle Forschung betreffen. Aber auch Eysenck (1991) formuliert in seiner Kriterienliste keine Kriterien für die Integration der taxonomischen und kausalen Befunde. Vielmehr trennt auch er diese beiden Bereiche bezüglich der Bewertung, so dass die Unverbindlichkeit, die in der Vernetzung dieser Befunde besteht, erhalten bleibt.

Simultane Integration verschiedener Datenebenen

Da die theoretische Integration taxonomischer und kausaler Befunde einen großen Spielraum für unterschiedliche Gewichtungen verschiedener Befunde lässt, fragt sich, ob die vielen verschiedenen Befunde überhaupt im nachhinein befriedigend in einem oder mehreren Gesamtsystemen der Persönlichkeit zusammengeführt werden können. Um diese Probleme zu umgehen, könnten Ansätze, die von vornherein inter- und (vorzugweise Treatment-bedingte) intraindividuelle Varianzen integrieren, sinnvoll sein. Hier könnte die dreimodale Faktorenanalyse (Tucker, 1964; Bartussek, 1973; Krolak-Schwerdt, 1991) ein nützliches Hilfsmittel sein, aber auch die methodischen Ansätze von Asendorpf (1991), Schmitz (1993, 2000) sowie Buse und Pawlik (1991) könnten hier Beispielcharakter haben.

In der Persönlichkeitsforschung sind nur wenige Studien verfügbar, bei denen in einem Datensatz interindividuelle, psychometrische und intraindividuelle, experimentelle Varianzen systematisch untersucht werden. Eine Ausnahme bildet beispielsweise die Studie von Zelenski und Larsen (1999), die jedoch auf einer für psychometrische Analysen zu kleinen Probandengruppe beruht. Aus einer allgemeinen Perspektive könnte man die oben dargestellten Probleme auch wie folgt subsumieren: Die Integration der Befunde aus verschiedenen Datenboxen (Cattell, 1957; Wittmann, 1988) sollte vorzugweise simultan innerhalb einzelner Studien erfolgen, damit der potentielle Einfluss der intraindividuellen, vorzugweise Treatment-bedingten Effekte auf die interindividuelle Varianz direkt untersucht werden kann und nicht im nachhinein zum Gegenstand unverbindlicher theoretischer Spekulationen wird.

Im Prinzip wird hier der gleiche integrative Gedanke verfolgt, wie bei der simultanen Analyse mehrerer Datenebenen mit Hilfe der Kommunalitätenanalyse oder kriterienorientierten Faktorenrotation (s.o.). Genauso wie Traits nicht zuerst über den psychometrischen Trait konzeptualisiert werden müssen, um im nachhinein mit biopsychologischen Korrelaten eine tiefere Erklärung zu erfahren, können auch die intraindividuellen, experimentell-psychologischen Effekte direkt oder simultan zur empirischen Erklärung der interindividuellen Varianz herangezogen werden. Diese Strategie hat den Vorteil, dass die Verknüpfung nicht erst im nachhinein über theoretische Zusatzannahmen erfolgen muss. Beispielsweise könnte man sich Faktorenlösungen für Fragebogen-Items vorstellen, bei denen die Achsen so rotiert werden, dass i.S. der konvergenten und diskriminanten Validität (Campbell & Fiske, 1959) wechselseitig vorhergesagte experimentelle Effekte besonders markant hervortreten und nicht vorhergesagte Effekte ausbleiben.

An diesem Denkmodell lassen sich die Probleme der Integration experimenteller und psychometrischer Forschung verdeutlichen: Es lassen sich Situationen denken, in denen eine nach psychometrischen Kriterien optimale Faktorenlösung (z.B. mit optimaler Einfachstruktur) zu einer Konfundierung verschiedener experimenteller Effekte führt. Andererseits kann eine Rotation, bei der experimentelle Effekte optimal getrennt werden können, zu einer suboptimalen Einfachstruktur bei der Beschreibung der interindividuellen Unterschiede führen. Auch hier stellt sich die Frage nach der relativen Gewichtung der Messebenen. Ein Beispiel dafür, dass Faktorenanalysen, die eine Beschreibung interindividueller Unterschiede mit optimaler Einfachstruktur leisten, zur Konfundierung oder Misallokation experimenteller Effekte führen können, findet sich beispielsweise in Wood und McCarthy (1984) für den Bereich der ereigniskorrelierten Potentiale.

Zur Verdeutlichung der Probleme, die sich bei simultaner Integration von Datenebenen ergeben können, sei noch ein Beispiel angeführt: Es ist allgemein bekannt, dass Cronbach (1957) eine Integration der korrelativen und experimentellen Psychologie forderte und dass er selbst versucht hat, diese Forderung mit dem Forschungsprogramm zu „Aptitude Treatment Interactions“ einzulösen. Bei diesen Forschungen werden interindividuelle Unterschiede in den Aptitudes systematisch mit intraindividuellen Treatment-Effekten in Verbindung gebracht. Es handelt sich um eine simultane Integration beider Datenarten in einem Design. Merkwürdigerweise werden Cronbachs spätere Bestandsaufnahmen und Schlussfolgerungen zu den Ergebnissen und Perspektiven dieses Forschungsprogramms (z.B. Cronbach, 1975) viel seltener zitiert als die programmatische Forderung von 1957: In seiner Bestandsaufnahme kommt Cronbach (1975) zu einer ernüchternden Einschätzung der Generalisierbarkeit der Befunde zu

den „Aptitude Treatment Interactions“. Gerade die Interaktionsverläufe waren kaum generalisierbar. Cronbach (1975, S. 199) beschreibt wie die Untersuchung von Interaktionen erster Ordnung bei mangelnder Generalisierbarkeit sich schnell zu einer unbewältigbaren Untersuchung von Interaktionen höherer Ordnung ausweitete: „In attempting to generalize from the literature, Snow and I have been thwarted by the inconsistent findings coming from roughly similar inquiries. [...] Some fraction of this inconsistency arises from statistical sampling error, but the remainder is evidence of unidentified interactions...“ und weiter: „Once we attend to interactions, we enter a hall of mirrors that extends to infinity.“ Die großen Schwierigkeiten, die Cronbach (1975) bei der Generalisierung von Interaktionseffekten feststellte, führten dazu, dass er weitreichende Generalisierungen, wie sie mit psychologischen Theorien häufig versucht werden, grundsätzlich in Frage stellte. Seine ernüchterte Bewertung der Forschungen zu „Aptitude Treatment Interactions“ drückt sich in folgender programmatischen Einschätzung der Aufgabe der Sozialwissenschaften aus (Cronbach, 1975, S. 126): „The goal of our work, I have argued here, is not to amass generalizations atop which a theoretical tower can someday be erected [...] The special task of the social scientist in each generation is to pin down the contemporary facts.“

Diese Empfehlung Cronbachs (1975) sollte sicherlich von Psychologen so ernst genommen werden wie die von 1957, denn die Fortschritte in der Psychologie sind nicht so rasant, dass sie jeden Zweifel über die Angemessenheit der Forschungsstrategie im Keim ersticken. Im Rahmen dieser Arbeit wird die zurückhaltende Einschätzung der Möglichkeiten der theoretischen Integration von Forschungen zu verschiedenen Datenebenen mit Cronbach (1975) geteilt. Allerdings wird hier die Auffassung vertreten, dass neben dem von Cronbach (1975) geforderten, eher konkret beschreibenden, weniger an Generalisierungen orientierten Forschungsprogramm, weiterhin ein an allgemeinen Gesetzmäßigkeiten orientiertes Forschungsprogramm in der Psychologie verfolgt werden sollte. Einem solchen, an allgemeinen Gesetzmäßigkeiten orientierten, Forschungsprogramm ist die vorliegende Arbeit verpflichtet. Wenn die Integration verschiedener Datenebenen zu einer unbewältigbaren Komplexität führt, wie beispielsweise zu der von Cronbach (1975) problematisierten Komplexität der Interaktionseffekte, dann kann es sich auch als forschungsstrategisch sinnvoll erweisen, die Komplexität und Breite durch getrennte Bearbeitung der Datenebenen zu reduzieren. Vielleicht sind die Probleme bei der Integration verschiedener Datenebenen einer der Gründe dafür, weshalb so wenig Forschung existiert, bei der die historisch gewachsenen Teildisziplinen der Psychologie integriert werden, obwohl die integrative Betrachtung der Phänomene allgemein erwünscht ist.

Diese Überlegungen verdeutlichen, dass auch die simultane Analyse experimenteller und psychometrischer Daten auf Probleme stößt: Zum einen kann die simultane Integration die Komplexität der Ergebnisse in einem Maße erhöhen, dass Schwierigkeiten bei der Replikation auftreten (s. Cronbach, 1975). Desweiteren kann es zu Effekten der Varianzmisallokation kommen, wenn die Komponenten, die die interindividuellen Unterschiede abbilden, mehrere Treatmenteffekte konfundiert erfassen (s. Wood & McCarthy, 1984). Außerdem sind Untersuchungsdesigns, die simultan interindividuell-psychometrische und intraindividuell-experimentelle Varianz in ausreichend großen Treatment- und Probandenstichproben beinhalten, extrem aufwendig und es wird nicht in allen Bereichen möglich sein, derartige Designs zu realisieren.

Zusammenfassung

Der kurze Blick auf die Möglichkeiten der Verbindung der experimentellen und biopsychologischen Forschung mit der psychometrischen Traitforschung, hat ergeben, dass derartige Integrationen fruchtbar sein können. Dennoch entstehen bei integrativer Betrachtung neue Schwierigkeiten: Im Falle einer theoretischen post hoc Integration experimenteller und psychometrischer Daten stellt sich das Problem der Kombination und relativen Gewichtung von Befunden unterschiedlichster Provenienz. Bei komplexen Designs zur simultanen Integration der verschiedenen Varianzquellen, werden die Unverbindlichkeiten der post hoc Integration von Befunden aus verschiedenen Ebenen vermieden. Andererseits können, wie im Fall von Cronbach (1975), bei simultanen Integrationen die Befunde so komplex werden, dass die Replikation erschwert wird. Dies spricht wiederum für eine getrennte Erforschung der experimentellen und psychometrischen Datenebene. Wenn man jedoch zu verbindlichen post hoc Integrationen von Befunden verschiedener Messebenen gelangen möchte, bedarf es eines Kataloges methodischer und vielleicht auch inhaltlicher Kriterien zur relativen Gewichtung der Befunde. Dabei muss jedoch bedacht werden, dass die Integration von Daten verschiedener Messebenen anhand theoretischer Erklärungen ein kreativer Prozess ist, der nur begrenzt systematisiert werden kann. Daraus ergibt sich, dass die Reduktion der Unverbindlichkeiten jeweils innerhalb der psychometrischen, experimentellen und biopsychologischen Teilbereiche auch in Zeiten, in denen der Wunsch nach Integration begründeterweise stark ist, eine sinnvolle, nicht zu vernachlässigende Aufgabe darstellt. Eine solide Arbeit in den jeweiligen Teilbereichen schafft zum einen die Voraussetzung für die weitergehende Integration mehrerer Messebenen und kann zum anderen vor verfrühten, übergeneralisierenden Integrationsversuchen bewahren. Auch kann durch diese Strategie vielleicht vermieden werden, dass das Feld der multimodal-kausalen Traittheorien schon besetzt wird, bevor die Befunde aus großen, inter- und intraindividuelle Varianz simultan integrierenden Studien in genügendem Umfang verfügbar sind.

2.2 Ziele der Arbeit

Um eine Potenzierung der Unverbindlichkeiten als Folge willkürlicher Gewichtungen und verfrühter Integrationen methodisch unzureichender Befunde verschiedener Datenebenen zu vermeiden, widmet sich die vorliegende Arbeit den Problemen, die der psychometrischen Traitforschung immanent sind. Die Debatten in der psychometrischen Traitforschung fokussieren, wie bereits angedeutet, auf Fragen der angemessenen Faktorenanzahl und Generalität. Debatten um die Vollständigkeit der Persönlichkeitsmodelle und in diesem Zusammenhang um die Dimensionsanzahl und -generalität müssen – auch wenn sie Jahrzehnte dauern – nicht unproduktiv sein. Man erhofft sich jedoch allmählich auch Lösungen und vor allem weiterführende Perspektiven.

Ein Beispiel dafür, dass sich auch psychometrische Strukturforschung sehr dynamisch entwickeln kann, findet sich im Intelligenzbereich: Sternberg und Powell (1982) fassen die Kontroversen in einem interessanten Entwicklungsmodell zusammen. In diesem Modell wird deutlich, wie verschiedene Positionen (z.B. Monismus und Pluralismus) in weiterführenden Modellen (z.B. hierarchische Modelle) integriert werden. Es können verschiedene Stufen der Entwicklung unterschieden werden, in denen immer umfassendere Struktur Aspekte der Intelligenz bis hin zum Facettenansatz integriert werden.

Demgegenüber hat die psychometrische Forschung zu Persönlichkeitsdimensionen oder Traits nur einen Typ von Strukturmodellen hervorgebracht: die hierarchischen Modelle. Im Rahmen der hierarchischen Modellvorstellung wird nun seit Jahrzehnten über die Anzahl der Dimensionen auf den verschiedenen Generalitätsebenen und über die Generalitätsgrade der Faktoren diskutiert (z.B. Costa & McCrae, 1992a; Eysenck, 1991, 1992).

Dabei war in den letzten Jahren kaum eine wesentliche Veränderung der Positionen bedeutender Proponenten von Traitmodellen festzustellen: „BIG-5“ Proponenten stellen immer wieder die BIG-5 fest (Costa & McCrae, 1992; Goldberg, 1990), „Giant Three“ Proponenten, die „Giant Three“ (Eysenck, 1991, 1997), eine Reihe von „Einzelkämpfern“ vertreten beharrlich Modelle mit sechs, sieben und anderen Faktorenzahlen (z.B. Andresen, 1995, 2000; Becker, 1999; Benet-Martínez & Waller, 1997; Brand, 1994). Wobei der Begriff „Einzelkämpfer“ lediglich darauf hinweisen soll, dass die Zahl der „Anhänger“, bei diesen weiteren Modellen meist relativ klein ist. Aus wissenschaftlicher Perspektive kann die Anzahl und der Einfluss der Proponenten für ein Modell durchaus ein historischer Zufall sein.

Solange es offenbar keine allgemein anerkannten Maßstäbe zur vergleichenden Bewertung der verschiedenen Modelle gibt, werden all die Modelle mit ihrem jeweiligen Geltungsanspruch weiterhin koexistieren. Man könnte die Vielfalt der Modelle natürlich auch begrüßen, etwa in der Erwartung, dass durch die Modell-Vielfalt das Konstrukt Persönlichkeit eine reichhaltigere Explikation erfährt. Da aber bisher kaum gezeigt werden konnte, dass die alternativen Modelle jeweils inkrementellen Erklärungswert besitzen, schrumpft diese Vielfalt zu einem wenig fruchtbaren „Modellpleonasmus“ zusammen. Die Vielfalt ist dann möglicherweise noch von heuristischem Wert.

Aus einer sozial-konstruktivistischen Perspektive (z.B. Westmeyer, 1995) ist die Koexistenz mehrerer Modelle zu einem Gegenstandsbereich natürlich unproblematisch und zu erwarten. Wenn jedoch die Proponenten zum gleichen Gegenstandsbereich auf der Basis wissenschaftlicher Übereinkünfte bezüglich Datenerhebung, -auswertung und Interpretation für ihre eigenen Modelle jeweils einen ‘Superioritätsanspruch‘ formulieren können (z.B. Eysenck, 1991, 1992; Costa & McCrae, 1992a, 1995), so müssen die entsprechenden Standards einen Mangel in der Verbindlichkeit der Bewertung aufweisen. Es drängt sich der Verdacht auf, dass die bestehenden Übereinkünfte, insbesondere bezüglich der Datenauswertung und Interpretation die unterschiedlichsten Konstruktionen bei gleicher Datenlage erlauben. Etwas optimistischer wäre die Vermutung, dass die Fragen, die zur Zeit in der Traitforschung diskutiert werden, vielleicht bereits so spezifisch sind, dass sie unterhalb des Auflösungsgrades der verwendeten Methodologie liegen.

Auch Eysenck (1991) hält offenbar übergeordnete Maßstäbe zur vergleichenden Bewertung psychometrischer Traitmodelle für erforderlich. Neben einigen Kriterien für kausale Traittheorien nennt Eysenck (1991, S. 790) acht Kriterien für die taxonomische Forschung, die helfen könnten, Unverbindlichkeiten bei der Bewertung verschiedener Modelle zu reduzieren:

- (1) Theory-based, and covering a wide field.
- (2) Theory formulated to permit testable deductions.
- (3) Theory deals with description and interaction of traits.
- (4) Theory has a hierarchical form.
- (5) Outcome of psychometric tests must be replicable.
- (6) First- and second-order factors must be clearly distinguished.
- (7) Meta-analysis of results must show convergence.
- (8) Outcome of studies using great variety of scales must show convergence.

Die genannten Kriterien beinhalten teilweise methodologische Forderungen (Replizierbarkeit, Konvergenz der Befunde bei unterschiedlichen Variablen), andere Kriterien wie Theorie-Orientierung und das Abdecken eines weiten Feldes (1) können nur schwer intersubjektiv einheitlich bewertet werden. Insgesamt stehen in Eysencks Liste allgemeine Kriterien wie Theorie-Orientierung, spezielle inhaltliche Struktur-Hypothesen (hierarchische Form) neben methodischen Anforderungen (Replizierbarkeit). So sinnvoll die genannten Kriterien jeweils sein mögen, bei der Kombination der Kriterien stellt sich zum einen das Problem der genauen Bedeutung allgemeiner Ziele wie „Theorie-Orientierung“ oder „Abdecken eines breiten Feldes“ und zum anderen das Problem der relativen Gewichtung der einzelnen Punkte. Auch ist nicht unbedingt einsichtig, warum Traittheorien Aussagen über die Interaktionen von Traits enthalten oder hierarchisch sein müssen. Interaktionen von Traits und Hierarchie können sich zwar als Ergebnis empirischer Untersuchungen ergeben, es erscheint jedoch fragwürdig, inhaltliche Befunde als Kriterien zur Bewertung von Traitmodellen vorab festzulegen. Darüber hinaus verweisen vier Punkte von Eysencks Liste wieder auf das Problem der Faktorenzahl und Generalität: Das Abdecken eines weiten Feldes (1), die hierarchische Form (4), die klare Unterscheidbarkeit von Faktoren erster und zweiter Ordnung (6) sowie die Konvergenz von Studien mit unterschiedlichen Skalen (8) hängen mit Fragen der Faktorenzahl und Generalität eng zusammen.

Die übrigen Punkte – außer dem rein methodischen Kriterium der Replizierbarkeit (5) und der kaum zuzuordnenden weil völlig allgemeinen Theorie-Orientierung (1) – betreffen den experimentellen Forschungsansatz: prüfbare Deduktionen (2) und Interaktion zwischen Traits (3). Somit ergibt sich für diese Liste von Kriterien, die in Kapitel 2.1 dargestellte Problematik der Gewichtung von Kriterien unterschiedlicher methodischer Provenienz. Somit erscheint die Liste in Eysenck (1991) insgesamt als ein erster, unvollständiger und teilweise problematischer Entwurf, der noch nicht als Standard für die Bewertung von Modellen verwendet werden kann, sondern lediglich eine Diskussionsbasis darstellt.

Bei einer Beschränkung auf die im Rahmen dieser Arbeit interessierenden an der psychometrischen Struktur orientierten Kriterien aus Eysenck (1991) kann festgehalten werden, dass sich diese auf die Fragen der Dimensionsanzahl und/oder Generalität beziehen. Dies kann – neben den bereits skizzierten Diskussionen – als Hinweis darauf gelten, dass diese Bereiche von besonderer Relevanz sind.

Sowohl bezüglich der Dimensionszahl als auch bezüglich der Generalität ist dabei ein hohes Maß an Unverbindlichkeit feststellbar: Der Bereich der Dimensionszahl betrifft eher die

empirisch-methodische Unverbindlichkeit, d.h. die Verwendung multivariater Verfahren. Hier steht eine äußerst große Vielzahl anerkannter Methoden zur Verfügung, so dass die Schwierigkeit in der Auswahl der methodisch überzeugendsten Kriterien besteht. Eine solche Situation trifft vor allem auf die Bestimmung der Faktorenzahl im Rahmen exploratorischer Faktorenanalysen zu, wo es eine Vielzahl – teilweise hochwertiger – Kriterien gibt, aber bestimmte althergebrachte und längst widerlegte „Faustregeln“ immer noch ihr Unwesen treiben (s. Costa, McCrae & Dye, 1991). Es könnte eingewendet werden, dass – zumindest aus einer sozial-konstruktivistischen Sicht – diejenigen methodischen Kriterien als angemessen gelten können, die sich sozial durchsetzen. Dem kann nach Westmeyer (1995, S. 750) folgendes entgegengehalten werden:

„Vom empirischen Gehalt einer psychologischen Theorie kann erst dann gesprochen werden, wenn sie in eine entsprechende methodische Umgebung eingebettet wurde, die alle methodischen Konstruktionen umfaßt, auf die bei einer empirischen Überprüfung der Theorie zurückgegriffen werden muss/kann.“

Nun stellt sich die Frage, wie die „entsprechende methodische Umgebung“ festgelegt werden kann. Man könnte beispielsweise festlegen, dass sich die methodische Umgebung der psychometrischen Traitmodelle aus den Beiträgen in multivariat und psychometrisch orientierten Methodenzeitschriften und Lehrbüchern ergibt. Es darf bezweifelt werden, dass die Faktorenanalysen, wie sie häufig im Traitbereich durchgeführt werden, den in der verfügbaren Literatur benannten methodischen Standards genügen. Vor allem wenn Methoden, die in vergleichenden Methodenstudien als besonders leistungsfähig identifiziert wurden (z.B. Parallelanalyse; Hubbard und Allen, 1987; Velicer et al., 2000), in den Analysen empirischer Studien kaum eingesetzt und nicht einmal diskutiert werden. Dementsprechend stellt beispielsweise Brocke (2000, S. 11) fest:

„In Bezug auf derartige Kriterien und die defizitäre empirische Basis vieler faktorenanalytischer Persönlichkeitssysteme ist in der scientific community eine bemerkenswerte Toleranz zu beobachten.“

Somit wird hier davon ausgegangen, dass eine Optimierung der Einbettung psychometrischer Traitmodelle in die methodische Umgebung möglich ist und dass diese Optimierung mit einer Zunahme des empirischen Gehalts der Traitmodelle verbunden ist.

Der zweite Bereich betrifft eher die *theoretisch-interpretative Unverbindlichkeit* und dort insbesondere Annahmen zur Repräsentativität der Modelle und zur Generalität der Traits. Zur Generalität von Traits erfährt man meist nur, dass sie sehr hoch ist (z.B. Costa et al., 1991), ohne dass über den Nutzen einer derartig hohen Generalität oder über deren methodisch-empirische Legitimation etwas ausgesagt wird (s. aber Ones & Viswesvaran, 1996). Zur Relativierung des Nutzens hoher Generalität siehe z.B. Block (1995), Mershon und Gorsuch (1988), Schneider, Hough und Dunnette (1996), sowie Saucier und Ostendorf (1999). Besonders markant wird die Unverbindlichkeit, wenn die wenigen Möglichkeiten, die zur Bewertung der Generalität von Traits bestehen, z.B. der Ordnungsgrad der Faktoren, negiert werden, indem von Faktoren erster Ordnung konsistent behauptet wird, sie hätten die Generalität von Faktoren zweiter Ordnung. Im Rahmen des Fünf-Faktoren-Modells (FFM) werden nämlich die fünf Faktoren erster Ordnung häufig i.S. von Faktoren zweiter Ordnung aufgefasst, zu denen im nachhinein Skalen als Surrogate für die eigentlichen Faktoren erster Ordnung entwickelt werden (z.B. Costa et al., 1991; Costa & McCrae, 1992b; Saucier & Ostendorf, 1999). Wenn man die Zuschreibung einer bestimmten Generalität zu einem Faktor als Interpretationshypothese auffasst, kann man mit Pawlik (1968) feststellen, dass die Überprüfung faktorenanalytischer Interpretationshypothesen offenbar zu wenig Beachtung findet.

Bemerkenswert ist, dass sowohl die *empirisch-methodische Unverbindlichkeit* bei multivariaten Analysen als auch die *theoretisch-interpretative Unverbindlichkeit* aufgrund willkürlicher Generalitätseinschätzungen sich auf für die Traitforschung zentrale Fragen bezieht. Denn sowohl bezüglich der Dimensionsanzahl als auch der Generalität von Traits konkurrieren die verschiedenen Traitmodelle miteinander (z.B. Cattell, 1972; Costa & McCrae, 1992b; Guilford, 1975; Eysenck, 1991).

Die Ziele der Arbeit ergeben sich direkt aus den oben skizzierten Defiziten psychometrischer Traitforschung:

Mit der vorliegenden Arbeit sollen Vorschläge zur Verminderung oder Vermeidung sowohl der *empirisch-methodischen Unverbindlichkeit* bei Modellevaluationen aufgrund defizitärer Verwendung multivariater Verfahren als auch der *theoretischen Unverbindlichkeit* aufgrund willkürlicher Generalitätseinschätzungen erarbeitet werden. Daher sollen mit der vorliegenden Arbeit die verfügbaren methodischen Standards multivariater Analysen, vor allem in Bezug auf die Bestimmung der Anzahl der Dimensionen, gesichtet und bewertet werden und neue Möglichkeiten der Bewertung der Generalität von Traits aufgezeigt werden. Diese Vorschläge

sollen dann auf einen empirischen Datensatz angewandt werden, um die Möglichkeiten und Konsequenzen der Vorschläge bewerten zu können.

Mit der Fokussierung auf die multivariaten Analysen und Ermittlung der Generalität wird nicht behauptet, dass dies die einzigen Bereiche seien, von denen eine Unverbindlichkeit in der psychometrischen Traitforschung ausgeht. Auch wurden die Unverbindlichkeiten innerhalb des Bereichs der psychometrischen Traitforschung von Problemen und Grenzen unterschieden, die bei einer Gesamtbetrachtung dieser Forschungstradition im Kontext anderer Disziplinen deutlich wurden. Der an den über die immanenten Probleme der psychometrischen Traitforschung hinausgehenden Fragen interessierte Leser sei auf die Arbeiten von Asendorpf (1991), Brocke (2000), Cattell (1957), Eysenck (1997), Schmitz (2000) und Wittmann (1988) verwiesen.

3. Zur empirisch-methodischen Unverbindlichkeit der Traitforschung

3.1 Bestandsaufnahme: Zum Einsatz multivariater Verfahren

Ein Hauptproblem bei der Bewertung und Diskussion unterschiedlicher Traitmodelle besteht in der Vielzahl multivariater Verfahren (z.B. konfirmatorische und exploratorische Faktorenanalyse) einerseits und in den vielen Möglichkeiten der Handhabung dieser Verfahren andererseits (z.B. Extraktions-, Rotationsmethoden, Fit-Indizes). Dieselbe Schwierigkeit besteht prinzipiell auch im Bereich der Intelligenzstrukturforschung. Dort waren die verschiedenen Modelle meist an verschiedene multivariate Analyseverfahren gekoppelt. Aber dort hat diese Problematik zu einer vitalen Auseinandersetzung mit der Entwicklung unterschiedlichster Modellstrukturen geführt (z.B. Carroll, 1993; Cattell, 1987; Guilford, 1967; Jäger, 1982; Spearman, 1927; Thurstone & Thurstone, 1941). Vermutlich hat die meist höhere Reliabilität und Interkorrelation von Intelligenztestdaten im Vergleich zu Items in Persönlichkeitstests zur größeren Verbindlichkeit und Prägnanz der verschiedenen Modellvorstellungen beigetragen.

Darüber hinaus mag aber auch ein besonders "freizügiger" Umgang mit den multivariaten Analyseverfahren (insbesondere der Faktorenanalyse) im Traitbereich zu einer relativen Unverbindlichkeit der Befunde und Interpretationen geführt haben. Die Stellung der Faktorenanalyse im Bereich der psychometrischen Traitforschung ist dabei besonders interessant: Häufig wird die Faktorenanalyse aufgrund der vermuteten Willkür der mit ihr zu erzeugenden Ergebnisse kritisiert, auch wenn diese Kritik meist nicht spezifisch für die Faktorenanalyse ist (s. Holz-Ebeling, 1995). Kritikwürdig erscheint jedoch weniger die Faktorenanalyse als solche, sondern vielmehr die Art ihrer Verwendung (Holz-Ebeling, 1995, S. 19):

Bedenklich erscheint aber, dass die historisch gewachsenen Konventionen der Verwendung von Faktorenanalysen Eingriffsstellen vorsehen, die die Einflussnahme auf das, was als Befund dargestellt wird, nicht nur ermöglichen, sondern legitimieren und als unproblematisch erscheinen lassen.

Da ist zunächst die Entscheidung für eine bestimmte Faktorenzahl. Mit erstaunlich großer Selbstverständlichkeit betrachten wir das Kriterium der 'Interpretierbarkeit' als eines der wichtigsten Auswahlkriterien.

McDonald (1985) nennt Gewohnheiten der Verwendung der Faktorenanalyse ohne formale oder explizite Basis die informelle "Folklore" der faktorenanalytischen Praxis. McDonald (1985) fokussiert dabei vor allem den laxen Umgang mit dem Begriff der Einfachstruktur, wenn Faktorladungen von .30 bereits als Nullladungen interpretiert werden können (zur Kritik

an diesem Vorgehen siehe auch Beauducel, 1997). Trotz der Kritik von McDonald (1985) wird die Nicht-Berücksichtigung von Ladungsbeträgen $\leq .30$ unabhängig von Überlegungen zur Signifikanz derartiger Ladungen, immernoch praktiziert und empfohlen (z.B. Streiner, 1994). Ein weiteres Beispiel für ein vollkommen willkürliches, methodisch nicht begründbares Vorgehen bei der Faktorenanalyse ist die Empfehlung, dass die extrahierten Faktoren insgesamt 50% der analysierten Gesamtvarianz erklären sollen, ohne dass bei dieser Empfehlung in Streiner (1994) die Differenzierung zwischen Gesamtvarianz und systematischer Varianz auch nur erwähnt wird. Die Liste der methodisch kaum begründbaren „Faustregeln“, die bei der Faktorenanalyse verwendet werden, könnte noch fortgesetzt werden. Dies soll jedoch hier nicht geschehen, da in Kapitel 3.2 die im Bereich der psychometrischen Persönlichkeitsforschung häufig verwendeten Vorgehensweisen ausführlich diskutiert werden. Es sei nur erwähnt, dass Vassend und Skrandal (1997) eine neue Variante vergleichbarer „Folklore“ für den Bereich der konfirmatorischen Faktorenanalyse befürchten.

Neben der spezifischen Kritik bestimmter Traditionen der unzureichenden Anwendung der Faktorenanalyse gibt es auch eine vergleichsweise allgemeine Kritik dahingehend, dass die methodischen Spielräume bei der Anwendung dieses Verfahrens groß sind und dass deswegen die Ergebnisse der exploratorischen Faktorenanalyse stark zu relativieren sind (z.B. Block, 1995). Viele Anwender nutzen vermutlich die genannten Spielräume, um ihre jeweiligen Positionen zu stützen. Leider wird die Möglichkeit, dass teilweise deutliche Unterschiede in der methodischen Qualität der verschiedenen Berechnungsalternativen geprüft und festgestellt wurden (z.B. Zwick & Velicer, 1986), sowohl von Kritikern als auch von Anwendern häufig vernachlässigt.

Die Kontroverse zwischen Block (1995) und Costa und McCrae (1995) ist ein Beispiel dafür, dass Kritiker und Protagonisten von faktorenanalytisch begründeten Traitmodellen bezüglich der Qualitätsunterschiede der verschiedenen Anwendungsmöglichkeiten der exploratorischen Faktorenanalyse kaum differenzieren. Der Gedanke, dass aus einer Qualitätsrangfolge für verschiedene Optionen und Varianten der Faktorenanalyse eine zunehmende Standardisierung der Anwendung und damit eine Erhöhung der Verbindlichkeit der Befunde folgen würde, wird jedenfalls in dieser Kontroverse nicht artikuliert. Die Kontroverse hat zwar die wissenschaftliche Bedeutung faktorenanalytischer Traitmodelle zum Gegenstand, blendet aber die methodische Qualität verschiedener Optionen und Varianten der Faktorenanalyse als Einflussgröße auf die wissenschaftliche Bedeutung der Modelle weitgehend aus. Dasselbe gilt übrigens auch für die Kontroverse zwischen Eysenck (1992) und Costa und McCrae (1992b).

Jedenfalls scheinen die bekannte Kritik an der faktorenanalytischen Praxis und die ebenfalls bekannten (zumindest publizierten) Möglichkeiten zur Verbesserung der Praxis (s. Kapitel 3.2), bisher kaum zu einer verbesserten, verbindlicheren Praxis geführt zu haben. Vielleicht auch, weil die Kritik an der faktorenanalytischen Praxis oft mit der Kritik an der Faktorenanalyse als solches verwechselt wird. Vielmehr scheint gerade eine diffus-kritische Bewertung der Faktorenanalyse zusammen mit der Vielzahl von Entscheidungsmöglichkeiten die Tendenz zu einem „freizügigen“ Umgang mit der Faktorenanalyse, d.h. die maximale Einflussmöglichkeit der Intentionen der Anwender auf die Ergebnisse noch zu befördern. So kreativ und die wissenschaftliche Intuition fördernd ein solcher „souveräner“ Umgang mit der Faktorenanalyse erscheinen mag, die daraus resultierende Unverbindlichkeit der Befunde kann viel zu der an Stagnation grenzenden Perpetuierung von Diskussionen zur Faktorenzahl und -generalität beigetragen haben. Denn offenbar konnten die Proponenten verschiedener Modelle hinreichende Möglichkeiten zur Stützung ihrer jeweiligen Positionen aus dem Reichtum der gewohnheitsmäßig legitimierten, „souveränen“ Anwendungspraxis schöpfen.

Die „kreative Souveränität“ bei der Anwendung der exploratorischen Faktorenanalyse hat sich auch auf die Publikationspraxis ausgewirkt. Die zur Bewertung einer Lösung erforderlichen numerischen Informationen werden nicht durchgängig berichtet. Oft werden die Informationen in einer Weise verbal zusammengefasst, dass der Leser die Angemessenheit der Schlussfolgerungen nicht nachprüfen kann. So berichtet Cattell (1994b) bei seinen exploratorischen Faktorenanalysen weder Eigenwerte noch Ladungsmuster; auch Carver und White (1994), Costa und McCrae (1988) sowie Costa et al. (1991) berichten keine Eigenwerte. Auch wenn nützliche Alternativen zur Ermittlung der Faktorenzahl verwendet werden, sollte dies nicht zum Anlass genommen werden, dem Leser konventionelle Informationen, wie die Eigenwerte einer Faktorenlösung vorzuenthalten wie das beispielsweise in Borkenau und Ostendorf, (1990) oder Goldberg (1990) erfolgt, die zwar Komparabilitäten aber keine Eigenwerte berichten. In einigen Studien werden die aufgeklärten Varianzen für die extrahierten Faktoren berichtet, aus denen, wenn es sich um die unrotierten Faktoren handelt, die Eigenwerte der extrahierten Faktoren ermittelt werden können (z.B. Weyers, Krebs & Janke, 1995). Allerdings können aus diesen Angaben die Eigenwerte der nicht-extrahierten Faktoren nicht ermittelt werden, so dass auch diese Angaben unvollständig sind. Manche Autoren beschränken sich auf die Angabe der Eigenwerte der extrahierten bzw. rotierten Faktoren (z.B. Church & Burke, 1994), woraus die Angemessenheit der Extraktionsentscheidung nicht optimal zu evaluieren ist. Dennoch sei angemerkt, dass in der Mehrzahl der Studien mit exploratorischen Faktorenanalysen immer noch mehr Eigenwerte als extrahierte Faktoren berichtet werden (z.B. Beauducel et al., 1999; Becker, 1999; Benet & Waller, 1995; Benet-Martínez & Waller, 1997;

Borkenau & Ostendorf, 1989, 1991; Cattell, 1994a; Church, Katigbak & Reyes, 1998; Saucier, 1997; Saucier & Ostendorf, 1999; Zuckerman, Kuhlman & Camac, 1988; Zuckerman et al., 1993). Man kann also nicht davon ausgehen, dass bezüglich des Berichtens von Eigenwerten inzwischen neue Standards bestehen.

Das defizitäre Berichten erforderlicher Informationen scheint den Charakter eines „Kavaliersdeliktes“ erhalten zu haben. Man denke sich zum Vergleich das Berichten varianzanalytischer Befunde, ohne Freiheitsgrade, F- und P-Werte. Die defizitäre Publikationspraxis wirkt sich vermutlich nachhaltig auf die Bewertung der Methode und der daraus resultierenden Ergebnisse aus, so dass eine Aufarbeitung der methodischen Standards zur Erhöhung der Verbindlichkeit der psychometrischen Traitforschung im Rahmen dieser Arbeit dringend geboten erscheint. Aus den genannten Gründen sollen – wie bereits dargelegt – im weiteren die Möglichkeiten im Rahmen multivariater Analysen und insbesondere der Faktorenanalyse zu verbindlicheren Bewertungen der Qualität unterschiedlicher Traitmodelle zu kommen ausgelotet werden (s. Kapitel 3.2).

Bezüglich der Beurteilung der Dimensionalität von Persönlichkeitsmerkmalen im Rahmen der probabilistischen Testtheorien wird hier folgendes festgehalten: Die Analysen können kompliziert werden und es ist nicht immer sicher, dass man Persönlichkeitsdimensionen eindimensional skalieren kann (s. z.B. Rost, Carstensen & von Davier, 1999). Inzwischen liegen auch probabilistische Testmodelle für mehrdimensionale Analysen vor (z.B. Carstensen, 2000; Embretson, 1991). Deren Anwendung könnte auch im Kontext von Persönlichkeitseigenschaften interessant sein. Damit würde jedoch eine neue Forschungsrichtung eingeschlagen, während die vorliegende Arbeit sich mit der Optimierung der bestehenden psychometrischen Traitforschung befasst. Auch lassen sich die typischen Fragen der Dimensionsanalyse in der Differentiellen Psychologie über Modelltests für probabilistische Modelle im allgemeinen nicht lösen (Fischer, 1996). Aus diesem Grund kommen, obgleich hier ein Innovationspotential besteht, diese Methoden in der vorliegenden Arbeit nicht zum Einsatz.

3.2 Vorschläge: Verbindlichere Traitforschung durch anspruchsvolleren Einsatz multivariater Verfahren²

Wie in Kapitel 2.2 ausgeführt, beruht ein anspruchsvollerer Einsatz multivariater Verfahren auf einer intensiveren Berücksichtigung der in der einschlägigen Methoden-Literatur benannten Verfahren. Die Anzahl der inzwischen entwickelten multivariaten, ordnungsanalytischen Verfahren ist allerdings so groß, dass im Rahmen dieser Arbeit kein Überblick gegeben werden kann. Allein das Spektrum der ordnungsanalytischen oder strukturanalytischen Verfahren, d.h. der Verfahren, mit denen die Struktur eines Datensatzes ohne Berücksichtigung von Prädiktor-Kriteriums-Relationen beschrieben wird, reicht von Verfahren auf Basis der Analyse logischer Strukturen (z.B. Feature Pattern Analyse; Feger, 1988, 2000), über neuronale Netzwerke (z.B. Independent Component Analysis; Bartlett et al., 1995), Clusteranalytische Verfahren (z.B. Moosbrugger & Frank, 1992), Multidimensionale Skalierung (z.B. Borg, 1981), exploratorische und konfirmatorische Faktorenanalysen (z.B. Gorsuch, 1983; Hodapp, 1984), bis hin zu strukturanalytischen Prüfungen auf der Basis multivariater probabilistischer Modelle (z.B. Carstensen, 2000). Es ließen sich vermutlich noch andere Verfahren in das Spektrum ordnungs- oder strukturanalytischer Verfahren einordnen. Typischer Weise wird bereits das hier genannte Spektrum von Verfahren von Lehrbüchern der multivariaten Analysetechnik nicht abgedeckt (z.B. Backhaus, Erichson, Plinke & Weiber, 1996). Im Rahmen dieser Arbeit sollen nur die faktorenanalytischen Verfahren diskutiert werden, da diese im Rahmen der psychometrischen Traitforschung am häufigsten eingesetzt wurden. Denn die hier intendierte Reduktion der Unverbindlichkeiten psychometrischer Traitforschung kann nicht durch die Exploration völlig anderer Verfahrensmöglichkeiten geleistet werden, sondern vor allem durch die Aufarbeitung der Vor- und Nachteile der Varianten der regelmäßig eingesetzten Verfahren. Auch wenn die Evaluation alternativer multivariater Verfahren für die Traitforschung der Zielsetzung der vorliegenden Arbeit entgegensteht, wäre eine derartige Evaluation grundsätzlich natürlich interessant.

Die vorliegende Ausarbeitung beschränkt sich somit auf die konfirmatorische und exploratorische Faktorenanalyse. Beide Verfahrensklassen wurden im Rahmen der psychometrischen Traitforschung regelmäßig eingesetzt, wobei die verbreiteten Traitmodelle, z.B. das FFM

² Die im folgenden dargestellten methodischen Standards beim Einsatz multivariater Verfahren können und sollen natürlich nicht einschlägige Lehrbücher (z.B. Bortz, 1999; Gorsuch, 1983; Harman, 1976; Hodapp, 1984; Revenstorff, 1976; Überla, 1968) ersetzen. Allerdings werden in den Lehrbüchern vor allem die Methoden selbst, weniger spezifische Empfehlungen und Probleme bei der Anwendung diskutiert. Hier sollen vor allem die Methoden und Anwendungsempfehlungen herausgearbeitet werden, die dazu beitragen können, die Verbindlichkeit faktorenanalytischer Befunde in der psychometrischen Traitforschung zu erhöhen.

(Goldberg, 1990), oder Cattells (1994a) und Eysencks (1991) Modelle primär auf der exploratorischen Faktorenanalyse beruhen, während konfirmatorische Faktorenanalysen eher zur Analyse einzelner Inventare eingesetzt wurden (z.B. Strobel, Beauducel, Debener & Brocke, in Druck; Haynes et al., 2000; Heubeck, Wilkinson & Cologon, 1998; Vassend & Skrandal, 1997). Aus diesem Grund werden die Varianten der exploratorischen Faktorenanalyse besonders detailliert betrachtet.

3.2.1 Zur konfirmatorischen Faktorenanalyse

Die konfirmatorische Faktorenanalyse (KFA) kann als Teil der Analyse linearer struktureller Beziehungen (Linear Structural Relationships; LISREL; Jöreskog & Sörbom, 1981, 1993; siehe aber auch Bentler, 1995) aufgefaßt werden. Obgleich die KFA als Bestandteil von Strukturgleichungsmodellen aufgefasst werden kann, wird hier nur auf die KFA eingegangen, da die Prüfung von Prädiktionen in der Regel erst in einem zweiten Schritt auf der Basis der Modelle erfolgt, die in der KFA ermittelt wurden. Außerdem können die großen, zur Zeit diskutierten Traitmodelle angemessener anhand der KFA beschrieben werden, als anhand von Prädiktionsmodellen. Zu den Problemen der über die KFA hinausgehenden Aspekte von Strukturgleichungsmodellen siehe Holling (1993).

Die Möglichkeit postulierte Modelle zu testen hat in Verbindung mit der hohen Flexibilität bei der Formulierung der Modelle zu einer hohen Verbreitung der KFA auch im Kontext der psychometrischen Traitforschung geführt. Von daher wird zunächst untersucht, inwieweit die Unverbindlichkeit der psychometrischen Traitforschung auf eine besondere Anwendungspraxis der KFA zurückgeführt werden kann. Vor diesem Hintergrund sollen dann die Spezifikationen festgelegt werden, mit denen die KFA zunächst in der vorliegenden Arbeit zur Überprüfung postulierter Modelle bzw. für den Vergleich von Modellen mit unterschiedlicher Faktorenanzahl verwendet werden soll.

KFA und Unverbindlichkeit psychometrischer Traitforschung

In den letzten Jahren sind kritische Einschätzungen zum Einsatz der KFA im Kontext der psychometrischen Traitforschung abgegeben worden. So kommen Church und Burke (1994) zu dem Schluss, dass es unwahrscheinlich ist, dass ökonomische (parsimonious) Persönlichkeitsmodelle in der KFA nach konventionellen Kriterien einen angemessenen Fit erreichen. Vassend und Skrandal (1997) bringen die schlechte Anpassungsgüte der Modelle an die Daten bei KFA mit der geringen Einfachstruktur in Verbindung, die diese Modelle bereits in den

exploratorischen Faktorenanalysen aufweisen. Sie argumentieren dahingehend, dass im Rahmen der exploratorischen Faktorenanalyse eine Aufweichung der Standards erfolgte, die nun im Rahmen der KFA zur Aufdeckung geringer Modell-Anpassungen für die anhand der exploratorischen Analyse entwickelten Modelle (in diesem Fall das FFM) führte.

Borkenau und Ostendorf (1990), erhielten ebenfalls eine geringe Modell-Anpassung an die Daten bei einer KFA von Fragebogen zum FFM. Da die exploratorischen Analysen und insbesondere die Komparabilitätskoeffizienten (Everett, 1983) in ihren Daten eine fünf-Faktoren-Lösung nahe legten, gelangen sie zu dem Schluss, dass die KFA die Güte der Lösungen nicht optimal abbildet. Sie sind der Auffassung, dass die KFA eine größere Anzahl von Faktoren nahe legt, während die Komparabilitätskoeffizienten zeigen, dass Lösungen mit einer größeren Anzahl von Faktoren nicht ausreichend replizierbar sind. Dem könnte man allerdings entgegenhalten, dass Lanning (1996) zeigen konnte, dass die Anzahl von Faktoren, die Anhand von Komparabilitätskoeffizienten als robust ausgewiesen werden, mit der Größe der untersuchten Stichprobe korreliert. Da Borkenau und Ostendorf (1990) nur 256 Probanden untersuchten, kann die Replizierbarkeit von nur fünf Faktoren, nur als erster Hinweis auf die Angemessenheit ihrer Argumentation dienen.

Allerdings konnten – in Übereinstimmung mit der Argumentation von Borkenau und Ostendorf (1990) – McCrae, Zonderman, Costa, Bond and Paunonen (1996) zeigen, dass mit den konventionellen Fit-Kriterien bei der KFA, Modelle abgelehnt werden können, die replizierbar sind und Modelle akzeptiert werden können, die nicht replizierbar sind. Die Stichproben in McCrae et al. (1996) waren meist größer als die von Borkenau und Ostendorf (1990), so dass der Nachweis der Replizierbarkeit von nur fünf Faktoren anhand von Komparabilitätskoeffizienten hier kaum auf zu geringe Stichprobengröße zurückgeführt werden kann. Der ungenügende Modellfit ergab sich in McCrae et al. (1996) für sehr viele konventionelle Fit-Indizes (z.B. Goodness of Fit Index, Normed Fit Index, Tucker-Lewis-Index, Comparative-Fit-Index), allerdings stellten McCrae et al. (1996) für Browne und Cudecks (1993) "Root Mean Square Error of Approximation" (RMSEA) bei einem Modell mit fünf korrelierten Faktoren eine akzeptable Anpassung an die Daten fest. Interessant ist an diesem obliquen Modell, dass es sowohl durch exploratorische als auch durch konfirmatorische Faktorenanalyse gestützt wird. McCrae et al. (1996) argumentieren allerdings bei Modellprüfungen für eine Verwendung der exploratorischen Faktorenanalyse mit Procrustes-Rotation, bei der inzwischen Möglichkeiten der zufallskritischen Absicherung der Lösungen bestehen (Paunonen, Jackson, Trzebinski & Forsterling, 1992).

Raykov (1998) gelangt zu der Einschätzung, dass die Verwendung der KFA in der psychometrischen Traitforschung problematisch war, weil eine unangemessene Nullhypothese zugrunde gelegt wurde. Er ist der Auffassung, dass die Nullhypothese der perfekten Reproduktion der Kovarianz- oder Korrelationsmatrix durch ein Modell, im Kontext psychometrischer Traits unrealistisch ist. Raykov (1998) begründet dies damit, dass Phänomene im Bereich der Persönlichkeitspsychologie extrem komplex und multifaktoriell bedingt sind und das jedes Modell, bereits vor der empirischen Überprüfung vor dem Hintergrund der extremen Komplexität des Gegenstandsbereiches als unzutreffend angesehen werden kann. Anstatt der perfekten Modellanpassung schlägt er daher vor zu prüfen, ob ein Modell eine vernünftige Approximation der Daten darstellt (Browne & Cudeck, 1993). Diese Prüfung kann anhand des RMSEA erfolgen, der im übrigen nur wenig von der Stichprobengröße abhängen soll. Der Index ist ein "Badness-of-Fit" Index, d.h. kleine Werte ($<.05$) entsprechen nach Browne und Cudeck einer hohen Modellgüte, während Werte zwischen $.05$ und $.08$ einem akzeptablen Fit entsprechen.

Das Problem der Argumentation von Raykov (1998) ist, dass die meisten Gegenstandsbereiche, die mit multivariaten Analysen untersucht werden, sehr komplex sind. Welcher Gegenstandsbereich darf nun eine erhöhte Komplexität reklamieren, der ihn von einer strengeren Modellprüfung bewahrt? Im Kontext der multivariaten Intelligenzforschung sind konventionelle Anpassungs-Indizes regelmäßig mit Erfolg verwendet worden (z.B. Beauducel, Brocke, Liepmann, 2001; Gustafsson, 1984; Oberauer, Süß, Schulze, Wilhelm & Wittmann, 2000), dabei waren die geprüften Modelle in den genannten Arbeiten durchaus komplex. Allerdings muss angemerkt werden, dass die Reliabilität von Leistungsmaßen in der Regel über der von Persönlichkeitsfragebogen liegt. Insofern sind die spezifischen Probleme der psychometrischen Traitforschung vielleicht eher auf eine unzureichende Reliabilität der verwendeten Maße zurückzuführen.

Zum in dieser Arbeit zentralen Problem der Unverbindlichkeit psychometrischer Traitforschung kann die KFA, weil sie in diesem Kontext vermutlich eher zu einer Zurückweisung von Modellen geführt hat, bisher nur in geringem Umfang beigetragen haben. In dieser Hinsicht hat sich die von vielen als für die Traitforschung unangemessen eingestufte Strenge der Modellprüfung als nützlich erwiesen. Hinzu kommt, dass – wie bereits erwähnt – die größeren konkurrierenden Traitmodelle nur selten anhand der KFA geprüft und verglichen wurden.

Auch wenn die KFA wegen häufiger Zurückweisung von Modellen bisher vermutlich eher weniger zur Unverbindlichkeit der Traitforschung beigetragen hat, besteht die Gefahr, dass bei

einer Neuformulierung spezifischer Standards für die Persönlichkeitsforschung eine derartige Unverbindlichkeit entsteht. So kritisieren Vassend und Skrandal (1997), dass die Probleme anhand der KFA das FFM mit dem NEO-PI (NEO Personality Inventory) nachzuweisen, beispielsweise in McCrae et al. (1996) eher zu einer Problematisierung der KFA geführt haben als zu einer Problematisierung des NEO-PI. Ein weiterer Aspekt der KFA, der durchaus zu einem, allerdings schwer zu ermittelnden, Ausmaß an Unverbindlichkeit geführt haben könnte, ist die große Anzahl von Indizes zur Bewertung der Modellgüte. Hier können Unverbindlichkeiten nur durch eine begründete Vorab-Auswahl vermieden werden.

Zur Verwendung der KFA in dieser Arbeit

Ein verfrühtes Aufgeben einer konservativen Prüfstrategie erscheint dem Ziel dieser Arbeit, die Verbindlichkeit der Traitforschung zu erhöhen, entgegenlaufend. Strenge cut-off Werte erscheinen auch indiziert, weil im Rahmen der KFA die Möglichkeit besteht Modelle sukzessive zu optimieren. Wenn die Unterscheidung zwischen rein confirmatorisch geprüften Modellen und post-hoc Anpassungen in den Publikationen verwischt wird, bestehen bei der KFA unangemessene Freiheitsgrade der Modelloptimierung. Um diese Freiheitsgrade aufzufangen, wird hier vorab festgelegt, wie die Modellprüfung erfolgen soll. Zunächst wird ein Modell mit optimaler orthogonaler Einfachstruktur, d.h. ein Modell, bei dem die postulierten Hauptladungen frei und die Nebenladungen auf null gesetzt werden, geprüft. In einem weiteren Modell werden Interkorrelationen zwischen den Faktoren zugelassen.

Bezüglich der Prüfung der Anpassungsgüte an die Daten ist folgendes zu bedenken: Für die Beibehaltung einer strengen Modellprüfung spricht, dass die meist verwendeten cut-off Werte (beispielsweise $>.90$ für den Comparative Fit Index; CFI), für die eine hinreichende Anpassung der Modelle an die Daten angenommen wird, nach Carlson und Mulaik (1993) zur Akzeptanz unangemessen spezifizierter Modelle geführt hat. Von den Autoren wird daher ein strengerer cut-off Wert für den CFI gefordert ($>.95$). Bemerkenswert dabei ist, dass der von Carlson und Mulaik (1993) als nicht streng genug eingeschätzte cut-off Wert von $.90$ für den CFI im Kontext der von McCrae et al. (1996) durchgeführten Analysen zu einer Ablehnung aller mit der KFA geprüften Fünf-Faktoren Modelle führte (nur für die exploratorische Maximum Likelihood Lösung war der CFI-Wert größer als $.90$). Aus der Literatur entsteht somit der Eindruck, dass der CFI mit einem cut-off Wert von $.90$ mitunter zu streng und mitunter nicht streng genug ist. Durch Zitieren der entsprechenden Arbeiten kann der Anwender die jeweilige Legitimation für seine Entscheidung finden, was zu einer im Sinne der Verbindlichkeit der Methoden problematischen Situation führt.

Andererseits kann die von McCrae et al. (1996) und Raykov (1998) beklagte unangemessen strenge Prüfung unter Umständen wenig informativ sein, insbesondere wenn überhaupt kein zu den Daten passendes Modell formuliert werden kann. Aus diesem Grund erscheint im Rahmen dieser Arbeit eine Bewertung der Modelle Anhand des von Raykov (1998) empfohlenen RMSEA indiziert. Um jedoch die Strenge bei der Testung nicht zu gering ausfallen zu lassen, wird in der vorliegenden Arbeit eine von Hu und Bentler (1999) vorgeschlagene Kombinationsstrategie gewählt. Hu und Bentler stellten in Vergleichen besonders wenige Fehler 1. und 2. Art für eine Kombination des RMSEA mit dem „Standardized Root Mean Squared Residual“ (SRMR) fest. Dabei schlugen sie einen cut-off Wert von .06 für RMSEA und von .09 oder .10 für SRMR vor. Allerdings empfehlen sie bei Verwendung dieser Indizes Stichproben mit über 500 Probanden zu untersuchen. Da Hu und Bentler (1999) die Komplexität der von ihnen geprüften Modelle manipulierten und beide Fehlerarten untersuchten, erscheint diese Strategie auch im vorliegenden Kontext psychometrischer Traits angemessen, sofern die Stichprobengröße mehr als 500 Probanden umfasst. In Tabelle 3.1 sind die Spezifikationen und Modellprüfungen, die hier im Rahmen der KFA durchgeführt werden, aufgeführt.

Tabelle 3.1

Im Rahmen dieser Arbeit durchgeführte KFA

Modelle	Prüfregeln
orthogonal, nur Hauptladungen frei gesetzt	Modell gilt als akzeptabel, wenn
oblique, nur Hauptladungen frei gesetzt	RMSEA \leq .06 und SRMR \leq .10

Um die Vergleichbarkeit mit anderen Modellen zu erhalten werden jedoch der χ^2 -Wert, der Goodness of Fit Index (GFI) und der CFI zusätzlich berichtet, auch wenn die Entscheidung über die Modellgültigkeit anhand der in Tabelle 3.1 dargestellten Strategie erfolgt.

3.2.2 Zur exploratorischen Faktorenanalyse

Im folgenden sollen wesentliche Variationen im Rahmen exploratorischer Faktorenanalysen behandelt werden, die durch unreflektierte oder willkürliche Verwendung zu ergebnisorientierten Permutationen von Analysen führen können. Es werden zunächst die wichtigsten Extraktionskriterien dargestellt, dann wird eine Auswahl der am besten begründeten Kriterien

vorgenommen, wobei vorab schon erwähnt sei, dass in der Anwendungspraxis nur ein kleiner Teil der verfügbaren Kriterien und – wie auch Matthews und Oddy (1993) feststellen – insbesondere nicht gerade die methodisch überzeugendsten Kriterien Verwendung finden. Nach der Auswahl der methodisch überzeugendsten Extraktionskriterien, wird versucht eine Auswahl optimaler Extraktions- und Rotationsmethoden durchzuführen.

3.2.2.1 Extraktionskriterien: Bestandsaufnahme und Auswahl

Den Extraktionskriterien wird hier besondere Aufmerksamkeit gewidmet, da die Verbindlichkeit der Entscheidung für eine bestimmte Anzahl zu extrahierender Faktoren, eine der wichtigsten Entscheidungen im Rahmen der exploratorischen Faktorenanalysen darstellt (Crawford, 1975). Vollständig kann die Darstellung deswegen schon nicht sein, weil bereits Vernon (zitiert nach Fruchter, 1954) 25 verschiedene Methoden zur Bestimmung der Faktorenanzahl nennt. Inzwischen sind – je nach zählweise – mindestens fünf bis zehn weitere Methoden hinzugekommen. Eine vollständige Darstellung würde wohl eine allein darauf ausgerichtete Abhandlung erfordern. In Horn und Engstrom (1979, S. 283) heißt es treffend: “Henry Kaiser once said ... that a solution to the number-of-factors problem in factor analysis is easy, that he used to make one up every morning before breakfast. But the problem, of course, is to find *the* solution, or at least a solution that others will regard quite highly if not as the best.“

Selbst wenn es immer noch nicht möglich sein sollte, das eine beste Extraktionskriterium zu identifizieren, so kann doch wenigstens eine Auswahl der besten Kriterien erfolgen. Die hier vorgenommene Auswahl orientiert sich vor allem an methodisch (z.B. in Simulationsstudien) besonders bewährten Extraktionskriterien. Bei der Darstellung sollen zum einen Extraktionskriterien berücksichtigt werden, die bisher häufig verwendet wurden und zum anderen Kriterien, die aufgrund des zugrundeliegenden Rationales besonders erfolgversprechend erscheinen. Die Darstellung gliedert sich in zwei Teile: Zunächst erfolgt die Bestandsaufnahme einschließlich Bewertung, dann erfolgt die Auswahl optimaler Kriterien.

Bestandsaufnahme

Eigenwerte > 1 Regel oder Guttman-Kaiser Kriterium

Dieses Kriterium läuft darauf hinaus, dass kein Faktor extrahiert werden darf, der weniger Varianz erklärt als eine einzelne Variable aufweist. Formal betrachtet ist die Varianz der einzelnen Variablen in der Faktorenanalyse meist auf eins standardisiert. Nun können sich

hinter der standardisierten Varianz von eins Varianzen von sehr unterschiedlicher Bedeutung für eine Population verbergen. Das Aggregationsniveau der zu faktorisierten Variablen kann sehr unterschiedlich sein, was zur Folge hat, dass das Extraktionskriterium sich entsprechend unterschiedlich auswirkt. Wenn einzelne Items im Trait-Kontext analysiert werden ist die Varianz eines einzelnen (oft binären) Items äußerst gering. Das Guttman-Kaiser Kriterium wird in solchen Fällen typischer Weise die Anzahl der zu extrahierenden Faktoren überschätzen (overextraction), da Faktoren extrahiert werden müssten, die kaum mehr Varianz erklären als ein einzelnes Item. Im Falle extrem aggregierter Skalen oder Faktoren kann dasselbe Kriterium zu einer Unterschätzung der Anzahl der zu extrahierenden Faktoren führen. Beispielsweise können Faktoren mit einem Eigenwert von 0.8 einen ganz bedeutenden Anteil einer Merkmalsvarianz in einer Population erklären. Neben dem Aggregationsniveau spielt auch die Reliabilität der Variablen eine Rolle: Warum sollen genau die Faktoren extrahiert werden, die etwas mehr Varianz erklären als die Gesamtvarianz einer einzelnen Variablen, wenn die Reliabilität der analysierten Variablen beispielsweise nicht größer als .70 ist?

Überdies ist das Guttman-Kaiser Kriterium stark von der Anzahl der analysierten Variablen abhängig. Gorsuch (1983) vermutet, dass dieses Kriterium nur dann gute Ergebnisse liefert, wenn die Daten so beschaffen sind, dass die Anzahl der zu analysierenden Variablen drei bis fünf mal so groß ist wie die Anzahl der zu extrahierenden Faktoren (was man vor der Analyse in der Regel jedoch nicht weiß).

Es gibt bereits eine große Zahl von Studien mit künstlichen Datensätzen (Plasmoden) zur Leistungsfähigkeit des Guttman-Kaiser Kriteriums: Cattell und Jaspers (1967) stellten in 6 von 8 Fällen eine Unterschätzung der Faktorenanzahl fest, Browne (1968) in 10 von 20 Fällen eine Überschätzung der Faktorenanzahl, Linn (1968) stellte in 5 von 16 Fällen eine Unterschätzung und in weiteren 5 dieser 16 eine Überschätzung der Faktorenanzahl fest. Weitere Arbeiten zu Problemen des Guttman-Kaiser Kriteriums finden sich in Hakstian, Rogers und Cattell (1982), Lee und Comrey (1979), Macrosson (1999), Revelle und Rocklin (1979), Velicer et al. (2000), Yeomans und Golder (1982), sowie Zwick und Velicer (1982, 1986), wobei überwiegend eine Überschätzung der Zahl der zu extrahierenden Faktoren festgestellt wurde und neben der Variablenanzahl auch die Kommunalität der Variablen einen Einfluss auf das Kriterium hat.

Überdies hat Cliff (1988) gezeigt, dass die mathematischen Überlegungen Kaisers, die zum Guttman-Kaiser Kriterium geführt haben, teilweise fehlerhaft sind. Hubbard und Allen (1987) empfehlen das Guttman-Kaiser Kriterium nicht mehr routinemäßig einzusetzen. Obwohl bereits eine Reihe von Arbeiten zu den Problemen des Guttman-Kaiser Kriteriums vorliegen,

wird dieses Kriterium noch immer ernsthaft diskutiert (z.B. Buley, 1995) und angewendet, wobei die hier aufgeführte Literatur in den entsprechenden Arbeiten nicht zitiert wird, was auf eine mangelhafte Rezeption der kritischen Arbeiten hindeutet. Allerdings findet dieses Kriterium weniger in der grundlagenorientierten Persönlichkeitsforschung Verwendung (siehe jedoch Costa et al., 1991) als vielmehr in anwendungsorientierten Traitstudien (z.B. Hojat, 1982; Kirkcaldy, Cooper, Eysenck, & Brown, 1994).

Bewertung: Das Guttman-Kaiser Kriterium ist, wie sich aus der Literatur ergibt, als solches kaum brauchbar. Dennoch kann es nur insofern zur problematischen Situation in der Traitforschung beigetragen haben, als seine Verwendung zu unreplizierbaren oder inkonsistenten Ergebnissen geführt haben mag. Aber eine systematische Nutzung der Probleme dieses Kriteriums zur Stützung eigener Positionen erscheint weitgehend ausgeschlossen. Eine Ausnahme könnte sich aus der systematischen Tendenz zur Überschätzung der Faktorenzahl – vor allem in Item-basierten Analysen – ergeben, wenn es darum geht, ein Modell mit einer größeren Anzahl postulierter Dimensionen gegen ein Modell mit weniger postulierten Dimensionen zu prüfen.

Scree-Test

Der Scree-Test geht in seiner ersten Form auf Cattell (1966) zurück. Es werden die Faktoren extrahiert, deren Eigenwert in einem Koordinatensystem von Faktorenzahl und Eigenwert über dem nahezu linearen Trend der kleinen Eigenwerte liegen (in der frühen Form wurde beim Scree-Test noch ein Faktor mehr extrahiert). Zu Problemen führt die Verwendung des Scree-Tests wenn mehrere „Sprünge“ in den Eigenwerten auftreten. Dann ist keine eindeutige Entscheidung möglich. Auch können durch gezerrte oder gestauchte Darstellungen kleine Sprünge überschätzt oder große Sprünge unterschätzt werden. Die Ergebnisse des Scree-Tests sind intersubjektiv nicht völlig einheitlich zu bewerten. Dies ergibt sich aus dem Umstand, dass die Größe von „Sprüngen“ im Eigenwertverlauf von verschiedenen Anwendern unterschiedlich bewertet werden kann. Daher kann der Scree-Test in dieser Form auch als „subjektiver“ Scree-Test bezeichnet werden (s. Gorsuch, 1983).

Darüber hinaus wurde von Gorsuch und Nelson (1981) ein „objektiver“ Scree-Test vorgeschlagen, bei dem die Steigung der ersten drei Eigenwerte mit der Steigung der nächsten drei Eigenwerte verglichen wird. Die Steigungen aller sukzessiv gebildeten (z.B. von den größten zu den kleinsten Eigenwerten) Dreiergruppen von Eigenwerten werden berechnet. Die größte Differenz zwischen den Steigungen tritt an dem Punkt auf, der die Anzahl der zu extrahierenden Faktoren markiert. Der „objektive“ Scree-Test war in Gorsuch und Nelson

(1981) dem "subjektiven" Scree-Test überlegen. Ein weiterer, weitgehend „objektiver“ Scree-Test wurde von Barrett und Kline (1982) entwickelt. Diese sogenannte „Autoscree“-Prozedur beinhaltet allerdings einige subjektive Aspekte bzw. willkürliche Setzungen (z.B. die besondere Behandlung der Steigungen, die größer als 20° sind). Inzwischen wurden weitere, vermutlich leistungsfähigere Varianten des „objektiven“ Scree-Tests entwickelt (Bentler & Yuan, 1998; Zoski & Jurs, 1993, 1996), für die allerdings noch keine Informationen zur Bewährung vorliegen.

Der "subjektive" Scree-Test hat in Cattell und Jaspers (1967) in 6 von 8 Fällen korrekte Ergebnisse erbracht. In zwei Fällen führte er zu einer Überschätzung der Faktorenanzahl. Tucker, Koopman und Linn (1969) stellten in 12 von 18 Fällen eine korrekte Bestimmung der Faktorenanzahl mit dem Scree-Test fest. In vier Fällen führte der Scree-Test zu einer Unterschätzung der Faktorenanzahl in zwei Fällen war das Ergebnis uneindeutig. Cliff (1970) stellte mit Monte Carlo Daten im allgemeinen korrekte Ergebnisse für den Scree-Test fest. Zwick und Velicer (1986) stellten nur in 57% der Fälle korrekte Ergebnisse fest. Wenn Fehler auftraten, waren es in 90% der Fälle Überschätzungen der Faktorenanzahl.

Bewertung: Der Scree-Test erbringt in Simulationsstudien meist etwas bessere Ergebnisse als das Guttman-Kaiser Kriterium. Es kommt aber auch beim Scree-Test zu einer fehlerhaften Einschätzung der Faktorenanzahl, wobei eine Tendenz zur Überschätzung der Faktorenanzahl besteht. Vielfach wird auch übersehen, dass für eine sinnvolle Verwendung des Scree-Tests eine sorgfältige Planung der Studie erforderlich ist, die sicherstellt, dass intendierte Faktoren deutlich repräsentiert bzw. „überdeterminiert“ werden (s. Horn & Engstrom, 1979). Wegen der moderaten Genauigkeit empfehlen Velicer et al. (2000) den Scree-Test niemals als alleiniges Extraktionskriterium zu verwenden. Aufgrund der subjektiven Aspekte kann der Scree-Test zu den Unverbindlichkeiten in der Traitforschung beigetragen haben. Bedauerlicher Weise wurde die Alternative des objektiven Scree-Tests bisher kaum durchgeführt, so dass eine Bewertung des Nutzens hier nicht erfolgen kann.

Minimale durchschnittliche quadrierte Partialkorrelation

Velicer (1976) schlug vor, die durchschnittliche quadrierte Partialkorrelation als Extraktionskriterium heranzuziehen. Die durchschnittliche quadrierte Partialkorrelation zwischen den Variablen wird sukzessive für alle Lösungen mit aufsteigender Faktorenanzahl ermittelt (für 1 bis m Faktoren). Die Anzahl von Faktoren, für die die durchschnittliche Partialkorrelation minimal ist, wird schließlich als optimale Lösung betrachtet. Die Verwendung der Partialkorrelationen impliziert folgendes: Solange gemeinsame Komponenten extrahiert werden, fallen die

Partialkorrelationen zwischen den Variablen (wobei die Korrelationen mit den Faktoren auspartialisiert werden) ab, da die Faktoren ja einen wesentlichen Anteil der Korrelationen zwischen den Variablen „erklären“. Von einem bestimmten Punkt steigen die Partialkorrelationen wieder an, weil die nun auspartialisierten Faktoren mit der Mehrzahl der Variablen gering oder überhaupt nicht korrelieren. Dieser Punkt ist der Übergang von den gemeinsamen Faktoren, die extrahiert werden sollen, zu den spezifischen Faktoren, die nicht extrahiert werden sollen. Velicer (1976) betont, dass auch Komponenten mit adäquater Reliabilität anhand dieses Kriteriums nicht extrahiert werden, wenn sie im wesentlichen spezifische Varianz abbilden.

Bewertung: Dieses Kriterium ist bisher selten verwendet worden, so dass eine Bewertung schwierig ist. In der vergleichenden Analysen von Zwick und Velicer (1986) wurden jedoch gute Eigenschaften für dieses Extraktionskriterium ermittelt. Barrett und Kline (1982) stellten Konvergenzen zwischen der minimalen quadrierten Partialkorrelation und ihrer Autoscreeprozedur fest. Velicer et al. (2000) führten neben der konventionellen Methode zwei neue Variationen dieses Kriteriums ein, die sich in ihren Simulationen als unterschiedlich leistungsfähig erwiesen. Nur die Variante, bei der die vierte Potenz der Partialkorrelationen zugrundegelegt wurde, erwies sich in Velicer et al. (2000) als etwa so leistungsfähig wie die weiter unten diskutierte Parallelanalyse. Da die konzeptuelle Legitimation für die Verwendung der vierten Potenz der Partialkorrelationen gering ist, spricht dieses Ergebnis weniger für die Verwendung dieses Kriteriums.

Faktor-Komparabilitäten

Diese von Everett (1983) vorgeschlagene Methode orientiert sich an der Replizierbarkeit von Faktorwerten in zufälligen Teilstichproben. Die Faktorwert-Koeffizienten (B-Gewichte) werden zunächst für zwei Teilstichproben ermittelt (dabei müssen die Lösungen in beiden Teilgruppen mit derselben Extraktions- und Rotationsmethode und mit der gleichen Faktorenanzahl gewonnen sein). Dann werden die Faktorwert-Koeffizienten auf die Gesamtstichprobe angewendet, um Faktorwerte zu schätzen. Diese Faktorwerte werden in der Gesamtstichprobe korreliert, um Faktor-Komparabilitätskoeffizienten zu berechnen. Komparabilitätskoeffizienten größer als .80 gelten nach Everett und Entekin (1980) als ausreichend. Koeffizienten größer .90 oder .95 gelten nach Everett (1983) als ausreichend.

Auch von Walkey und McCormick (1985) wurde eine Methode zur Bestimmung der Anzahl der zu extrahierenden Faktoren mit Hilfe der Replizierbarkeit von Faktoren vorgeschlagen. Allerdings wird die Ähnlichkeit der Faktoren durch einen Index ermittelt, bei dem ein Cutt-

Off-Wert für nicht bedeutsame Ladungen erforderlich ist. Walkey und McCormick (1985) verwenden .30, .40 und .50 als Cutt-Off-Werte. Die Festsetzung dieses Wertes bringt zusätzliche subjektive Entscheidungsspielräume mit sich, weshalb im weiteren nur noch die Methode von Everett (1983) diskutiert werden soll.

Gegen eine Verbindung der Frage nach der Dimensionalität einer Lösung mit der Frage nach der Replizierbarkeit (in Teilgruppen) haben Matthews und Oddy (1993) sowie Lanning (1996) argumentiert. Da die Replizierbarkeit der Faktoren auch von der Stichprobengröße abhängt, hängt die Entscheidung für eine bestimmte Faktorenanzahl ebenfalls von der Stichprobengröße ab. Die Anzahl der aufgrund der Komparabilitäten zu extrahierenden Faktoren korrelierte in Lanning (1996) zwischen .56 (bei Mindestkomparabilität von .85 pro Faktor) und .62 (bei Mindestkomparabilität von .90 pro Faktor) mit der Stichprobengröße.

Bewertung: Konzeptuell erscheint die Berücksichtigung der Replizierbarkeit von Faktoren bei der Entscheidung für eine bestimmte Faktorenanzahl interessant. Andererseits entsteht dadurch eine Abhängigkeit der Faktorenanzahl von der Stichprobengröße. Man könnte auch argumentieren, dass eine größere Stichprobe eine bessere Legitimation darstellt, um eine größere Anzahl von Faktoren zu extrahieren. Aber eine a priori Abhängigkeit der Anzahl zu extrahierender Faktoren von der Stichprobengröße führt, selbst wenn konzeptuell einige Argumente dafür sprechen, zu einer unübersichtlichen Situation. Untersuchungen mit verschiedenen Stichprobengrößen können dann noch schwieriger verglichen werden. Daher sollten Komparabilitätskoeffizienten auf keinen Fall als alleiniges Extraktionskriterium verwendet werden.

Komparabilitätskoeffizienten wurden, mitunter im Kontext des FFM zur Ermittlung der Anzahl zu extrahierender Faktoren eingesetzt (z.B. Borkenau & Ostendorf, 1990; ein analoges Vorgehen auf der Basis von Kongruenzkoeffizienten findet sich in Di Blas & Forzi, 1998; Goldberg, 1990). Inwieweit die Befunde zum FFM von der Stichprobengröße abhängen, bleibt offen. Mit großer Wahrscheinlichkeit sind diese Effekte für die Befunde zum FFM nicht entscheidend, aber es bleibt offen, wie gut das FFM repliziert werden kann, wenn andere Extraktionskriterien verwendet werden.

Lawleys Chi²-Test

Lawley (1940) entwickelte eine Formel, um die Residuen in der Maximum Likelihood Faktorenanalyse auf Signifikanz zu prüfen. Der Test setzt voraus, dass keine Faktoren identische Eigenwerte haben und dass eine genügende Zahl von Iterationen durchgeführt

wurde. Der Test kann in konfirmatorischen und exploratorischen Analysen eingesetzt werden. Im konfirmatorischen Fall wird davon ausgegangen, dass die postulierten Faktoren die systematischen Anteile der Korrelationsmatrix erklären, im exploratorischen Fall, wird der Test sukzessive für jeden weiteren extrahierten Faktor durchgeführt, solange bis die Residuen nicht mehr signifikant sind. Die sukzessive Testung ist jedoch eine Form multipler Testung, die zufallskritisch korrigiert werden müsste (Gorsuch, 1983).

Zur Qualität des Tests von Lawley finden sich widersprüchliche Ergebnisse: In Browne (1968) waren die Schätzungen der Faktorenzahl zufriedenstellend, bei Linn (1968) gab es in zwei Stichproben Unterschätzungen und in fünf Stichproben Überschätzungen der Faktorenzahl. Nur in einer Stichprobe wurde eine korrekte Schätzung der Faktorenzahl festgestellt. Harris und Harris (1971) sowie Schönemann und Wang (1972) stellten eher eine Tendenz zur Überschätzung der Faktorenzahl fest. Der Test ist im Fall der Komponentenanalyse offenbar verwendbar (Geweke & Singleton, 1980), hat aber im Falle reduzierter Kommunalitäten (z.B. bei klassischer Faktorenanalyse) eine zu geringe Teststärke (Schönemann, 1981). Beauducel (2001a) stellte für den Chi²-Test eine deutliche Abhängigkeit der Ergebnisse von der Stichprobengröße fest.

Bewertung: Der Test ist bei Verwendung der Maximum Likelihood Faktorenanalyse trotz einiger Ungenauigkeiten verwendbar. Bei anderen Extraktionsmethoden können sich jedoch erhebliche Ungenauigkeiten ergeben (Browne, 1968), weshalb eine Verwendung in diesen Fällen nicht empfohlen werden kann. In diesen Fällen ist der als nächstes dargestellte Bartlett-Test eher zu empfehlen. Grundsätzlich besteht die Problematik, dass Signifikanztests nur prüfen können, ob systematische Varianz oder Zufallsvarianz vorliegt, aber keine Aussagen über die Spezifität von Faktoren erlauben. In diesem Sinne erlauben Signifikanztests nur die Festlegung der Obergrenze der Zahl zu extrahierender Faktoren, was aber bereits eine nützliche Eingrenzung darstellt. Andererseits ist auch hier auf den Einfluss der Stichprobengröße hinzuweisen.

Bartletts Chi²-Test

In Analogie zu Lawley's (1940, 1941) Signifikanztest für die Maximum Likelihood Faktorenanalyse entwickelte Bartlett (1950, 1951) einen Signifikanz-Test für die Residuen einer Korrelationsmatrix nachdem Faktoren extrahiert wurden. Es wird die Nullhypothese getestet, dass die nach einer Extraktion verbleibenden Eigenwerte gleich sind. Die Eigenwerte werden sukzessive ausgeschlossen bis die Nullhypothese auf Gleichheit mit dem Chi²-Test nicht mehr

abgelehnt wird. Die Faktoren der bis zu diesem Punkt ausgeschlossenen Eigenwerte werden extrahiert.

Der Bartlett-Test soll sensitiv für die Stichprobengröße sein. Gorsuch (1973) vermutet, dass eine größere Teststärke bei größeren Stichproben zu einer Extraktion einer größeren Zahl von Komponenten führen könnte. Horn und Engstrom (1979) schlugen vor, die Abhängigkeit von der Stichprobengröße durch eine Anpassung des Alpha-Niveaus zu kompensieren. Zwick und Velicer (1980) stellten präzise Schätzungen der Faktorenzahl für den Fall fest, dass die Stichprobengröße deutlich über der Variablenanzahl liegt. Zwick und Velicer (1986) stellten demgegenüber trotz Anpassung des Alpha-Niveaus eine Abhängigkeit von der Stichprobengröße, aber auch eine sehr variable Güte der Schätzungen der Anzahl der zu extrahierenden Faktoren fest.

Bewertung: Dieser Test wurde bisher in der Traitforschung und auch sonst kaum eingesetzt, so dass keine Aussagen über die Qualität dieses Signifikanztests in der Anwendungspraxis möglich sind. Die offene Situation in der Traitforschung bezüglich der Anzahl zu extrahierender Faktoren ist jedenfalls nicht auf häufigen Einsatz des Bartlett-Tests zurückführbar.

Aufschlußreich ist die Argumentation, mit der Zwick und Velicer (1986) sich gegen den generellen Einsatz des Bartlett-Tests aussprechen: Die Bedingung der gleichen Eigenwerte, die der Bartlett-Test prüft, kann nur für Eigenwerte, die Fehlervarianz repräsentieren, gelten. Die meisten Forscher würden aber kleine oder triviale Komponenten nicht mit ihrer impliziten Definition einer zu extrahierenden Komponente vereinbaren können (Zwick & Velicer, 1986, S. 440). Der Bartlett-Test sei als Signifikanztest zwar korrekt, aber wegen seiner strikten Orientierung an der Fehlervarianz wenig geeignet. Hier wird deutlich, dass häufig - zumindest implizit - zwei verschiedene Ziele bei der Faktorenextraktion verfolgt werden. Zum einen soll die systematische Varianz von der Fehlervarianz getrennt werden, zum anderen soll die triviale von der nicht-trivialen Varianz getrennt werden. Beide Ziele erscheinen sinnvoll, aber es stellt sich die Frage, ob man die Information zur Signifikanz und zur Trivialität von vornherein in der Zahl zu extrahierender Faktoren einschmelzen sollte. Vielleicht wäre es hilfreich, Fehlervarianz und Trivialität getrennt zu bewerten. In dieser Hinsicht könnte auch Cattells (zit. nach Gorsuch, 1983, S. 171) Empfehlung hilfreich sein, im Zweifelsfall lieber einige Faktoren zuviel zu extrahieren und sie fallen zu lassen, wenn sie nicht repliziert werden können, als von vornherein zu wenige Faktoren zu extrahieren. Für den Fall der Trennung der Ermittlung der Fehlervarianz von der Ermittlung der Bedeutsamkeit oder Trivialität eines Faktors könnte der Bartlett-Test vielleicht von Interesse sein, wenngleich die Abhängigkeit von der Stichproben-

größe trotz Korrekturvorschlägen von Horn und Engstrom (1979) weiterhin als Problem besteht. Mit einer ähnlichen Intention schlägt Gorsuch (1983) vor, den Bartlett-Test zur Ermittlung der Obergrenze der Anzahl zu extrahierender Faktoren fest in Standard-Programmpaketen zu implementieren.

Extraktionskriterien anhand rotierter Faktoren

Crawford (1975) stellte ein Kriterium zur Entscheidung über die Anzahl zu extrahierender Faktoren auf, das sich an der Interpretierbarkeit der Faktoren orientiert, wobei die Interpretierbarkeit hier nicht die Konformität zu einem inhaltlichen Modell meint. Die Interpretierbarkeit setzt sich i.S. von Crawford (1975) aus zwei Aspekten zusammen: Menge erklärter Varianz und Einfachstruktur. Nur Faktoren, die ein genügendes Maß an erklärter Varianz aufweisen und die zugleich eine hinreichende Einfachstruktur haben, sollten nach Crawford interpretiert werden. Auf der Basis dieser Überlegungen formuliert Crawford (1975) einen Index der Interpretierbarkeit, in dem sowohl die erklärte Varianz als auch die Einfachstruktur berücksichtigt wird. Es werden sukzessive zwei Faktoren und alle größeren Anzahlen von Faktoren extrahiert. Für jede dieser Lösungen wird der Index der Interpretierbarkeit berechnet. Crawford empfiehlt dann die Festlegung der Faktorenanzahl auf den Wert, bei dem das erste Maximum im Index der Interpretierbarkeit auftritt. In den empirischen Überprüfungen, die Crawford (1975) durchgeführt hat, erwies sich die Equamax-Rotation bei Faktoren auf der Basis von quadrierten multiplen Korrelationen zur initialen Kommunalitätenschätzung als besonders geeignet. Bei der Hauptkomponentenanalyse mit anschließender Varimax-Rotation war der Index der Interpretierbarkeit nicht zu verwenden. In den empirischen Analysen stellte sich weiterhin heraus, dass die Ergebnisse mit dem Index der Interpretierbarkeit denen von Guttman (1954) „weak lower bound“ ähnlich waren.

Ein weiteres Extraktionskriterium auf der Basis rotierter Faktoren wurde von Revelle und Rocklin (1979) entwickelt. Das Verfahren wird als „very simple structure“ (VSS) Methode bezeichnet. Als Kriterium dient die Reproduzierbarkeit der Korrelationsmatrix aus einer Ladungsmatrix, bei der die jeweils größte Ladung aller Variablen beibehalten wird und alle anderen Ladungen auf null gesetzt werden. Revelle und Rocklin (1979) schlagen ihre Methode auch für die Fälle vor, in denen für jede Variable mehr als eine Hauptladung angenommen wird (wobei dann die zwei größten Ladungen beibehalten und die übrigen auf null gesetzt werden). Es werden wie bei Crawford (1975) sukzessive zwei Faktoren und alle größeren Anzahlen von Faktoren extrahiert. Dann wird die Anzahl von Faktoren beibehalten, für die die Reproduzierbarkeit der Korrelationsmatrix unter o.g. Bedingungen maximal ist. Revelle und Rocklin (1979) zeigten anhand künstlicher Datensätze, dass ihre Methode dem Guttman-Kaiser

Kriterium, dem Kriterium von Montanelli und Humphreys (1976) und zwei Varianten des Chi²-Tests überlegen ist.

Bewertung: Sowohl das Verfahren von Crawford (1975) als auch das von Revelle und Rocklin (1979) ermöglicht die Entscheidung zugunsten einer bestimmten Faktorenanzahl auf der Basis der Information zur Einfachstruktur. Diese Extraktionskriterien haben beide den Vorteil, dass im wesentlichen die Eigenschaften der Ladungsmatrix optimiert werden, die meist im Anschluß an die Analyse interpretiert werden. Eine wesentliche Einschränkung ergibt sich – zumindest für das Verfahren von Crawford – dann, wenn Einfachstruktur nicht das primäre Ziel bei der Rotation ist, oder wenn aus theoretischen Gründen keine optimale Einfachstruktur erwartet werden kann (z.B. Hofstee et al., 1992). Es muss aber eingeräumt werden, dass bei den meisten Rotationen im Traitbereich Einfachstruktur angestrebt wurde, oder zumindest die Interpretation der Ladungsmatrix nur die Aspekte der Einfachstruktur betraf. In den meisten dieser Fälle wäre das Verfahren von Crawford (1975) oder das von Revelle und Rocklin (1979) einsetzbar gewesen. Umso erstaunlicher ist, dass sich in der gesichteten Literatur kein Beispiel für die Verwendung des Verfahrens von Crawford findet und sich nur ein Beispiel für eine Verwendung des Verfahrens von Revelle und Rocklin (Parkes, 1985) fand. Der Wert der Verfahren kann daher nicht genau abgeschätzt werden. Im Vergleich zwischen dem Verfahren von Crawford und dem von Revelle und Rocklin deutet sich jedoch eine etwas größere Flexibilität des Verfahrens von Revelle und Rocklin (1979) an, da dieses Verfahren nicht nur auf Einfachstruktur beschränkt ist und auch bei obliquer Rotation eingesetzt werden kann (Parkes, 1985). Andererseits setzt auch die Methode von Revelle und Rocklin eine Kenntnis darüber voraus, ob Einfachstruktur erwartet wird oder nicht. Dies wird nicht immer im voraus zu beantworten sein. Es ist allerdings auch möglich, die Entscheidung für Einfachstruktur oder komplexere Strukturen (z.B. mit zwei Hauptladungen pro Variable), nach der Extraktion und Rotation der Faktoren durchzuführen. Insgesamt sind die beiden in diesem Kapitel dargestellten Verfahren wohl in unangemessener Weise in der bisherigen Forschung vernachlässigt worden. Andererseits verringert die Entscheidungsfreiheit darüber, welcher Grad an Einfachstruktur den Berechnungen zugrunde gelegt werden soll, die Verbindlichkeit der Befunde.

Extraktion mit variierender Faktorenzahl

Innerhalb eines Intervalls können auch alle möglichen Lösungen mit unterschiedlicher Faktorenzahl extrahiert werden (Wrigley, 1960; Howard & Gordon, 1963). In der ursprünglichen Form von Wrigley (1960) wurde die Methode so verwendet, dass sie der Unterscheidung zwischen trivialen und nicht-trivialen Faktoren dienen sollte. Die Lösungen wurden rotiert und dann wurde die Zahl “trivialer” Faktoren ermittelt. Als triviale Faktoren

galten diejenigen, die nicht wenigstens zwei oder drei Ladungen über einem bestimmten Mindestwert (z.B. .30) aufwiesen. Eine verbindlichere Definition von Hauptladungen kann aus Fürntratt (1969) gewonnen werden, der forderte, dass die quadrierten Ladungen dividiert durch die Kommunalität größer oder gleich .50 sein sollten ($a^2/h^2 \geq .50$). Dann wurde die Anzahl trivialer Faktoren von der Gesamtzahl extrahierter Faktoren abgezogen, um die Anzahl letztlich zu extrahierender Faktoren festzulegen.

Inzwischen wird diese Methode auch in einer offeneren Variante durchgeführt, indem auf die explizite Bestimmung trivialer Faktoren mit Hilfe von Schwellenwerten verzichtet wird (z.B. Zuckerman, Kuhlman, Thornquist und Kiers, 1991) und andere Kriterien zur abschließenden Bestimmung der Faktorenzahl herangezogen werden (s.u.). Die Kombination von Faktoren bei fallender Faktorenzahl bzw. die Aufspaltung von Faktoren bei steigender Faktorenzahl wird dabei interpretiert.

Bewertung: Ein Vorteil dieser Methode ist, dass ein Datensatz relativ umfassend beschrieben werden kann und dass einseitige Interpretationen unwahrscheinlicher werden. Dennoch ist diese Methode auch mit einer Reihe von Schwierigkeiten behaftet, die im folgenden aufgelistet werden:

- Zunächst ist ein Intervall festzulegen, innerhalb dessen die Faktoren extrahiert werden sollen. Mindestens zur Festlegung der Obergrenze der zu extrahierenden Faktoren, meist aber auch zur Festlegung der Untergrenze sind Kriterien erforderlich.
- Des Weiteren muss bedacht werden, dass die Mindestanzahl substantieller Ladungen pro Faktor nach Velicer und Fava (1998) nicht mehr auf drei sondern auf vier festgelegt werden muss. Dies wäre nicht weiter problematisch, wenn Velicer und Fava (1998) nicht auch festgestellt hätten, dass die Anzahl erforderlicher Ladungen pro Faktor mit der Höhe der Ladungen deutlich schwankt, so dass mitunter eine deutlich größere Anzahl substantieller Ladungen pro Faktor für die Identifikation erforderlich ist. Vor diesem Hintergrund erscheint es kaum möglich, ein klares Kriterium für die minimale Anzahl von Hauptladungen zu formulieren. Auch der Einfluss der Stichprobengröße wäre hier noch zu prüfen.
- Unklar ist ferner, wie die sukzessive Kombination von Faktoren mit abnehmender Faktorenzahl interpretiert werden kann. Zuckerman et al. (1991) interpretieren die Faktoren in den Lösungen mit kleinerer Faktorenzahl als generellere Faktoren (höherer Ordnung). Diese Interpretation erscheint dann gerechtfertigt, wenn mit einer kleineren Anzahl von Faktoren

nahezu die gleiche Varianz erklärt werden kann. Letzteres hängt natürlich vom Eigenwertverlauf ab. Wenn die erklärte Varianz einer Lösung mit einem Faktor weniger stark abnimmt, ist es jedoch wahrscheinlich, dass die Faktoren der Lösung, die einen Faktor weniger aufweist, keine höhere Generalität aufweisen. Cattell (1972) kritisierte das Vorgehen, bei dem, durch in der Faktorenzahl reduzierte Faktorenanalysen generellere Faktoren gewonnen werden sollen. In vergleichenden Analysen mit aufwärtsgerichteten, hierarchischen Faktorenanalysen zeigte sich, dass die in reduzierten Faktorenanalysen gewonnenen generelleren Faktoren, die in hierarchischen Analysen gewonnenen Faktoren nur sehr ungenau abbilden (s. Cattell, 1972).

- Ein weiteres Problem ist die mögliche Entscheidung für eine Lösung aus einer Kaskade von Lösungen. Denn mitunter wurde bei sukzessiven Analysen kein festes Kriterium für die Mindestanzahl und -höhe substanzieller Ladungen verwendet (z.B. Andresen, 2000). Wenn eine Entscheidung zugunsten einer Lösung nicht aufgrund formaler Kriterien getroffen wird, stellt sich das Problem, welche theoretische Bedeutung eine Folge von Lösungen haben kann. Wenn eine Entscheidung zugunsten einer Lösung erfolgt, stellt sich die Frage nach den Kriterien für die Entscheidung. Zum einen kann die Einfachstruktur der Lösungen Berücksichtigung finden (s. Crawford, 1975, Revelle & Rocklin, 1979, wobei die oben dargestellten Ansätze nicht „interpretativ“ bewertend, sondern mit objektivierbaren Indexwerten operieren). Zum anderen können – und das ist im Rahmen dieser Arbeit von besonderer Bedeutung – die Interpretierbarkeit der Lösungen oder andere theoretische Überlegungen eine Rolle spielen. Dabei ergibt sich folgendes Problem: Es ist sicherzustellen, dass die interpretierten Faktoren keine Zufallsfaktoren sind. Die meisten Extraktionskriterien dienen aber gerade der Identifikation der Faktoren mit systematischer Varianz und der Abgrenzung der Faktoren mit Zufallsvarianz. Die Entscheidung darüber, ob ein Faktor Zufallsvarianz oder systematische Varianz erklärt oder nicht, kann nicht auf der Basis der theoretischen Interpretierbarkeit erfolgen. Andererseits fragt sich, inwiefern Faktoren bei der Extraktion ignoriert werden dürfen, die nachweislich systematische Varianz erklären. Selbstverständlich müssen nicht alle Faktoren einer Lösung interpretiert werden. Zwischen Extraktion und Interpretation ist jedoch strikt zu unterscheiden: Wenn ein Faktor systematische Varianz erklärt, darf er bei der Extraktion nicht ignoriert werden, weil sonst jeder Faktor, der den theoretischen Vorstellungen eines Forschers nicht entspricht, eliminiert werden könnte. Ein derartiges Vorgehen würde den empirischen Gehalt der auf der Faktorenanalysen beruhenden Theorien vermutlich weiter reduzieren. Wenn eine Immunisierung der theoretischen Vorstellungen gegen die Empirie vermieden werden soll, müssen Faktoren auch dann, wenn

sie nicht interpretiert werden können, extrahiert und dargestellt werden, wenn sie systematische gemeinsame Varianz erfassen.

Die variierende Extraktion von Faktoren mit anschließender Auswahl einer Lösung ist insgesamt als problematisches Extraktionskriterium einzustufen, weil die Bevorzugung von Lösungen sehr leicht nach subjektiven Kriterien erfolgen kann und Aussagen über die systematische gemeinsame Varianz in einem Datensatz mit Aussagen zur Validität kontaminiert werden. Auch Cattell (1972) kommt zu einer negativen Bewertung der Festlegung der Faktorenanzahl in Abhängigkeit von den eigenen theoretischen Vorstellungen. In welcher Weise durch Verwendung dieses Kriteriums unfruchtbare Diskussionen entstehen, kann an folgendem Beispiel gezeigt werden: Nach der Bewertung ihrer sukzessiven Analysen entscheiden sich Zuckerman et al. (1991) für die Fünf-Faktorenlösung, während Eysenck (1991) aus derselben Analyse die Drei-Faktorenlösung als angemessenste Lösung hervorhebt. Zwischen beiden Positionen ist kaum eine wissenschaftliche Abwägung möglich: Für die Fünf-Faktorenlösung spricht der größere Grad an Differenziertheit und für die Drei-Faktorenlösung spricht die Übereinstimmung mit den vielfältigen Befunden zu Eysencks Persönlichkeitstheorie (z.B. Eysenck & Eysenck, 1985). Es ist nun eine Frage der jeweiligen Präferenzen, welche Lösung bevorzugt wird. Der pragmatische Ausweg, die Extraktion von drei oder fünf Faktoren von der jeweiligen Zielsetzung der Untersuchung abhängig zu machen, ist zwar für die diagnostische Praxis unter Umständen hinreichend, erscheint aber für die Beantwortung von Fragen im Rahmen der Grundlagenforschung zum Traitbereich unbefriedigend.

Alternative: Die hierarchische Faktorenanalyse gibt, eine integrative, konzeptuell überzeugendere Darstellung von Lösungen mit unterschiedlicher Faktorenzahl, ohne das eine Willkürlichkeit bei der Festlegung der Faktorenzahlen entsteht. Bei der (aufwärtsgerichteten) hierarchischen Faktorenanalyse wird die Konfundierung von Hierarchieaspekten mit dem Problem der Anzahl zu extrahierender Faktoren vermieden. Die hierarchische Faktorenlösung hat überdies den Vorteil, dass nicht die Validität (i.S. einer besonders prägnanten oder theoretisch plausiblen Lösung) als Kriterium für den Anteil systematischer gemeinsamer Varianz eines Faktors verwendet werden muss. Vielmehr kann auf jeder Hierarchieebene eine Unterscheidung der Zufalls- bzw. spezifischen Varianz von der gemeinsamen systematischen Varianz vorgenommen werden. Grundsätzlich ist in Fällen uneindeutiger Anzahl zu extrahierender Faktoren zu empfehlen, die größere Anzahl der Faktoren zu extrahieren und die sparsameren Lösungen anschließend in hierarchischen Analysen zu gewinnen. Als Extraktionskriterium kann die sukzessive Analyse mit variierender Faktorenzahl unter dieser

Perspektive jedoch nicht mehr verstanden werden, sondern eher als Vorprüfung im Hinblick auf eine mögliche Hierarchie von Faktoren.

Dass die hierarchische Faktorenanalyse als Alternative zur sukzessiven Extraktion unterschiedlicher Faktorenzahlen eingesetzt werden kann, bedeutet im übrigen nicht, dass diese beiden Vorgehensweisen – wie beispielsweise in Watson, Clark und Harkness (1994) – als äquivalent anzusehen sind. Dies ist bereits daran zu erkennen, dass bei hierarchischen Faktorenanalysen, die Faktoren höherer Ordnung stark von der Rotation der Faktoren erster Ordnung abhängen. Auch wenn die Ladungsmatrizen verschiedener schiefwinkliger Lösungen mitunter sehr ähnlich erscheinen, bestehen oft große Unterschiede zwischen den Interfaktorkorrelationen (Beauducel & Kersting, in Druck; Carroll, 1993). Demgegenüber werden bei sukzessiven Analysen mit unterschiedlicher Faktorenzahl alle Faktoren stets nur einmal rotiert, so dass kein kumulativer Einfluss der Rotationsmethoden auf die Faktoren höherer Ordnung besteht. Dieser eventuelle Vorteil wird jedoch dadurch aufgewogen, dass bei den sukzessiven Analysen die allgemeineren Faktoren notwendigerweise die Varianz speziellerer, nicht extrahierter Faktoren ausblenden, während diese Varianzen bei hierarchischen Analysen integriert werden.

Parallelanalyse

Horn (1965) schlug vor, die Eigenwerte der empirischen Korrelationsmatrix mit den Eigenwerten einer Matrix von Zufallskorrelationen zu vergleichen, wobei die Zahl der Fälle und die Variablenanzahl in der Zufallsmatrix der empirischen Matrix entsprechen muss. Wenn die Eigenwerte der empirischen Korrelationsmatrix größer sind als die der entsprechenden Zufallsmatrix können die Faktoren als bedeutsam angesehen werden. Dabei müssen die Eigenwerte einer großen Zahl von Zufallsdaten gemittelt werden, um Schwankungen in den Zufallsdaten zu minimieren.

Die Parallelanalyse mit gemittelten Eigenwerten von 50 Zufallsdatensätzen war in Zwick und Velicer (1986), wo eine Reihe von Extraktionskriterien (Scree-Test, Guttman-Kaiser Kriterium, Bartlett's Chi² und Velicer's Minimum Durchschnitts-Partialkorrelation) untersucht wurde, das beste Verfahren. Wenn überhaupt Fehler auftraten, war eine schwache Tendenz zur Überschätzung der Anzahl zu extrahierender Faktoren zu erkennen. Die Parallelanalyse schätzte in 97-100% der Fälle in einem Bereich von ± 1 um die tatsächliche Zahl zu extrahierender Faktoren. Auch in den Simulationen von Velicer et al. (2000) sowie in Hubbard und Allen (1987) erbrachte die Parallelanalyse mit die besten Ergebnisse.

Ein Nachteil der Parallelanalyse ist der große Aufwand, der darin besteht, dass eine große Zahl von Zufallskorrelationsmatrizen einer Faktorenanalyse unterzogen werden muss. Um hier den Aufwand zu reduzieren, gibt es inzwischen eine Reihe von Verfahren, die eine Schätzung der Eigenwerte ohne vorherige Berechnung der Eigenwerte über Zufallszahlen erlauben (z.B. Cota, Longman, Holden & Fekken, 1993; Lautenschlager, 1989).

Inzwischen wurden auch einige Probleme der Parallelanalyse thematisiert (Glorfeld, 1995; Turner, 1998). Glorfeld (1995) empfiehlt die Verwendung des 95. Perzentils anstatt des Mittelwertes der Zufallseigenwerte bei der Parallelanalyse. Sein Argument ist, dass ein empirischer Eigenwert, der gerade noch größer ist als der Mittelwert der entsprechenden Zufallseigenwerte immer noch kleiner ist als 50% der entsprechenden Zufallseigenwerte. Erst wenn der empirische Eigenwert größer sei als 95% der Zufallseigenwerte, könnte er im zufallskritischen Sinne als bedeutsam angesehen werden. Das Argument von Glorfeld (1995) ist überzeugend, wenn man die Parallelanalyse als quasi inferenzstatistisches Verfahren auffasst. Es darf jedoch bezweifelt werden, ob es angemessen ist, die Parallelanalyse als quasi inferenzstatistisches Verfahren aufzufassen. Dem entgegen steht die Beobachtung von Turner (1998) sowie Beauducel (2001b), dass die Parallelanalyse zu einer Unterschätzung der Anzahl zu extrahierender Faktoren führt, wenn die empirischen Daten einen großen ersten Eigenwert aufweisen. Beauducel (2001b) stellte fest, dass die Parallelanalyse bei Lösungen mit schiefwinkliger Einfachstruktur, die einen hohen ersten Eigenwert aufwiesen, die Anzahl zu extrahierender Faktoren unterschätzte. Wenn das von Glorfeld (1995) angeführte Argument zutreffend wäre, müsste die Parallelanalyse eher zu einer Überschätzung der Anzahl zu extrahierender Faktoren führen. Zur Reduzierung der Unterschätzung der Faktorenzahl bei großem ersten Eigenwert schlägt Turner (1998) vor, bei der Simulation von Zufallsdaten bestimmte Eigenschaften der empirischen Daten mitzusimulieren. Dieses Vorgehen ist allerdings sehr komplex und eröffnet subjektive Entscheidungsspielräume (welche Aspekte der empirischen Daten werden simuliert, welche nicht). Beauducel (2000) schlug vor, die Parallelanalyse auf der Basis der Eigenwertsdifferenzen durchzuführen, um die in Turner (1998) genannten Probleme zu umgehen. Allerdings bestehen für diese neue Methode bisher keine Erfahrungswerte, weshalb sie in diesem Kontext nicht weiter verfolgt wird.

Bewertung: Auch die Verwendung der ökonomischen Varianten der Parallelanalyse führen mit hoher Wahrscheinlichkeit in vielen Fällen zu besseren Ergebnissen als andere Extraktionskriterien. Allerdings führt die ursprüngliche, aufwendige Methode, bei der die Eigenwerte der Zufallsdaten nicht geschätzt sondern durch eine große Zahl von Simulationen ermittelt werden, vermutlich zu den besten Ergebnissen, weshalb nur in Ausnahmefällen von der ursprünglichen

Methode abgewichen werden sollte. Die Parallelanalyse ist leider in der bisherigen Traitforschung sehr selten eingesetzt worden (s. aber Matthews & Oddy, 1993). Dieses Extraktionskriterium sollte in Zukunft größere Beachtung finden. Es sollte jedoch bedacht werden, dass auch beim Einsatz dieser Methode die Anzahl der zu extrahierenden Faktoren nur mit einer Genauigkeit von ± 1 geschätzt werden kann. Da die Parallelanalyse bei großem ersten Eigenwert offenbar eher zu einer Unterschätzung der Faktorenzahl führt (Turner, 1998; Beauducel, 2001b), andererseits bisher keine Fälle von Überschätzung der Faktorenzahl anhand der Parallelanalyse berichtet wurden, sollte die Parallelanalyse vermutlich eher in ihrer konventionellen Form, d.h. über den Mittelwert der Zufallseigenwerte (Horn, 1965) und nicht über den von Glorfeld (1995) vorgeschlagenen 95. Perzentil verwendet werden. Dafür spricht auch, dass die günstigen Ergebnisse in den Simulationsstudien von Zwick und Velicer (1986) und Velicer et al. (2000) für eine anhand des Mittelwertes der Zufallseigenwerte durchgeführte Parallelanalyse gewonnen wurden.

Auswahl optimaler Extraktionskriterien

Im folgenden soll die Auswahl der optimalen Extraktionskriterien als Grundlage für die weiteren Analysen erfolgen. Die Auswahl erfolgt auf der Basis der bereits dargestellten Bewertungen.

Eigenwerte > 1 Regel oder Guttman-Kaiser Kriterium

Dieses Kriterium erwies sich in den vergleichenden Studien als zu ungenau. Die Menge der kritischen Literatur ist so groß (s.o.), dass dieses Kriterium auf keinen Fall für die hier geplanten Faktorenanalysen als geeignet angesehen werden kann. Die Probleme mit diesem Kriterium sind so massiv, dass auch eine über diese Arbeit hinausgehende Empfehlung ausgesprochen werden muss, dieses Kriterium nicht mehr als Standardkriterium zu verwenden (s. Hubbard & Allen, 1987).

Scree-Test

Der subjektive Scree-Test erwies sich zwar als deutlich genauer als das Guttman-Kaiser Kriterium, er ist jedoch aufgrund der seiner Subjektivität nicht geeignet die Verbindlichkeit der Datenanalysen zu erhöhen. Aus diesem Grund wird dieses Kriterium in dieser Arbeit nicht verwendet. Für die Varianten objektiver Scree-Tests liegen bisher noch nicht genügend Erfahrungen vor, weshalb diese hier ebenfalls nicht eingesetzt werden. Da die Eigenwerte der Lösungen dargestellt werden, besteht für Interessenten die Möglichkeit, die Anzahl von Faktoren, die sich bei Verwendung des subjektiven Scree-Tests ergeben würde, abzuschätzen.

Minimale durchschnittliche quadrierte Partialkorrelation

Dieses Kriterium erscheint zwar grundsätzlich interessant, die konventionelle Methode, bei der die quadrierten Partialkorrelationen verwendet wurde, war jedoch in Velicer et al. (2000) weniger leistungsfähig als die Parallelanalyse. Nur die Variante, bei der die vierte Potenz der Partialkorrelationen verwendet wurde, erwies sich als ebenso leistungsfähig wie die Parallelanalyse. Da jedoch die Verwendung der vierten Potenz der Partialkorrelationen vergleichsweise arbiträr erscheint, wurde dieses Kriterium für die weiteren Analysen nicht herangezogen.

Faktor-Komparabilitäten

Dieses Kriterium erscheint, insofern als Replizierbarkeit eine der wichtigsten Eigenschaften empirischer Befunde darstellen sollte, für die durchzuführenden Analysen geeignet. Auch fußen einige Studien in der psychometrischen Traitforschung auf diesem Kriterium (z.B. Borkenau & Ostendorf, 1990), so dass auch aus Vergleichsgründen die Verwendung dieses Kriteriums interessant erscheint. Allerdings wurde für dieses Kriterium eine Abhängigkeit von der Stichprobengröße ermittelt. Daher wird dieses Kriterium nicht als alleiniges Extraktionskriterium verwendet werden. Da in Lanning (1996) die Abhängigkeit der Anzahl extrahierter Faktoren von der Stichprobengröße bei einer minimalen Komparabilität von .90 größer war als bei einer minimalen Komparabilität von .85, wird hier ein Schwellenwert von .85 verwendet.

Lawleys Chi²-Test und Bartletts Chi²-Test

Beide Signifikanztests erwiesen sich – vor allem wegen der deutlichen Abhängigkeit der Ergebnisse von der Stichprobengröße – als zu ungenau. Da sie in der psychometrischen Traitforschung kaum Verwendung finden, werden diese Tests auch nicht zu Vergleichszwecken eingesetzt.

Extraktionskriterien anhand rotierter Faktoren

Die Extraktionskriterien anhand rotierter Faktoren erfordern, wenn auch keine Einfachstruktur, so doch zumindest klare Hypothesen über den Grad erwarteter Einfachstruktur. Mit der erforderlichen Festlegung, ob nur die größten Ladungen oder auch die Zweit- oder Drittgrößten Ladungen bei der Ermittlung der Faktorenzahl berücksichtigt werden sollen, entsteht ein Maß an Willkürlichkeit oder Subjektivität, das bei den vorliegenden Analysen gerade vermieden werden soll. Aus diesem Grund und weil sie auch zu Vergleichszwecken nicht eingesetzt werden müssen, werden diese Kriterien hier nicht verwendet.

Extraktion mit variierender Faktorenzahl

Dieses Kriterium führt – zusammen mit dem Scree-Test – zum größten Einfluss subjektiver Vorstellungen auf die Ergebnisse in den Analysen. Aus diesem Grunde wird dieses Kriterium in der vorliegenden Arbeit nicht verwendet. Die Information bezüglich hierarchischer Strukturen sollen hier durch aufwärtsgerichtete hierarchische Faktorenanalysen gewonnen werden.

Parallelanalyse

Da die Parallelanalyse bei den meisten Simulationsstudien die besten Ergebnisse erbrachte, wird dieses Verfahren hier eingesetzt. Um die von Turner (1998) und Beauducel (2001b) berichteten Probleme der Parallelanalyse zu umgehen wird die Parallelanalyse hier nicht als alleiniges Kriterium eingesetzt.

Zusammenfassung

Bei den modellspezifischen Analysen einzelner Fragebogen (siehe Kapitel 6.2) werden nur dann Faktorladungsmuster berichtet, wenn entweder die Parallelanalyse oder die Komparabilitäten eine Faktorenzahl indizieren, die mit der intendierten Struktur übereinstimmt. Da die modellspezifischen Analysen lediglich dem Ziel dienen, das Vorliegen der jeweils intendierten Struktur in den Fragebogen zu prüfen, ergibt sich nicht die Notwendigkeit einer weiterführenden Entscheidungsregel.

Bei den induktiven, modellübergreifenden Analysen (siehe Kapitel 6.3) ist die Variablenanzahl im Verhältnis zur Probandenanzahl so groß, dass die Komparabilitäten, die eine Teilung der Stichprobe erfordern, nicht zur Entscheidungsfindung herangezogen werden können. Dies hängt damit zusammen, dass in den getrennten modellübergreifenden Faktorenanalysen für die Teilstichproben die Anzahl der Probanden kleiner ist als die Anzahl der Variablen. Daher wird die Festlegung der Faktorenanzahl bei den modellübergreifenden Analysen nur anhand der Parallelanalyse getroffen. Aus Vergleichsgründen werden dennoch die Komparabilitäten berichtet. Allerdings können die Komparabilitäten nur für die Faktoren erster Ordnung berechnet werden, da die Faktorenanalysen höherer Ordnung anhand der Korrelationsmatrizen der Faktoren durchgeführt werden. Dies hat zur Folge, dass keine Faktorwerte für die Faktoren höherer Ordnung ermittelt werden können, die die Grundlage für die Berechnung der Komparabilitäten darstellen.

3.2.2.2 Extraktionsmethoden: Bestandsaufnahme und Auswahl

Häufig wurde die Auffassung vertreten, verschiedene Extraktionsmethoden wie beispielsweise Hauptkomponentenanalyse oder Hauptachsenanalyse würden nur zu minimal unterschiedlichen Ergebnissen führen (Velicer & Jackson, 1990). Auch ergaben sich in den wenigen empirischen Vergleichsstudien im Bereich der psychometrischen Traitforschung bisher nur minimale Unterschiede bezüglich verschiedener Extraktionsmethoden (z.B. Goldberg, 1990). Im Vergleich zur Diskussion der Extraktionskriterien ist die Zahl der Publikationen zur Diskussion der verschiedenen Extraktionsmethoden insgesamt eher gering. Dies mag dazu geführt haben, dass die bestehende Diskussion zu Unterscheiden zwischen den Extraktionsmethoden noch weniger wahrgenommen wurde als die der Extraktionskriterien, so dass auch in diesem Bereich die Freiheitsgrade für den Anwender der exploratorischen Faktorenanalyse sehr groß sind. Da das Ziel dieser Arbeit u.a. die Reduktion dieser Freiheitsgrade ist, soll im folgenden auf die Arbeiten, die sich mit dem Vergleich der Extraktionsmethoden befassen, eingegangen werden.

Einen umfassenden Überblick über verschiedene Extraktionsmethoden leistet Wittmann (1978). Dort werden 12 (!) verschiedene Extraktionsmethoden und deren Unterschiede beschrieben. In der Regel werden nur wenige dieser Modelle in der Literatur diskutiert und in einschlägigen Programmpacketen angeboten. Die in der Regel verfügbaren Methoden erstrecken sich von der Hauptkomponentenanalyse über die Hauptachsenanalyse mit Kommunalitäteniteration, die Alpha-Faktorenanalyse (Kaiser & Caffrey, 1965), Image-Analyse (Guttman, 1953) bis zur Maximum-Likelihood-Faktorenanalyse (Lawley & Maxwell, 1963). Die Tatsache, dass nur eine geringe Zahl der bekannten Extraktionsmethoden in der Forschung angewandt wird, deutet auf ein möglicherweise erhebliches ungenutztes Potential der exploratorischen Faktorenanalyse hin. Es wird auch im Rahmen dieser Arbeit nicht möglich sein, eine vollständige Darstellung des Potentials der unterschiedlichen Extraktionsmethoden zu geben. Die vorliegende Betrachtung soll jedoch über die in konventionellen Analysen festzustellende, meist unbegründete Beschränkung auf Hauptkomponenten oder klassischer Faktorenanalyse bzw. Hauptachsenanalyse mit iterativer Kommunalitätenschätzung hinausgehen und zusätzlich Alpha-, Image- und Maximum-Likelihood-Faktorenanalyse einbeziehen, wobei der Schwerpunkt auf der Auswahl von für die geplanten Analysen optimalen Extraktionsmethoden liegt.

Studien zu Unterschieden zwischen Extraktionsmethoden

Die meisten Vergleiche zwischen Extraktionsmethoden wurden anhand von Simulationsstudien durchgeführt. Dabei wurde meist die Hauptkomponentenanalyse (d.h. Analyse einer vollstän-

digen Korrelationsmatrix mit Einsen in der Hauptdiagonale; Hotelling, 1933) mit der klassischen Faktorenanalyse (d.h. Hauptachsenanalyse mit iterativer Kommunalitätenschätzung) verglichen. Velicer (1977) verglich neben der Hauptkomponentenanalyse und der klassischen Faktorenanalyse auch die Image-Analyse und die Maximum-Likelihood-Faktorenanalyse anhand simulierter Daten. Er stellte die größten Ähnlichkeiten zwischen den Ergebnissen der Image-Analyse und der Maximum-Likelihood-Faktorenanalyse und die größten Unterschiede zwischen den Ergebnissen der Hauptkomponentenanalyse und der Maximum-Likelihood-Faktorenanalyse fest. Insgesamt waren die Unterschiede zwischen den Extraktionsmethoden in Velicer (1977) gering und traten im allgemeinen auf dem letzten extrahierten Faktor auf, was als Hinweis auf eine zu große Zahl extrahierter Faktoren gedeutet wurde. Im Einklang damit stellten auch Velicer und Jackson (1990) nur für den Fall einer zu großen Anzahl extrahierter Faktoren nennenswerte Unterschiede zwischen Hauptkomponentenanalyse und klassischer Faktorenanalyse fest. Dies verdeutlicht, dass die oben dargestellte Problematik optimaler Extraktionskriterien mit der Problematik optimaler Extraktionsmethoden interagieren könnte. Dieses Problem kann hier nur genannt aber nicht gelöst werden, da bisher noch keinerlei systematische Untersuchungen zu diesem Interaktionsproblem vorliegen. Hier kann nur der Versuch unternommen werden, die verfügbaren Studien zu den Extraktionskriterien und Extraktionsmethoden sukzessive für die Entwicklung von Vorschlägen in beiden Bereichen heranzuziehen.

Snook und Gorsuch (1989) ermittelten im Gegensatz zu Velicer und Jackson (1990), dass Hauptkomponentenanalyse und Hauptachsenanalyse mit Kommunalitäteniteration auch dann unterschiedliche Ergebnisse erbringen können, wenn die korrekte Anzahl an Faktoren extrahiert wird. Die Hauptkomponentenanalyse führte in ihren Simulationen zu einer Überschätzung der Ladungshöhe.

Kallina und Hartmann (1976) stellten in einer Simulationsstudie aus ihrer Sicht vernachlässigbare Unterschiede zwischen den Ergebnissen der Hauptkomponentenanalyse und der klassischen Faktorenanalyse fest. Sie plädieren aus diesem Grund für die Verwendung der weniger aufwändigen Hauptkomponentenanalyse. Allerdings berichten sie für einige Lösungen mit geringer Einfachstruktur nennenswerte Unterschiede zwischen den Methoden, die allerdings jeweils auf einen extrem varianzschwachen Faktor zurückgeführt wurden, der nach Kallina und Hartmann (1976) ohnehin nicht extrahiert werden sollte. Kallina und Hartmann (1976) berichten allerdings nicht die Unterschiede zwischen Hauptkomponentenanalyse und klassischer Faktorenanalyse, die sich in diesen Lösungen bei der Rotation mit einem Faktor

weniger ergeben würden. Insgesamt erbrachten die Analysen von Kallina und Hartmann (1976) jedoch nur geringe Unterschiede.

Widaman (1993) konnte die Ergebnisse von Snook und Gorsuch (1989) replizieren. Er verglich darüber hinaus die Robustheit von Ladungen auf der Basis von Hauptkomponentenanalyse und klassischer Faktorenanalyse und stellte fest, dass die auf der Basis der klassischen Faktorenanalyse geschätzten Ladungen robuster gegenüber Veränderungen des Variablen-settings waren. Die Frage der Robustheit bzw. Generalisierbarkeit von Ladungsmatrizen über verschiedene Variablengruppen wurde bereits von Thurstone (1947) als wesentliches Kriterium für die Bedeutung faktorenanalytischer Ergebnisse genannt. Das Ergebnis von Widaman (1993), dass die Generalisierbarkeit von Ladungsmatrizen nicht pauschal von den untersuchten Variablen abhängt, sondern auch von der gewählten Extraktionsmethode, ist jedoch von erheblicher Bedeutung.

Beauducel (2001a) untersuchte die Generalisierbarkeit von Ergebnissen auf der Basis unterschiedlicher Extraktionsmethoden anhand einer Simulationsstudie, wobei nicht nur Hauptkomponentenanalyse und Hauptachsenanalyse, sondern auch Alpha- und Maximum-Likelihood-Faktorenanalyse einbezogen wurden. Es wurde ein anderes Vorgehen gewählt als in Widaman (1993). In Widaman (1993) wurde für zwei nachweisbare Faktoren, der Variablenpool variiert. In Beauducel (2001a) wurde ein Faktor zunächst in einem Variablenkontext mit ausgeprägter Einfachstruktur nachgewiesen. Dann wurden die Markiervariablen des Faktors in einem anderen Variablen-Kontext untersucht. Die Einfachstruktur des neuen Variablen-Kontextes war gegenüber dem vorherigen deutlich reduziert. Die Reduktion der Einfachstruktur wurde in der Weise vorgenommen, dass mehrere Faktoren mit den Markiervariablen des ursprünglichen Faktors überlappen, so dass der ursprüngliche Faktor in dem neuen Kontext „aufgelöst“ wurde. Es wurde nun untersucht, mit welcher Extraktionsmethode der „aufgelöste“ Faktor am ehesten identifiziert werden konnte. Es zeigte sich, dass die Maximum-Likelihood-Faktorenanalyse eine hohe Sensitivität für den „aufgelösten“ Faktor hatte, allerdings waren die Ergebnisse der Maximum-Likelihood Faktorenanalyse bei geringer Kommunalität und Stichprobengröße instabil. Die Hauptachsenanalyse und die Alpha-Faktorenanalyse waren zwar etwas weniger sensitiv gegenüber dem „aufgelösten“ Faktor, zeichneten sich aber durch eine größere Robustheit der Lösungen bei geringer Kommunalität und Stichprobengröße aus. Mit der Hauptkomponentenanalyse konnte der „aufgelöste“ Faktor in keinem Fall identifiziert werden.

Wittmann (1978) stellte keine Unterschiede zwischen der Hauptkomponentenanalyse und der Image-Analyse bei der Rotation auf die richtige Faktorenanzahl fest. Bei einer Rotation auf eine falsche Faktorenanzahl ergaben sich jedoch Unterschiede zwischen den Modellen. Wittmann (1978) stellte darüber hinaus fest, dass die Hauptkomponentenanalyse deutlich empfindlicher auf eine falsche Rotationshypothese reagierte als die kommunalitätskalierte Image-Analyse. Buley (1995) stellte in Simulationsstudien insgesamt besonders hohe Identifikationsraten für die Image-Analyse im Vergleich zur Hauptkomponentenanalyse, Alpha-, Maximum-Likelihood-, klassischen Faktorenanalyse, sowie im Vergleich zu den Faktorenanalysen nach dem Prinzip der kleinsten Quadrate fest. Für Besonderheiten der Image-Analyse spricht auch, dass die auf Basis der Image-Analyse ermittelten Faktorwerte in Fava und Velicer (1992) etwas geringer mit den Faktorwerten anderer Extraktionsmethoden (Hauptkomponentenanalyse, Hauptachsenanalyse) korrelierten.

Auch Schweizer, Boller und Braun (1996) verglichen die Leistungsfähigkeit verschiedener Extraktionsmethoden bei der Identifikation von Faktoren oder Komponenten anhand simulierter Daten. Sie verglichen die Hauptkomponentenanalyse, Hauptachsenanalyse mit und ohne Kommunalitäteniteration, Alpha-Faktorenanalyse, Unweighted-Least-Square-Faktorenanalyse, Harris-Komponentenanalyse und die Maximum-Likelihood-Methode. Die Hauptkomponentenanalyse bewährte sich in den Simulationen von Schweizer et al. (1996) am besten, wobei sie zu keinen wesentlich höheren Identifikationsraten als die übrigen Verfahren führte. Es muss allerdings angemerkt werden, dass die Anzahl der Fälle in den Simulationen von Schweizer et al. (1996) für Faktorenanalysen sehr klein waren (100, 150, 200), was vor allem für die Maximum-Likelihood-Faktorenanalyse zu Problemen führte. Die kleinen Fallzahlen schränken die Aussagekraft der Ergebnisse vor allem deswegen ein, weil in den Simulationen ein deutlicher Effekt für die Fallzahl auftritt. Es ist durchaus möglich, dass die kleinen Fallzahlen zu einer Instabilität bei der Kommunalitätenschätzung führten, so dass die Hauptkomponentenanalyse, bei der dieser Schritt entfällt, für die kleinen Fallzahlen vergleichsweise besonders robust erschien. Das Problem der Generalisierbarkeit faktorenanalytischer Ergebnisse wurde in Buley (1995) und Schweizer et al. (1996) nicht untersucht.

Gründe für die Unterschiede zwischen Extraktionsmethoden

Gorsuch (1983) diskutiert die Unterschiede zwischen Hauptkomponentenanalyse und klassischer Faktorenanalyse im Hinblick auf die analysierten Korrelationsmatrizen: Bei einer geringen Anzahl analysierter Variablen und geringen Interkorrelationen zwischen den Variablen sind die Unterschiede zwischen den Verfahren groß, bei großer Variablenanzahl und hohen Interkorrelationen sind die Unterschiede zwischen den Verfahren gering. Gorsuch (1983) bringt

diese Beobachtung mit dem relativen Anteil, den die Hauptdiagonale an einer Korrelationsmatrix hat, in Verbindung. Bei kleinen Korrelationsmatrizen, ist der relative Anteil der Hauptdiagonalen groß, bei großen Korrelationsmatrizen spielt die Hauptdiagonale insgesamt eine geringere Rolle. Bei hohen Interkorrelationen sind die Kommunalitäten insgesamt höher und der Unterschied zwischen der Schätzung von eins in der Hauptkomponentenanalyse und anderen Kommunalitätenschätzungen fällt entsprechend geringer aus. Damit sind wesentliche Gründe für das Erhalten ähnlicher und unähnlicher Lösungen mit den beiden Verfahren genannt.

Beauducel (2001a) greift das Argument von Gorsuch (1983) auf, um die geringere Generalisierbarkeit der anhand der Hauptkomponentenanalyse gewonnenen Ladungsmuster zu erklären: Da die Hauptkomponentenanalyse in hohem Maße auch die Einsen in der Hauptdiagonale erklärt, ist ihre Sensitivität für den nicht diagonalen Teil der Korrelationsmatrix vermutlich geringer. Die Extraktionsmethoden wie die klassische, Alpha- und Maximum-Likelihood-Faktorenanalyse, die kleinere Kommunalitätenschätzungen verwenden, erhöhen vermutlich die relative Bedeutung der nicht-diagonalen Elemente der Korrelationsmatrix für die Lösung und somit die Sensitivität für die relevante Information.

Auswahl optimaler Extraktionsmethoden

Da eine mangelnde Robustheit gegenüber wechselnden Variablen-Kontexten ein entscheidender Kritikpunkt an der faktorenanalytischen Forschung ist (z.B. Block, 1995), erhält dieser Aspekt im Rahmen dieser Arbeit ein besonders großes Gewicht. Demgegenüber erhalten die grundsätzlich interessanten Ergebnisse von Schweizer et al. (1996) zur Identifikation von Faktoren ein geringeres Gewicht. Die Ergebnisse von Beauducel (2001a) bezüglich der Generalisierbarkeit von Ladungsmatrizen bestätigten die Ergebnisse von Widaman (1993): Während die Identifikation des „aufgelösten“ Faktors mit der Hauptkomponentenanalyse nicht möglich war, konnte der Faktor mit der klassischen Faktorenanalyse bei einer leichten Überextraktion identifiziert werden. Eine hohe Sensitivität bei der Identifikation des „aufgelösten“ Faktors war auch für die Maximum-Likelihood-Faktorenanalyse feststellbar, aber die Ergebnisse dieser Methode waren bei niedrigen Kommunalitäten und geringer Stichprobengröße sehr instabil (Beauducel, 2001a). Im Bereich der psychometrischen Traitforschung treten häufig geringe Kommunalitäten auf, weshalb diese Methode für diesen Forschungsgegenstand vermutlich weniger geeignet ist. Eine akzeptable Sensitivität für den „aufgelösten“ Faktor war sowohl für die klassische als auch für die Alpha-Faktorenanalyse feststellbar. Da sich die Ergebnisse dieser beiden Methoden in Beauducel (2001a) kaum unterschieden und im übrigen kaum Befunde zur Alpha-Faktorenanalyse verfügbar sind, liegt

es nahe, bei vermutlich gleicher Güte das gebräuchlichere der beiden Verfahren und somit die Hauptachsenanalyse zu präferieren.

Es muss allerdings einschränkend festgestellt werden, dass sich die Image-Analyse in den Simulationen von Wittmann (1978) und Buley (1995) bewährt hat, während sie in den Studien von Widaman (1993) und Beauducel (2001a) zur Generalisierbarkeit faktorenanalytischer Ergebnisse nicht berücksichtigt wurde. Somit stellt sich die Frage, ob die Image-Analyse neben der klassischen Faktorenanalyse als leistungsfähige Methode eingesetzt werden sollte. Diese Frage kann jedoch nur anhand einer Simulation der Generalisierbarkeit der Ergebnisse der Image-Analyse beantwortet werden. Daher wurde im folgenden eine entsprechende Simulation für die Image-Analyse durchgeführt.

Simulationsstudie zur Generalisierbarkeit der Ergebnisse der Image-Analyse

Die hier vorgenommene Simulationsstudie orientiert sich am Vorgehen von Beauducel (2001a), d.h. es wurde ein Faktor mit vier Markiertvariablen zunächst in einem für die Identifikation des Faktors vorteilhaften Kontext nachgewiesen. Dieser Kontext bestand aus drei weiteren Faktoren mit jeweils vier Markiertvariablen. Die Methode der Datengenerierung folgt dem Vorgehen von Beauducel (2001a), damit die dort berichteten Ergebnisse zur Hauptkomponentenanalyse, klassischen, Alpha- und Maximum-Likelihood-Faktorenanalyse mit den hier festgestellten Ergebnissen für die Image-Analyse verglichen werden können. Die Datengenerierung für diesen vorteilhaften Kontext wird im folgenden beschrieben.

Methode der Datengenerierung für den vorteilhaften Variablenkontext

Mit Hilfe von SPSS für Windows 9.0 (1999) wurden vierfaktorielle Lösungen mit eindeutiger Einfachstruktur generiert. Die Markiertvariablen wurden durch gezielte Aggregation von Zufallsvariablen erzeugt. Dieses Verfahren wurde in Schweizer et al. (1996) sowie Beauducel (2001a) verwendet (für die Grundlagen siehe Knuth, 1981). Systematische Korrelationen zwischen Variablen können bei diesem Verfahren dadurch generiert werden, dass eine oder mehrere Zufallsvariablen in mehreren aggregierten Variablen gemeinsam auftreten. Die Formel für die Aggregation lautet für die ersten vier Variablen V_1 bis V_4 wie in Beauducel (2001a):

$$V_i = \frac{1}{\sqrt{6}} c_1 + \frac{\sqrt{2}}{\sqrt{3}} s_i + \frac{1}{\sqrt{6}} c_5, \quad \text{for } i = 1 \text{ to } 4 \quad (1)$$

Die Variable c_1 ist den Variablen V_1 bis V_4 gemeinsam, die Variablen s_1 bis s_4 stellen die jeweils spezifischen Varianzanteile der Variablen V_1 bis V_4 dar (diese sind hier und im weiteren stets unkorreliert). Variable c_5 ist allen Variablen (V_1 bis V_{16}) gemeinsam. Dieselben Gewichte

wurden auch für die Variablen V_5 bis V_8 verwendet, nur dass die gemeinsame Variable c_1 durch die gemeinsame Variable c_2 ausgetauscht wurde. Die Variablen V_5 bis V_8 markieren daher einen anderen Faktor. Entsprechend wurde die gemeinsame Variable c_3 für die Variablen V_9 bis V_{12} und die gemeinsame Variable c_4 für die Variablen V_{13} bis V_{16} verwendet.

Die Gewichte der Variablen wurden so gewählt, dass die Variablen, die einen gemeinsamen Faktor bilden, aufgrund von jeweils zwei gemeinsamen Variablen etwa zu .33 interkorrelieren, während die Variablen, die unterschiedliche Faktoren markieren, aufgrund der gemeinsamen Variable c_5 etwa zu .17 interkorrelieren. Es wurden bewusst relativ geringe Interkorrelationen zwischen den Variablen generiert, da im Bereich der psychometrischen Traitforschung keine hohen Korrelationen zwischen einzelnen Items zu erwarten sind.

Auf der Basis der Gewichte in Formel (1) wurden 500 vierfaktorielle Datensätze auf der Basis von 200 Fällen und 500 vierfaktorielle Datensätze auf der Basis von 1000 Fällen generiert.

Ergebnisse für den vorteilhaften Variablenkontext

Für die Image-Analyse gibt es kaum Literatur zur Bestimmung der Anzahl der zu extrahierenden bzw. zu rotierenden Faktoren (s. Hakstian & Muller, 1970). Aus diesem Grund wird hier die Parallelanalyse verwendet. Da Velicer et al. (2000) bereits auf der Basis von fünf Datensätzen zu stabilen Schätzungen der Zufallseigenwerte kamen, wurden die mittleren Zufallseigenwerte hier nicht auf der Basis von 500 Datensätzen, sondern auf der Basis von 100 Datensätzen geschätzt. Die Größe der Eigenwerte in Tabelle 3.2 darf nicht wie bei Eigenwerten der Korrelationen zwischen Variablen interpretiert werden, denn die Image-Kovarianzmatrix stellt eine stark reduzierte Kovarianzmatrix dar.

Die Parallelanalyse spricht im Trend für die Extraktion von vier Faktoren, da der mittlere Eigenwert des vierten Faktors größer ist (.21, siehe 2. Spalte in Tabelle 3.2) als der mittlere Eigenwert der entsprechenden Zufallsfaktoren (.08, siehe 5. Spalte in Tabelle 3.2). Der minimale Eigenwert des vierten Faktors der vierfaktoriellen Lösung ist mit .06 jedoch kleiner als der mittlere Zufallseigenwert von .08 (siehe Tabelle 3.2).

Insgesamt war in zwei der 500 Lösungen der Eigenwert des vierten Faktors kleiner als der Mittelwert der entsprechenden Zufallsfaktoren. D.h. in zwei von 500 Fällen indizierte die Parallelanalyse, dass nur drei Faktoren extrahiert werden sollten. Auch war das Maximum des fünften Eigenwertes der vierfaktoriellen Lösungen mit .11 größer als der Mittelwert der entsprechenden Zufallseigenwerte von .06, was für die Extraktion von fünf Faktoren spricht.

Dieser Fall trat immerhin in 60 der 500 Lösungen, also in 12% der Lösungen auf. Des weiteren fällt auf, dass die Eigenwerte der vierfaktoriellen Lösungen und die der Zufallslösungen etwa ab dem zehnten Faktor wieder ansteigen.

Tabelle 3.2

Mittelwerte, Minima und Maxima der Image-Eigenwerte für 500 vierfaktorielle Lösungen und 100 Zufallslösungen auf der Basis von 200 Fällen

Faktor	Vierfaktorielle Lösungen			Zufalls- Lösungen
	Mittelwert	Minimum	Maximum	Mittelwert
1	2.76	1.95	3.79	.25
2	.47	.26	.72	.18
3	.33	.15	.53	.12
4	.21	.06	.38	.08
5	.04	.01	.11	.06
6	.02	.00	.05	.03
7	.01	.00	.03	.02
8	.00	.00	.02	.01
9	.00	.00	.01	.00
10	.01	.00	.03	.00
11	.02	.01	.04	.01
12	.04	.02	.07	.02
13	.07	.03	.11	.04
14	.10	.05	.15	.07
15	.15	.09	.22	.12

Dieser Anstieg ist bei Eigenwerten auf der Basis von Korrelationsmatrizen der Variablen nicht zu verzeichnen und ist somit ein Spezifikum der Image-Analyse. Bei der hier durchgeführten Parallelanalyse kann dieser Anstieg ignoriert werden und führt somit nicht zu einer Überschätzung der Anzahl der zu extrahierenden Faktoren, allerdings kann dieser Anstieg im Fall echter Daten dann zu Problemen führen, wenn der Eigenwertverlauf insgesamt wenig markant ist. Als nächstes wurde die Parallelanalyse für die Daten auf Basis von 1000 Fällen durchgeführt (siehe Tabelle 3.3).

Hier spricht die Parallelanalyse für die Extraktion von vier Faktoren, da auch der kleinste Eigenwert der 500 Eigenwerte des vierten Faktors größer ist als der entsprechende Zufallseigenwert. Allerdings war in vier der 500 Lösungen der fünfte Eigenwert größer als der entsprechende Zufallseigenwert. Die Parallelanalyse spricht also in den Daten auf der Basis von 1000 Fällen deutlicher für eine vierfaktorielle Lösung als in den Daten auf der Basis von 200 Fällen. Aber auch hier zeigt sich der Anstieg der Eigenwerte ungefähr ab dem zehnten Faktor. Die Uneindeutigkeit der Ergebnisse der Parallelanalyse hängt sicher mit der geringen

numerischen Höhe der Korrelationen ($r \approx .33$) zusammen, die für Variablen, die einen gemeinsamen Faktor konstituieren, gewählt wurde.

Tabelle 3.3

Mittelwerte, Minima und Maxima der Image-Eigenwerte für 500 vierfaktorielle Lösungen und 100 Zufallslösungen auf der Basis von 1000 Fällen

Faktor	Vierfaktorielle Lösungen			Zufalls-Lösungen
	Mittelwert	Minimum	Maximum	Mittelwert
1	2.62	2.20	3.01	.05
2	.31	.22	.40	.03
3	.24	.18	.31	.02
4	.19	.13	.27	.01
5	.00	.00	.01	.00
6	.00	.00	.00	.00
7	.00	.00	.01	.00
8	.00	.00	.01	.00
9	.01	.00	.02	.00
10	.02	.01	.02	.00
11	.02	.01	.03	.00
12	.03	.02	.04	.00
13	.04	.03	.05	.00
14	.05	.04	.07	.00
15	.06	.04	.09	.01

Man könnte einwenden, dass höhere Korrelationen zu eindeutigeren Lösungen führen würden. Da sich jedoch die hier verwendeten Extraktionsmethoden im Bereich der geringen Korrelationen, die in der Regel zwischen Traititems bestehen, bewähren müssen, wird ein weniger markantes, aber realistischeres Simulationsergebnis hier bevorzugt. Außerdem wurden dieselben geringen Korrelationen auch in Beauducel (2001a) verwendet, wo sie für die Hauptachsenanalyse, Alpha- und Maximum-Likelihood-Faktorenanalyse vergleichsweise eindeutigere Ergebnisse erbrachten.

Trotz der nicht ganz eindeutigen Ergebnisse auf Basis der Parallelanalyse wurden daher im folgenden die mittleren Varimax-Ladungen für die 500 vierfaktoriellen Lösungen auf der Basis von 200 Fällen berichtet (s. Tabelle 3.4). Die Einfachstruktur der mittleren Ladungen ist, obwohl die Hauptladungen nicht sehr hoch sind, prägnant. Die mittleren Varimax-Ladungen der 500 Lösungen auf der Basis von 1000 Fällen unterscheiden sich nur minimal von den in Tabelle 3.4 dargestellten Ladungen und wurden daher hier nicht dargestellt.

Die einzigen Unterschiede sind: Der Mittelwerte der Hauptladungen beträgt bei den Lösungen auf der Basis von 1000 Fällen .39 und die größte Standardabweichung beträgt .03. Die Lösungen variieren somit auf der Basis von 1000 Fällen deutlich weniger. Trotz der Uneindeutigkeiten bei der Parallelanalyse spricht das Ladungsmuster eindeutig für die Extraktion von vier Faktoren.

Tabelle 3.4

Mittelwerte und Standardabweichungen der Varimax-Ladungen der 500 vierfaktoriellen Lösungen auf der Basis von 200 Fällen

Variable	Faktor 1	Faktor 2	Faktor 3	Faktor 4
V1	.41 (.07)	.13 (.06)	.14 (.06)	.14 (.06)
V2	.41 (.07)	.14 (.06)	.14 (.06)	.14 (.06)
V3	.41 (.07)	.14 (.06)	.14 (.06)	.14 (.06)
V4	.41 (.07)	.13 (.06)	.14 (.06)	.14 (.06)
V5	.14 (.06)	.41 (.07)	.14 (.06)	.14 (.06)
V6	.13 (.06)	.41 (.07)	.13 (.06)	.14 (.06)
V7	.13 (.06)	.41 (.07)	.14 (.06)	.14 (.06)
V8	.14 (.06)	.41 (.07)	.14 (.06)	.14 (.06)
V9	.14 (.06)	.14 (.06)	.41 (.06)	.14 (.06)
V10	.14 (.06)	.14 (.06)	.41 (.07)	.14 (.06)
V11	.14 (.06)	.14 (.06)	.41 (.06)	.14 (.06)
V12	.14 (.06)	.14 (.06)	.41 (.06)	.13 (.06)
V13	.14 (.06)	.14 (.06)	.14 (.06)	.41 (.07)
V14	.14 (.06)	.14 (.06)	.14 (.06)	.40 (.07)
V15	.14 (.06)	.14 (.06)	.14 (.07)	.41 (.07)
V16	.14 (.06)	.14 (.06)	.14 (.06)	.41 (.07)

Anmerkungen. Ladungen > .30 wurden fett gedruckt. Da die Varimax-Ladungen Korrelationen sind, wurden sie vor der Mittelung Fisher-Z transformiert. Dann wurden die Mittelwerte der Fisher-Z transformierten Ladungen zurücktransformiert.

Allerdings betrug der Mittelwert der Hauptladungen auf Basis der klassischen Faktorenanalyse in Beauducel (2001a) immerhin .53 während die mittleren Nebenladungen .13 betragen. Die Prägnanz der Lösungen war auch bei den übrigen Extraktionsmethoden größer. Dennoch konnten auch für die Image-Analyse die vierfaktoriellen Lösungen aufgezeigt werden. Anders ausgedrückt: Ein Faktor konnte im Kontext von drei weiteren Faktoren aufgezeigt werden. Als nächstes wird untersucht, ob es möglich ist, einen der in Tabelle 3.4 aufgezeigten Faktoren im Kontext von drei anderen stark überlappenden Faktoren zu identifizieren.

Methode der Datengenerierung für den überlappenden Variablenkontext

In dieser zweiten Simulationsstudie wurden die vier Markiertvariablen V_1 bis V_4 genau auf dieselbe Weise wie in der vorherigen Simulationsstudie generiert (siehe Gleichung 1). Die erste

Gruppe von neuen Kontext-Variablen NV_5 bis NV_8 wurde anhand folgender Gleichung gebildet:

$$NV_i = \frac{1}{3}c_2 + \frac{2}{3}s_i + \frac{2}{3}c_1, \quad \text{for } i = 5 \text{ to } 8 \quad (2)$$

Dabei ist anzumerken, dass die gemeinsame Variable c_1 auch eine gemeinsame Variable von V_1 bis V_4 ist. Der Erwartungswert der Interkorrelationen von NV_5 bis NV_8 beträgt .56. Der Erwartungswert der Korrelationen von NV_5 bis NV_8 mit V_1 bis V_4 beträgt .27. Die nächste Gruppe neuer Kontextvariablen NV_9 bis NV_{12} wurde auf analoge Weise generiert, nur dass anstatt der gemeinsamen Variablen c_1 und c_2 die gemeinsamen Variablen c_3 und c_5 verwendet wurden:

$$NV_i = \frac{1}{3}c_3 + \frac{2}{3}s_i + \frac{2}{3}c_5, \quad \text{for } i = 9 \text{ to } 12 \quad (3)$$

Der Erwartungswert der Interkorrelationen zwischen NV_9 bis NV_{12} beträgt wieder .57 und die Korrelation mit V_1 bis V_4 beträgt wieder .27, diesmal aufgrund der gemeinsamen Variablen c_5 . Da NV_9 bis NV_{12} jedoch keine gemeinsamen Variablen mit NV_5 bis NV_8 enthalten, beträgt der Erwartungswert der Korrelationen von NV_9 bis NV_{12} mit NV_5 bis NV_8 null. Beide Gruppen von Variablen korrelieren also mit V_1 bis V_4 ohne miteinander zu korrelieren. Somit besteht die Möglichkeit, dass der Faktor, der in der ersten Simulationsstudie durch V_1 bis V_4 repräsentiert wurde, in diesem beiden Faktoren „aufgelöst“ wird, d.h. nicht mehr identifiziert werden kann.

Es wurde noch eine dritte Gruppe neuer Kontextvariablen NV_{13} bis NV_{16} generiert. Hier wurden beide gemeinsame Variablen von V_1 bis V_4 , nämlich c_1 und c_5 , allerdings mit kleinem Gewicht einbezogen:

$$NV_i = \frac{1}{\sqrt{19}}c_4 + \frac{4}{\sqrt{19}}s_i + \frac{1}{\sqrt{19}}c_1 + \frac{1}{\sqrt{19}}c_5, \quad \text{for } i = 13 \text{ to } 16 \quad (4)$$

Der Erwartungswert beträgt für die Interkorrelationen dieser Variablen .16, für die Korrelationen mit V_1 bis V_4 .19. Trotz der geringen Korrelationen erschweren auch diese Variablen vermutlich die Identifikation des durch V_1 bis V_4 repräsentierten Faktors. Wie in der vorherigen Studie wurden jeweils 500 Datensätze auf der Basis von 200 sowie 1000 Fällen generiert.

Ergebnisse für den überlappenden Variablenkontext

Zunächst wurde wieder eine Parallelanalyse auf der Basis der Image-Eigenwerte durchgeführt. Anhand der in Tabelle 3.5 dargestellten Eigenwerte erkennt man, dass die Parallelanalyse auf der Basis von 200 Fällen zu uneindeutigen Ergebnissen führte.

Tabelle 3.5

Mittelwerte, Minima und Maxima der Image-Eigenwerte für 500 Lösungen und 100 Zufallslösungen auf der Basis von 200 Fällen

Faktor	Vierfaktorielle Lösungen			Zufalls- Lösungen
	Mittelwert	Minimum	Maximum	Mittelwert
1	3.27	2.16	4.30	.25
2	1.80	1.22	2.59	.18
3	.15	.05	.32	.12
4	.07	.05	.18	.08
5	.04	.01	.08	.06
6	.02	.00	.05	.03
7	.01	.00	.02	.02
8	.00	.00	.01	.01
9	.00	.00	.01	.00
10	.00	.00	.02	.00
11	.02	.00	.05	.01
12	.03	.01	.06	.02
13	.06	.03	.09	.04
14	.08	.05	.14	.07
15	.12	.06	.22	.12

Die Parallelanalyse sprach in 32% der Fälle für die Extraktion von 2 Faktoren, da bei 32% der Lösungen der dritte Zufallseigenwert größer war als der Eigenwert des dritten Faktors, in 46% der Lösungen für die Extraktion von drei Faktoren und in 22% der Lösungen für die Extraktion von vier oder mehr Faktoren. Auch hier war wieder der problematische Anstieg der Eigenwerte für die letzten Faktoren zu verzeichnen.

Auch die Parallelanalyse auf der Basis von 1000 Fällen erbrachte nicht viel eindeutigere Ergebnisse (s. Tabelle 3.6). Eine genauere Betrachtung der Eigenwerte ergab hier, dass die Parallelanalyse in 2% der Fälle für die Extraktion von zwei Faktoren, in 72% der Fälle für die Extraktion von drei Faktoren und in 24% der Fälle für die Extraktion von vier Faktoren sprach. Insgesamt fällt auf, dass der vierte Eigenwert so gering ausfällt, dass er keine nennenswerten Varianzanteile mehr repräsentieren kann (s. Tabellen 3.4 und 3.5).

Allerdings wurde in Beauducel (2001a) festgestellt, dass eine Überextraktion von einem Faktor mitunter die Identifikation eines „aufgelösten“ Faktors erleichtern kann. Da es hier nicht um die Etablierung von Faktoren geht, sondern um die Untersuchung, ob der durch die Variablen

V_1 bis V_4 repräsentierte Faktor im vorliegenden Variablen Kontext identifiziert werden kann, wurden hier die zwei-, drei- und vierfaktoriellen Lösungen berechnet.

Tabelle 3.6

Mittelwerte, Minima und Maxima der Image-Eigenwerte für 500 Lösungen und 100 Zufallslösungen auf der Basis von 1000 Fällen

Faktor	Vierfaktorielle Lösungen			Zufalls- Lösungen
	Mittelwert	Minimum	Maximum	Mittelwert
1	3.18	2.58	3.59	.05
2	1.71	1.43	2.01	.03
3	.06	.02	.13	.02
4	.01	.00	.04	.01
5	.00	.00	.01	.00
6	.00	.00	.00	.00
7	.00	.00	.00	.00
8	.00	.01	.00	.00
9	.00	.01	.00	.00
10	.01	.02	.01	.00
11	.02	.01	.03	.00
12	.02	.01	.04	.00
13	.03	.02	.05	.00
14	.04	.03	.06	.00
15	.05	.04	.08	.01

In Tabelle 3.7 sind die mittleren Varimax-Ladungen für die zwei- und dreifaktoriellen Lösungen auf der Basis von 200 Fällen dargestellt. Die Lösungen auf der Basis von 1000 Fällen unterscheiden sich nur dadurch, dass kleinere Standardabweichungen für die Ladungen auftreten (die maximale Standardabweichung beträgt .04 bei der zweifaktoriellen und .07 bei der dreifaktoriellen Lösung) und dass in der dreifaktoriellen Lösung die mittleren Hauptladungen auf dem dritten Faktor nur .19 betragen.

In der zweifaktoriellen Lösung laden die Variablen V_1 bis V_4 in etwa gleicher Höhe auf den beiden Faktoren. Der durch die Variablen V_1 bis V_4 repräsentierte Faktor, wird – wie zu erwarten war – in den beiden Faktoren, die durch die stark mit V_1 bis V_4 überlappenden Variablengruppen gebildet werden, „aufgelöst“. Dies entspricht den Ergebnissen in Beauducel (2001a) für die zweifaktoriellen Lösungen auf der Basis der Hauptachsenanalyse, der klassischen Faktorenanalyse, der Maximum-Likelihood-Faktorenanalyse und der Alpha-Faktorenanalyse. In der zweifaktoriellen Lösungen unterscheiden sich die Extraktionsmethoden in Bezug auf die Identifikation des durch die Variablen V_1 bis V_4 repräsentierten Faktors nicht.

Tabelle 3.7

Mittelwerte und Standardabweichungen der Varimax-Ladungen der 500 zwei- und dreifaktoriellen Lösungen auf der Basis von 200 Fällen

Variable	zweifaktorielle Lösung		dreifaktorielle Lösung		
	Faktor 1	Faktor 2	Faktor 1	Faktor 2	Faktor 3
V1	.37 (.08)	.36 (.07)	.33 (.08)	.32 (.08)	.23 (.11)
V2	.37 (.08)	.36 (.07)	.33 (.08)	.32 (.08)	.23 (.11)
V3	.37 (.08)	.36 (.07)	.32 (.08)	.32 (.08)	.23 (.11)
V4	.37 (.08)	.36 (.07)	.33 (.08)	.32 (.07)	.23 (.11)
NV5	.67 (.07)	.01 (.06)	.66 (.07)	-.01 (.06)	.13 (.06)
NV6	.67 (.07)	.00 (.06)	.66 (.07)	-.01 (.06)	.13 (.07)
NV7	.67 (.07)	.00 (.06)	.65 (.07)	-.01 (.05)	.13 (.06)
NV8	.67 (.07)	.00 (.06)	.65 (.07)	-.01 (.06)	.13 (.07)
NV9	.01 (.06)	.67 (.06)	-.01 (.06)	.65 (.07)	.14 (.06)
NV10	.01 (.06)	.67 (.06)	-.01 (.06)	.65 (.07)	.13 (.07)
NV11	.01 (.06)	.66 (.06)	-.01 (.06)	.65 (.06)	.13 (.07)
NV12	.00 (.06)	.67 (.07)	-.01 (.06)	.65 (.07)	.13 (.06)
NV13	.22 (.08)	.21 (.07)	.17 (.07)	.17 (.07)	.22 (.10)
NV14	.21 (.08)	.21 (.07)	.16 (.07)	.17 (.07)	.22 (.10)
NV15	.21 (.08)	.21 (.07)	.16 (.08)	.16 (.07)	.21 (.10)
NV16	.21 (.08)	.21 (.07)	.16 (.07)	.17 (.07)	.21 (.11)

Anmerkungen. Ladungen > .30 wurden fett gedruckt. Da die Varimax-Ladungen Korrelationen sind, wurden sie vor der Mittelung Fisher-Z transformiert. Dann wurden die Mittelwerte der Fisher-Z transformierten Ladungen zurücktransformiert.

Auch in der dreifaktoriellen Lösung konstituieren die Variablen V_1 bis V_4 keinen eigenen Faktor, sondern teilen ihre Varianz auf die drei Faktoren auf. Während jedoch die Aufteilung der Varianzen von V_1 bis V_4 auf die drei Faktoren in Beauducel (2001a) für die Hauptachsenanalyse und die Maximum-Likelihood-Faktorenanalyse nahezu gleich ist, sind die mittleren Ladungen auf dem dritten Faktor in der Image-Analyse deutlich reduziert. Dies ist insofern von Bedeutung als der dritte Faktor der einzige Faktor ist, bei dem die Variablen V_1 bis V_4 Hauptladungen aufweisen. Somit ist der Varianzanteil, mit dem die Variablen V_1 bis V_4 über Nebenladungen in die ersten beiden Faktoren eingehen, größer als der Varianzanteil, mit dem sie Hauptladungen auf dem dritten Faktor konstituieren. Also tritt auch in der Lösung mit drei Faktoren kein Faktor auf, der dem durch die Variablen V_1 bis V_4 repräsentierten Faktor entspricht. In Beauducel (2001a) wurde gezeigt, dass bei Hauptachsenanalyse mit iterativer Kommunalitätenschätzung und Maximum-Likelihood-Faktorenanalyse der dritte Faktor in deutlichem Umfang durch die Variablen V_1 bis V_4 repräsentiert wird, so dass diese Verfahren eine bessere Identifikation des „aufgelösten“ Faktors erlauben.

Die geringe Höhe der Ladungen auf dem dritten Faktor ist ein Hinweis darauf, dass die Extraktion von vier Faktoren wenig angemessen ist. Dennoch wurden die vierfaktoriellen Lösungen berechnet, um sicherzustellen, dass nicht etwa ein vierter Faktor nach der Rotation nennenswerte Anteile der Variablen V_1 bis V_4 repräsentiert. Die ersten drei Varimax-rotierten Faktoren entsprechen weitestgehend den in Tabelle 3.7 dargestellten Faktoren der dreifaktoriellen Lösung. Die größten mittleren Varimax-Ladungen, die auf dem vierten Faktor auftreten liegen bei .10. Somit tritt auch bei der Extraktion und Rotation von vier Faktoren kein Faktor auf, der primär durch die Variablen V_1 bis V_4 markiert wird. Es war also auch mit Hilfe einer leichten Überextraktion bei der Image-Analyse nicht möglich, den durch V_1 bis V_4 markierten Faktor im überlappenden Variablenkontext zu identifizieren.

Dieser Befund steht in deutlichem Gegensatz zu dem Befund von Beauducel (2001a), dass bei der Extraktion von vier Faktoren mit Hilfe von Hauptachsenanalyse mit iterativer Kommunalitätenschätzung, Maximum-Likelihood- sowie Alpha-Faktorenanalyse der vierte Faktor nur durch Hauptladungen der Variablen V_1 bis V_4 markiert wird. Insofern erwies sich die Image-Analyse bei der Identifikation von Faktoren in einem Kontext überlappender Variablen der Hauptachsenanalyse mit iterativer Kommunalitätenschätzung, der Maximum-Likelihood- und der Alpha-Faktorenanalyse als unterlegen. Der Grund für die deutlichere Abhängigkeit der Ergebnisse der Image-Analyse vom Variablenkontext besteht vermutlich daher, dass die Images der Variablen, jeweils durch die übrigen Variablen konstituiert werden. Vermutlich besteht ein hohes Maß an Kontext-Abhängigkeit der Image-Faktoren, weil bei der Image-Analyse nur die Varianzanteile von Variablen berücksichtigt werden, die durch lineare Regression von den übrigen Variablen vorhergesagt werden können. Der o.g. Befund ist insofern bemerkenswert, als für die Hauptkomponentenanalyse, die im Gegensatz zur Image-Analyse die meiste spezifische Varianz in die Analyse einbezieht, ebenfalls ungünstige Ergebnisse bei der Generalisierung von Faktoren auftraten (Widaman, 1993; Beauducel, 2001a). Offenbar führt sowohl eine extrem starke Betonung der spezifischen Varianz wie in der Hauptkomponentenanalyse als auch eine extrem starke Betonung der gemeinsamen Varianz wie in der Image-Analyse zu Problemen bei der Generalisierbarkeit von Faktoren. Die Hauptachsenanalyse, aber auch die Alpha- und die Maximum-Likelihood-Faktorenanalyse weisen offenbar einen für die Generalisierbarkeit von Faktoren günstigeren mittleren Grad an Berücksichtigung gemeinsamer und spezifischer Varianz auf. Natürlich müssen weitere Analysen mit weiteren Datensätzen folgen, um die hier dargestellten Ergebnisse zu erhärten. Die vorliegenden Ergebnisse sind jedoch im Einklang mit Velicer und Fava (1998), die eine etwas ungünstigere Reproduktionrate von Faktorenlösungen für die Image-Analyse im Vergleich zur Hauptkomponenten- und Maximum-Likelihood-Faktorenanalyse feststellten. Eine

zusätzliche Berücksichtigung der Image-Analyse als Extraktionsmethode im Kontext der vorliegenden Arbeit erscheint somit auf Basis der vorliegenden Befunde nicht indiziert.

3.2.2.3 Rotationsmethoden: Bestandsaufnahme und Auswahl

Wie bei den Extraktionsmethoden wird auch bei den Rotationsmethoden mitunter die Auffassung vertreten, dass verschiedene Verfahren zu sehr ähnlichen Ergebnissen führen. Diese Auffassung wird durch einige Studien im Bereich psychometrischer Traits gestützt, bei denen Rotationsmethoden verglichen wurden (Goldberg, 1990; Ostendorf, 1990). Allerdings beziehen sich die Studien von Goldberg (1990) und Ostendorf (1990) allein auf den Adjektiv-Ansatz, so dass die Ergebnisse auf der Basis von ausformulierten Items durchaus davon abweichen können.

Goldberg (1990) und Ostendorf (1990) berichten auch, dass orthogonale und oblique Rotationsverfahren praktisch zu denselben Ergebnissen führten. Allerdings berichtet Ostendorf (1990) immerhin Korrelationen um .25 zwischen einigen oblique rotierten Faktoren des FFM. Die Frage ob derartige Korrelationen, die bei der Gruppengröße in Ostendorf (1990) sicher statistisch bedeutsam sind, orthogonale Rotation legitimieren, darf zumindest als offen angesehen werden. Becker (1999), der die Oblimin-Rotation (mit $\Delta=0$) verwendet, stellte immerhin eine Korrelation von .37 zwischen Extraversion und Openness fest. Digman (1997) schlägt Faktoren höherer Ordnung für die Faktoren des FFM vor, was nur durch Interkorrelationen zwischen den Faktoren des FFM legitimiert werden kann. Daher hält Digman (1997) orthogonale Rotationen im Kontext des FFM für problematisch. Die Problematisierung orthogonaler Rotation erfolgte natürlich bereits viel früher und ohne Bezug zum FFM, beispielsweise in Cattell und Dickman (1962, S. 395): „Unfortunately, in nature, [...] , orthogonal factors are as uncommon as a straight tree.“

Da in der psychometrischen Traitforschung regelmäßig sowohl oblique (z.B. Becker, 1999; Carver & White, 1994; Cattell, 1994a; Digman, 1997) als auch orthogonal (z.B. Costa & McCrae, 1992; Costa & McCrae, 1995; Di Blas & Forzi, 1998) rotiert wird, kann die Frage nach der angemessenen oder optimalen Faktorenrotation nicht ohne weiteres als beantwortet angesehen werden. Auch sind sowohl für oblique als auch für orthogonale Faktorenrotation viele verschiedene Methoden verfügbar, die – zumindest in Methodenstudien – teilweise zu deutlich unterschiedlichen Ergebnissen führten (z.B. Hakstian & Abell, 1974). Daher wird im folgenden ein Vergleich von Rotationsmethoden durchgeführt.

Zum Vergleich von Rotationsmethoden

Der Vergleich verschiedener Rotationsmethoden kann natürlich nicht die Vielzahl von Rotationsmethoden einbeziehen, die bisher entwickelt wurden. Hier sollen nur die Klassen von Rotationsmethoden betrachtet werden, die sich entweder im Rahmen von Methodenstudien besonders bewährt oder aber in der empirischen Anwendungspraxis, insbesondere in der psychometrischen Traitforschung eine besondere Rolle spielen. Einen Überblick über viele, heute teilweise kaum noch gebräuchliche Kriterien, findet sich beispielsweise in Gorsuch (1983) und Harman (1967).

Wie bei den Extraktionskriterien und -methoden stellt sich zunächst die Frage nach der methodologischen Qualität verschiedener Rotationsverfahren. Bei den Extraktionskriterien und -methoden konnte die vergleichende Bewertung auf einer methodologischen Basis erfolgen, die von inhaltlichen Kriterien relativ unabhängig ist. So ist beispielsweise die Forderung in der Parallelanalyse, dass die Eigenwerte der rotierten Komponenten bzw. Faktoren größer sein sollten als die Eigenwerte auf der Basis entsprechender Zufallskorrelationen weitgehend unabhängig von inhaltlichen Überlegungen. Bei den Rotationsmethoden ist die Abhängigkeit der methodologischen Bewertung von inhaltlichen Kriterien jedoch größer. Dies hat vermutlich zwei Ursachen: Zum einen bleibt die erklärte Varianz der Faktorenlösungen bei der Rotation gleich, d.h. für eine Faktorenlösung gibt es im Prinzip unendlich viele Möglichkeiten der Rotation, die alle dieselbe Legitimation auf der Basis einer gegebenen Kovarianz- oder Korrelationsmatrix haben. Zum anderen wurde die Faktorenrotation mit dem Ziel eingeführt, die Interpretation von Faktorladungsmatrizen zu erleichtern (s. Thurstone, 1947). Der Ausgangspunkt für die Frage nach der optimalen Faktorenrotation war somit nicht die Empirie (denn die empirische Legitimation der Lösung wird durch die Rotation nicht tangiert) sondern die Interpretierbarkeit. Die Interpretation einer Faktorladungsmatrix kann jedoch nicht ohne inhaltliche Vorannahmen erfolgen.

Ein allgemeiner inhaltlicher Bezugsrahmen, der keine methodologische Selbstverständlichkeit darstellt, ist Thurstones (1947) Kriterium der Einfachstruktur. Das Kriterium, das – zumindest nominell – mit den meisten Rotationsmethoden erreicht werden soll, ist optimale Einfachstruktur. Thurstone (1947, S. 335) stellte fünf Kriterien auf, denen eine Ladungsmatrix genügen muss, wenn sie eine optimale Einfachstruktur haben soll (zitiert nach Pawlik, 1967, S.180):

- 1) Jede Zeile der Faktormatrix soll mindestens eine Nullladung haben.
- 2) Zu jeder der k Spalten der Faktormatrix soll es eine andere Kombination von mindestens k der m Variablen geben, die linear unabhängig sind und in diesem Faktor Nullladungen besitzen.
- 2a) Jede Spalte der Faktormatrix soll mindestens k Nullladungen enthalten, wobei die Kombination dieser k Variablen von Spalte zu Spalte eine andere sein soll.
- 3) Für jede Kombination zweier Spalten der Faktormatrix soll es mehrere Variablen geben, die nur in einem der beiden Faktoren zu Null geladen sind.
- 4) Für jede Kombination zweier Spalten der Faktormatrix soll es eine große Zahl von Variablen geben, die in beiden Faktoren zu Null geladen sind. (Dieses Kriterium gilt erst ab $k > 5$.)
- 5) Für jede Kombination zweier Spalten der Faktormatrix soll es nur wenige Variablen geben, die in beiden von Null verschiedene Ladungen besitzen.

Bemerkenswert ist Thurstones deutliche Fokussierung auf die Nullladungen, lange bevor Campbell und Fiske (1959) den Aspekt der diskriminanten Validität als bedeutsam herausstellten. Der Versuch Thurstones durch die Erhöhung der Anzahl der Nullladungen in den Faktorladungsmatrizen die Prägnanz und Interpretierbarkeit der Lösungen zu erhöhen sollte nicht unabhängig von Thurstones (1938) Auseinandersetzung mit Spearmans (1927) Ideen im Rahmen der Intelligenzstrukturforschung gesehen werden. Erst durch die Rotation zur Einfachstruktur konnte Thurstones Loslösung von Spearmans Annahme eines Generalfaktors im Rahmen der Faktorenanalyse methodisch vollzogen werden. Obwohl das Konzept der Einfachstruktur, zusammen mit dem Konzept der faktoriell „reinen“ Markiertvariablen, nur eine inhaltliche Herangehensweise an die Interpretation von Faktorladungsmatrizen darstellt, stellt dieses Konzept die Grundlage fast für die ganze psychometrische Tradition im Rahmen der exploratorischen Faktorenanalyse und in der psychometrischen Traitforschung dar.

Bisher wurden nur wenige Alternativen zum Konzept der Einfachstruktur entwickelt und diskutiert. Eine Alternative besteht in der kriteriumsorientierten Testkonstruktion bzw. damit zusammenhängend in der kriteriumsorientierten Faktorenrotation (z.B. Eysenck, 1950), bei der die Frage nach der Strukturierung auf pragmatische Art beantwortet wird (zu Problemen dieses Konzepts siehe Holz-Ebeling, 1995; Pawlik, 1968). Eine weitere Alternative besteht in der Rotation zur Erhöhung der Invarianz zwischen Lösungen. Cattell und Cattell (1955) schlugen dazu eine Rotationsmethode vor, bei der die Proportionalität der Ladungsprofile derselben Variablen in verschiedenen Lösungen maximiert wird. Es stellt sich jedoch die Frage, ob dieses Kriterium zur Faktorenrotation ausreichend ist und ob es nicht eher eine Ergänzung denn eine

Alternative zum Konzept der Einfachstruktur darstellt. Inzwischen wird die Frage der Faktoreninvarianz auch im Rahmen von Zielrotationen auf Lösungen geprüft, die in der Regel eine hohe Einfachstruktur aufweisen (z.B. Paunonen, 1997). Rotationsmethoden zur Erreichung ausgeglichener Verlaufskurven anstatt von Einfachstruktur wurden von Tucker (1960; zit. nach Pawlik, 1968) vorgeschlagen. Keine dieser alternativen Rotationsmethoden spielte jedoch in der psychometrischen Traitforschung bisher eine besondere Rolle. Da im Rahmen dieser Arbeit eher eine methodisch begründete Einschränkung der Freiheitsgrade bei der Faktorenanalyse angestrebt wird, erscheint eine Exploration alternativer Ziele der Faktorenrotation hier nicht indiziert.

Wie bereits erwähnt, beziehen sich die von Thurstone (1947) aufgestellten Kriterien der Einfachstruktur alle auf die Anzahl der Nullladungen in den Ladungsmatrizen. Daher ist es nicht erstaunlich, dass in der Folge von Thurstone häufig die Anzahl der Nullladungen als Kriterium für den Grad der Einfachstruktur herangezogen wurde. Als Nullladungen wurden meist Ladungen vom Betrag $\leq .10$ definiert (z.B. Cattell & Dickman, 1962; Cattell & Muerle, 1960; Hakstian & Abell, 1974; Thurstone, 1940). Allerdings sollten bei sehr großen Probandengruppen kleinere Ladungsbeträge als Nullladungen definiert werden (Cattell, 1978). Dann wird für jeden Faktor bzw. für die gesamte Ladungsmatrix die Zahl der so definierten Nullladungen ausgezählt (z.B. Cattell, 1994a; Hakstian & Boyd, 1972; Horn, 1963; Ostendorf, 1990). Im übrigen entwickelte Bargmann (1955) auf der Basis der Anzahl der Nullladungen pro Faktor einen Test für die Signifikanz der Einfachstruktur von Faktoren. Dieses Vorgehen berücksichtigt natürlich nicht die vielen Kombinationen von Nullladungen auf Faktoren, die Thurstone fordert. Allerdings ist bei einer großen Zahl von Nullladungen pro Faktor die Wahrscheinlichkeit groß, dass alle von Thurstone genannten Kriterien erfüllt sind.

Das Kriterium der Anzahl der Nullladungen pro Faktor wird zur Zeit selten verwendet. Dabei gibt es allerdings keine Arbeit, in der dieses Kriterium als fehlerhaft identifiziert wurde. Zwar wurden Thurstones Kriterien von Kaiser (1958) und Harman (1967) als nicht genügend objektiv kritisiert, jedoch trifft diese Kritik auf die vereinfachte Form dieses Kriteriums, die Auszählung der Nullladungen nicht zu (da dieses Kriterium natürlich bei jedem Forscher zum gleichen Ergebnis führt). Der Grund für die seltene Auszählung der Nullladungen ist vermutlich das Vertrauen, dass von Anwendern den sogenannten „analytischen“ Rotationsmethoden entgegengebracht wird. Bei den analytischen Rotationsmethoden wird ein Rotationskriterium maximiert, das mit der Einfachstruktur im Sinne von Thurstone zusammenhängt. Die Enge dieses Zusammenhangs ist jedoch oft nur schwer zu beurteilen. Jennrich und Sampson (1966, S. 313), die mit der direkten Oblimin-Rotation eine der am weitesten verbreiteten analytischen Rotationsmethoden entwickelten, beschrieben die Situation wie folgt:

“There are a number of popular oblique rotation schemes [...]. Most of the simplicity criteria used in these schemes are sufficiently complex that it is difficult, if not impossible, to tell in terms of reference structure or the primary loading matrix what “simplicity“ really means.“ Dennoch werden als nächstes auch die erfolgversprechend erscheinenden analytischen Rotationsmethoden diskutiert.

Methoden der orthogonalen Rotation zur Einfachstruktur

Am bekanntesten und am weitesten verbreitet ist wahrscheinlich die Varimax-Rotation von Kaiser (1958), bei der die Varianz der quadrierten Faktorladungen maximiert wird. Dieses Kriterium ist sicherlich sehr nützlich für die Forschungspraxis, sollte jedoch nicht unreflektiert verwendet werden, wie übrigens Kaiser (1970) selbst, mit viel selbstkritischem Humor bemerkt.

Eine Reihe weiterer Rotationsmethoden, wie Quartimax (Carroll, 1953) und Equamax (Saunders, 1962) können zusammen mit der Varimax-Rotation als Varianten der Orthomax-Rotation (Harman, 1960; Harris & Kaiser, 1964) aufgefaßt werden. Die Methoden unterscheiden sich nur in Bezug darauf, ob die Varianz der quadrierten Ladungen primär über die Spalten bzw. Faktoren (Quartimax) oder primär über die Zeilen bzw. Variablen (Varimax) oder über Zeilen und Spalten in gleichem Maße (Equamax) maximiert wird. Das allgemeine Orthomax-Kriterium erlaubt jedoch noch andere Gewichtungsverhältnisse. Die verschiedenen Rotationsmethoden, die das Orthomax-Kriterium bietet, sind bisher leider nur selten angewendet und untersucht worden. Ein Grund dafür könnte in der extrem hohen Verbreitung der Varimax-Rotation, eines Spezialfalls der Orthomax-Rotation liegen.

Hakstian und Boyd (1972) haben allerdings Quartimax, Varimax, Equamax und darüber hinausgehende Varianten der Orthomax-Kriteriums untersucht und kommen zu folgendem Ergebnis: Wenn der Parameter w (mitunter auch als „ γ “ bezeichnet) des Orthomax-Kriteriums null (wie bei Quartimax) oder kleiner ist, dann ist die Einfachstruktur in der Regel suboptimal. Darüber hinaus verteilt sich bei diesen Lösungen die Varianz in sehr ungleichen Anteilen auf die Faktoren.

Wenn der Parameter w eins (wie bei Varimax), die Hälfte der Faktorenzahl (wie bei Equamax) oder mehr beträgt, ist die Einfachstruktur der Lösungen in der Regel hoch. Allerdings unterscheiden sich die Orthomax-Lösungen mit $w \geq 1$ teilweise erheblich voneinander: Je größer w , desto gleichmäßiger verteilt sich die Varianz auf die Faktoren. Hakstian und Boyd (1972) sehen keine Möglichkeit auf der Basis der Evaluation der Einfachstruktur zu einer

allgemeinen Präferenz einer bestimmten Form der Orthomax-Rotation zu kommen. Sie schlagen daher vor, die Entscheidung zwischen den verschiedenen, in Bezug auf die Einfachstruktur gleichwertigen Lösungen, auf Basis der Interpretierbarkeit der Faktoren vorzunehmen. Sie sind der Auffassung, dass der Vergleich verschiedener Orthomax-Lösungen zu einer weniger „blinden“ Maximierung der Einfachstruktur führt als die einfache Verwendung der Spezialfalls der Varimax-Rotation.

So überzeugend die Analysen von Hakstian und Boyd (1972) auch sind, so läuft die Auswahl eines Orthomax-Kriteriums anhand der subjektiven Bewertung der Einfachstruktur verschiedener Lösungen dem strategischen Ziel dieser Arbeit, d.h. der Erhöhung der Verbindlichkeit psychometrischer Forschung deutlich entgegen. Eine Möglichkeit zu einer konzeptuell orientierten Auswahl eines geeigneten Orthomax-Kriteriums ergibt sich aus Crawford und Ferguson (1970), die eine weitere, aus dem Orthomax-Kriterium ableitbare Rotationsmethode vorschlagen: Die orthogonale Variante der Parsimax-Rotation. Wie bereits erwähnt, wurde die Equamax-Rotation entwickelt, um die Varianz der Zeilen (Variablen) und Spalten (Faktoren) in gleichem Maß zu maximieren. Allerdings wurde das Kriterium empirisch ermittelt (Saunders, 1962). Demgegenüber gelangten Crawford und Ferguson (1970) auf analytischem Wege zu einer Gleichgewichtung der Zeilen- und Spaltenvarianzen, wobei die analytische Lösung hier vermutlich genauer ist, als eine notwendigerweise auf unvollständiger Empirie beruhende Näherungslösung.

Die Gleichsetzung der Zeilen- und Spalten-Varianz ist ein konzeptuell attraktiver Spezialfall des Orthomax-Kriteriums, bei dem Variablen und Faktoren bei der Maximierung der Einfachstruktur das gleiche Gewicht zukommen. Wenn man dieses Vorgehen kritisieren wollte, könnte man anregen, dass die Faktoren gegenüber den Variablen bei der Maximierung der Varianzen der quadrierten Ladungen ein größeres Gewicht bekommen sollten, da ja in der Regel die Faktoren interpretiert werden und seltener die Interpretation der Ladungen einzelner Variablen im Vordergrund steht. Aber gerade bei der häufig verwendeten Varimax-Rotation spielt die Varianz der Variablen eine größere Rolle als die Varianz der Faktoren. Wenn also über Alternativen zum Parsimax-Kriterium nachgedacht wird, sollten eher Kriterien erwogen werden, bei denen die Varianz der quadrierten Ladungen über Faktoren mehr maximiert wird als über Variablen. Ein Kriterium, das nach Crawford und Ferguson (1970) allein die Varianz der quadrierten Ladungen über die Faktoren maximiert, ist das „Factor-Parsimony“ Kriterium, bei dem fast ausschließlich die Varianz der quadrierten Ladungen über die Faktoren maximiert wird. Dieses Kriterium weist aber ähnliche Probleme auf wie das Quartimax-Kriterium, bei dem allein die Varianz der quadrierten Ladungen über die Variablen maximiert wird. So wie

das Quartimax-Kriterium zu General-Faktoren führen kann und die Varianz unter den Faktoren sehr ungleich verteilt, verteilt das Factor Parsimony Kriterium die Varianz unter den Variablen ungleich und kann zu sehr unterschiedlich komplexen Variablen führen, was die Interpretation ebenfalls erschwert. Daher sind das Equamax und das Parsimax Kriterium einem derartigen Kriterium vorzuziehen, wobei das Parsimax-Kriterium als die konsequentere Umsetzung der mit dem Equamax-Kriterium verfolgten Idee der Gleichgewichtung der Zeilen- und Spaltenvarianzen angesehen werden kann.

Auch Schweizer (2001) versucht, Defizite der Varimax-Rotation anhand der gleichgewichteten Berücksichtigung verschiedener Regeln, die wiederum verschiedene Aspekte der Einfachstruktur repräsentieren, zu überwinden. Seine Simulationsstudien deuten darauf hin, dass ein kombiniertes Rotationskriterium bei weniger als 20 Variablen zu einer optimalen Einfachstruktur führt. Allerdings ist die Variablenanzahl bei psychometrischen Traitstudien sehr groß, weshalb das Rotationskriterium von Schweizer (2001) im vorliegenden Kontext weniger geeignet erscheint.

Methoden der obliquen Rotation zur Einfachstruktur

Hakstian (1971) sowie Hakstian und Abell (1974) untersuchten verschiedene Methoden der obliquen Rotation zur Einfachstruktur. Dabei wurden die Lösungen in Bezug auf die Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ und in Bezug auf die Ähnlichkeit mit visuell rotierten Lösungen verglichen. Als besonders leistungsfähig erwiesen sich die Harris-Kaiser-Typ II-Rotationen (Harris & Kaiser, 1964), Promax- (Hendrickson & White, 1964) und die direkte Oblimin-Rotation (Jennrich & Sampson, 1966). Beauducél (1997) stellte jedoch fest, dass die Harris-Kaiser-Typ II Rotation der Promax- und direkten Oblimin-Rotation hinsichtlich der Optimierung der Einfachstruktur unterlegen war, weshalb diese Methode in der vorliegenden Arbeit keine weitere Berücksichtigung fand.

Die Promax-Rotation baut direkt auf einer Varimax- bzw. Orthomax-Rotation auf, die Ladungen einer Orthomax-Rotation werden potenziert und dann als Zielmatrix für eine Procrustes-Rotation verwendet. Auf diese Weise erzeugt die Promax-Rotation oblique Lösungen aus den Orthomax-Lösungen. Der Grad der Potenzierung der Ladungen ist durch einen Parameter einzustellen und bestimmt den Grad der Schiefwinkligkeit der Lösungen.

Die direkte Oblimin-Rotation basiert für den Fall, dass der frei wählbare Parameter (oft mit „ Δ “ oder „ γ “ bezeichnet) null gesetzt wird (das Oblimin-Kriterium entspricht dann dem Quartimin-Kriterium), auf der Minimierung des über alle Faktorenpaare berechneten Kreuz-

produkts der quadrierten Ladungen. Dies kommt einer Minimierung der Korrelation zwischen den Ladungen aller Faktorenpaare sehr nahe. Wenn der Parameter größer null gesetzt wird, wird zusätzlich das Produkt der Summen der quadrierten Ladungen maximiert. Wenn der Parameter kleiner null gesetzt wird, wird der Betrag des Produkts zusätzlich minimiert. Der Parameter steuert somit die Gleichmäßigkeit der Verteilung der Ladungen auf die Faktoren.

Hakstian und Abell (1974) stellten fest, dass der jeweils frei zu wählende Parameter bei Harris-Kaiser-, Promax-, und direkter Oblimin-Rotation die Güte der Lösungen stark beeinflusst. Sie geben für jede der drei Methoden den Wertebereich des Parameters an, für den optimale Lösungen zu erwarten sind. Trendafilov (1994) problematisierte, dass das Ergebnis der Promax-Rotation von der subjektiven Entscheidung für einen bestimmten Parameter abhängt. Er schlug daher eine alternative Variante der Konstruktion einer Zielmatrix für die Promax-Rotation vor (Promaj-Rotation), bei der die Wahl verschiedener Parameter vermieden wird. Trendafilov (1996) entwickelte schließlich auch eine orthogonale Rotationsmethode (Orthomaj), die die Varimax-Vorrotation bei der Promaj-Rotation ersetzen kann. Trendafilov (1994, 1996) überprüft die Zunahme der Einfachstruktur bei Promaj- und Orthomaj-Rotation anhand von Bentlers (1977) Index der faktoriellen Einfachheit. Beauducel (1996) stellte jedoch fest, dass der Index der faktoriellen Einfachheit nur gering mit dem klassischen Maß für die Güte der Einfachstruktur, d.h. mit der Anzahl der Variablen mit Ladungen nahe null zusammenhängt. Insofern stellt sich die Frage, ob mit den Verfahren von Trendafilov (1994, 1996) tatsächlich primär die Einfachstruktur im Sinne Thurstones maximiert wird. Diese Frage stellt sich jedoch grundsätzlich für alle analytischen Rotationsverfahren und soll daher in einem eigenen Abschnitt behandelt werden.

Analytische Rotationsmethoden und Einfachstruktur im Sinne Thurstones

Es stellt sich die Frage, inwieweit Varimax- bzw. Orthomax-, Oblimin- und andere analytische Kriterien Thurstones Prinzip der Einfachstruktur entsprechen, für das sie eine objektivierende Konkretisierung darstellen sollen. Grundsätzlich erfolgt beispielsweise durch die Orthomax-Kriterien eine Maximierung der Varianz der quadrierten Ladungen, was in der Regel auch die von Thurstone thematisierte Anzahl der Nullladungen maximiert. Die Ladungen werden aus dem mittleren Bereich „herausgedrückt“, mittlere Ladungen sollen entweder zu Nullladungen oder zu Hauptladungen werden. Der wesentliche Unterschied der bei den Orthomax- und Oblimin-Methoden verwendeten Rotationskriterien zu Thurstones Kriterien ist, dass ein cut-off Wert für die Definition von Nullladungen vermieden wird.

Dies ist einerseits ein Vorteil dieser Verfahren, weil dadurch die Überlegungen, ob bei unterschiedlicher Stichprobengröße unterschiedlich rotiert werden muss (Cattell, 1978), entfallen. Andererseits ist das auch ein Nachteil, weil zur Maximierung der Varianz der quadrierten Ladungen beispielsweise auch Ladungsbeträge minimiert werden, die vor und nach der Minimierung statistisch bedeutsam sind. Es stellt sich die Frage, warum der Betrag substanzieller, statistisch bedeutsamer Ladungen verändert werden soll, nur um die Varianz der quadrierten Ladungen zu maximieren (z.B. Varimax) oder das Kreuzprodukt der quadrierten Ladungen zu minimieren (z.B. Quartimin)? Bei Thurstone ist das Ziel eindeutig benannt: Wenn die Zahl der Nullladungen in verschiedenster Kombination maximiert wird, dann wird auch die Zahl der Variablen, die bei der Interpretation der Faktoren berücksichtigt werden muss, minimiert. Dadurch vereinfacht sich die Interpretation der Faktoren, was das primäre Ziel der Faktorenrotation ist. Wenn nun auch Ladungsbeträge zur Maximierung der Ladungsvarianz verändert werden, ohne dass sie dadurch zu statistisch unbedeutenden Nullladungen werden, dann wird die Interpretation der Faktoren im Grunde nicht vereinfacht.

Die Gefahr, die bei der Verwendung der Rotationsmethoden der Orthomax- aber auch der Oblimin-Familie besteht, ist aber die: Dass Ladungen, die vor und nach der Rotation statistisch bedeutsam waren, durch die Rotation im Betrag soweit verringert werden, dass sie vom interpretierenden Anwender quasi als Nullladungen behandelt werden. McDonald (1985) beklagt beispielsweise, dass Faktorladungen von .30 bereits als Nullladungen interpretiert werden. Teilweise werden noch größere Ladungen als Nullladungen interpretiert bzw. nicht berichtet, die bei der verwendeten Stichprobengröße mit Sicherheit statistisch bedeutsam waren (z.B. Carver & White, 1994). Die in jüngerer Zeit zu beobachtende Vernachlässigung der Auszählung der Ladungsbeträge $\leq .10$ und die seltene Verwendung des Bargmann-Tests kann als Aufweichung der methodischen Standards angesehen werden. Auch die Argumentation von Gorsuch (1983), der es als Vorteil ansieht, wenn substanzielle Ladungen, d.h. Ladungen, die keine Nullladungen sondern statistisch bedeutsame Ladungen sind, möglichst gering gehalten werden, wenn sie keine Hauptladungen sind, erscheint vor diesem Hintergrund problematisch. Noch einmal pointiert: Die durch die statistisch bedeutsamen Nebenladungen abgebildeten empirischen Beziehungen sind, genau wie die Hauptladungen relevante empirische Phänomene, die nicht interpretativ eliminiert werden dürfen. Von daher erscheinen Rotationen, die ein derartiges Vorgehen nahelegen, problematisch. Vassend und Skrandal (1997, S. 163) beschreiben diesen Sachverhalt folgendermaßen: "Thus, the common EFA [EFA=exploratory factor analysis; der Verf.] practice has led investigators to interpret their results *as though* they had simple structure when they might be said not to have one at all."

Ein weiteres Problem der analytischen Rotationsmethoden ist, dass mit all diesen Methoden einerseits eine Maximierung der Einfachstruktur angestrebt wird, sich die Lösungen jedoch teilweise erheblich unterscheiden (z.B. Hakstian & Abell, 1974). Diese Unterschiede sind insofern nicht verwunderlich als die verschiedenen analytischen Methoden unterschiedliche Aspekte von Thurstones (1947) Kriterien der Einfachstruktur maximieren. Beispielsweise wird Thurstones erstes Kriterium besonders durch das Quartimax-Kriterium maximiert, während das zweite und dritte Kriterium der Einfachstruktur besonders durch das Quartimin-Kriterium maximiert wird (für eine Zuordnung von Rotationskriterien zu Thurstones Kriterien siehe auch Revenstorff, 1976). Im Grunde ist das Problem folgendes: Thurstones ursprüngliche, verbal-deskriptive Kriterien der Einfachstruktur wurden von vielen als zu unverbindlich eingestuft (z.B. Kaiser, 1958). Daraufhin wurden analytische Kriterien, d.h. mathematische Formalisierungen der Einfachstruktur vorgenommen (z.B. Jennrich & Sampson, 1966; Kaiser, 1958). Der Bezug der verschiedenen analytischen Rotationskriterien zueinander und zu Thurstones ursprünglichen Kriterien wurde jedoch unzureichend diskutiert, vermutlich da die Entwickler der Kriterien jeweils mit ihrer Entwicklung nicht eine bestimmte Form der Einfachstruktur intendierten, sondern Einfachstruktur schlechthin optimal zu erfassen suchten. Jedenfalls wird in den einschlägigen Arbeiten (z.B. Hendrickson & White, 1964; Jennrich & Sampson, 1966) in der Regel nicht diskutiert, inwieweit die jeweilige Formalisierung lediglich einen Aspekt der Einfachstruktur konkretisiert. Die Spezifität einzelner Rotations-Kriterien wie Quartimax, Varimax oder Equamax in Bezug auf die Maximierung der Einfachstruktur wird jedoch sehr deutlich, wenn man erkennt, dass es sich dabei um Spezialfälle allgemeinerer Rotationskriterien wie beispielsweise des Orthomax-Kriteriums handelt. Dass auch die Orthomax- und Oblimin-Familie als Spezialfälle eines noch allgemeineren quartischen Rotationskriteriums aufgefaßt werden können, wurde erst von Clarkson und Jennrich (1988) gezeigt. Aus der Verallgemeinerung von Clarkson und Jennrich (1988) ergibt sich, dass bei orthogonaler Rotation zwei verschiedene Terme mit unterschiedlichem relativen Gewicht (entspricht Orthomax) und bei obliquen Rotation vier verschiedene Terme mit unterschiedlichem relativen Gewicht maximiert werden können. Die Entscheidung über die relativen Gewichte der verschiedenen Kriterien ist dabei im Prinzip dem Anwender überlassen. Insgesamt kann also festgehalten werden, dass die mathematischen Formalisierungen, die seit Carroll (1953) entwickelt wurden, zwar Konkretisierungen von Thurstones Beschreibung der Einfachstruktur darstellen, aber dass durch die Vielzahl der Konkretisierungen jeweils unterschiedlicher Aspekte der Einfachstruktur die eingangs beklagte Unverbindlichkeit der verbalen Beschreibungen der Einfachstruktur (Kaiser, 1958) nun durch eine Unverbindlichkeit der mathematischen Konkretisierungen komplementiert wurde. Vermutlich ist diese Entwicklung ein Hinweis darauf, dass mathematische Formalisierungen konzeptuelle Stringenz nicht auto-

matisch herbeiführen können. Konzeptuell könnte es zweckmäßig sein, wie bei der Parsimax-Rotation von Crawford und Ferguson (1970) für die beiden Kriterien im orthogonalen Fall, alle vier quartischen Kriterien bei obliquen Rotation gleich zu gewichten um zu einem möglichst ausgewogenen Rotationskriterium zu kommen³.

Damit wäre der Eingangs problematisierte Aspekt der Minimierung bedeutsamer Nebenladungen und der damit einhergehenden Vernachlässigung derartiger Nebenladungen bei der Interpretation allerdings nicht ausgeräumt. Aus diesem Grund wird im folgenden auf eine weitere Variante von Rotationsmethoden eingegangen, bei der primär die Anzahl der Nullladungen maximiert wird. Insofern sich diese Methoden auf die von Thurstone thematisierten Nullladungen konzentrieren, sind sie konzeptuell sparsamere Annäherungen an das Prinzip der Einfachstruktur.

Rotationsmethoden zur Maximierung der Anzahl der Nullladungen

Cattell und Muerle (1960) vertraten die Auffassung, dass die analytischen Rotationsmethoden Thurstones Kriterien der Einfachstruktur nicht gerecht werden und entwickelten die Maxplane-Rotation, bei der die Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ maximiert werden sollte. Allerdings erwies sich die Maxplane-Rotation auch in bezug auf das explizit maximierte Kriterium weder als besonders leistungsfähig noch als anderen Verfahren überlegen (z.B. Hakstian, 1971; Wittmann & Hampel, 1976).

Katz und Rohlf (1974) entwickelten die Functionplane-Rotation, eine Variante der Maxplane-Rotation, die das Problem lokaler Optima, das bei der Maxplane-Rotation auftritt, umgehen soll. Allerdings kritisierte Hakstian (1971), dass bei der Maxplane-Rotation die Einfachstruktur in den Bezugsvektoren maximiert wird und nicht im Faktorenmuster. Diese Kritik wurde von Katz und Rohlf (1974) nicht aufgefangen. Katz und Rohlf (1975) stellten eine Weiterentwicklung ihrer Methode (Primary Product Functionplane) vor und wendeten diese sowie die Functionplane-Methode auf „Harmans 24 psychologische Variablen“ (ein „klassischer“ Datensatz zum Vergleich von Rotationsmethoden) an. Dabei stellten sie für Primary Product Functionplane eine große Anzahl (46) von Variablen mit Ladungsbeträgen $\leq .10$ fest. Da mit Maxplane bei diesem Datensatz nur 32 Ladungsbeträge $\leq .10$ waren, kann dies als erster Hinweis auf deutliche Verbesserung der Rotationsmethode gegenüber Maxplane gewertet werden.

³ Eine analoge Überlegung beschreibt Schweizer (2001) für orthogonale Rotationskriterien

Auch Kiers (1994) war der Auffassung, dass optimale Einfachstruktur mit einer großen Zahl von Nullladungen einher geht. Er entwickelte die Simplimax-Rotation, bei der allerdings auch ein Parameter frei gewählt werden muss. Der Parameter p entspricht bei Kiers (1994) der Anzahl der Nullladungen, die für eine Ladungsmatrix angenommen werden. Damit erfordert dieser Parameter ein beachtliches Vorwissen über die Faktorenstruktur. Darüber hinaus ist die Streubreite von Werten für die Anzahl der Nullladungen von der Größe der Faktormatrix abhängig. Wenn die Faktormatrix viele Elemente aufweist, hat man mehr Freiheitsgrade bei der Parameterwahl als bei wenigen Elementen. Dagegen erscheint die Streubreite der Werte der Parameter bei Promax-, Oblimin- oder Harris-Kaiser-Rotation gering. Insofern stellt die Simplimax-Rotation keine optimale Alternative zu den bestehenden Rotationsverfahren dar. Außerdem liegen in Kiers (1994) bei Simplimax-Rotation auf der Basis von „Harmans 24 psychologischen Variablen“ nur 40 Ladungsbeträge $\leq .10$ und somit weniger als 46, die in Katz und Rohlf (1975) mit Primary Product Functionplane erreicht wurden. Zwar konnten Katz und Rohlf für das 26-Variablen-Box-Problem eine optimale Lösung erreichen, dieser Datensatz ist jedoch mit den üblichen Datensätzen kaum zu vergleichen und stellt sehr spezifische Anforderungen an Rotationsmethoden (Cureton & Mulaik, 1975).

Hyball, eine leistungsfähige Rotationsmethode, die sich ebenfalls an der Maximierung der Anzahl der Variablen in den Hyperebenen orientiert, wurde von Rozeboom (1991a, b) vorgeschlagen. Bei der Hyball-Rotation können viele Parameter frei gewählt werden, um Optimierungen zu erreichen. Die Freiheitsgrade sind bei der Hyball-Rotation sehr groß, die Rotation kann vom Anwender auch auf einzelne Faktorenachsen beschränkt werden. Die Leistungsfähigkeit der Methode zeigt sich beispielsweise daran, dass bei der Rotation von „Harmans 24 psychologischen Variablen“ in einer Variante 47 Ladungsbeträge $\leq .10$ waren, somit noch eine mehr als mit Primary Product Functionplane. Leider wurde die Leistungsfähigkeit von Hyball darüber hinaus nur an einem unpublizierten Datensatz und an einem Datensatz auf der Basis einer von Thurstone (1938) publizierten Korrelationsmatrix untersucht. Da die Extraktionsmethode für die Faktoren von Rozeboom (1991a) nicht beschrieben wurde, kann auch für diesen Datensatz keine vergleichende Bewertung vorgenommen werden.

Beauducel (1997) stellte mit der Trasid-Rotation ebenfalls eine Rotationsmethode vor, die sich an der Maximierung der Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ orientiert. Bei dieser Methode wird nach einer Varimax-Vorrotation die Transformationsmatrix iterativ verändert, wobei immer, wenn eine Zunahme der Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ auftritt, die Lösung als optimale Lösung gespeichert wird. Der einzige Parameter der frei gewählt wird, ist der Ladungsbetrag von dem ab eine Ladung als Nullladung zählt. Dieser

Freiheitsgrad erwies sich als notwendig, da von Cattell (1978) gezeigt wurde, dass bei großen Stichproben kleinere Schwellenwerte für die Nullladungen gewählt werden sollten als bei kleinen Stichproben (die Schwelle sollte sich idealerweise an der Signifikanzgrenze für Ladungen orientieren). Im übrigen enthält das Programm die Option, die Maximierung für die durch die Quadratwurzel aus der Kommunalität dividierten Ladungsbeträge durchzuführen. Diese Option ist vermutlich nur dann interessant, wenn man den Bargmann-Test (1955) für die Signifikanz der Einfachstruktur durchführen möchte. Die Leistungsfähigkeit der Trasid-Rotation wurde von Beauducel (1997) für neun publizierte Datensätze untersucht. Die Trasid-Rotation führte in den meisten Fällen zu einer größeren Anzahl an Ladungsbeträgen $\leq .10$ als die Harris-Kaiser-, Promax- und Oblimin-Rotation. Nur in zwei der neun Datensätze (Thurstones Box-Probleme) wurde bei Trasid-Rotation lediglich eine gleichgroße Anzahl an Ladungsbeträgen $\leq .10$ erreicht, was vermutlich mit der Eindeutigkeit der Datenstrukturen zusammenhängt. Auch für die Datensätze, bei denen Vergleiche mit neueren Rotationsmethoden möglich waren, erwies sich die Trasid-Rotation als den anderen Verfahren in Bezug auf die Anzahl der Variablen in den Hyperebenen als überlegen. Beispielsweise waren bei Trasid-Rotation zur Maximierung der Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ für „Harmans 24 psychologischen Variablen“ 51 Ladungsbeträge $\leq .10$, mehr als bei Hyball und Primary Product Functionplane. Die Trasid-Rotation und Primary Product Functionplane können auch anhand des Datensatzes von Overall und Klett (1972) verglichen werden: Mit Primary Product Functionplane waren bei diesem Datensatz 31 und bei Trasid-Rotation 36 Ladungsbeträge $\leq .10$. Auch Thurstones-Box-Probleme wurden in Bezug auf die Anzahl der Ladungsbeträge $\leq .10$ mit Trasid optimal rotiert, wobei für das 26-Variablen Box-Problem die gewichtete Varimax-Vorrotation von Cureton und Mulaik (1975) gewählt wurde.

Auswahl optimaler Rotationsmethoden

Die Entscheidung, ob oblique oder orthogonal rotiert werden sollte, kann im Rahmen dieser Arbeit folgendermaßen begründet werden: Da oblique Rotationsmethoden die orthogonale Rotation als Spezialfall einschließen, ermöglicht die Verwendung obliquer Rotationsmethoden auch die Prüfung, ob Orthogonalität eintritt. Wenn optimale Einfachstruktur mit einer orthogonalen Lösung erreicht wird, werden auch die obliquen Rotationsmethoden orthogonale Lösungen generieren. Da somit die Annahme der Orthogonalität von Persönlichkeitsfaktoren nur über oblique Rotationstechniken geprüft werden kann, werden im Rahmen dieser Arbeit oblique Rotationsmethoden verwendet. Die konsistente Verwendung orthogonaler Rotationen, wie sie beispielsweise im Rahmen des FFM erfolgt (z.B. Costa & McCrae, 1992; Costa & McCrae, 1995; McCrae & Costa, 1987) setzt eine große Zahl erfolgreicher Überprüfungen der Orthogonalität anhand schiefwinkliger Rotationen voraus. Leider liegen diesbezüglich nur wenige

Studien im Rahmen des FFM vor (z.B. Goldberg, 1990; Ostendorf, 1990) und auch Studien, die die Annahme der Orthogonalität in Frage stellen (z.B. Digman, 1997), was ein weiterer Grund für die Verwendung obliquen Rotationsmethoden ist.

Um innerhalb der Vielzahl obliquen Rotationsmethoden zu einer optimalen Auswahl zu gelangen, werden folgende Kriterien angelegt:

- 1) Die Rotationsmethoden sollten eine möglichst geringe Anzahl frei wählbarer Parameter haben, um den Anteil subjektiver Entscheidungen zu verringern.
- 2) Für die eingesetzten Rotationsmethoden sollte bereits eine Zahl empirischer Bewährungen vorliegen.
- 3) Die ausgewählten Rotationsmethoden sollten Thurstones (1947) Prinzip der Einfachstruktur möglichst gut repräsentieren. Das Prinzip der Maximierung der statistisch nicht bedeutsamen Ladungen (Nullladungen) wird dabei aus oben genannten Gründen gegenüber anderen Prinzipien vorgezogen, u.a. da bei Verwendung dieses Prinzips die wenigsten Zusatzannahmen zu Thurstones ursprünglicher Kennzeichnung erforderlich sind.

Bei einer Anwendung der genannten Kriterien auf die verfügbaren Rotationsverfahren ergibt sich folgendes:

Promax & Oblimin

Im Vergleich von Rotationsmethoden erwiesen sich sowohl Promax als auch direktes Oblimin als brauchbare Rotationsmethoden (Hakstian, 1971). Zum ersten Kriterium: In beiden Methoden ist nur ein Parameter frei wählbar. Zum zweiten Kriterium: Insgesamt zeichnet sich bezüglich der empirischen Bewährung bei diesen Verfahren eine leichte Überlegenheit der Oblimin-Rotation gegenüber der Promax-Rotation bezüglich der Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ ab (Beauducel, 1997; Hakstian & Abell, 1974). Allerdings nur, wenn man davon absieht, dass Oblimin-Lösungen in seltenen Fällen nicht sinnvoll konvergieren. Zum dritten Kriterium: Die Umsetzung des Prinzips der Einfachstruktur erfolgt bei beiden Verfahren über ein abgeleitetes mathematisches Prinzip. Promax beruht in der Regel auf der Varimax-Vorrotation. Für Varimax gilt, dass nicht nur die Zahl der Nullladungen maximiert und die Verteilung der Nullladungen optimiert wird, sondern dass auch die Verteilung der bedeutsamen Ladungen mit dem Ziel der Erhöhung der Ladungsvarianz modifiziert wird. Dieser Aspekt wird bei der Promax-Rotation durch die Potenzierung der Varimax-Ladungen weiter verstärkt, obwohl er nicht direkt aus Thurstones (1947) Kennzeichnungen der Einfachstruktur abgeleitet werden kann. Auch die direkte Oblimin-Rotation basiert auf einem Kriterium, das nur einen

indirekten Bezug zur Einfachstruktur im Sinne von Thurstone aufweist. Bei Oblimin-Rotation werden genau wie bei der Maximierung der Varianzen, teilweise substanzielle Ladungen in ihren Beträgen verringert, obgleich sie dabei statistisch bedeutsam bleiben. Es stellt sich, wie bei den Orthomax-Kriterien, die Frage wie sinnvoll die Reduktion von Ladungsbeträgen ist, wenn es sich in jedem Fall um bedeutsame Ladungen handelt. Zumindest macht Thurstone (1947) nur über Nullladungen Aussagen und es besteht die Gefahr, dass kleinere, bedeutsame Ladungen vom Anwender wie Nullladungen interpretiert werden.

Promaj

Zum ersten Kriterium: Die Promaj-Rotation enthält keinen frei wählbaren Parameter was i.S. des Kriteriums optimal ist. Zweites Kriterium: Die empirische Überprüfung der Promaj-Rotation ist bisher unzureichend. Zum dritten Kriterium: Die Promaj-Rotation (Trendafilov, 1994) wurde nicht in Bezug auf die Maximierung der Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ geprüft und entwickelt, sondern in Bezug auf Bentlers (1977) Index der faktoriellen Einfachheit. Dieser Index bildet jedoch die Einfachstruktur im Sinne Thurstones nur bedingt ab (Beauducel, 1996). Bewertung: Die Promaj-Rotation kommt aufgrund der unzureichenden Erfüllung des zweiten und dritten Kriteriums in der vorliegenden Arbeit nicht zum Einsatz.

Primary Product Functionplane

Zum ersten Kriterium: Es kann ein Parameter gewählt werden, der die Größe der Ladungsbeträge, die als Nullladungen angesehen werden, angibt (Katz & Rohlf, 1974, 1975). Dieser Parameter muss wählbar sein, da er der von der Stichprobengröße abhängigen Signifikanz von Nullladungen (Cattell, 1978) angepaßt werden muss. Zweites Kriterium: Die wenigen empirischen Überprüfungen von Primary Product Functionplane ergaben, dass diese Rotationsmethode zwar brauchbare Ergebnisse erbringt, dass aber die Leistungsfähigkeit etwas geringer ist als die der Trasid-Rotation (Beauducel, 1997). Drittes Kriterium: Mit der Anzahl der Ladungsbeträge $\leq .10$ wird ein Kriterium geprüft, dass mit Thurstones (1947) ursprünglichem Konzept der Einfachstruktur kompatibel ist. Bewertung: Aufgrund des zweiten Kriteriums ist die Trasid-Rotation dem Primary Product Functionplane vorzuziehen, allerdings ist Primary Product Functionplane nach den hier angelegten Kriterien als sehr leistungsfähig einzuschätzen.

Simplimax

Zum ersten Kriterium: Die Simplimax-Rotation von Kiers (1994) hat einen frei zu wählenden Parameter mit einem extrem großen Wertebereich, der von der Größe der Ladungsmatrix

abhängt. Zweites Kriterium: Die Simplimax-Rotation erwies sich in den wenigen empirischen Prüfungen in Bezug auf die Anzahl der Variablen mit Ladungsbeträgen $\leq .10$ eher als anderen Rotationsverfahren unterlegen. Drittes Kriterium: Auch hier steht die Maximierung der Anzahl der Nullladungen im Vordergrund. Bewertung: Die Simplimax-Rotation stellt aufgrund der ersten beiden Kriterien eine suboptimale Variante der Rotation zur Einfachstruktur dar.

Hyball

Erstes Kriterium: Die Anzahl frei wählbarer Parameter ist bei der Hyball-Rotation am größten. Anhand dieser Methode soll in einem interaktiven Prozess eine optimale Lösung erzeugt werden. Zweites Kriterium: Obwohl einige ermutigende Ergebnisse vorliegen (Rozeboom, 1991a) ist die empirische Bewährung der Hyball-Rotation bisher noch unzureichend geprüft. Für den einen Datensatz bei dem ein Vergleich mit Trasid möglich war, erwies sich die Trasid-der Hyball-Rotation als überlegen. Drittes Kriterium: Hyball orientiert sich ebenfalls an der Maximierung der Anzahl der Ladungsbeträge $\leq .10$ und somit an einem klassischen Kriterium der Einfachstruktur. Bewertung: So verdienstvoll der interaktive Ansatz von Rozeboom (1991a, b) auch sein mag, so läuft er doch dem Ziel dieser Arbeit, nämlich der Reduktion von Unverbindlichkeiten bei der Analyse entgegen.

Trasid

Erstes Kriterium: Bei der Trasid-Rotation muss der Schwellenwert für die als nicht bedeutsam anzusehenden Ladungsbeträge (Nullladungen) gesetzt werden. Zweites Kriterium: Die empirische Bewährung wurde an neun publizierten Datensätzen vergleichend geprüft. Dabei war die Trasid-Rotation anderen Methoden in Bezug auf die Anzahl der Ladungsbeträge $\leq .10$ überlegen (Beauducel, 1997). Drittes Kriterium: Mit Trasid wird die Anzahl der Ladungsbeträge $\leq .10$ und somit ein klassisches Kriterium der Einfachstruktur maximiert. Bewertung: Auch die Trasid-Rotation weist einen wählbaren Parameter auf. Allerdings sind die Freiheitsgrade bei der Wahl begrenzt, da sie sich an der Signifikanz der Nullladungen bei gegebener Gruppengröße orientiert. Insgesamt spricht sowohl die empirische Bewährung als auch das maximierte Kriterium für die Verwendung dieser Methode im Rahmen dieser Arbeit.

Die dargestellten Bewertungen der Rotationsmethoden in Bezug auf die drei Kriterien wurden in Tabelle 3.8 zusammengefasst. Es zeigt sich, dass auf der Basis dieser Kriterien die Verwendung der Trasid-Rotation begründet erscheint. An dem in Tabelle 3.8 dargestellten Überblick zeigt sich darüber hinaus, dass in Bezug auf die genannten Kriterien die Functionplane, bzw. Primary Pattern Functionplane-Rotation der Trasid-Rotation am nächsten kommt. Lediglich die empirische Überlegenheit beim direkten Vergleich an zwei Datensätzen

und die etwas breitere empirische Überprüfung der Trasid-Rotation führen zum Vorzug der Trasid-Rotation.

Um die Freiheitsgrade bei den Analysen zu reduzieren, orientiert sich die Festlegung des Schwellenwertes für Nullladungen an der Signifikanzgrenze der Ladungen. Cattell (1978) schlägt vor, bei Datensätzen mit über 300 Probanden Ladungsbeträge $\leq .05$ als Schwelle für Nullladungen anzusehen. Pawlik (1968) nennt (ebenfalls mit Bezug auf Cattell) diese Schwelle für Probandengruppen mit 500 und mehr Fällen. Da in der vorliegenden Arbeit Datensätze mit über 500 Probanden analysiert werden sollen, werden hier Ladungsbeträge $\leq .05$ als Nullladungen angesehen, so dass dieser Schwellenwert bei der Trasid-Rotation verwendet wird.

Tabelle 3.8

Vergleichende Bewertung der obliquen Rotationsmethoden

Rotationsmethode	wählbare Parameter	Empirische Bewährung	Kriterium der Einfachstruktur
Promax	+/-	+/-	+/-
Oblimin	+/-	+/-	+/-
Promaj	+	-	+/-
Functionplane	+/-	+/-	+
Simplimax	-	-	+
Hypall	-	+/-	+
Trasid	+/-	+	+

Da die Trasid-Rotation jedoch bisher selten eingesetzt wurde, erscheint die zusätzliche Verwendung einer verbreiteteren Rotationsmethode zu Vergleichszwecken sinnvoll. Da die direkte Oblimin-Rotation zu den verbreitetsten Rotationsverfahren zählt und sich in empirischen Überprüfungen recht gut bewährt hat (Beauducel, 1997; Hakstian & Abell, 1974), erscheint die direkte Oblimin-Rotation als Grundlage für eine vergleichende Untersuchung zweckmäßig. Gegenüber der Promax-Rotation hat die direkte Oblimin-Rotation den Vorteil, dass bei der Einstellung von null für den wählbaren Parameter (γ oder Δ), das Oblimin-Kriterium eine besonders einfache Struktur hat: In diesem Fall erhält man das direkte Quartimin-Kriterium, d.h. es erfolgt eine direkte Minimierung der Kreuzprodukte der Ladungen im Faktorenmuster. Außerdem konnte gezeigt werden, dass bei dieser Einstellung eine perfekte Einfachstruktur in jedem Fall erreicht wird, wenn dies mit den Daten möglich ist (s. Hakstian & Abell, 1974). Die

hier geplanten direkten Oblimin-Rotationen sollen ausschließlich mit $\Delta=0$ durchgeführt werden, so dass die subjektiven Freiheitsgrade auch bei diesen Analysen deutlich reduziert werden. Darüber hinaus ist $\Delta=0$ die Defaulteinstellung bei den meisten Statistik-Programm-Paketen (z.B. SPSS für Windows 9.0) und wurde vermutlich auch am häufigsten in empirischen Analysen verwendet, so dass die Vergleichbarkeit mit anderen Lösungen auf der Basis der Oblimin-Rotationen mit $\Delta=0$ am ehesten hergestellt werden kann. Im übrigen wird durch die Einbeziehung der Oblimin-Rotation der Vergleich zwischen einem analytischen Rotationskriterium und einer primär an der Maximierung der Nullladungen orientierten Rotation möglich.

3.2.2.4 Probanden- und Variablenanzahl in exploratorischen Faktorenanalysen

Gorsuch (1983) empfiehlt, Faktorenanalysen mit mindestens 100 Probanden durchzuführen. Diese grobe Faustregel muss allerdings durch folgende Informationen ergänzt werden: Cliff (1970) stellte in einer Simulationsstudie fest, dass von vier Faktoren, die in einer Stichprobe mit 600 Fällen nachweisbar waren, nur zwei bis drei Faktoren in einer Stichprobe mit 200 Fällen nachweisbar waren. Beauducel (2001a) stellte in Simulationsstudien fest, dass bei einer Stichprobengröße von 200 Fällen teilweise sehr große Streuungen bei den geschätzten Faktorladungen auftraten, was auf eine Instabilität der Lösungen hindeutet.

Weitere Faustregeln beziehen sich auf das Verhältnis der Anzahl der Probanden zur Anzahl der Variablen. Gorsuch (1983) empfiehlt, dass mindestens fünf Probanden pro Variable untersucht werden sollten. Nunnally (1978) empfiehlt, dass sogar 10 Probanden pro Variable untersucht werden sollten. Die Ergebnisse von Barrett und Kline (1981) legen nahe, dass die Anzahl der Probanden pro Variable allerdings weniger entscheidend ist, wenn die Anzahl untersuchter Probanden insgesamt groß ist. Darüber hinaus stellten Velicer und Fava (1998) in ihren Simulationsstudien fest, dass die Regeln, nach denen die Anzahl der Fälle mit der Anzahl der Variablen verknüpft werden (z.B. mindestens 5 Fälle pro Variable), unzutreffend sind. Sie stellten im Gegensatz zu diesen Regeln fest, dass eine Erhöhung der Variablenanzahl bei konstanter Faktoren- und Probandenzahl zu einer Optimierung der Lösungen führt. Velicer und Fava (1998) fassen ihre Ergebnisse dahingehend zusammen, dass Schwächen in einem der drei Bereiche Variablenanzahl, Probandenzahl und Ladungshöhe durch Stärken in einem anderen der drei Bereiche kompensiert werden können.

In Bezug auf die Variablenanzahl legen die Ergebnisse von Velicer und Fava (1998) nahe, dass es vorteilhaft ist, eine große Anzahl von Variablen pro Faktor zu analysieren. Dabei darf die Anzahl der Variablen jedoch nicht die Anzahl der Probanden übersteigen. Sie kommen zu dem Ergebnis, dass ein Minimum von drei Variablen pro Faktor unzureichend ist und empfehlen, dass mindestens fünf Variablen pro Faktor analysiert werden sollten.

Da im vorliegenden Kontext der vergleichenden Untersuchung verschiedener Traitmodelle die Anzahl der untersuchten Variablen eher groß ist, wird die Anzahl der Variablen pro Faktor vermutlich weniger zum Problem werden. Für die geplanten Analysen folgt jedoch, dass mindestens fünf Markiertvariablen für jeden postulierten Faktor eines Modells verfügbar sein müssen. Darüber hinaus folgt aus den Darstellungen, dass mindestens so viele Probanden untersucht werden müssen, wie Variablen analysiert werden sollen. Das bedeutet, dass im vorliegenden Kontext, in dem die Inventare für einzelne Modelle zwischen 100 und 241 Items haben, mindestens 500 Probanden untersucht werden müssen, um genügend große Variablengruppen modellübergreifend analysieren zu können.

3.2.2.5 Faktorenanalyse binärer Daten

Viele Persönlichkeitsfragebogen und Modelle basieren auf binären Items (z.B. der Eysenck-Personality-Questionnaire Revised, EPQ-R; Eysenck & Eysenck, 1991; Ruch, 1999), so dass eine Auseinandersetzung auch mit dem Einfluss binärer Daten auf die Ergebnisse von Faktorenanalysen erforderlich ist. Bereits Ferguson (1941) stellte fest, dass bei eindimensionalen Tests auf der Basis binärer Daten die Matrix der Phi-Koeffizienten einen Rang größer als eins hat, wenn die Items bezüglich ihrer Schwierigkeit stark streuen. Solche zusätzlichen, nur durch die Streuung der Itemschwierigkeiten⁴, bedingten Faktoren werden auch als Schwierigkeitsfaktoren (z.B. Gourlay, 1951) oder allgemeiner als verteilungsbedingte Faktoren (Corballis, 1968) bezeichnet. Für das Problem der verteilungsbedingten Faktoren wurden bereits viele Lösungsvorschläge unterbreitet (z.B. Mislavy, 1986), wobei keiner der Vorschläge sich bisher allgemein durchgesetzt hat. Es sei auch festgehalten, dass es nicht grundsätzlich sinnlos sein muss, wenn Faktoren auch Aspekte der Verteilungen, der Auftretenshäufigkeiten von Merkmalen repräsentieren. Dabei ist das Auftreten verteilungsbedingter Faktoren nicht an die Analyse binärer Daten gebunden, es ist bei derartigen Daten nur besonders wahrscheinlich,

⁴ Im Kontext von Trait-Items bezeichnet man diese auch als Popularitäten. Da die Debatte zu Schwierigkeitsfaktoren jedoch im Kontext Leistungsmessungen entstanden ist, wird hier im weiteren der Terminus „Schwierigkeit“ i.S. von Popularität verwendet.

da sich bei binären Daten jede Abweichung vom Mittelwert direkt in einer veränderten Verteilung und damit auch in veränderter Varianz niederschlägt. Weil verteilungsbedingte Faktoren meist unerwünscht sind, und die Interpretation erschweren, werden zunächst die Möglichkeiten gesichtet, derartige Faktoren zu vermeiden.

Wahl eines angemessenen Korrelationskoeffizienten

Verteilungsbedingte Faktoren können möglicherweise durch die Wahl eines angemessenen Korrelationskoeffizienten vermieden werden. Bereits Wherry und Gaylord (1944) stellten fest, dass die Entstehung von Schwierigkeitsfaktoren mit der Verwendung des Phi-Koeffizienten zusammenhängt und empfahlen die Verwendung des tetrachorischen Korrelationskoeffizienten bei binären Daten. Allerdings stellte Gourlay (1951) auch Schwierigkeitsfaktoren bei der Analyse von tetrachorischen Korrelationen fest.

In der Folge wurde eine Reihe von Korrelationskoeffizienten vorgeschlagen, anhand derer Schwierigkeitsfaktoren vermieden werden sollen. So haben Jäger (1976), Gebert (1977) und Boesser (1979) versucht, den Einfluss der Schwierigkeiten auf den Phi-Koeffizienten durch Korrekturformeln zu vermeiden. Eine Zusammenfassung und Bewertung der Entwicklungen findet sich in Zysno (1997). Die einfachste Korrekturformel ist dabei die von Jäger (1976), bei der die Phi-Koeffizienten durch ihren möglichen Maximalwert dividiert werden. Lienert und Sporer (1982) schlagen die Verwendung des nullfeld-korrigierten Yule-Q-Koeffizienten vor. Dieser Koeffizient soll gegenüber Boessers (1979) korrigiertem Phi-Koeffizienten den Vorteil haben, dass er den Zusammenhang selten auftretender Symptome nicht unterschätzt. Andererseits können bei Verwendung dieses Koeffizienten in Faktorenanalysen Ladungen nicht mehr als Korrelationen interpretiert werden, was die Vergleichbarkeit mit publizierten Befunden erschwert. Die nullfeld-korrigierten Yule-Q-Koeffizienten sind wegen Problemen der Vergleichbarkeit wenig für die Faktorenanalysen geeignet, während die tetrachorischen Koeffizienten Schwierigkeitsfaktoren nicht sicher vermeiden und wie die korrigierten Phi-Koeffizienten mitunter keine optimalen Korrelationsmatrizen als Grundlage für die Faktorenanalyse ergeben (Revenstorff, 1980). Gorsuch (1983) ist der Auffassung, dass Phi- und Punktbiseriale Korrelationskoeffizienten eine bessere Grundlage für Faktorenanalysen darstellen als tetrachorische und korrigierte Phi-Koeffizienten. Im übrigen ist Gorsuch (1983) der Ansicht, dass Phi-, Punktbiseriale- und Produkt-Moment-Korrelationskoeffizienten problemlos in einer Analyse gemischt werden können.

Image-Analyse

Kaiser (1970) schlug vor, verteilungsbedingte Faktoren in binären Daten durch die Verwendung der Image-Analyse zu vermeiden. Da das Image einer Variable sich aus der Prädiktion durch die jeweils übrigen Variablen ergibt, gehen spezifische Varianzanteile, die sich durch extreme Verteilungen ergeben, weniger in die Analysen ein. Somit werden verteilungsbedingte Faktoren unwahrscheinlicher. Allerdings erbrachte die Image-Analyse keine günstigen Ergebnisse bezüglich der Abhängigkeit faktorenanalytischer Ergebnisse vom Variablenkontext (siehe Kapitel 3.2.2.2).

Nichtlineare Faktorenanalyse

McDonald und Ahlawat (1974) konnten zeigen, dass Schwierigkeitsfaktoren mit nichtlinearen Zusammenhängen zwischen den Variablen in Verbindung stehen. Darüber hinaus vertreten sie die Auffassung, dass Schwierigkeitsfaktoren nicht notwendig Artefakte darstellen müssen und dass anhand der linearen Faktorenanalyse nicht ermittelt werden kann, ob es sich um Artefakte handelt oder nicht. McDonald und Ahlawat (1974) stellten darüber hinaus fest, dass die Annahme, dass Artefakt-Faktoren durch eine Abminderung der Phi-Koeffizienten bei Schwierigkeitsunterschieden zwischen Items entstehen würden, unzutreffend sei. Damit die Annahme zutreffend sei, muss nach McDonald und Ahlawat (1974) zusätzlich gelten, dass die Daten dem Normal-Ogiven-Modell entsprechen (für ein Anwendungsbeispiel mit Persönlichkeitstraits siehe Waller, Tellegen, McDonald & Lykken, 1996).

Vermeidung binärer Daten durch „radiale Bündel“ oder „Miniskalen“

Eine weitere Möglichkeit die Verteilungen für die Faktorenanalysen zu optimieren, besteht in der Verwendung radialer Bündel („radial parcels“) zur Voraggregation der Daten. Durch eine Voraggregation könnte man binäre Daten bereits im Vorfeld vermeiden. Zugleich würden die Analysen nicht auf die meist sehr unreliablen Einzelitems gestützt. Cattell (1994b) verwendete diese Methode folgendermaßen: Zunächst wurde eine Hauptkomponentenanalyse der Items zur Bestimmung der Anzahl zu rotierender Faktoren durchgeführt. Nachdem Cattell (1994b) sich für eine bestimmte Faktorenzahl entschieden hatte, berechnete er den Kosinus der Winkel zwischen allen Variablenpaaren in dieser Lösung ($\cos \alpha_{ab} = r_{ab}/h_a h_b$). Dann wurden die Items mit dem größten Kosinus paarweise zu Bündeln zusammengefaßt. Cattell betont, dass das Zusammenfassen anhand des Kosinus der Winkel zu deutlich anderen Ergebnissen führt als das Zusammenfassen von Variablen auf der Basis von Korrelationen. Dann berechnete er den mittleren Vektor für jedes Bündel und faßte auf Basis des Kosinus zwischen den mittleren Vektoren der Bündel die Bündel erneut paarweise zusammen. Auf diese Weise gelangte er zu Bündeln, die sich aus vier Variablen zusammensetzen und verwendete diese als Grundlage für

anschließende Faktorenanalysen. Barrett und Kline (1982) weisen darauf hin, dass es auch möglich ist, eine erneute Faktorenanalyse bereits anhand der Bündel, die sich aus dem Zusammenfassen von zwei Items ergeben, durchzuführen.

Gegen die Methode kann eingewendet werden, dass sie bereits eine optimale Entscheidung über die Anzahl zu rotierender Faktoren voraussetzt und dass verteilungsbedingt bei der Ermittlung der Anzahl der Faktoren in einer Matrix bereits Fehler auftreten können. Dem Argument könnte dadurch begegnet werden, dass die Faktorenanalyse anhand einer angemesseneren Korrelationsmatrix durchgeführt wird. Somit stellt sich dann wieder die Frage nach der Auswahl eines angemessenen Korrelationskoeffizienten. Andererseits besteht die Möglichkeit, dass durch die reliableren Bündel und besseren Verteilungen Primärfaktoren angemessener identifiziert werden können als bei der Analyse von Einzelitems.

Gorsuch (1983) schlug vor, Miniskalen zu bilden, beispielsweise, indem die beiden Items mit der höchsten Schwierigkeit mit den beiden Items mit der geringsten Schwierigkeit zusammengefasst werden. Die Aggregate haben dann deutlich geringere Schwierigkeitsunterschiede. Im Gegensatz zur Methode der radialen Bündelung werden bei dieser Methode die Schwierigkeitsunterschiede direkt minimiert. Andererseits kann diese Methode nur dann eingesetzt werden, wenn die Binnenstruktur einer Itemmenge bereits weitgehend bekannt ist, da bei beliebigem Zusammenfassen von Items verschiedener Subskalen keine sinnvollen Ergebnisse zu erwarten sind. Insofern eignet sich diese Methode nicht für eine induktive, wohl aber für eine deduktive Forschungsstrategie.

Diskussion der dargestellten Vorschläge

Die Verwendung neuerer alternativer faktorenanalytischer Methoden wie der nicht-linearen Faktorenanalyse (z.B. McDonald & Ahlawat, 1974) und anderer Varianten (z.B. Mislavy, 1986) eröffnet der multivariaten Forschung neue Möglichkeiten. Diese neuen Möglichkeiten sind zwar grundsätzlich zu begrüßen, führen aber zunächst nicht zu einer Erhöhung der Verbindlichkeit der psychometrischen Traitforschung, da zum einen für die neuen Methoden neue Anwendungsstandards (z.B. bezüglich der Extraktionskriterien) entwickelt werden müssen und zum anderen die alternativen Methoden in der Regel auch alternative Formen der Modellbildung nach sich ziehen müssten. Man denke beispielsweise an die inhaltliche Interpretation nicht-linearer Effekte in Faktorenanalysen. Aus diesem Grund werden hier Lösungsansätze bevorzugt, die enger an der bestehenden Forschungspraxis orientiert sind. Ein pragmatisches Vorgehen findet man beispielsweise in Ruch (1999), der für die P-Skala des EPQ-R eine große

Schwierigkeitsstreuung feststellt und daher für diese Skala eine Analyse anhand der nach Jäger (1976) korrigierten Phi-Koeffizienten durchführt.

Seit langem werden auch Faktorenanalysen binärer Daten im Traitbereich ohne besondere Vorkehrungen wie korrigierte Phi-Koeffizienten, tetrachorische Korrelationen oder Bündelungsmethode durchgeführt (z.B. Corulla, 1988; Eysenck, Daum, Schugens & Diehl, 1990; Eysenck, Eysenck & Barrett, 1985; Zuckerman, Eysenck & Eysenck, 1978). Insofern erscheint es zweckmäßig, den Einfluss binärer Daten auf die Ergebnisse bei herkömmlichem Vorgehen genauer zu untersuchen, bevor neue Varianten der Faktorenanalyse entwickelt werden.

Bei der Diskussion zum Auftreten verteilungsbedingter Faktoren wurde bisher kaum berücksichtigt, dass die Ermittlung der Anzahl zu extrahierender bzw. zu rotierender Faktoren anhand der unterschiedlichsten Methoden erfolgen kann (s. Kapitel 3.2.2.1). Selbst wenn einige Eigenwerte durch eine große Streuung der Schwierigkeiten größer werden, muss das nicht notwendig zur Extraktion entsprechender Faktoren führen. Bernstein und Teng (1989) stellten beispielsweise in einer Simulationsstudie fest, dass bei der Verwendung der Eigenwerte > 1 Regel eher verteilungsbedingte Faktoren extrahiert werden, als bei Verwendung des Scree-Tests. Ausgehend von diesem Ergebnis kann vermutet werden, dass unterschiedliche Extraktionskriterien und -methoden auf Schwierigkeitsunterschiede zwischen Items unterschiedlich reagieren. Daher sollte geprüft werden, wie groß die Unterschiede bezüglich der Anfälligkeit für verteilungsbedingte Faktoren bei den verschiedenen Extraktionskriterien im Rahmen der konventionellen Faktorenanalyse sind. Der Fokus liegt auf den Extraktionskriterien und -methoden, da die Rotationsmethoden nur die Umverteilung ggf. verteilungsbedingter Ladungsmuster bewirken können und somit hier weniger relevant sind.

Da bezüglich der Extraktionskriterien und -methoden in den Kapiteln 3.2.2.1 und 3.2.2.2 bereits eine begründete Auswahl erfolgte, erscheint hier eine Überprüfung des Einflusses von Schwierigkeitsunterschieden auf die Lösungen für die ausgewählten Methoden besonders relevant. Aus diesem Grund wird im folgenden der Einfluss der Schwierigkeitsunterschiede auf die Ergebnisse der Parallelanalyse untersucht. Die Untersuchung erfolgte in Anlehnung an Bernstein und Teng (1989) anhand einer Simulationsstudie, bei der große Schwierigkeitsunterschiede zwischen Items generiert wurden.

Simulationen zur Problematik von Schwierigkeitsfaktoren

Im folgenden wurde der Einfluss von Schwierigkeitsunterschieden bei binären Daten auf die Extraktion von Faktoren für die Parallelanalyse bei verschiedenen Extraktionsmethoden untersucht.

Methode der Datengenerierung

Mit Hilfe von SPSS für Windows 9.0 (1999) wurden zweifaktorielle Lösungen mit eindeutiger Einfachstruktur generiert. Es wurden zweifaktorielle Lösungen erzeugt, um den Einfluss der Schwierigkeitsunterschiede in einer gut überschaubaren Datenlage zu prüfen. Ein Minimum von zwei Faktoren war nötig, da auch geprüft werden sollte, ob sich Schwierigkeitsvariationen in den Items eines Faktors auch auf Faktoren ohne nennenswerte Schwierigkeitsvariationen auswirken. Außerdem mussten genügend viele Items pro Faktor generiert werden, um eine deutliche Schwierigkeitsvariation abbilden zu können. Daher wurden hier acht Markiertvariablen pro Faktor generiert. Die Markiertvariablen wurden – wie in Kapitel 3.2.2.2 – zunächst durch gezielte Aggregation normalverteilter, z-transformierter Zufallsvariablen erzeugt. Die Formel für die Aggregation der ersten acht Variablen V_1 bis V_8 lautet:

$$V_i = \frac{1}{\sqrt{3}} c_1 + \frac{\sqrt{2}}{\sqrt{3}} s_i, \quad \text{for } i = 1 \text{ to } 8 \quad (5)$$

Die Variable c_1 ist den Variablen V_1 bis V_8 gemeinsam, die Variablen s_1 bis s_8 stellen die jeweils spezifischen Varianzanteile der Variablen V_1 bis V_8 dar (diese sind auch hier stets unkorreliert). Dieselben Gewichte wurden auch für die Variablen V_9 bis V_{16} verwendet, nur dass die gemeinsame Variable c_1 durch die gemeinsame Variable c_2 ausgetauscht wurde. Die Variablen V_9 bis V_{16} markieren daher den zweiten Faktor.

Um zum einen binäre Daten und zum anderen Schwierigkeitsvariationen zu erzeugen, wurden die 16 normalverteilten Variablen in einem zweiten Schritt in die binären Variablen B_1 bis B_{16} umkodiert. Für die ersten acht Variablen erfolgte die Umkodierung so, dass eine große Schwierigkeitsvariation in den Items entstand. Für die Variablen V_9 bis V_{16} wurde die Umkodierung so durchgeführt, dass sich für B_9 bis B_{16} eine mittlere Schwierigkeit von .50 ergab. Auf diese Weise wurden jeweils 100 Datensätze mit 200 und mit 1000 Fällen als Grundlage für die Faktorenanalysen generiert.

Ergebnisse

Die mittleren Itemschwierigkeiten, die sich in den 100 Datensätzen nach der Umkodierung in binäre Daten ergaben, sind in Tabelle 3.9 dargestellt. Die Items B_1 , B_2 , B_7 und B_8 haben

extreme Schwierigkeiten, die bereits zu deutlichen Varianzeinschränkungen führen. Auch könnten die Variablen B_1 bis B_3 einen Faktor für leichte und die Variablen B_6 bis B_8 einen Faktor für schwierige Items bilden. Die Möglichkeiten für das Auftreten von Schwierigkeitsfaktoren sind in den vorliegenden Daten somit gegeben.

Die mittlere Interkorrelation (Phi) der binären Items B_1 bis B_8 beträgt etwa .16, die der Items B_9 bis B_{16} beträgt etwa .21, die Korrelationen zwischen den Items, die jeweils einen Faktor konstituieren sollen, sind somit eher gering. Daher sollte der Einfluss verteilungsbedingter Faktoren gegenüber dem der beiden intendierten Faktoren relativ groß sein.

Tabelle 3.9
Mittlere Schwierigkeiten und Standardabweichungen der umkodierten Items

Item	Schwierigkeit	Standard- abweichung
B_1	.87	.33
B_2	.81	.35
B_3	.71	.45
B_4	.62	.49
B_5	.38	.49
B_6	.28	.45
B_7	.19	.39
B_8	.12	.33
B_{9-16}	.50	.50

Als nächstes werden die Eigenwerte der unrotierten zweifaktoriellen Lösungen, für Hauptkomponentenanalyse, Hauptachsenanalyse, Alpha-, Maximum-Likelihood- und Image-Analyse dargestellt. Für die fünf Extraktionsmethoden werden darüber hinaus die entsprechenden Eigenwerte auf der Basis von 16 binären, gleichverteilten Zufallsvariablen berichtet (siehe Tabelle 3.10).

Für die Hauptkomponentenanalyse sind sowohl bei Analysen mit 200 als auch mit 1000 Fällen die ersten beiden mittleren Eigenwerte der zweifaktoriellen Lösung größer, die übrigen Eigenwerte eindeutig kleiner als die der Zufallslösung. Allerdings war in 2 der 100 Simulationen für 200 Fälle der dritte Eigenwert der zweifaktoriellen Lösung größer als der mittlere Zufallseigenwert. In keiner der 100 Simulationen mit 1000 Fällen war der dritte Eigenwert größer als der Zufallseigenwert. Die Parallelanalyse deutet somit trotz großer Schwierigkeitsstreuung auf die Extraktion von zwei Faktoren hin. Bei Stichproben um 200 Probanden könnten sich allerdings in seltenen Fällen Ungenauigkeiten ergeben.

Tabelle 3.10

Mittlere Eigenwerte der 100 unrotierten Lösungen

1000 Fälle										
Fak- tor	HKA- 2 Fakt	HKA- Zufall	FA- 2 Fakt	FA- Zufall	AL- 2 Fakt	AL- Zufall	ML- 2 Fakt	ML- Zufall	IM- 2 Fakt	IM- Zufall
1	2.54	1.22	1.89	.42	1.86	.38	1.54	1.07	1.14	.05
2	2.17	1.18	1.48	.37	1.51	.36	1.25	1.04	.76	.03
3	1.01	1.14	.33	.33	.32	.33	1.10	1.01	.02	.02
4	.95	1.11	.27	.29	.27	.30	1.01	.98	.01	.01
5	.91	1.08	.24	.26	.24	.27	.88	.88	.00	.01
6	.88	1.06	.20	.23	.21	.25	.65	.82	.00	.00
7	.86	1.03	.17	.20	.18	.22	.54	.73	.00	.00
8	.83	1.01	.15	.17	.15	.19	.39	.64	.00	.00
9	.81	.99	.13	.15	.13	.16	.32	.59	.00	.00
10	.79	.96	.10	.12	.11	.13	.28	.51	.01	.00
200 Fälle										
Fak- tor	HKA- 2 Fakt	HKA- Zufall	FA- 2 Fakt	FA- Zufall	AL- 2 Fakt	AL- Zufall	ML- 2 Fakt	ML- Zufall	IM- 2 Fakt	IM- Zufall
1	2.62	1.52	2.13	.93	2.09	.87	1.47	1.08	1.29	.25
2	2.16	1.40	1.63	.80	1.67	.77	1.36	1.04	.83	.17
3	1.20	1.32	.70	.71	.68	.71	1.26	.97	.13	.12
4	1.11	1.24	.59	.63	.59	.65	1.12	.88	.07	.08
5	1.04	1.18	.51	.56	.53	.59	.95	.80	.04	.05
6	.97	1.12	.44	.49	.46	.52	.78	.74	.02	.03
7	.91	1.06	.38	.43	.39	.47	.62	.66	.01	.02
8	.86	1.01	.31	.37	.33	.40	.52	.59	.00	.01
9	.81	.96	.27	.31	.28	.33	.44	.53	.00	.00
10	.76	.90	.22	.25	.23	.27	.39	.46	.01	.00

Anmerkung. Es wurden nur die ersten 10 Eigenwerte dargestellt. HKA=Hauptkomponentenanalyse; FA=Hauptachsenanalyse; AL= Alpha Faktorenanalyse; ML= Maximum Likelihood Faktorenanalyse; IM= Image Analyse; 2 Fakt = 2 Faktoren-Lösung; Zufall= Eigenwerte auf der Basis von Zufallsvariablen.

Anhand der Eigenwerte > 1 Regel würde man in 100% der Analysen für 200 Fälle und in 39% der Analysen für 1000 Fälle drei Faktoren extrahieren, was darauf hindeutet, dass diese Regel eher dazu führt, dass verteilungsbedingte Faktoren extrahiert werden. Für die Hauptachsenanalyse, Alpha- und Image-Analyse ist der dritte mittlere Eigenwert etwa so groß wie der Zufallseigenwert, was darauf hindeutet, dass in etwa 50% der Fälle ein verteilungsbedingter Faktor extrahiert würde. Allerdings ist bei der Image-Analyse der Eigenwert des dritten Faktors der zweifaktoriellen Lösung so klein, dass man dennoch lediglich zwei Faktoren extrahieren würde. Anhand der Maximum-Likelihood Faktorenanalyse würde man in den meisten Fällen vier oder mehr Faktoren und in einigen Fällen drei Faktoren extrahieren. Bei einer Parallelanalyse auf Basis der Eigenwerte der Maximum-Likelihood-Faktorenanalyse bilden sich verteilungsbedingte Faktoren somit am deutlichsten ab.

Die Ergebnisse sprechen dafür, die Parallelanalyse auf der Basis der Eigenwerte der Hauptkomponentenanalyse mit mehr als 200 Fällen durchzuführen. Als nächstes wurde geprüft, ob im Ladungsmuster der zweifaktoriellen Lösungen ein problematischer Einfluss der Schwierigkeitsunterschiede feststellbar ist.

In Tabelle 3.11 sind die mittleren Ladungen der 100 zweifaktoriellen Lösungen auf Basis von 200 Fällen für die Hauptkomponentenanalyse und die Hauptachsenanalyse dargestellt. Es zeigt sich, dass die Faktoren die beiden Variablengruppen eindeutig repräsentieren. Die Streuung der Schwierigkeiten führt dazu, dass die Ladungen des zweiten Faktors geringfügig streuen. Zum einen sind die mittleren Ladungen für die Items mittlerer Schwierigkeit am größten, zum anderen sind die Standardabweichungen der Ladungen über die 100 Lösungen für den zweiten Faktor größer als für den ersten Faktor. Diese Effekte beeinträchtigen in diesem Fall allerdings nicht die Interpretation der Lösungen.

Tabelle 3.11

Mittelwerte und Standardabweichungen der Varimax-Ladungen der 100 zweifaktoriellen Lösungen auf der Basis von 200 Fällen

Variable	HKA		FA	
	Faktor 1	Faktor 2	Faktor 1	Faktor 2
B1	-.01 (.04)	.46 (.11)	-.01 (.03)	.35 (.10)
B2	-.01 (.04)	.50 (.10)	-.01 (.04)	.39 (.09)
B3	.01 (.04)	.54 (.10)	.01 (.03)	.44 (.09)
B4	.01 (.04)	.55 (.09)	.00 (.04)	.45 (.09)
B5	.00 (.04)	.55 (.10)	.00 (.03)	.44 (.10)
B6	.01 (.05)	.55 (.09)	.01 (.04)	.44 (.09)
B7	-.01 (.04)	.51 (.11)	-.01 (.03)	.40 (.10)
B8	-.01 (.04)	.49 (.10)	-.01 (.03)	.35 (.09)
B9	.56 (.04)	.01 (.10)	.47 (.04)	.01 (.08)
B10	.55 (.04)	-.02 (.10)	.46 (.04)	-.01 (.08)
B11	.56 (.04)	-.01 (.09)	.46 (.04)	-.01 (.07)
B12	.57 (.04)	.01 (.08)	.48 (.04)	.01 (.07)
B13	.55 (.04)	.01 (.09)	.46 (.04)	.01 (.07)
B14	.55 (.05)	-.01 (.09)	.46 (.05)	.00 (.08)
B15	.56 (.04)	-.01 (.10)	.46 (.04)	-.01 (.09)
B16	.56 (.05)	-.01 (.09)	.47 (.05)	-.01 (.08)

Anmerkungen. Ladungen > .30 wurden fett gedruckt. HKA=Hauptkomponentenanalyse; FA= Hauptachsenanalyse. Da die Varimax-Ladungen Korrelationen sind, wurden sie vor der Mittelung Fisher-Z transformiert. Dann wurden die Mittelwerte der Fisher-Z transformierten Ladungen zurücktransformiert.

Diskussion

Die vorliegende Simulationsstudie ist nicht so umfassend, dass bereits weitreichende Konsequenzen gezogen werden können. Allerdings konnte anhand der vorliegenden Studie klar gezeigt werden, dass verteilungsbedingte Faktoren, auch unter Bedingungen, die sie begünstigen, bei der Verwendung bestimmter Extraktionskriterien und -methoden nicht auftreten müssen. Verteilungsbedingte Faktoren entstehen nicht automatisch bei Faktorenanalysen binärer Items mit großen Schwierigkeitsstreuungen. Da die Debatte zu verteilungsbedingten Faktoren bisher nicht die Ergebnisse der Debatte zu angemessenen Extraktionskriterien berücksichtigt hat, sind die vorliegenden Ergebnisse zu dieser Frage relevant.

Die Ergebnisse der Simulationsstudie sprechen dafür, dass die Parallelanalyse am besten auf der Basis der unrotierten Eigenwerte einer Hauptkomponentenanalyse durchgeführt wird, selbst wenn im Anschluss an die Entscheidung für eine bestimmte Faktorenanzahl eine Hauptachsenanalyse durchgeführt wird. Da die simulierte Streuung der Schwierigkeiten groß war und die mittleren Interkorrelationen zwischen den Variablen, die einen Faktor konstituieren, in einem auch für Trait-Items niedrigen Bereich von .16 bis .21 lagen, begünstigte die hier durchgeführte Simulation das Auftreten verteilungsbedingter Faktoren. Das Ergebnis, dass – bei genügend großer Stichprobe – dennoch anhand der Parallelanalyse auf der Basis der Hauptkomponentenanalyse keine verteilungsbedingten Faktoren extrahiert wurden, ist ein Hinweis darauf, dass die Verwendung der Parallelanalyse die Wahrscheinlichkeit der Extraktion verteilungsbedingter Faktoren verringert. Natürlich wären zur weiteren Klärung dieser Frage weitere Simulationsstudien mit anderen Variablen und Faktorenzahlen erforderlich, diese können im Rahmen dieser Arbeit jedoch nicht durchgeführt werden.

Zusammenfassung zur Faktorenanalyse binärer Daten

Insgesamt erbrachte die Auseinandersetzung mit der Faktorenanalyse binärer Daten eine weitere Bestätigung der ausgewählten Extraktionskriterien und der Empfehlung große Stichproben zu verwenden. Da auch Gorsuch (1983) Phi-Koeffizienten gegenüber tetrachorischen Korrelationen und korrigierten Phi-Koeffizienten empfiehlt, ergeben sich für die Durchführung der exploratorischen Faktorenanalysen hier keine weiteren Modifikationen.

Wenn die überwiegende Zahl der Items eine mittlere Schwierigkeit aufweist, was beim Einsatz elaborierter Verfahren im Trait-Bereich zu erwarten ist, werden die Phi-Koeffizienten von Items mit extremen Häufigkeitsverteilungen eher gering ausfallen. Es könnte als Vorteil angesehen werden, dass Phi-Koeffizienten auf diese Weise zu einer Gewichtung der Korrelationen zugunsten von Items mit mittlerer Schwierigkeit führen.

3.2.2.6 Liste der begründeten Spezifikationen für die exploratorischen Faktorenanalysen

Für die modellspezifischen Analysen einzelner Fragebogen (siehe Kapitel 6.2) ergibt sich zusammenfassend folgendes: Zunächst wird eine Hauptkomponentenanalyse zur Ermittlung der Eigenwerte durchgeführt. Dann erfolgt Anhand der Parallelanalyse und der Komparabilitäten die Entscheidung über die Anzahl der zu extrahierenden bzw. zu rotierenden Faktoren. Entweder ergibt sich aus den Extraktionskriterien eine den jeweiligen Modellen entsprechende Faktorenanzahl oder nicht. Wenn die Kriterien nicht in Bezug auf die intendierte Faktorenanzahl konvergieren, ist dies ein Hinweis auf mangelnde Modellgültigkeit.

Bei den modellübergreifenden Analysen (siehe Kapitel 6.3) erfolgt aufgrund der großen Variablenzahl und der mit hierarchischen Analysen verbundenen Einschränkungen die Festlegung der Faktorenanzahl allein Anhand der Parallelanalyse.

Auf der Basis der Extraktionsentscheidung erfolgt eine Hauptachsenanalyse mit iterativer Reduktion der Kommunalitäten (und quadrierten multiplen Korrelationen als Anfangsschätzer der Kommunalitäten). Zuletzt erfolgt eine Trasid-Rotation (nach Varimax-Vorrotation), bei der Ladungsbeträge $\leq .05$ als Nullladungen zählen, was einer Hyperebenendicke von $.10$ entspricht und eine direkte Oblimin-Rotation mit $\Delta=0$ (s. Tabelle 3.12). Auch die hierarchischen Analysen werden auf der Basis der Trasid- und Oblimin-Rotation durchgeführt.

Tabelle 3.12

Im Rahmen dieser Arbeit durchgeführte exploratorische Faktorenanalysen

Extraktionskriterium	Extraktionsmethode	Rotationsmethode
Parallelanalyse*	Hauptachsenanalyse mit Kommunalitäteniteration	Trasid mit Beträgen $\leq .05$ (Hyperebenendicke = $.10$)
Komparabilitäten $> .85^*$		Direktes Oblimin mit $\Delta=0$

* Bei den modellübergreifenden Analysen wird die Festlegung der Faktorenanzahl allein anhand der Parallelanalyse durchgeführt.

Die analysierte Probandenanzahl muss mindestens so groß sein wie die Anzahl analysierter Variablen, so dass im vorliegenden Kontext, bei dem insgesamt viele Variablen analysiert werden müssen, mindestens 500 Probanden untersucht werden müssen.

4. Zur theoretisch-interpretativen Unverbindlichkeit der Traitforschung

4.1 Zur Repräsentativität von Traitmodellen

Die Frage der Repräsentativität von Traitmodellen für die Beschreibung von Persönlichkeitseigenschaften soll in dieser Arbeit gestreift werden, da sie mit der Frage nach der Generalität von Traits verknüpft ist. Die Verknüpfung der Generalität von Traits mit der Repräsentativität von Traitmodellen ergibt sich aus folgender Überlegung: Ein Modell, das eine bestimmte Anzahl von sehr generellen Traitdimensionen aufweist, wird in der Regel als repräsentativer gelten können, als ein Modell, das dieselbe Anzahl weniger genereller Traitdimensionen aufweist. Diese Annahme ist darin begründet, dass eine hohe Generalität eines Traits im Sinne einer hohen transsituationalen Konsistenz des Traits interpretiert werden kann. Eine hohe transsituationale Konsistenz bedeutet, dass ein Trait für viele Situationen relevant ist. Insofern ist ein Modell, das sich aus einer bestimmten Anzahl genereller Traits zusammensetzt, für eine größere Zahl von Situationen relevant als ein Modell, das sich aus derselben Zahl weniger genereller Traits zusammensetzt. Dies setzt allerdings voraus, dass die generellen Traits nicht in höherem Maße korrelieren als die weniger generellen Traits. Denn höhere Korrelationen zwischen den generelleren Traits würden dafür sprechen, dass die generelleren Traits in höherem Maße für überlappende Situationen relevant sind. Dies würde wiederum dazu führen, dass die Gesamtzahl an Situationen, für die das Modell relevant ist, geringer ist, als man aufgrund der Generalität der einzelnen Traits erwarten würde. Somit spricht nur bei gleicher Interkorrelation der Traits von zwei Modellen eine höhere Generalität der Traits eines Modells für eine höhere Repräsentativität des Modells.

Obgleich der Zusammenhang zwischen der Generalität einzelner Traits und der Repräsentativität von Modellen in der dargestellten Weise indirekt ist, erscheint eine Einbettung der Frage nach der Generalität einzelner Traits in die allgemeinere Frage der Repräsentativität von Modellen für das weitergehende Verständnis von Generalitätskonzepten relevant. Da die Schwierigkeiten bei der Ermittlung und Beurteilung der Repräsentativität von Traitmodellen denen der Ermittlung und Beurteilung der Generalität von Traits vermutlich sehr ähnlich sind, sollen im folgenden vor allem die Schwierigkeiten bei der Bewertung der Repräsentativität von Modellen diskutiert werden. Weitergehende Beiträge zur Repräsentativität von Traitmodellen finden sich in Brocke (2000) sowie in Saucier und Goldberg (1998).

Zum Problem der Grundgesamtheit relevanter Items

Bei der Beurteilung der Repräsentativität von Traitmodellen geht es um die Frage, wie umfassend die Persönlichkeit des Menschen anhand von Traitmodellen beschrieben werden

kann. Um die Repräsentativität eines Modells bewerten zu können, muss zunächst die Grundgesamtheit von Merkmalen oder Verhaltensweisen, für die ein Persönlichkeitsmodell repräsentativ sein soll, angegeben werden (s. Brocke, 2000). Es müsste also vor der Itemgenerierung und vor den empirischen Untersuchungen festgelegt werden, für welchen Bereich von Personen, Situationen und Zeitpunkten ein Modell gelten soll. Wenn hier keine expliziten Festlegungen vorgenommen werden, bleiben die Aussagen über die Repräsentativität der Modelle notwendig spekulativ.

Eine Art der Festlegung des intendierten Geltungsbereiches erfolgte über die Sedimentationshypothese von Klages (1926) und die daraus abgeleitete repräsentative Erfassung von Adjektiven. Diese stellt zwar eine Möglichkeit der Definition einer Grundgesamtheit und damit der Beurteilung der Repräsentativität dar, aber die Repräsentativität in Bezug auf die Adjektive einer Sprache ist nicht zu verwechseln mit der Repräsentativität eines Modells und der darin postulierten Traits für Reaktionen in traitrelevanten Situationen, denen Menschen im Laufe ihres Lebens begegnen. Dies gilt insbesondere, da sich vermutlich manche Persönlichkeitsmerkmale eher in anderen Wortarten (z.B. Henss, 1998) oder komplexeren Kombinationen von Ausdrücken und Sätzen niederschlagen.

Auch wenn die Sprache als „natürliche“ Grundlage für die Untersuchung der Persönlichkeit herangezogen wird, bleibt zu klären, wie sichergestellt werden kann, dass aus der Sprache alle relevanten Persönlichkeitsmerkmale gewonnen wurden. Allein bei der Auswahl von Adjektiven zur Persönlichkeitsbeschreibung wurden unterschiedliche Schwerpunkte gesetzt. So wurden Adjektive, die soziale Evaluationen, soziale Rollen und physische Merkmale betreffen, häufig eliminiert, obwohl manche Autoren deren Einbeziehung als zweckmäßig ansehen (s. Tellegen, 1993). Saucier und Goldberg (1998) stellen ebenfalls fest, dass Bereiche wie Attraktivität, Größe oder auch Religiosität als relevant für die Persönlichkeitsbeschreibung angesehen werden sollten. Sie kommen daher in Bezug auf das FFM, bei dem diese Adjektive in der Regel nicht einbezogen wurden, zu dem Schluss, dass es Persönlichkeit beschreibt, wie sie konventionell verstanden wird. Allerdings ist der Begriff der Konvention selbst wieder unscharf.

Mitunter werden Fragebogen auf der Basis ausformulierter Items (d.h. kurze Sätze oder Satzteile) explizit an Ergebnissen des Adjektiv-Ansatzes ausgerichtet, um so an die über den Adjektiv-Ansatz konzeptualisierte Repräsentativität anzuknüpfen (z.B. Caprara, Barbaranelli, Borgogni & Perugini, 1993; Costa & McCrae, 1992a). Natürlich stellen die ausformulierten Items, die ein anhand von Adjektiven gewonnenes Modell abbilden, keine repräsentative

Stichprobe aus allen relevanten ausformulierbaren Items dar, denn die Anzahl formulierbarer Aussage- oder Frage-Sätze ist im Gegensatz zur Zahl der Adjektive prinzipiell unbegrenzt. Darüber hinaus führen die ständigen Veränderungen der Lebensbedingungen (z.B. Massenmedien, Internet, etc.) zu ständig sich verändernden Situationen, denen Menschen ausgesetzt sein können, so dass immer neue Items formuliert werden können. Wenn nun ein Fragebogen auf Basis ausformulierter Items genau den Kontext-Bereich abbildet, der von einem Adjektiv-Fragebogen abgebildet wird (z.B. Costa & McCrae, 1992a), dann fragt sich, welcher wissenschaftliche Erkenntnisgewinn damit verbunden ist, da der Kontext-Bereich weiterhin allein über die Adjektive definiert bleibt. Das Ergebnis bedeutet nur, dass es möglich ist, aus der sehr viel breiteren Population ausformulierter Sätze eine Menge zu ziehen, mit der man den Bereich, der durch Adjektive abgedeckt wird, ebenfalls abdecken kann. Ein derartiges Ergebnis ist zumindest für die Frage nach der Repräsentativität von Traitmodellen wenig interessant.

Das Problem der Festlegung des Geltungsbereichs wird besonders prägnant, wenn versucht wird, eine Taxonomie von Situationen (z.B. van Heck, Perugini, Caprara & Fröger, 1994) zu entwickeln, um die Repräsentativität der Traits in Bezug auf die traitspezifischen Reaktionen in traitrelevanten Situationen beurteilen zu können. Natürlich erscheint der Versuch zu einer repräsentativen Taxonomie von Situationen zu kommen zunächst wenig erfolgversprechend, insbesondere wenn man interkulturelle Unterschiede bedenkt. Aber selbst wenn eine solche Taxonomie nicht repräsentativ für alle möglichen Situationen sein kann, würde ein sinnvoller Ausschnitt bereits helfen, die Repräsentativität von Modellen für bestimmte Situationsbereiche und die transssituationale Konsistenz von Traits zu bewerten. Solange die Situationen nicht repräsentativ erfasst werden können, bleiben die Aussagen über Repräsentativität und Generalität im Traitbereich sehr spekulativ.

Beispiele für die postulierte Repräsentativität von Traitmodellen

Trotz der Unsicherheit darüber, ob und wie eine Vollständigkeit der Erfassung menschlicher Eigenschaften überhaupt sichergestellt werden kann, wird von manchen Traitforschern eine Vollständigkeit ihrer Modelle postuliert. Beispielsweise heißt es in Wiggins und Trapnell (1997, S. 756-757) zur Repräsentativität des FFMs: "Interest in the five factor model derived mainly from the claim that five dimensions might provide an adequate preliminary taxonomy for *all* nontrivial personality traits...". Interessant ist dabei, dass die Klassifikation vorläufig sein soll und zugleich die Annahme vertreten wird, sie könne alle nicht-trivialen Traits umfassen. Auch fragt sich, was genau „triviale“ von „nicht-trivialen“ Traits unterscheiden soll. Ein weiteres Beispiel für Spekulationen zur Repräsentativität, das sich zwar nicht auf die Modelle, sondern direkt auf die zugrundeliegenden Faktorenlösungen bezieht, findet sich in

Eysencks (1991, S. 782) Kommentar zur Fünf-Faktorenlösung von Zuckerman et al. (1988): “There is nothing in the data to elevate a five factor solution over solutions containing more or fewer factors; the most comprehensive is certainly the 3-factor solution, which also fits in best with the Royce and Powell meta-analysis.”

Wenn tatsächlich die Fünf-Faktorenlösung in Zuckerman et al. (1988) gleichwertig neben den anderen Lösungen steht, fragt sich, wie die Drei-Faktorenlösung dann zugleich “most comprehensive” (am umfassendsten) sein kann, da mit den drei Faktoren sicherlich etwas weniger Varianz erklärt wird, als mit den fünf Faktoren. Anhand beider Beispiele wird deutlich, wie der Grad, in dem Traits als generell und Modelle als “umfassend” oder repräsentativ angesehen werden, offenbar als Gütekriterium verstanden wird, obgleich über diese “Breite” letztlich nur spekuliert werden kann und auch der Nutzen der hohen Generalität teilweise in Frage gestellt werden kann (siehe Block, 1995; Hough, 1992; Schneider, Hough & Dunnette, 1996).

Diese zwei Beispiele sollten verdeutlichen, dass die Frage nach der Repräsentativität von Traitmodellen einerseits stark mit der Bewertung der Modelle verknüpft wird und andererseits auf einer sehr unsicheren Grundlage erfolgt. Die Dezidiertheit der Aussagen zur Repräsentativität steht somit hinter den Möglichkeiten ihrer rationalen Bewertung zurück. Die Frage nach der Repräsentativität von Traitmodellen ist vermutlich noch schwieriger zu beantworten als die Frage nach der Generalität von Traits. Im Rahmen dieser Arbeit soll lediglich versucht werden, die Beurteilung der Generalität von Traits auf eine verbindlichere Grundlage zu stellen. Wegen des dargestellten Zusammenhangs zwischen der Generalität von Traits und der Repräsentativität von Modellen kann jedoch vermutet werden, dass eine verbindlichere Ermittlung und Bewertung der Generalität von Traits auch als erster Schritt zu einer verbindlicheren Ermittlung und Bewertung der Repräsentativität von Traitmodellen beitragen kann.

4.2 Zum Problem der Generalität von Traits

Zur Ermittlung der Generalität von Traits wurde bisher nur eine Methode diskutiert: Die Bewertung des Ordnungsgrades von Faktoren. Dabei haben Faktoren zweiter Ordnung, die sich aus Faktoren erster Ordnung zusammensetzen, vermutlich eine entsprechend höhere Generalität als die Faktoren erster Ordnung. Diese herkömmliche Ermittlung der Generalität im Rahmen hierarchischer Faktorenanalysen wird allerdings negiert, wenn beispielsweise im Rahmen des FFM Faktoren, die als Faktoren erster Ordnung auftreten, eine Generalität von

Faktoren höherer Ordnung zugeschrieben wird (z.B. Costa et al., 1991). Die Loslösung vom Ordnungsgrad der Faktoren bei der Interpretation der Generalität war bereits wesentlicher Bestandteil von Cattells (1972) Kritik an Eysencks Persönlichkeitsfaktoren. Cattell konzediert, dass bei der Extraktion von zu wenigen Faktoren die Entstehung von Primärfaktoren vermieden werden kann. Die Annäherung an Faktoren zweiter Ordnung, die bei einer absichtlichen „Unterfaktorisierung“ entsteht, ist nach Cattell (1972) aber äußerst ungenau. Cattell (1972, S. 183) bezeichnet diese Art von Faktoren zweiter Ordnung daher als „biased definitions of the second order factors“, oder als „pseudo second orders“ (S. 169). Ein Problem dabei ist, dass die Frage nach der Anzahl der zu extrahierenden Faktoren offenbar mit der Generalität der Primärfaktoren interagiert. Eine Entscheidung über die Anzahl der zu extrahierenden Faktoren ist auf der Grundlage verschiedener Verfahren, mit teilweise sehr unterschiedlicher methodischer Qualität möglich (s. Kapitel 2). Wenn methodisch defizitäre Extraktionskriterien verwendet wurden, kann kaum etwas darüber ausgesagt werden, ob eine „Unterfaktorisierung“ vorliegt, die nach Cattell (1972) dazu führt, dass echte Primärfaktoren von vornherein vermieden werden. Insofern sind die Aussagen über die Generalität von Faktoren anhand des Ordnungsgrades von Faktoren mit großer Unsicherheit behaftet. Da darüber hinaus die klassische, aufsteigende hierarchische Faktorenanalyse nicht konsistent eingesetzt und zur Bewertung der Generalität von Traits herangezogen wird, entsteht für die Bewertung der Generalität von Faktoren eine sehr große Unverbindlichkeit.

Diese Unverbindlichkeit trägt vermutlich dazu bei, dass sich in der Literatur wenig fundierte Aussagen über die Generalität von Traits finden. Ein Beispiel für Spekulationen, die neben der Repräsentativität der Modelle auch die Generalität der Traits betreffen, findet sich in Costa et al. (1991, S. 887): “There is growing agreement among personality psychologists from many different perspectives that, at the broadest level, there are five major dimensions of personality [...]” Auf die Generalität der Traits bezogen heißt es in McCrae und Costa (1995, S. 457): “In fact, however, the FFM has consistently been described as the structure of personality at the highest level of generality...“. Weiterhin: Die Fünf Faktoren „...are both necessary and reasonably sufficient for describing at a global level the major features of personality“ (McCrae & Costa, 1986, S. 1001).

Die Tatsache, dass diese Beispiele alle aus dem Bereich des FFM der Persönlichkeit stammen, soll hier nicht als besondere Kritik gerade an diesem Modell aufgefasst werden. Möglicherweise tritt im Kontext des FFM der weitreichende Anspruch deutlicher zutage als bei anderen Modellen, aber dies ist nicht der Hauptaspekt der vorliegenden Kritik. Im wesentlichen soll

hier die geringe empirische Fundiertheit solcher Aussagen zur Generalität problematisiert werden, unabhängig vom Traitmodell, auf das sie sich beziehen.

Weitreichend und spekulativ sind die o.g. Feststellungen in Bezug auf die Generalität der Traits insofern als mit „the broadest level“ oder „a global level“ offenbar die Stufe höchster Generalität bzw. transsituationaler Konsistenz gemeint ist. Dies impliziert bereits, dass die fünf Dimensionen des FFM nahezu orthogonal sein müssen, da keine Dimensionen höherer Ordnung denkbar sein sollen. Unklar ist die erste Feststellung bezüglich „there are five major dimensions“. Hier bleibt offen, ob „fünf Hauptdimensionen und keine mehr“ gemeint ist, oder „unter anderem fünf Hauptdimensionen“. Wie auch immer man diese Statements bewerten möchte, sie verwenden Begriffe, die die Generalität der Faktoren (broadest, global) beschreiben, ohne dass sie sich dabei auf eine klare Operationalisierung oder gar Quantifizierung der Generalität von Faktoren stützen. Im übrigen werden die vorstehenden Aussagen zur Generalität der Faktoren des FFM durch die Analysen von Digman (1997), der zwei generellere Faktoren über der Ebene des FFM nachgewiesen hat, in Frage gestellt.

Da der empirische Gehalt allgemeiner Aussagen über die Generalität von Faktoren meist eher gering ist, soll mit der vorliegenden Arbeit der Versuch unternommen werden, eine vergleichende, möglichst auch quantitative Beurteilung einzelner Traits in Bezug auf ihre Generalität vorzunehmen (s. Kapitel 4.3). Eine dabei zu beantwortende Frage lautet: Welche Traits können in andere integriert werden? Diese Frage ist zumindest prinzipiell empirisch beantwortbar, indem gemeinsame Faktorenanalysen mit den Markiertvariablen für verschiedene Modelle durchgeführt werden.

Modellspezifische und modellübergreifende Faktoren

Wenn ein Trait beispielsweise als Faktor erster Ordnung in einen anderen eingeht, dann ist als wahrscheinlich anzusehen, dass letzterer ein höheres Generalitätsniveau aufweist. Einen Faktor eines Modells, der in gemeinsamen Analysen verschiedener Modelle die Varianz der Markiertvariablen von Faktoren anderer Modelle erklärt, könnte man unter dem Gesichtspunkt der Generalität als „modellübergreifenden Faktor“ bezeichnen. Davon abzugrenzen sind Faktoren, die entweder nur bei Analysen in einem modellspezifischen Variablenpool auftreten, oder aber in gemeinsamen Analysen im wesentlichen durch die Variablen eines Modells bzw. eines zugehörigen Inventars markiert werden. Modellspezifische Faktoren, die auch in modellübergreifenden Analysen auftreten, sind für den Modellvergleich von besonderer Bedeutung, da sie Varianzen abbilden, die von den anderen analysierten Modellen nicht oder nur in

geringem Umfang erfasst werden, während die modellübergreifenden Faktoren auf gemeinsame Varianzen der Modelle hinweisen.

Die Unterscheidung von modellübergreifenden Faktoren einerseits und modellspezifischen Faktoren andererseits kann natürlich nur zu sehr groben Differenzierungen der Generalität von Traits führen und hängt mit der in Kapitel 3.2.2 geschilderten Problematik der Extraktionskriterien und überdies mit der verwendeten Rotationsmethode zusammen. Daher sollen in Kapitel 4.3 Möglichkeiten einer genaueren Quantifizierung bestimmter Aspekte der Generalität von Traits vorgeschlagen werden. Zunächst werden aber die bestehenden Begriffe und Konzepte der Generalität genauer betrachtet.

Generalitätskonzepte und mit Generalität verwandte Konzepte

Möglicherweise hängt der geringe Konsens bezüglich der Generalität verschiedener Traits auch mit der Vielfalt von Begriffen und Konzepten zusammen, die mehr oder weniger implizit mit Generalität in Zusammenhang gebracht werden. Daher erscheint es zweckmäßig, die verschiedenen mit Generalität verwandten Konzepte zu sichten.

Generalität

Der Begriff der Generalität beschreibt, wenn er auf Persönlichkeitstraiten angewendet wird, meist den *Grad der Vielfalt der unterschiedlichen Situationen oder Stimuli, für die der jeweilige Trait relevant ist* (s. Amelang & Bartussek, 2001; Goldberg, 1990; Jäger, 1982, Jäger, Süß & Beauducel, 1997; Michel & Conrad, 1982). Herrmann (1969) bezeichnet diesen Sachverhalt als die „Weite“ der Dimensionen (der Begriff der „Generalität“ wird bei ihm für die Vielfalt der Probandengruppen, für die ein Konstrukt relevant ist, verwendet). Prinzipiell kann über Personen, Zeitpunkte, Situationen, ggf. auch Reaktionen generalisiert werden (siehe auch die „Datenbox“ von Cattell, 1957; s. auch Wittmann, 1988).

In dieser Hinsicht kann der Begriff der Generalität auf alle verschiedenen Arten der Generalisierung angewendet werden, auch auf Generalisierungen, die eher im testtheoretischen Sinne zu verstehen sind (s. Generalisierbarkeitstheorie, Cronbach, Gleser, Nanda & Rajaratnam, 1972). Dabei ist jedoch zwischen dem Konzept der Generalisierbarkeit von Meßwerten über verschiedene Stichproben und Situationen hinweg, wie es in der Generalisierbarkeitstheorie untersucht wird und der psychometrischen Generalität von Faktoren zu unterscheiden. Die Generalisierbarkeitstheorie ist vor allem dann von Nutzen, wenn die Variation der Meßwerte in Bezug auf Items, Probanden und Meßgelegenheiten simultan untersucht werden soll. Vor allem die Zuverlässigkeit (Dependability) von Meßwerten bei verschiedenen Meßgelegenheiten wird

anhand der Generalisierbarkeitstheorie ermittelt (Burns, 1998). Demgegenüber bezieht sich das psychometrische Konzept der Generalität auf die unterschiedliche Generalität aggregierter Meßwerte innerhalb derselben Meßgelegenheit. Es handelt sich bei der psychometrischen Generalität auch nicht um ein Konzept, das mit dem Meßfehler-Konzept der klassischen Testtheorie konkurrieren würde, während die Generalisierbarkeit als ein Konzept angesehen wird, das präzisere Information liefern soll als die Reliabilitätskennwerte der klassischen Testtheorie. Aus ähnlichen Gründen bezeichnet Wittmann (1988) die mit der Generalisierbarkeitstheorie anvisierten Konzepte als Konstruktreliabilität („construct reliability“). Eine psychometrisch generellere Skala kann höhere oder geringere Reliabilitäten aufweisen als eine weniger generelle Skala, d.h. die Information über die Generalität kann die Information über die Reliabilität nicht ersetzen. Insofern ist die hier untersuchte Generalität von Faktoren weniger ein Aspekt der Konstruktreliabilität als der Konstruktvalidität.

Man könnte das Konzept der psychometrischen Generalität vielleicht in den breiten Rahmen der Generalisierbarkeitstheorie integrieren, etwa wenn man untersuchen würde, ob generellere Aggregate eine höhere Generalisierbarkeit über Meßgelegenheiten und Probanden aufweisen, als weniger generelle Aggregate. Dies ist jedoch nicht das Ziel der vorliegenden Arbeit, zumal die Generalität als psychometrisches Konzept noch nicht konkret genug elaboriert und quantifiziert wurde, um in den formalen Rahmen der Generalisierbarkeitstheorie integrierbar zu sein. Im Rahmen dieser Arbeit sollen überhaupt erst einmal Konkretisierungen und Quantifizierungen der psychometrischen Generalität vorgenommen werden. Die Bezüge der hier entwickelten Quantifizierungen zur Generalisierbarkeitstheorie könnten Gegenstand weiterführender Arbeiten sein.

Funktionale Äquivalenz

Die Generalität von Traits hängt eng mit Allports (1937) Konzept der „funktionalen Äquivalenz“ zusammen. Traits können nach Allport Situationen und Reaktionen funktional äquivalent machen, beispielsweise können verschiedene Situationen dieselbe Reaktion oder dieselbe Klasse von Reaktionen nach sich ziehen. Wenn ein Trait viele verschiedene Situationen funktional äquivalent macht, dann hat er auch eine hohe Generalität. Wenn jedoch ein Trait viele verschiedene Reaktionen funktional äquivalent macht, dann bedeutet dies, dass ein reichhaltiges Repertoire an Reaktionen für den Ausdruck traitspezifischen Verhaltens zur Verfügung steht. Dies muss nicht zu einer hohen Generalität des Traits i.S. hoher trans-situationaler Konsistenz führen. Beispiel: Der Trait Spinnenangst ist hochgradig situationspezifisch, d.h. eine Person würde nur dann Spinnenangst erleben, wenn sie eine Spinne sieht. Die transsituational hochspezifische Spinnenangst kann nun zu sehr vielen verschiedenen Reaktio-

nen führen: Feuchte Hände, schnellerer Herzschlag, Schreien, Töten der Spinne, Weglaufen, etc. So können viele verschiedene Verhaltensweisen durch denselben Trait ausgelöst werden, obwohl dieser sehr situationsspezifisch ist bzw. eine geringe Generalität aufweist.

Das Beispiel wurde angeführt, um die Tragfähigkeit von Allports (1937) Konzept zu verdeutlichen. Diese Differenzierung ist jedoch in der bisherigen empirischen Forschung kaum umgesetzt worden. Im Adjektiv-Ansatz, der den meisten großen Traitmodellen zugrunde liegt, werden die Situationen nicht weiter spezifiziert („virtuelle Omnibus-Situationen“; Brocke, 2000). Auch die Reaktionen, die die Probanden bei ihrer mentalen Aggregation zugrunde legen, werden nicht erfasst. Insofern kann die Differenzierung von Situations- und Reaktionsgeneralität anhand von Adjektiven nicht erfolgen. Möglicherweise gehen bei Fremdbeurteilungen die Reaktionsanteile der beurteilten Person stärker ein als bei Selbstbeurteilungen, da die beurteilenden Personen ihre Einschätzungen aus den Reaktionen der beurteilten Personen ableiten müssen. Dies ist allerdings spekulativ, da ja die Reaktionen der beurteilten Personen wiederum aus deren wahrgenommenen Situationen folgen. Inwieweit die Probanden bei Selbsteinschätzungen eher über die eigenen Reaktionen oder über die erlebten Situationen aggregieren, bleibt offen. Der Ansatz ausformulierter Items böte die Möglichkeit, Situationen und Reaktionen getrennt zu erfassen und vielleicht zu permutieren, um die verschiedenen Aspekte der Generalität i.S. von Allport differenzieren zu können. Eine Möglichkeit sich der dargestellten Differenzierung zu nähern, besteht auch im TinSit Fragebogen von van Heck et al. (1994). Wie auch immer, die meisten Trait-Inventare auf Basis ausformulierter Items präsentieren Situationen und Reaktionen in vollkommen unsystematischer Form, so dass auch bei einer Diskussion der auf den verfügbaren Item-Inventaren beruhenden Trait-Modelle Situations- und Reaktionsgeneralisierung in der Regel nicht differenziert werden kann.

Transsituationale Konsistenz

Der hier verwendete Begriff der Generalität ist weitgehend synonym mit dem Begriff der transsituationalen Konsistenz, wie er in der Interaktionismus-Debatte verwendet wurde (z.B. Mischel, 1968). Der einzige Unterschied ergibt sich daraus, dass Mischel mit transsituationaler Konsistenz explizit die Generalisierung über Situationen meinte, während – wie sich aus der Darstellung zur funktionalen Äquivalenz – ergibt, in die Generalität empirischer Traits in unbekanntem Ausmaß vermutlich auch Generalisierungen über Reaktionen eingehen.

Aggregationsniveau

Unter einem Aggregationsniveau versteht man die Anzahl einzelner Messungen, die in eine Summe von Messungen (=Aggregat) eingehen. Wenn eine Skala ein hohes Aggregations-

niveau aufweist, kann sie, da sie sich aus vielen Einzelmessungen zusammensetzt, auch ein hohes Generalitätsniveau aufweisen. Dieser Zusammenhang ist jedoch keineswegs zwingend. Es ist beispielsweise möglich, dass ein Aggregat sich aus sehr vielen spezifischen, extrem hoch korrelierenden Einzelmaßen zusammensetzt und somit das Generalitätsniveau insgesamt recht gering ist. Oder aber es ist möglich, dass ein einzelnes Maß extrem komplexe Sachverhalte widerspiegelt, auch wenn es sich formal nicht um ein Aggregat handelt. Beispielsweise könnte die Anzahl an Versicherungen, die ein Vertreter in einem Jahr verkauft, äußerst komplexe Information enthalten. Die Aggregation einzelner Messungen erfolgt bei solchen Maßen implizit, d.h. es müssen sehr viele Bedingungen realisiert sein, damit eine bestimmte Anzahl verkaufter Versicherungen zustande kommt. In dieser Hinsicht enthält das Einzelmaß, als Konsequenz der Bedingungen, die Messung der vielen Einzelbedingungen. Da die Anzahl von Bedingungen, die realisiert sein müssen, damit eine Variable einen bestimmten Wert annimmt, meist unbekannt ist, folgt aus dem Aggregationsniveau strenggenommen nichts für die Generalität der Aggregate. In Abhängigkeit von den Maßen, über die aggregiert wird, können hoch-aggregierte Maße sehr speziell und wenig aggregierte Maße sehr generell sein.

Ein weiteres Beispiel für diese Problematik ergibt sich aus dem Phänomen der mentalen Aggregation, die bei der Bearbeitung von Trait-Items, insbesondere von Adjektiven, relevant wird. Über wie viele Bedingungen aggregiert ein Proband seine Einschätzungen? Spiegelt sich in seiner Einschätzung die ganze Breite seiner Erlebnisse über einen längeren Zeitraum wieder? Wie groß sind die interindividuellen Unterschiede in der Sorgfalt, Breite und Genauigkeit mentaler Aggregationen? Wie wird die Situations- und die Reaktionskomponente gewichtet? Aus all diesen Unsicherheiten folgt, dass die Aussage über das Aggregationsniveau eher technischen Charakter hat und nicht direkt i.S. der Generalität interpretiert werden kann. Nur unter der Voraussetzung, dass die Generalität der aggregierten Maße gleich ist, steht ein höheres Aggregationsniveau für eine höhere Generalität. Allerdings hängt der Bezug des Aggregationsniveaus zur Generalität noch von einer weiteren Bedingung, nämlich der Homogenität der aggregierten Variablen ab.

Homogenität und interne Konsistenz

Die Höhe der Interkorrelationen zwischen den aggregierten Einzelmaßen spielt eine große Rolle für den potentiellen Generalitätszuwachs bei der Aggregation. Die Höhe der Interkorrelation der Einzelmaße ist proportional zur internen Konsistenz oder Homogenität eines Aggregates.

Wenn man die Generalität von Aggregaten beurteilen möchte, muss man neben dem Aggregationsniveau und der Generalität der Einzelmaße auch die Homogenität der Aggregate berücksichtigen. Die sechs Kreise in Abbildung 4.1 entsprechen hier den Varianzen sechs standardisierter Items. Die Homogenität wird durch den Grad an Überlappung zwischen den Kreisen angezeigt. Dabei wird zur Vereinfachung angenommen, dass die Reliabilität aller Items eins ist, d.h. der Grad an Überlappung ist hier kein Maß für die Reliabilität (für Zusammenhänge zwischen Reliabilität und Aggregation siehe Wittmann, 1988). Es wird deutlich, dass bei einem konstanten Grad an Überlappung die insgesamt abgedeckte Fläche der Kreise mit zunehmender Itemzahl steigt.

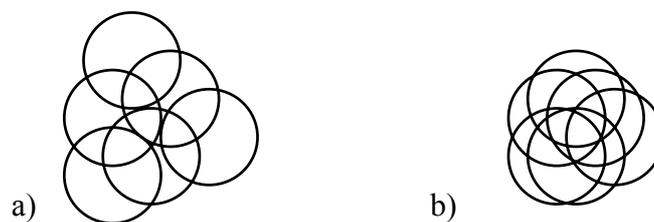


Abbildung 4.1. Veranschaulichung der Interaktion von Aggregation und Homogenität

Wenn man die Diagramme so interpretiert, dass die abgedeckte Fläche, welche die in die Skalen eingehenden Varianzen darstellt, dann zeigt sich unmittelbar, dass die Breite der Situationen, die eine Skala abbildet, d.h. ihre Generalität bei einer gegebenen Homogenität auch von der Anzahl der Items der Skala, also vom Aggregationsniveau abhängt.

Wie der Vergleich der Diagramme a) und b) zeigt, gibt es darüber hinaus eine Interaktion zwischen der Zunahme der Generalität mit der Zahl aggregierter Items und der Homogenität. Bei geringer Homogenität führt dieselbe Zunahme im Aggregationsniveau zu einer größeren Zunahme der Generalität als bei großer Homogenität. Im Extremfall einer Homogenität von 1 ändert eine Aggregation einer beliebig großen Anzahl von Items nichts am Generalitätsniveau: Die sechs Kreise würden in diesem Fall alle übereinandergelegt werden.

Die Steigerung der Generalität mit steigendem Aggregationsniveau ist also umso größer, je geringer die Homogenität der Skala ist. Eine weniger homogene Skala kann mit weniger Items ein hohes Generalitätsniveau erreichen als eine homogenere Skala. Dies ergibt sich auch aus den Formeln, die Mulaik (1972) für die Korrelationen zwischen Prädiktoren und komplexen Kriterien ableitet: Wenn die Inter-Itemkorrelationen gering sind, dann können – bei mittlerer Item-Prädiktor-Korrelation – durch Skalenverlängerung (=Erhöhung des Aggregationsniveaus) deutlich größere Steigerungen der Prädiktor-Kriteriumskorrelationen erreicht werden als wenn

die Inter-Itemkorrelationen hoch sind. Aus dem Dargestellten folgt, dass man unter bestimmten, noch zu spezifizierenden Bedingungen, Aussagen über den relativen Zuwachs an Generalität in Abhängigkeit von Aggregationsniveau und Homogenität machen kann, etwa derart: Eine Steigerung des Aggregationsniveaus führt bei geringer Homogenität und gegebener Reliabilität der dazukommenden Einzelmessungen zu einer Zunahme der Generalität. Aber man kann keine Aussagen über den (absoluten) Betrag der Generalität machen, da man die Generalität der Einzelmaße nicht kennt. Aus einem ähnlichen Grund unterscheidet Humphreys (1981) zwischen Homogenität und Unidimensionalität. Ein Aggregat mehrerer mehrdimensionaler Maße kann eine hohe Homogenität aufweisen, wenn die Einzelmaße sich alle aus ähnlichen oder denselben Dimensionen zusammensetzen. Dennoch kann es sich bei dem Aggregat inhaltlich um ein sehr generelles Maß handeln.

Aus der Unkenntnis der Generalität der Einzelmaße folgt auch eine Einschränkung für die Aussagen zur relativen Generalität: Wenn zu einem Einzelmaß, das aufgrund der (im Abschnitt Aggregationsniveau, siehe oben) geschilderten impliziten Aggregation von Bedingungen, sehr generell ist, mehrere, sehr spezielle Maße hinzuaggregiert werden, dann kann das Generalitätsniveau bei der Aggregation sogar abnehmen. Eine geringe Korrelation zwischen den Einzelmaßen schützt dagegen nicht, denn ein generelles Maß wird mit einem spezifischen auch eher gering korrelieren (s. Wittmann, 1988). Genauso müssen mehrere, sehr spezifische Maße nicht hoch korrelieren. Dieses Problem ist besonders gravierend, wenn eine geringe Zahl von Einzelmessungen aggregiert wird, und in den Einzelmessungen nur ein einzelnes, sehr generelles Maß enthalten ist. Ist jedoch die Zahl der aggregierten Einzelmaße extrem hoch, kann die Einbuße an Generalität durch Aggregation spezifischer Einzelitems möglicherweise kompensiert werden, wenn die Interkorrelationen der Einzelmaße moderat sind und die moderaten Korrelationen nicht durch geringe Reliabilität der Einzelmaße zustande kommen.

Für eine Proportionalität von Generalität und Aggregationsniveau erweist sich somit die Bedingung, dass das Aggregationsniveau der einzelnen aggregierten Maße ähnlich ist, als zentral. Diese Bedingung kann, wenn das Format und der Inhalt der Einzelmessungen ähnlich ist (wie bspw. bei Adjektiv-Skalen), durchaus erfüllt sein. Natürlich sind diese Relativierungen Aufgrund einer hypothetischen, unterschiedlichen Generalität der Einzelmessungen so grundsätzlicher Art, dass sie nur schwer in Konsequenzen für die Forschungspraxis umgesetzt werden können. Mit Sicherheit sprechen diese Relativierungen jedoch dafür, mit Generalitätsbehauptungen, wie sie in Kapitel 4.1 für Traitmodelle dargestellt wurden, vorsichtig zu sein.

Trotz der genannten Relativierungen könnte man aus dem Dargestellten ableiten, dass bei annähernd gleichem Aggregationsniveau heterogene Aggregate eher eine hohe Generalität aufweisen sollten als homogene Aggregate. Zwar kann, wenn das Generalitätsniveau der einzelnen Items unbekannt ist, keine sichere Aussage getroffen werden, aber es ist wahrscheinlicher, dass ein Maß, das sich aus vielen verschiedenen Maßen zusammensetzt, eine hohe Generalität aufweist. Denn ein homogenes Maß kann nur dann eine hohe Generalität aufweisen, wenn bereits die Items eine hohe Generalität aufweisen. Diese Bedingung muss bei einem heterogenen Maß nicht erfüllt sein, damit es eine hohe Generalität aufweist. Aus diesem Argument könnte man ableiten, dass man nur hinreichend heterogene Aggregate bilden müsste, um generelle Merkmale zu erfassen. Umgekehrt könnte man die Reliabilitäts-korrigierte Homogenität eines Merkmals als Maß für dessen Generalität verwenden. Bei Traitskalen könnte beispielsweise die mittlere Inter-Itemkorrelation durch die mittlere Retest-Reliabilität der Items dividiert werden. Dann könnte man diese Reliabilitäts-korrigierte Korrelation in entsprechend abgewandelte Formeln zur Schätzung der internen Konsistenz einsetzen und auf diese Weise mit vergleichsweise geringem Aufwand eine Schätzung der Generalität ermitteln.

Auch wenn derartige Maße einen Aspekt der Generalität abbilden könnten, so ergibt sich dabei doch das konzeptuelle Problem, dass derartige Maße dann maximale Generalität anzeigen, wenn die interne Konsistenz null ist, d.h. wenn ein Aggregat kein einheitliches Merkmal erfasst. Den Nachweis eines Minimums an inhaltlicher Homogenität und damit minimaler interner Konsistenz der Items eines Aggregates kann man als Aspekt der konvergenten Validität ansehen (auch wenn der Aspekt der Methodeninvarianz hier außer acht bleibt). Wenn man nun Heterogenität als Aspekt der Generalität auffasst, dann gelten Maße mit geringer interner Konsistenz und somit geringer konvergenter Validität als Maße hoher Generalität. Es erscheint nicht sinnvoll, Generalität so zu konzeptualisieren, dass sie mit Aspekten konvergenter Validität inkompatibel ist. Auch die Forderung einer wenigstens signifikanten Korrelation zwischen den Items eines Aggregates als Grundlage für minimale konvergente Validität, führt hier nicht weiter, da die Signifikanz von Korrelationen von der Stichprobengröße abhängt. Die einfache Konzeptualisierung von Generalität über die Heterogenität eines Aggregates erscheint somit problematisch. Aus diesem Grund werden in Kapitel 4.3 kompliziertere, aber konzeptuell weniger problematische Zugänge zur Quantifizierung der Generalität dargestellt.

Faktoren erster, zweiter und höherer Ordnung

Im Prinzip gelten alle Überlegungen, die zum Begriff des Aggregationsniveaus ausgeführt wurden, auch für den Ordnungsgrad von Faktoren, denn Faktoren sind das Ergebnis einer gewichteten Aggregation von Variablen. Deswegen können prinzipiell auch einzelne Variablen

eine höhere Generalität aufweisen als die Faktoren, in die sie eingehen. Ein fiktives Beispiel dafür findet sich in Block (1995, S. 189): Man denke sich 100 Variablen, von denen 99 interkorrelieren, weil sie vollständig reliable Manifestationen der latenten Fähigkeit des „Schuhe-Zubindens“ sind. Variable 100 sei eine vollständig valide aber unbekannte Manifestation der latenten Variable „allgemeine Intelligenz“. Eine Faktorenanalyse dieser Variablen würde zu einem Faktor führen, der nahezu die gesamte gemeinsame Varianz des „Schuhe-Zubindens“ erklärt. Das einzige, unbekannte Maß für allgemeine Intelligenz würde in diesem Variablenkontext mit hoher Wahrscheinlichkeit als unbedeutend betrachtet werden. Dennoch wäre vermutlich die Generalität dieses Maßes für allgemeine Intelligenz größer als die des Faktors für die Fähigkeit des „Schuhe-Zubindens“.

Aus dem Beispiel folgt nicht nur, dass einzelne Variablen größere Generalität als die resultierenden Faktoren aufweisen können, sondern auch, dass man aus der erklärten Varianz eines Faktors nicht direkt auf seine Generalität schließen kann. Dennoch kann die erklärte Gesamtvarianz aller Faktoren einer Lösung interessant sein, wenn beschrieben werden soll, wie umfassend ein (Trait-)Modell die Daten beschreibt (z.B. Becker, 1999).

Methodenkritisch muss allerdings angemerkt werden, dass man im Beispiel von Block (1995) keine Aussage über die Generalität des in die Lösung nicht eingehenden Maßes machen kann. Man würde in diesem Fall eher versuchen, den Faktor, den eine einzelne Variable möglicherweise repräsentiert, bei einer weiteren Analyse mit mehreren Variablen (mindestens fünf Variablen, siehe Kapitel 3) abzudecken.

Im übrigen gilt die Annahme, dass Faktoren zweiter Ordnung einer Lösung genereller sind als die dazugehörigen Faktoren erster Ordnung, nur unter der Bedingung, dass die Faktoren erster Ordnung eine ähnliche Generalität aufweisen. Ansonsten kann der Fall eintreten, dass ein hochgradig genereller Primärfaktor mit einem weiteren, spezielleren Primärfaktor zu einem Faktor zweiter Ordnung verschmilzt, der eine geringere Generalität aufweist als der generellere der beiden Faktoren erster Ordnung. Neben der vergleichbaren Generalität für die Faktoren erster Ordnung gilt, dass man von einer vergleichbaren Generalität der Einzelmessungen bzw. Items ausgehen muss, um vom Ordnungsgrad der Faktoren auf die Generalität der Faktoren schließen zu können. Denn nur bei einer vergleichbaren Generalität der Items kann auch eine vergleichbare Generalität der Faktoren erster Ordnung auf kontrollierbare Weise erreicht werden.

Zusammenfassung

Eine Zusammenfassung der dargestellten Generalitätskonzepte bzw. mit Generalität verwandter Konzepte findet sich in Tabelle 4.1. Die ersten beiden Konzepte, funktionale Äquivalenz und transssituationale Konsistenz sind eher theoretisch-begrifflicher Art, während Aggregationsniveau, Ordnungsgrad von Faktoren und Heterogenität eher technisch-methodische Konzepte sind.

Für die methodischen Konzepte ergibt sich aus dem bisher dargestellten, dass die Generalität eines Aggregates oder Faktors von der Anzahl der aggregierten Items (Aggregationsniveau), der Homogenität der Einzelmaße, sowie dem Generalitätsniveau der Einzelmaße bzw. Items abhängt. Da die Generalität der Einzelmaße meist unbekannt ist, sind die Aussagen zur Generalität von Traits, die sich auf Aggregationsniveau, Ordnungsgrad der Faktoren und Homogenität bzw. Heterogenität der Maße gründen, notwendig spekulativ und mit einem erheblichen Maß an Unsicherheit behaftet. Generalität kann auch nicht allein über die Heterogenität von Aggregaten definiert werden, da dann eine geringe konvergente Validität eine Voraussetzung für hohe Generalität wäre.

Tabelle 4.1

Dem Konzept der „Generalität“ verwandte Konzepte

funktionale Äquivalenz	Wenn Traits viele verschiedene Situationen funktional äquivalent machen, weisen sie eine hohe Generalität auf
transssituationale Konsistenz	weitgehend synonym mit Generalitätsniveau
Aggregationsniveau	eher technischer Begriff, der sich auf die Anzahl der Variablen, die in eine Skala/einen Faktor eingehen, bezieht; Vermutung: hohes Aggregationsniveau kann, aber muss nicht, hohe Generalität implizieren
Faktor 1., 2., 3. Ordnung	Vermutung: höhere Ordnung kann zu höherem Generalitätsniveau führen
Heterogenität (Homogenität)	technisch: interne Konsistenz (Cronbachs Alpha, Split Half) Reliabilitätsmaß; Vermutung: hohe Heterogenität (geringe Homogenität) kann hohe Generalität implizieren; Beispiel: Strukturelle Testtheorie (Cattell & Tsujioka, 1964)

Heterogenität, Aggregationsniveau und Ordnungsgrad der Faktoren können somit nicht allein zu einer sinnvollen Schätzung der Generalität von Traits beitragen. Da die Generalität von Traits nicht direkt aus den konventionellen statistischen Merkmalen von Aggregaten ermittelt werden kann, werden im folgenden grundsätzlich andere Zugänge der Quantifizierung von Generalität vorgeschlagen.

Mit den im folgenden entwickelten Indizes zur Generalitätsermittlung wird keinesfalls eine Methodenvielfalt in die psychometrische Traitforschung eingeführt, die die Freiheitsgrade bei der Modellbildung erhöhen würde. Vielmehr setzen die neu vorgeschlagenen Methoden bei den – nach den Spezifikationen in Kapitel 3 – durchgeführten Ergebnissen der Faktorenanalysen an. Gegebene Faktoren sind also der Input für die Generalitäts-Indizes, und es wird auf der Basis dieser Indizes lediglich möglich, weitere Informationen zu bestehenden Faktorenlösungen zu erhalten. Insofern steht die Verwendung dieser Indizes nicht den Zielen dieser Arbeit entgegen. Wie in den Kapiteln 2 und 4.2 dargelegt wurde, führen die Freiheitsgrade bei der Generalitätszuschreibung zu Konstrukten zu einer deutlichen Verminderung der Verbindlichkeit von Modellvergleichen in der psychometrischen Traitforschung. Insofern können Indizes, anhand derer die Generalität von Faktoren evaluiert werden kann, zu einer Steigerung der Verbindlichkeit der Diskussionen beitragen.

4.3 Vorschläge zur vergleichenden Generalitätsbewertung von Traits

Im folgenden soll der Versuch unternommen werden, die Generalität von Traits anhand einer vergleichenden Betrachtung einzuschätzen. Die Beurteilung der relativen Generalität von Traits ist dabei eine für theoretische Diskussionen relevante Beurteilungsdimension, wenn es beispielsweise darum geht, bestimmte Traits als anderen untergeordnete Traits einzustufen.

Für die vergleichende Ermittlung der Generalität wird davon ausgegangen, dass die von den Proponenten eines Modells, das den Trait beinhaltet, angebotenen Operationalisierungen die besten verfügbaren Operationalisierungen des jeweiligen Traits darstellen. Diese Operationalisierungen (Skalen, Faktorwerte) können nun in Bezug auf ihre Generalität verglichen werden. Dies ist besonders dann informativ, wenn von verschiedenen Autoren verschiedene Operationalisierungen zu sehr ähnlichen Konstrukten angeboten werden. Zum Beispiel kann der Extraversionsfaktor wie er sich aus dem EPQ-R (Eysenck & Eysenck, 1991; deutsche Fassung: Ruch, 1999) gemäß der PEN-Theorie ergibt, mit dem über das NEO-PI-R (Costa & McCrae, 1992a; deutsche Fassung: Ostendorf & Angleitner, in Druck) erfassten Extraversionsfaktor des

FFM verglichen werden. Ein Beispiel für relevante Generalitätsvergleiche verschiedener Faktoren wäre auch der von Psychotizismus mit Offenheit für Erfahrungen, da Offenheit für Erfahrungen nach Eysencks (1992) Vorstellungen eine geringere Generalität als Psychotizismus aufweisen sollte, während dies nach Costa und McCrae (1992b) nicht der Fall sein sollte. Im folgenden werden hier als Inklusions-Indizes bezeichnete Kennwerte vorgeschlagen, mit deren Hilfe Vergleiche der Generalität ausgewählter Traits durchgeführt werden können.

4.3.1 Inklusions-Indizes für Skalen

Ogleich die Bewertung der Generalität im Rahmen dieser Arbeit für Faktoren erfolgen soll, da die meisten Trait-Modelle aufgrund von Faktorenanalysen konzipiert wurden, soll zunächst ein Inklusions-Index für Skalen vorgestellt werden. An diesem Index kann zum einen die Denkweise veranschaulicht werden, zum anderen wäre es denkbar, die Generalität von Testskalen, wie sie in der diagnostischen Praxis oft verwendet werden, anhand dieses Indexes zu vergleichen.

Konzeptuell orientiert sich der Inklusions-Index an der Brunswik-Symmetrie (s. Hammond, 1966), wie sie von Wittmann (1988) als Grundlage für optimale Prädiktionen gesehen wird. Um zwei Tests zu vergleichen, wird jeder Test einmal als Prädiktor und einmal als Kriterium verwendet. Wittmann (1988) geht davon aus, dass Prädiktionen dann maximal sein können, wenn Prädiktor und Kriterium das gleiche Aggregationsniveau bzw. das gleiche Generalitätsniveau aufweisen. Aggregationsniveau und Generalitätsniveau können konzeptuell nicht gleichgesetzt werden (s.o.). Zur Vereinfachung sollen hier dennoch zunächst Aggregationsniveau und Generalitätsniveau gleichgesetzt werden, bevor Möglichkeiten der Evaluation der Äquivalenz von Aggregations- und Generalitätsniveau vorgeschlagen werden.

Sofern man das Prinzip der Brunswik-Symmetrie als gültig voraussetzt, kann man daraus ableiten, dass das Generalitätsniveau von zwei Maßen maximal ähnlich ist, wenn die Prädiktor-Kriteriums-Korrelation von zwei Maßen maximal ist. Es wird lediglich die Schlußrichtung umgekehrt. In Wittmann (1988) wird ein ähnliches Aggregationsniveau angestrebt, um eine optimale Prädiktion zu erreichen, hier soll eine optimale Prädiktion – zumindest bei homogenen Konstrukten – als Hinweis auf ein ähnliches Aggregationsniveau dienen.

Bei heterogenen Konstrukten mit Binnenstruktur hängt die höchste Korrelation mit den Prädiktoren auch von der Übereinstimmung der Zusammensetzung der Prädiktoren mit der Zu-

sammensetzung des Kriteriums ab. Je mehr die Zusammensetzung der Prädiktoren der Zusammensetzung des Kriteriums entspricht, desto höher die Korrelation. Dabei impliziert auch eine ähnliche Zusammensetzung von Prädiktor und Kriterium eine ähnliche Breite erfasster Konstrukte und somit ähnliche Generalität. Insofern haben die hier dargestellten Überlegungen auch Gültigkeit für Konstrukte mit Binnenstruktur bzw. Skalen, die sich aus Subskalen zusammensetzen. Allerdings müssen die Maße, deren Generalität verglichen werden soll, ein einheitliches Merkmal repräsentieren, auch wenn dieses Merkmal eine gewisse Heterogenität aufweisen kann. Für sozial definierte Kriterien oder Maße, die keine Homogenität aufweisen, können die hier verwendeten Indizes nicht zur Generalitätsschätzung verwendet werden, da das Prinzip der Brunswik-Symmetrie bei willkürlichen Aggregaten nicht gelten kann¹.

Aggregationsbasierter Inklusions-Index für Skalen

Wenn man das Generalitätsniveau eines Tests 1 im Vergleich zu einem Test 2 ermitteln möchte, kann man diesen Test 1 mit allen möglichen Teilaggregaten von Test 2 korrelieren. Test 2 muss in alle möglichen Teilaggregate zerlegt werden, damit man die Korrelation für alle möglichen Aggregationsniveaus prüfen kann (s. Abbildung 4.2). Wenn der Gesamtwert von Test 2 höher mit dem Gesamtwert von Test 1 korreliert als alle möglichen Teilaggregate von Test 2, dann ist das Aggregationsniveau des Test 2 Gesamtwertes dem Aggregationsniveau des Test 1 Gesamtwertes näher als das Aggregationsniveau aller möglichen Teilaggregate von Test 2. Dies ist ein Hinweis darauf, dass Test 2 höchstens dasselbe Generalitätsniveau wie Test 1 aufweist, auf keinen Fall ein höheres.

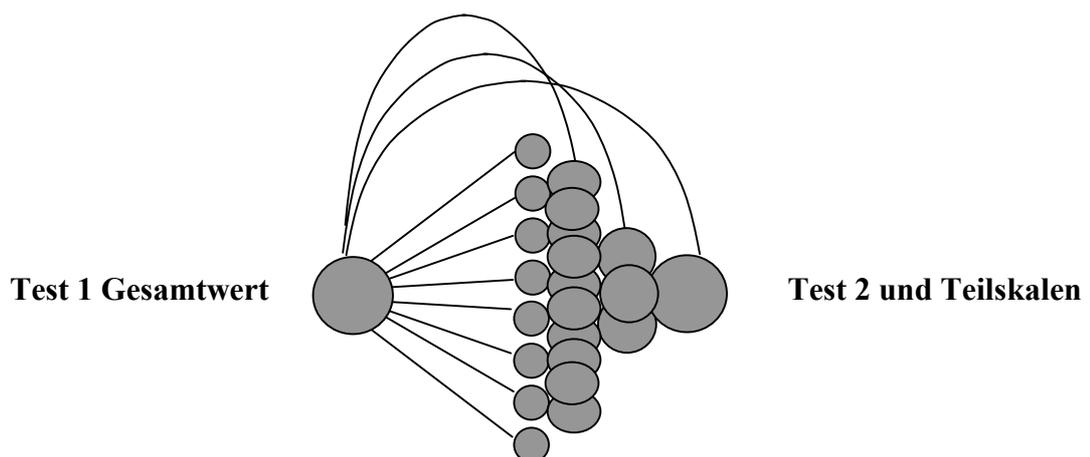


Abbildung 4.2. Korrelationen von Test 2 und seiner Teilaggregate mit dem Test 1 Gesamtwert

¹ Wenn beispielsweise ein Maß der Lebenszufriedenheit mit der Fähigkeit des „Schuhe-Zubindens“ aggregiert wird, hängt die Korrelation dieses Aggregates mit beliebigen Prädiktoren nicht vom Aggregationsniveau der Prädiktoren ab.

Wenn ein einzelnes Item von Test 2 am höchsten mit Test 1 korreliert, dann ist dies ein Hinweis darauf, dass ein einzelnes Item von Test 2 bereits ein sehr hohes Generalitätsniveau aufweist und dass Test 2 vermutlich ein höheres Generalitätsniveau aufweist als Test 1. Allerdings ist in diesem Fall, die Eingangs eingeführte Voraussetzung der Äquivalenz von Aggregationsniveau und Generalitätsniveau vermutlich verletzt, was die genannte Aussage in Frage stellt (siehe unten). Um den Generalitätsvergleich vollständig durchzuführen, müssen in einem zweiten Schritt alle Teilkombinationen von Test 1 mit dem Gesamtwert von Test 2 verglichen werden. Auf diese Weise kann die vergleichende Bewertung der Generalität anhand des wechselseitigen Einschlusses der Varianzen der beiden zu vergleichenden Tests erfolgen.

Auf analoge Weise kann geprüft werden, ob die eingangs zur Vereinfachung eingeführte Äquivalenz von Aggregationsniveau und Generalität gegeben ist. Aggregationsniveau und Generalitätsniveau können dann als äquivalent behandelt werden, wenn die einzelnen Items, die den Tests zugrunde liegen, ein sehr ähnliches Generalitätsniveau aufweisen. Die Items haben naturgemäß alle dasselbe Aggregationsniveau (nämlich das geringste, das möglich ist). Wenn Aggregationsniveau und Generalitätsniveau äquivalent sind, d.h. wenn Aggregationsniveau als Generalitätsniveau interpretiert werden kann, dann sollte ein gleiches Aggregationsniveau, das bei den Items gegeben ist, zu einem gleichen Generalitätsniveau der Items führen. Nach dem Prinzip der Brunswik-Symmetrie müsste ein gleiches Generalitätsniveau der Items dazu führen, dass – sofern ähnliche Inhalte getestet werden – die höchsten Korrelationen eines Items mit einem anderen Item auftreten und nicht mit Aggregaten aus Items (die Korrelationen mit den Aggregaten müssen ggf. part-whole korrigiert werden). Wenn also die Items ihre maximale Korrelation nicht mit Aggregaten, sondern mit anderen Items haben, dann spricht dies dafür, dass die Items ein ähnliches Generalitätsniveau aufweisen. Wie bereits erwähnt, ist ein vergleichbares Generalitätsniveau der Items die Voraussetzung dafür, dass man das Aggregationsniveau im Sinne eines Generalitätsniveaus interpretieren kann. Für diese Prüfung ist die Unterscheidung, ob die Aggregate Test 1 oder 2 entstammen, unerheblich. Sie kann allerdings auch für die Tests getrennt durchgeführt werden. Wenn einige Items mit unterschiedlich großen Aggregaten maximal korrelieren während andere Items mit einzelnen Items maximal korrelieren, kann von einer Ähnlichkeit der Generalität der Items nicht mehr ausgegangen werden und das Aggregationsniveau kann dann nicht direkt i.S. des Generalitätsniveaus interpretiert werden.

Eine Möglichkeit, das Ergebnis der dargestellten Vergleiche zu quantifizieren, ergibt sich aus folgenden Schritten:

1) Es werden sukzessive die Gesamtskala von Test 1 und alle möglichen Kombinationen von Teilskalen von Test 1 mit der Gesamtskala von Test 2 korreliert. Als Teilskala zählen auch einzelne Items, denn die Möglichkeit, dass ein einzelnes Item mit einer Skala am höchsten korreliert, ist theoretisch nicht auszuschließen. Die Anzahl der Teilskalen, die für eine Skala auf diese Weise gebildet werden kann, ergibt sich, da die Reihenfolge der Aggregation keine Rolle spielt, aus $2^n - 1$, wobei n die Anzahl der Items in der Skala ist. Dies führt bei langen Skalen zu einer sehr großen Zahl an Teilskalen. Die maximale Korrelation der Gesamtskala von Test 2 mit einer der Skalen von Test 1 wird festgehalten. Dabei sind auch Reliabilitätsunterschiede zwischen den Items bzw. Aggregaten zu beachten. Um diese zu berücksichtigen, könnte man die Berechnungen anhand minderungskorrigierter Korrelationen durchführen, was natürlich Reliabilitätsschätzer für alle Items voraussetzt.

2) Es werden die Gesamtskala und alle möglichen Kombinationen von Teilskalen von Test 2 mit der Gesamtskala von Test 1 korreliert. Die maximale Korrelation der Gesamtskala von Test 1 und einer Skala von Test 2 wird festgehalten.

3) Von der maximalen quadrierten Korrelation r_{12}^2 einer Skala von Test 1 mit der Gesamtskala von Test 2 wird die maximale quadrierte Korrelation r_{21}^2 einer Skala von Test 2 mit der Gesamtskala von Test 1 abgezogen (Formel 6). Wenn der Index größer null ist, schließt Test 1 eher Test 2 ein und wenn er kleiner null ist, schließt Test 2 eher Test 1 ein.

$$r_{12}^2 - r_{21}^2 = \text{Inklusion} \quad (6)$$

Beispiel: Die Gesamtskalen von zwei Tests 1 und 2 korrelieren miteinander .58. Die maximale Korrelation einer Subskala von Test 2 mit der 20 Items umfassenden Gesamtskala von Test 2 beträgt .70 für eine Subskala aus 10 Items (s. Abbildung 4.3).

Die maximale Korrelation einer Subskala von 18 Items des Tests 2 mit der Gesamtskala von Test 1 beträgt .60. In diesem Beispiel würde also Test 1 eher Test 2 einschließen als umgekehrt, da man bereits mit der Hälfte der Items von Test 1 die Hälfte der Varianz von Test 2 erklären kann, während man mit keiner Kombination von Teilskalen und auch nicht mit der Gesamtskala von Test 2 die Hälfte der Varianz von Test 1 erklären kann.

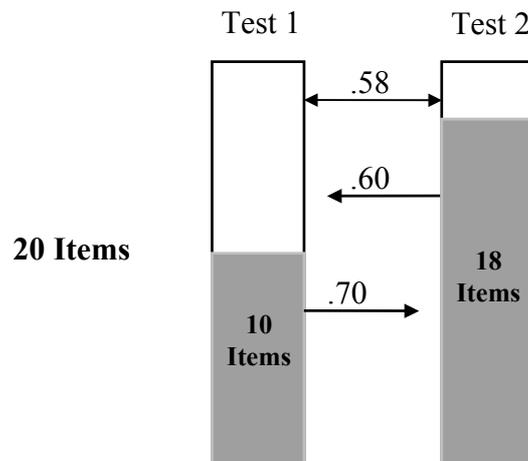


Abbildung 4.3. Beispiel für die wechselseitige Inklusion zweier Tests

Um diesen Sachverhalt möglichst einfach quantitativ auszudrücken, kann gemäß Formel 6 die Differenz zwischen der maximalen quadrierten Korrelation r_{12} des Subsets von Test 1 mit Test 2 und der maximalen quadrierten Korrelation r_{21} des Subsets von Test 2 mit Test 1 gebildet werden:

$$.70^2 - .60^2 = .13$$

Da der Index größer null ist, schließt Test 1 eher Test 2 ein. Die Bedeutung des Inklusions-Indexes lässt sich an dem Spezialfall veranschaulichen, dass ein Test 2 tatsächlich ein Bestandteil (Subtest) eines anderen (Test 1) ist. In diesem Fall wird Test 2 zu 1.0 mit sich selbst, d.h. mit dem entsprechenden Subtest von Test 1 korrelieren, während kein Subset von Test 2 mit dem Gesamttest 1 entsprechend hoch korrelieren wird. Der Inklusions-Index dient also der Beurteilung der Frage, wie gut man einen Test oder Fragebogen als Subtest des jeweils anderen interpretieren kann. Je eher ein Test der Subtest eines anderen sein kann, desto eher hat er im Sinne des Inklusions-Index eine geringere Generalität. Es liegt auf der Hand, dass mit diesem Index nicht alle Aspekte der Generalität geschätzt werden können, sondern nur der Aspekt der wechselseitig erklärten Varianzen.

Als nächstes stellt sich die Frage, wie zu verfahren ist, wenn die maximalen quadrierten Korrelationen der Subsets beider Tests nahezu gleich groß sind, aber auf der Basis unterschiedlich großer Itemmengen zustande kommen. Man kann dies ignorieren und dennoch von einer vergleichbaren Generalität der Skalen ausgehen. Wenn das Verhältnis zwischen den Itemmengen in den Subsets jedoch extrem ist, erscheint dies unangemessen. Wenn im Extremfall

beispielsweise ein Subset von 20 Items von Test 1 mit Test 2 eine maximale quadrierte Korrelation von .40 aufweist und ein Subset von zwei Items aus Test 2 dieselbe maximale quadrierte Korrelation mit Test 1 aufweist, dann fragt sich, ob Test 2 nicht als genereller einzustufen ist, weil ja eine deutlich geringere Anzahl von Items von Test 2 ausreicht, um einen nennenswerten Teil von Test 1 zu erklären. Dabei spielt die Gesamtanzahl der Items in den Tests natürlich eine Rolle. Wenn Test 2 insgesamt nur zwei Items umfasst, so weist er allein deswegen natürlich noch keine höhere Generalität als Test 1 auf, selbst wenn mit den beiden Items genauso viel Varianz von Test 1 erklärt wird, wie mit dem Subset von 20 Items aus Test 1 an Test 2.

Zur Berücksichtigung der Itemanzahl kann das Verhältnis der Itemanzahl der maximal mit dem jeweils anderen Test korrelierenden Subsets zur Itemzahl des zugehörigen Gesamttests betrachtet werden. Für dasjenige der beiden Subsets (Index s in Formel 7), das die größere Korrelation (Index g in Formel 7) mit dem jeweils anderen Test aufweist, gilt dabei folgende Beziehung: Wenn die Itemanzahl des Subsets (n_{sg} in Formel 7) dividiert durch die Itemanzahl des Gesamttests (n_{gg} in Formel 7) klein ist, liegt eine große Inklusion vor. Zusätzlich gilt, dass die Inklusion als geringer einzuschätzen ist, wenn die Itemzahl des Subsets des anderen Tests (n_{sk}) nur wenige Items im Verhältnis zum anderen Gesamttest (n_{gk}) umfasst.

Beispiel: Die quadrierte Korrelation des Subsets von Test 1 mit Test 2 betrage .40. Zugleich betrage die quadrierte Korrelation des Subsets von Test 2 mit Test 1 .35. Aus der Differenz der quadrierten Korrelationen könnte man gemäß Formel 6 schließen, dass eher Test 1 den Test 2 einschließt als umgekehrt. Nun beinhaltet Subset 1 alle 10 Items (n_{sg} in Formel 7) von Test 1 (n_{gg}), während Subset 2 nur 5 Items (n_{sk}) eines insgesamt 20 Items umfassenden Tests (n_{gk}) beinhaltet. Ein Inklusions-Index sollte so ausgelegt sein, dass ein solches Verhältnis zu einer Minderung der allein anhand der Korrelationen ermittelten Inklusion führt.

Um die Itemanzahl in den Subsets im Verhältnis zu den jeweiligen Gesamttests einzubeziehen, wird der in Formel (7) dargestellte Inklusions-Index vorgeschlagen. Der Index ist so aufgebaut, dass er einen Wertebereich zwischen -1 und +1 annehmen kann (um diesen Wertebereich zu erhalten wird im Nenner zum Quotienten der Itemanzahlen 1 addiert). Auch ist beim Einsetzen von Werten zu beachten, dass die Itemanzahlen des Tests mit der größeren quadrierten Korrelation (n_{sg} , n_{gg}) anders verrechnet werden, als die Itemanzahlen des Tests mit der kleineren Korrelation (n_{sk} , n_{gk}).

$$\frac{r_{12}^2 - r_{21}^2}{\left(\frac{n_{sg}}{n_{gg}} \cdot \frac{n_{gk}}{n_{sk}}\right) / \left(\frac{n_{sg}}{n_{gg}} + \frac{n_{gk}}{n_{sk}}\right) + 1} = \text{Inklusion} \quad (7)$$

wobei:

r_{12} = Korrelation von Subset 1 mit Test 2

r_{21} = Korrelation von Subset 2 mit Test 1

n_{sg} = Anzahl der Items im Subset mit der größeren Korrelation

n_{gg} = Gesamtzahl der Items im Test mit der größeren Korrelation

n_{sk} = Anzahl der Items im Subset mit der kleineren Korrelation

n_{gk} = Gesamtzahl der Items im Test mit der kleineren Korrelation

Zur Veranschaulichung: Wenn man die Werte gemäß des letzten Beispiels in Formel (7) einsetzt erhält man anstatt eines Inklusions-Index von .05, der sich aus der Differenz der quadrierten Korrelationen ergeben würde (Formel 7), folgendes:

$$\frac{.40 - .35}{\left(\frac{10}{10} \cdot \frac{20}{5}\right) / \left(\frac{10}{10} + \frac{20}{5}\right) + 1} = \frac{.05}{\frac{4}{5} + 1} = 0.02\bar{7}$$

Die Verhältnisse bezüglich der Itemanzahlen wirken sich zwar in Formel 7 auf die Höhe des Inklusions-Indexes aus, aber sie können das Verhältnis nicht zugunsten des jeweils anderen Tests umkehren. Die Differenz zwischen den Korrelationen in den Subsets bleibt trotz der Korrektur ausschlaggebend. Man könnte sich zwar auch Indizes vorstellen, die aufgrund der Itemanzahlen, sogar zu einer Umkehrung kommen. Es fragt sich jedoch, ob eine solche Umkehrung der Verhältnisse aufgrund der Itemanzahlen legitim ist. Außerdem könnte eine stärkere Betonung der Itemanzahl im Vergleich zu den Korrelationen dazu führen, dass lange Skalen, für die ja bereits statistisch eine größere Chance besteht, dass ein Subset mit einer hohen Korrelation mit einem anderen Test auftritt, zusätzlich dadurch „bevorzugt“ werden, dass die Subsets mit größerer Wahrscheinlichkeit eher kleine Anteile der Gesamtskala darstellen, als bei einem kurzen Test.

Um die Wirkweise des zweiten Inklusions-Indexes (Formel 7) eingehender darzustellen, wurde zunächst eine Beispielrechnung für die Differenzen der quadrierten Korrelationen zwischen -1 und 1 durchgeführt. Der Wertebereich des Nenners von Formel (7), der den Einfluss der Itemanzahlen wiedergibt, liegt zwischen 0 und 1. Abbildung 4.3 stellt den Einfluss der Differenz der quadrierten Korrelationen (Zähler von Formel 7) und der Itemanteile (Nenner von Formel 7) auf den Inklusions-Index dar. Es zeigt sich, dass der Itemanteil einen mindernden Einfluss auf den Inklusions-Index hat: Wenn das Subset, das die größere

Korrelation erbringt, nur aus einem kleinen Teil der Items des Gesamttests besteht und zugleich das Subset, das die geringere Korrelation erbringt, sich aus allen Items des Tests zusammensetzt, dann tendiert der Nenner von Formel 7 gegen eins. Unter günstigen Bedingungen der Itemanzahl entspricht der Inklusions-Index somit der Differenz der quadrierten Korrelationen und nimmt somit Werte zwischen -1 und +1 an (s. Abbildung 4.3). Wenn jedoch die Itemanzahl des Subsets mit der größeren Korrelation der Itemanzahl des Gesamttests entspricht und zugleich das Subset mit der kleineren Korrelation sich nur aus einem minimalen Anteil des zugehörigen Gesamttests zusammensetzt, dann tendiert der Nenner von Formel 7 gegen zwei. Unter diesen ungünstigsten Bedingungen für die Inklusion nimmt der Inklusions-Index somit nur noch Werte zwischen -0.5 und +0.5 an.

Auf diese Weise wird berücksichtigt, dass ein Test vermutlich genereller ist, wenn er bereits auf der Basis von wenigen Items hoch mit einem anderen korreliert. Dennoch ist der oben dargestellte Index natürlich nicht die einzige Möglichkeit das Inklusionsverhältnis von Tests oder Skalen zu ermitteln. Die Ermittlung der maximalen Korrelation aus allen möglichen Korrelationen von allen möglichen Subskalen oder Subsets eines Tests mit einem anderen Test kann, bei sehr langen Skalen, auch mit Hilfe von Computer-Programmen sehr langwierig sein.

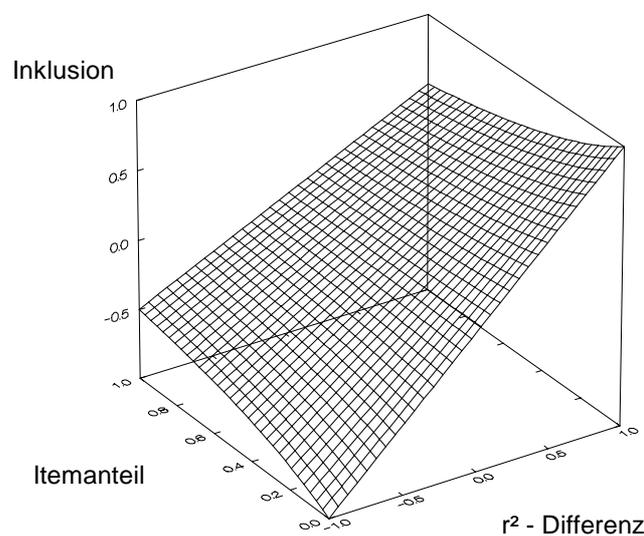


Abbildung 4.3. Inklusions-Index in Abhängigkeit vom Itemanteil und der r^2 -Differenz

Regressionsbasierter Inklusions-Index für Skalen

Eine andere Möglichkeit das wechselseitige Inklusionsverhältnis von Skalen zu bestimmen, basiert auf wechselseitiger multipler Regression. Man kann wechselseitig multiple Regressionen mit allen Items des einen Tests auf das jeweilige Aggregat des anderen Tests

durchführen. Die Differenz der kreuzvalidierten quadrierten multiplen Korrelationen könnte analog zu Formel (6) als Inklusions-Index dienen. Die Kreuzvalidierung ist hierbei zentral, da ansonsten die Generalität längerer Skalen, bei denen mehr Freiheitsgrade für die Maximierung bestehen, überschätzt würde.

Bei der Ermittlung der Inklusion durch wechselseitige multiple Regression ist die Einschränkung zu machen, dass die Items der jeweiligen Skalen in differenzierterer Weise kombiniert werden, als es bei einer einfachen Skalenbildung erfolgt. Vor allem negative Regressionsgewichte stellen (bei korrekt gepolten Items) gegenüber der einfachen Aggregation eine massive Veränderung dar. Um der Skalenbildung näher zu kommen, könnte man alle negativen B-Gewichte im nachhinein auf null setzen und die multiple Regression dann neu bestimmen. Auch empfiehlt es sich wohl, mit einem Backward-Algorithmus alle nicht-signifikanten Items zu eliminieren (der Backward-Algorithmus ist hier sinnvoller als der Forward-Algorithmus, weil dabei zunächst von der tatsächlichen Gesamtskala ausgegangen wird). Diese Methode hat, gegenüber der zuvor dargestellten, den Vorteil, dass sie mit herkömmlicher Statistik-Software zu bewältigen ist, aber sie hat auch den Nachteil, dass die differenzierte Gewichtung bei der multiplen Regression nicht der einfachen Aggregation entspricht. Auch die Möglichkeit, eine Regression über Einheitsgewichtung (mit -1 , 0 und 1) herzustellen, bildet in diesem Fall keine Alternative, da die Items einer Skala in der Regel, positiv miteinander korrelieren werden und somit alle Items das Einheitsgewicht von „eins“ bekommen würden.

Im übrigen wird mit einem regressionsbasierten Inklusions-Index eher das wechselseitige Inklusionspotential der Skalen ermittelt als die tatsächliche Inklusion, da die Regressionsgewichte im Sinne der Maximierung der Vorhersage optimiert werden. Dem könnte man entgegenhalten, dass auch bei der Subset-Bildung eher das Inklusionspotential erfasst wird, da ja nicht die Subsets ausgewertet werden, sondern die Gesamtskala. Die Subsets sind jedoch stets einfacher Bestandteil der Gesamtskala, was die Interpretierbarkeit der Befunde einfacher gestaltet als bei der Regressionsmethode. Allerdings wird auch das wechselseitige, maximale Inklusionspotential mit dem Regressionsindex nicht eindeutig erfasst, da zwar die Gewichte zur Prädiktion der jeweils anderen Skala optimiert sind, die Skala, die als Kriterium eingeht, sich jedoch aus gleichgewichteten Items zusammensetzt.

Bestimmung des Inklusionspotentials anhand des kanonischen Korrelationsindex

Ein Index, der das maximal mögliche Inklusionspotenzial eindeutiger erfasst, ist der allgemeine kanonische Korrelationsindex von Stewart und Love (1968). Dieser Index unterscheidet sich von der herkömmlichen kanonischen Korrelation (Hotelling, 1936) und auch von Cohen's

(1982) Set-Korrelationen dadurch, dass er asymmetrisch ist. Asymmetrie bedeutet hier, dass der Korrelationsindex für die beiden Vorhersagerichtungen unterschiedlich ausfallen kann. Die Eigenschaft der Asymmetrie ist die Voraussetzung dafür, Unterschiede in der wechselseitigen Inklusion, als Hinweis auf Generalitätsunterschiede zwischen Skalen interpretieren zu können. Da kanonische Korrelationen und Set-Korrelationen die symmetrischen Beziehungen zwischen Variablen-Gruppen abbilden, können sie hier nicht direkt verwendet werden. Stewart und Love (1968) bezeichnen den hier mit Inklusion bezeichneten Sachverhalt mit Redundanz und wählten einen informationstheoretischen Hintergrund für ihren Index. Der von ihnen vorgeschlagene Index wird dadurch berechnet, dass die mittleren erklärten Varianzen der kanonischen Variablen (d.h. die Summe der Ladungsquadrate dividiert durch die Anzahl der Variablen im dazugehörigen Variablen-set) mit der zugehörigen quadrierten kanonischen Korrelation multipliziert werden. Diese Berechnung wird dann für alle kanonischen Variablen durchgeführt und das Ergebnis wird gemittelt.

Dieser Index spiegelt das maximale wechselseitige Inklusionspotenzial zwischen zwei Itemmengen wieder und nicht das tatsächliche Inklusionsverhältnis zwischen zwei Skalen. Der Index ist allerdings gerade wegen dieser Eigenschaft interessant: Das wechselseitige Inklusionspotential von Itemmengen kann als Hinweis auf die jeweilige Repräsentativität von Itemmengen gelten. Allerdings ist auch hier eine Kreuzvalidierung notwendig, damit der Index nicht direkt von der Größe des Itempools abhängt. Darüber hinaus würde ein direkter Vergleich der Itempools, auf denen aktuelle Traittheorien basieren, beispielsweise anhand modellspezifischer Persönlichkeitstests, anhand des kanonischen Indexes von Stewart und Love (1968) extrem große Probandenzahlen erfordern, da die Itempools meist mehrere hundert Items umfassen. Wenn man für die Berechnung der kanonischen Korrelationen als Grundlage für den Index pro Prädiktor beispielsweise 10 Probanden veranschlagt (z.B. Tabachnick & Fidell, 1989), dann wären bei Verwendung der 240 Items des NEO-PI-R (Costa & McCrae, 1992) bereits 2400 Probanden erforderlich, bei einer Kreuzvalidierung idealer Weise sogar 4800 Probanden! Derartige Untersuchungen erscheinen zwar grundsätzlich von Bedeutung, können aber aus Aufwandsgründen im Rahmen dieser Arbeit nicht durchgeführt werden. Dies gilt auch für die hier entwickelten Inklusions-Indizes, die auf der Basis multipler Regression ermittelt werden könnten.

4.3.2 Inklusions-Indizes für Faktorwerte

Oftmals bilden Skalen die in den jeweiligen Modellen postulierten Faktoren insofern unzureichend ab, als die Items unterschiedlich hoch auf den Faktoren laden, während sie in den Skalen gleichgewichtet werden. Da die Konstrukte und deren Binnenstrukturierung in der Traitforschung meist erst mit Hilfe ordnungsanalytischer Verfahren, insbesondere der Faktorenanalyse erschlossen werden, stellen Faktoren oder Faktorwerte in der Regel auch konstruktnähere Indikatoren dar als Skalen (z.B. Brocke, 2000). In diesem Sinne ist ein Vergleich der Generalität mit Hilfe von Faktoren aussagekräftiger als mit Hilfe von Skalen. Bevor jedoch ein Generalitätsvergleich sinnvoll durchgeführt werden kann, muss geprüft werden, ob und inwieweit die Faktoren, für die der Generalitätsvergleich durchgeführt werden soll, den von den Modellautoren intendierten Faktoren entsprechen.

Zur Modellkonformität der zu vergleichenden Faktoren

Eine Möglichkeit die Modellkonformität der Faktoren zu überprüfen besteht darin, die intendierte Modellstruktur oder die Struktur einzelner Faktoren zunächst mit Hilfe der KFA zu prüfen und nur dann eine weitere Auswertung vorzunehmen, wenn das Modell insgesamt oder die postulierten Faktoren anhand der Daten bestätigt werden können. Dieses Vorgehen stößt an seine Grenzen, wenn die KFA als Mittel zur Überprüfung von Traitstrukturen grundsätzlich in Frage gestellt wird (s. Kapitel 3.2.1) bzw. wenn die Modellgültigkeit bisher ausschließlich anhand exploratorischer Faktorenanalysen nachgewiesen wurde. Ein weiteres Problem besteht darin, dass nicht klar ist, in welcher Weise die Erwartungen zur intendierten Struktur in die KFA eingehen sollten, man könnte alle intendierten Hauptladungen freisetzen, oder aber – wenn eine Ladungsmatrix der Autoren zur Verfügung steht – diese als Ausgangsmatrix verwenden.

Eine weitere Möglichkeit der Bewertung der Modellkonformität der Faktoren, mit denen ein Generalitätsvergleich durchgeführt werden soll, besteht in der Berechnung von Faktorkongruenzkoeffizienten (Wrigley & Neuhaus, 1955) zwischen der Ladungsmatrix der empirisch ermittelten Faktoren und einer von den Autoren eines Modells als modellkonform eingeschätzten Ladungsmatrix. Wenn die Kongruenzkoeffizienten niedrig ausfallen, kann – je nach Modellannahmen – durch eine orthogonale (Schönemann, 1966) oder schiefwinklige Procrustes-Rotation (Hurley & Cattell, 1962) die Modellkonformität nachträglich erhöht werden. Für den orthogonalen Fall liegen inzwischen auch einige Hinweise zur zufallskritischen Einschätzung der Kongruenzkoeffizienten vor (s. Paunonen, 1997). Das Problem der zufallskritischen Bewertung ist damit allerdings noch nicht gelöst (vor allem nicht für den

schiefwinkligen Fall). Wenn jedoch die Erhöhung der Modellkonformität nicht im Sinne der Modellbestätigung interpretiert wird, sondern nur als Hilfsmittel, um möglichst modellkonforme Faktoren herzustellen, die dann mit anderen Faktoren in Bezug auf die Generalität verglichen werden, kann die Procrustes-Rotation hier dennoch von Nutzen sein.

Der Vergleich einer empirischen Ladungsmatrix mit einer vorgegebenen Ladungsmatrix auf der Basis von Kongruenzkoeffizienten und/oder die konfirmatorische Überprüfung einer vorgegebenen Ladungsmatrix stellen Möglichkeiten für die Absicherung der Modellkonformität als Voraussetzung für den Generalitätsvergleich von Faktoren dar. Wenn die Modellkonformität der Faktoren ermittelt bzw. hergestellt wurde, kann der Generalitätsvergleich auf der Basis von Inklusionsindizes durchgeführt werden.

Inklusions-Index für Faktorwerte auf der Basis multipler Regression

Inklusions-Indizes für Faktorwerte können auf der Basis multipler Regression erstellt werden. Dabei sind allerdings eine Reihe von Problemen zu lösen, weshalb derartige Indizes vergleichsweise komplex werden. Aus diesem Grund werden regressionsbasierte Indizes in dieser Arbeit nicht weiter verfolgt. Um jedoch einen Eindruck von den Möglichkeiten und Problemen regressionsbasierter Indizes zu vermitteln, werden im folgenden die Schritte, die bei der Ermittlung regressionsbasierter Inklusions-Indizes durchlaufen werden müssen, dargestellt.

- Zunächst werden getrennte Faktorenanalysen für die beiden Inventare durchgeführt, in denen die zu vergleichenden Faktoren und die zugehörigen Faktorwerte ermittelt werden. Die methodische Umsetzung der Faktorenanalysen einschließlich der Methode zur Bestimmung der Faktorwerte muss als Grundlage für den Generalitätsvergleich natürlich in beiden Analysen gleich sein.
- Als nächstes wird die Modellkonformität der Faktoren anhand von Kongruenzkoeffizienten überprüft und ggf. mit Hilfe von Procrustes-Rotation erhöht. Wenn trotz einer Procrustes-Rotation keine befriedigende Modellkonformität der Faktoren hergestellt werden kann, ist der Generalitätsvergleich auf der Basis von Faktorwerten nicht sinnvoll. Wenn die Modellautoren auch Test-Skalen anbieten, kann in solchen Fällen ein Generalitätsvergleich auf der Basis von Skalen durchgeführt werden (s.o.), obgleich die Schwierigkeit die Faktoren zu replizieren auch die Validität der Skalen einschränkt.
- Im nächsten Schritt werden die Faktorwerte des einen Faktors durch die Items des jeweils anderen Faktors wechselseitig in einer multiplen Regression vorhergesagt. Um einerseits alle

Items von vornherein einzubeziehen, andererseits unbedeutende Anteile zu vermeiden, bietet sich hier ein Vorgehen nach dem Backward-Algorithmus an. Die Differenz der quadrierten multiplen Korrelationen aus den wechselseitigen Vorhersagen ist ein Maß für das Inklusionsverhältnis der Faktoren.

- Natürlich müssen die wechselseitigen multiplen Regressionen kreuzvalidiert werden, um Zufallseinflüsse zu minimieren, vor allem auch um den Einfluss einer unterschiedlichen Anzahl von Items bei der multiplen Regression zu minimieren.

Bevor die Differenz der quadrierten multiplen Korrelationen als Index der Inklusion verwendet werden kann, ist aber noch einiges zu beachten:

Die multiple Regression kann in der Regel schon aus technischen Gründen nicht mit allen Items eines Inventars durchgeführt werden, weil die Itemzahl teilweise extrem groß ist (z.B. 241 im NEO-PI-R) und die Probandenzahl ein vielfaches (das 7-10fache) der Prädiktoranzahl sein muss (Cohen & Cohen, 1983). Außerdem besteht mitunter das Problem, dass die Vorhersage des jeweils anderen Faktors wesentlich durch Items erfolgt, die den modellkonformen Faktor, der sich in Faktorenanalysen der Items ergibt, nicht definieren, d.h. kaum auf diesem laden. Oder allgemeiner ausgedrückt, die Kombination der Beta-Gewichte, die zur optimalen Vorhersage des zweiten Faktors anhand des Itemmaterials des ersten Faktors führt, kann mehr oder weniger stark von der Kombination der Beta-Gewichte abweichen, die sich bei der Berechnung der modellkonformen Faktorwerte ergibt². Eine Möglichkeit, sowohl die Anzahl der Variablen in den Regressionsgleichungen zu verringern als auch die Komplexität der Kombinationen von Beta-Gewichten etwas einzuschränken, besteht darin, bei der Vorhersage der Faktorwerte des jeweils anderen Faktors nur die Variablen zu verwenden, die einen statistisch bedeutsamen Beitrag zur Vorhersage leisten (z.B. Backward-Regression). Bei der Berechnung wird dann für die Beta-Gewichte der Variablen, die nicht in die Regressionsgleichung aufgenommen wurden, der Wert null eingesetzt.

Auch wenn diese Form der Reduktion der Anzahl der Variablen durchgeführt wird, kann die Kombination der verbleibenden Beta-Gewichte bei der Vorhersage des jeweils anderen Faktors mehr oder weniger stark von den Beta-Gewichten abweichen, die zur Berechnung des

² Die b- oder Beta-Gewichte (je nachdem ob standardisiert wurde) zur Berechnung der Faktorwerte aus den standardisierten Variablen bieten sich als Vergleichsmaßstab an, weil auch für die Variablen, die den Faktor nicht konstituieren, Beta-Gewichte zur Schätzung dieser Faktorwerte ermittelt werden können. Auf die Bedeutung der Beta-Gewichte zur Berechnung der Faktorwerte aus den Variablen für die Interpretation von Faktoren wurde von Cattell (1962) und Becker (1999) hingewiesen.

jeweiligen Original-Faktors erforderlich waren. Aus diesem Grund kann die Differenz der kreuzvalidierten, quadrierten multiplen Korrelationen nicht allein als Index für den Generalitätsvergleich von Faktoren herangezogen werden. Die Abweichungen der Kombination der Beta-Gewichte bei der Vorhersage des jeweils anderen Faktors von der Kombination der Beta-Gewichte, die bei der Berechnung der Faktorwerte für den modellkonformen Ausgangsfaktor auftritt, müssen berücksichtigt werden. Diese Berücksichtigung der Modellkonformität der Beta-Gewichte bei der multiplen Regression führt dazu, dass Inklusions-Indizes auf dieser Basis vergleichsweise komplex und wenig eindeutig interpretierbar sind.

Ein weiterer Nachteil der Verrechnung der Beta-Gewichte ergibt sich aus deren Instabilität und der Notwendigkeit von Kreuzvalidierungen. Auch wenn mit Hilfe von Kreuzvalidierungen angemessene Schätzungen der multiplen Korrelationen erfolgen können, fragt sich, wie man zu einer angemessenen Schätzung der kreuzvalidierten Beta-Gewichte kommt (beispielsweise durch Mittelung), vor allem wenn ein Item bei der Kreuzvalidierung mal einen nicht signifikanten und mal einen signifikanten Beitrag leistet. Auch wenn dieses Problem befriedigend gelöst werden könnte, verbleibt trotz Kreuzvalidierung eine schwer durchschaubare Interdependenz der multiplen Korrelationen von der Anzahl der Prädiktoren. Ein weiterer Nachteil besteht in der Metrik eines regressionsbasierten Inklusions-Indexes für Faktorwerte. Durch die Verrechnung der Beta-Gewichte kann die Metrik des Indexes nicht i.S. statistisch erklärter Varianz interpretiert werden. Aus diesen Gründen wird hier ein alternatives Vorgehen zur Ermittlung der wechselseitigen Inklusion von Faktoren anhand der Korrelationen zwischen Items und Faktorwerten vorgeschlagen.

Inklusions-Index für Faktoren mit Hilfe der Korrelationen von Items und Faktorwerten

Zunächst werden die Korrelationen der Faktorwerte eines Faktors mit allen Items, die in die Faktorenanalyse zur Ermittlung dieses Faktors eingingen, berechnet. Die Items, die in die Faktorenanalyse zur Ermittlung eines Faktors eingehen, werden im folgenden auch als die dem Faktor „zugehörigen“ Items bezeichnet. Wenn eine Hauptkomponentenanalyse durchgeführt wurde, entsprechen diese Korrelationen den Faktorladungen der Items, wenn eine Faktorenanalyse (z.B. Hauptachsenanalyse mit iterativer Kommunalitätenschätzung) durchgeführt wurde, entsprechen die Ladungen nicht exakt den Korrelationen der Faktorwerte mit den Items. Des weiteren werden die Korrelationen der Faktorwerte eines zweiten, in der Generalität mit dem ersten Faktor zu vergleichenden Faktors, mit denselben Items berechnet.

Die i bis n Korrelationen zwischen den Faktorwerten eines Faktors 1 und den zugehörigen Items der Gruppe 1, werden mit r_{i11} bezeichnet, wobei der erste Index i ein Laufindex ist, um

jede einzelne Korrelation zu kennzeichnen, der zweite Index 1 die Itemgruppe und der dritte Index den Faktor (1 oder 2) angibt. Die Korrelationen zwischen den Faktorwerten des Faktors 2 mit denselben, diesem Faktor nicht zugehörigen Items der Gruppe 1 werden mit r_{i12} bezeichnet. Nun kann man die mittlere quadrierte Korrelation von Itemgruppe 1 mit den zugehörigen Faktorwerten des Faktors 1 mit der mittleren quadrierten Korrelation von Itemgruppe 1 mit den Faktorwerten des Faktors 2 vergleichen. Man kann beispielsweise die Differenz D dieser mittleren quadrierten Korrelationen bilden (Formel 8).

$$D = \frac{\sum_i^n r_{i11}^2}{n} - \frac{\sum_i^n r_{i12}^2}{n} \quad (8)$$

Man vergleicht somit die statistische Vorhersage der Itemvarianzen der Itemgruppe 1 durch die Faktorwerte des anhand von Itemgruppe 1 berechneten Faktors 1 mit der statistischen Vorhersage der Itemvarianzen von Itemgruppe 1 durch den Faktor 2, der nicht aus den Items der Gruppe 1 gebildet ist. Es ist sehr unwahrscheinlich, dass die Differenz D negativ wird, da die Items in der Regel mit den Faktorwerten des zugehörigen Faktors höher korrelieren werden, als mit den Faktorwerten eines Faktors, der aus einer anderen Faktorenanalyse hervorgeht.

Zur Berechnung der wechselseitigen Inklusion muss die Differenz D für beide zu vergleichenden Faktoren und deren Itemgruppen ermittelt werden. Die Differenz der beiden Differenzen gibt dann die wechselseitige Inklusion an: Der Faktor, dessen Faktorwerte die Varianz der Items des jeweils anderen Faktors im Durchschnitt besser erklärt, ist derjenige, der den jeweils anderen Faktor eher einschließt. Formal ergibt sich für diesen Inklusions-Index I_{f1} folgendes:

$$I_{f1} = \left(\frac{\sum_i^n r_{i11}^2}{n} - \frac{\sum_i^n r_{i12}^2}{n} \right) - \left(\frac{\sum_j^o r_{j22}^2}{o} - \frac{\sum_j^o r_{j21}^2}{o} \right) \quad (9)$$

Dabei ist o die Anzahl der Items in Itemgruppe 2, die der Faktorenanalyse zur Berechnung von Faktor 2 zugrunde liegen. Wenn I_{f1} positiv ist, kann mehr Varianz der Itemgruppe 2 durch den Faktor 1 erklärt werden als umgekehrt. Wenn I_{f1} negativ ist, kann mehr Varianz der Itemgruppe 1 durch den Faktor 2 erklärt werden als umgekehrt. Um den Begriff der Inklusion zu verwenden, kann man auch feststellen, dass bei einem positiven Index Faktor 1 eher Faktor 2 einschließt, und wenn I_{f1} negativ ist, Faktor 2 eher Faktor 1 einschließt.

Allerdings muss bei dem in Formel 9 dargestellten Inklusions-Index folgendes Problem bewältigt werden: Hohe durchschnittliche quadrierte Korrelationen des Faktorwertes 2 mit den Items der Itemgruppe 1 können auch durch die Items der Gruppe 1 bedingt sein, die nur eine geringe oder Null-Korrelation mit dem Faktor 1 aufweisen. Es sind dies die Items der Gruppe 1, die Null-Ladungen auf Faktor 1 aufweisen bzw. die in der Hyperebene (siehe z.B. Gorsuch, 1983) des Faktors 1 liegen. In diesem Fall würden die Korrelationen lediglich die Unterschiedlichkeit der Faktoren widerspiegeln, nicht aber die Inklusion des jeweils anderen Faktors.

Diese Situation sei an einem Beispiel verdeutlicht: Ein Faktor wird durch fünf Extraversions-Items markiert und 10 Neurotizismus-Items haben Nebenladungen nahe null auf diesem Faktor (bzw. liegen in der Hyperebene). Wenn man nun die Generalität des Extraversionsfaktors nicht mit einem anderen extraversionsähnlichen Faktor, sondern mit einem Neurotizismusfaktor vergleicht, dann könnte dieser leicht mit den 10 Items, die Nullladungen auf dem Extraversionsfaktor aufweisen, hoch korrelieren und so insgesamt mehr Varianz der Items erklären als der zugehörige Extraversionsfaktor. Dies würde nun nicht dafür sprechen, dass der Neurotizismusfaktor den Extraversionsfaktor einschließt, da die hohe durchschnittliche Korrelation gerade durch die Items zustande kommt, die nicht substantziell auf dem Extraversionsfaktor laden.

Um diesen Fall zu vermeiden, kann man eine Gewichtung der Korrelationen der Items mit dem Varianzanteil vornehmen, den die Items mit dem zugehörigen Faktorwert gemeinsam haben. Je höher ein Item mit den Faktorwerten eines zugehörigen Faktors korreliert, desto höher sollte das Gewicht der Korrelation dieses Items mit den Faktorwerten des jeweils anderen Faktors sein. Daher wird der Inklusions-Index anhand gewichteter Mittelwerte berechnet, wobei die quadrierten Korrelationen zwischen den Items und den ihnen zugehörigen Faktorwerten die Gewichte für die Mittellung der quadrierten Korrelationen der Items mit den Faktorwerten sowohl des zugehörigen als auch des jeweils anderen Faktors darstellen (Formel 10). Diese Gewichtung hat zur Folge, dass hohe quadrierte Korrelationen der Faktorwerte mit den nicht zugehörigen Items nur dann ein hohes Gewicht bekommen, wenn die Items hoch mit dem ihnen zugehörigen Faktor korrelieren.

$$I_{f2} = \frac{\sum_i^n r_{i11}^2 r_{i11}^2 - \sum_i^n r_{i11}^2 r_{i12}^2}{\sum_i^n r_{i11}^2} - \frac{\sum_j^o r_{j22}^2 r_{j22}^2 - \sum_j^o r_{j22}^2 r_{j21}^2}{\sum_j^o r_{j22}^2} \quad (10)$$

Auf das oben genannte Beispiel bezogen hätte dies zur Folge, dass die substanziellen quadrierten Korrelationen des Neurotizismusfaktors mit den 10 Neurotizismus-Items mit Werten nahe null gewichtet würden, da die Korrelationen der 10 Neurotizismus-Items mit dem zugehörigen Faktor, der aus den insgesamt 15 Items konstituiert wird, nahe null ist. Die Gewichtung hat somit die Funktion bei der Berechnung der Inklusion die inhaltliche Bedeutung der Faktoren zu berücksichtigen. Zwar wird man in der Regel ähnliche Faktoren vergleichen, aber auch geringere inhaltliche Verschiebungen könnten zu dem o.g. Problem führen, weshalb der Inklusions-Index I_{12} (Formel 10) sicherlich angemessener ist als der in Formel 9 ermittelte Inklusions-Index.

Ein Aspekt des in Formel 12 dargestellten Indexes besteht darin, dass die quadrierten Korrelationen der einen Faktor konstituierenden Items mit dem jeweils anderen Faktor größer ausfallen können als die Korrelationen der Items mit dem zugehörigen Faktor. Ob dies für einen Inklusions-Index zugelassen werden sollte, ist eine inhaltliche Frage. Wenn man zulässt, dass ein Faktor an einer Itemgruppe, an der er nicht ermittelt wurde, Varianz erklärt, die der aus der Itemgruppe ermittelte Faktor nicht erklärt, dann lässt man inhaltliche Bedeutungsverschiebungen zwischen den Faktoren mit in den Generalitätsvergleich einfließen. Dann kann der Fall eintreten, dass ein Faktor Varianz der ihn nicht konstituierenden Items erklärt, die in der Vergleichslösung von einem anderen, zusätzlichen Faktor erklärt wird. Es ist durchaus als sinnvoll anzusehen, beim Generalitätsvergleich auch in dieser Form Bedeutungsunterschiede zwischen den Faktoren zu berücksichtigen. Man könnte den Standpunkt vertreten, dass Generalitätsunterschiede mit, zumindest geringfügigen, inhaltlichen Verschiebungen einhergehen müssen. Um den Generalitätsvergleich auf einer möglichst eindeutigen Grundlage durchzuführen, kann es jedoch auch als Vorteil gelten, wenn die Faktoren wechselseitig nur die Varianz der sie nicht konstituierenden Items erklären, mit denen diese in die zugehörigen Faktoren eingehen. Dies würde bedeuten, dass der jeweils andere Faktor nur die Varianz erklären darf, die Bestandteil des zu vergleichenden Faktors ist. Die Unterschiede zwischen den Faktoren und den Varianzen, die sie erklären, würde auf diese Weise von vornherein auf die Varianz beschränkt, die die beiden Faktoren erklären.

Bei einem Inklusions-Index, der diese Bedingung erfüllt, stellen die Korrelationen der Items mit dem sie konstituierenden Faktor die zulässige Obergrenze für die Korrelationen dieser Items mit dem jeweils anderen Faktor dar. Diese Bedingung kann für die quadrierten Korrelationen der Itemgruppe 1 mit den Faktorwerten des Faktors 2 folgendermaßen ausgedrückt werden:

$$r_{i12}^{*2} = \begin{cases} r_{i12}^2 = r_{i11}^2 & \text{für } r_{i12}^2 \leq r_{i11}^2 \\ r_{i12}^2 = r_{i11}^2 & \text{für } r_{i12}^2 > r_{i11}^2 \end{cases} \quad (11)$$

Entsprechend gilt für die quadrierten Korrelationen des Itemgruppe 2 mit den Faktorwerten des Faktors 1 folgendes:

$$r_{i21}^{*2} = \begin{cases} r_{i21}^2 = r_{i22}^2 & \text{für } r_{i21}^2 \leq r_{i22}^2 \\ r_{i21}^2 = r_{i22}^2 & \text{für } r_{i21}^2 > r_{i22}^2 \end{cases} \quad (12)$$

Wenn man die in Formel 11 und 12 ermittelten quadrierten Korrelationen in Formel 10 einsetzt, ergibt sich der angepasste Inklusions-Index I_{f3} in Formel (13).

$$I_{f3} = \frac{\sum_i^n r_{i11}^2 r_{i11}^2 - \sum_i^n r_{i11}^2 r_{i12}^{*2}}{\sum_i^n r_{i11}^2} - \frac{\sum_j^o r_{j22}^2 r_{j22}^2 - \sum_j^o r_{j22}^2 r_{j21}^{*2}}{\sum_j^o r_{j22}^2} \quad (13)$$

Der Index nimmt Werte zwischen -1 und $+1$ an, wobei Werte >0 auf eine höhere Generalität des durch die Itemgruppe 1 gebildeten Faktors hindeuten.

Inklusions-Indizes auf Basis der Verteilungen der Varianzen

Ein konzeptuelles Problem, der in Formel 10 und 13 dargestellten Inklusions-Indizes I_{f2} und I_{f3} , aber auch der übrigen bisher dargestellten Inklusions-Indizes, hängt mit deren alleiniger Orientierung an der wechselseitig erklärten Varianz zusammen. Im Prinzip sollten generellere Faktoren an einer größeren Itemmenge Varianz erklären als weniger generelle Faktoren, weshalb auch vermutet werden kann, dass generellere Faktoren mehr Varianz aufklären als weniger generelle Faktoren. Andererseits besteht zugleich die Möglichkeit, dass die generelleren Faktoren an den einzelnen Items jeweils weniger Varianz erklären. Insofern kann nicht ausgeschlossen werden, dass ein weniger genereller Faktor, der große Mengen relativ spezifischer Varianz erklärt, insgesamt mehr Varianz erklärt als ein generellerer Faktor, der insgesamt wenig Varianz an sehr vielen Items erklärt. Ein Faktor, der wenig Varianz an vielen Items erklärt, kann somit unter Umständen einen kleineren Inklusions-Index erhalten als ein Faktor, der viel Varianz an wenigen Items erklärt. Dieser Fall ist zwar unwahrscheinlich, wenn Faktoren, die ähnliche Inhaltsbereiche abdecken, verglichen werden, er kann aber bei Faktoren, die auf eher unterschiedlichem Material beruhen, auftreten. Aus diesem Grund könnte sich der

Gedanke der wechselseitigen Inklusion als unvollständig für den Generalitätsvergleich von Faktoren erweisen.

Um diesen Nachteil der Inklusions-Indizes zu kompensieren, kann man den Aspekt der Verteilung der erklärten Varianzen in die Indizes einfließen lassen. Wenn ein Faktor eine hohe Generalität aufweist, sollte er die Varianz möglichst vieler Items erklären. Die erklärte Varianz verteilt sich somit auf viele Items, während sich die erklärte Varianz eines weniger generellen Faktors vermutlich auf weniger Items verteilt. Neben der Höhe der wechselseitig erklärten Varianzen, die in den Inklusions-Indizes betrachtet wird, kann demnach die Verteilung der Varianzen Information über die Generalität von Faktoren enthalten.

Die durch die Faktoren erklärten Varianzen werden in Formel 10 und 13 über die quadrierten Korrelationen zwischen den Faktorwerten und den Items geschätzt. Man kann somit die Verteilung der quadrierten Korrelationen zwischen den Items und Faktorwerten über die Varianz der quadrierten Korrelationen erfassen³. Somit kann die Varianz der quadrierten Korrelationen als zusätzliches Maß für die Generalität herangezogen werden. Dabei ist zu beachten, dass eine geringe Varianz der quadrierten Korrelationen für eine breite Verteilung der erklärten Varianz und somit für eine hohe Generalität steht.

Es macht allerdings wenig Sinn, allein die Varianz der quadrierten Korrelationen der Items mit den Faktorwerten als Generalitätsindex zu verwenden. Dies ergibt sich daraus, dass für einen Faktor, der mit keinem Item einer Gruppe korreliert, die Varianz der quadrierten Korrelationen null wäre, was auf eine maximale Generalität hindeuten würde. Um derartig absurde Ergebnisse zu vermeiden, erscheint es zweckmäßig, die Information über die Varianz der quadrierten Ladungen mit der Information über die Höhe der quadrierten Ladungen zu kombinieren. Dies kann in diesem Rahmen durch eine Kombination der Varianzen mit den oben dargestellten Inklusions-Indizes I_{f2} und I_{f3} erfolgen.

Die Varianz der quadrierten Ladungen kann dabei vergleichend betrachtet werden, wobei sowohl die quadrierten Korrelationen der Items mit den Faktorwerten der Faktoren, die sie konstituieren, als auch quadrierten Korrelationen der Items mit den Faktorwerten der Faktoren, die sie nicht konstituieren, zugrundegelegt werden können. Des Weiteren sollte sich die Ermittlung der Varianz auf die Items konzentrieren, die hohe Ladungen auf dem jeweiligen Faktor aufweisen. Je höher die Ladungen, desto bedeutsamer sollte eine geringe Varianz der

³ Der Begriff der Varianz wird hier somit in zwei verschiedenen Bedeutungen verwendet: Zum einen kennzeichnet die erklärte Varianz die Summe der quadrierten Korrelationen zwischen Item und Faktorwerten, und zum anderen wird hier die Varianz dieser quadrierten Korrelationen als Generalitätsmaß diskutiert.

quadrierten Ladungen sein. Insofern sollte die Varianz mit der Höhe der jeweiligen Ladungen gewichtet werden.

Wie in den obigen Darstellungen, wird der Index für zwei fiktive Itemgruppen 1 und 2 formuliert. Für die mit der Höhe der quadrierten Korrelationen gewichtete Varianz der quadrierten Korrelationen zwischen den Items der Itemgruppe 1 mit den Faktorwerten des Faktors 1, den sie konstituieren, ergibt sich somit folgendes:

$$s_{11}^2 = \frac{\sum_i^n r_{i11}^2 \left(r_{i11}^2 - \frac{\sum_i^n r_{i11}^2}{n} \right)^2}{\sum_i^n r_{i11}^2} \quad (14)$$

Analog zu Formel 14 können die Varianzen der quadrierten Korrelationen für beide zu vergleichenden Faktoren in beiden Itemgruppen berechnet werden. Es ergeben sich somit vier zu berechnende Varianzen (s_{11}^2 , s_{12}^2 , s_{21}^2 , s_{22}^2). Hier sollen diese Varianzen in die in Formel 10 und 13 dargestellten Inklusions-Indizes I_{f2} und I_{f3} Eingang finden. Für den durch die Varianzen ergänzten Inklusions-Index I_{f2} ergibt sich der in Formel 15 dargestellte Index I_{f4} . Es gibt verschiedene Möglichkeiten, eine Gewichtung der quadrierten Korrelationen mit den Varianzen der quadrierten Korrelationen vorzunehmen. Da kleine Varianzen für hohe Generalität stehen sollen, wurde in Formel 15 eine Division durch die jeweiligen Varianzen vorgenommen.

$$I_{f4} = \frac{\sum_i^n \left(\frac{r_{i11}^2}{s_{11}^2} - \frac{r_{i12}^2}{s_{12}^2} \right) r_{i11}^2}{\sum_i^n r_{i11}^2} - \frac{\sum_i^n \left(\frac{r_{i22}^2}{s_{22}^2} - \frac{r_{i21}^2}{s_{21}^2} \right) r_{i22}^2}{\sum_i^n r_{i22}^2} \quad (15)$$

Für den Inklusions-Index I_{f3} , bei dem die Korrelationen der Items mit dem sie konstituierenden Faktor die zulässige Obergrenze für die Korrelationen dieser Items mit dem jeweils anderen Faktor darstellen (Formel 13), ergibt sich bei zusätzlicher Division durch die jeweiligen Varianzen folgendes:

$$I_{f5} = \frac{\sum_i^n \left(\frac{r_{i11}^2}{s_{11}^2} - \frac{r_{i12}^2}{s_{12}^2} \right) r_{i11}^2}{\sum_i^n r_{i11}^2} - \frac{\sum_i^n \left(\frac{r_{i22}^2}{s_{22}^2} - \frac{r_{i21}^2}{s_{21}^2} \right) r_{i22}^2}{\sum_i^n r_{i22}^2} \quad (16)$$

Die Inklusions-Indizes I_{f_4} und I_{f_5} , die die Varianzen der quadrierten Korrelationen berücksichtigen, haben gegenüber den Inklusions-Indizes, die allein auf der wechselseitig erklärten Varianz beruhen, den Nachteil, dass sie als kombinierte Maße, die verschiedene Aspekte der Generalität integrieren, weniger eindeutig interpretiert werden können. Während die in Formel 10 und 13 dargestellten Inklusions-Indizes I_{f_2} und I_{f_3} einen Wertebereich zwischen null und eins aufweisen, können die in Formel 15 und 16 dargestellten Indizes Werte deutlich über eins annehmen, wenn die Varianzen der quadrierten Ladungen sehr klein werden. Um die vergleichende Bewertung der Generalität von Faktoren im folgenden auf eine optimale Grundlage zu stellen, erscheint es insofern sinnvoll, sowohl die Inklusions-Indizes I_{f_2} und I_{f_3} als Maße für die wechselseitig erklärte Varianz als auch die Inklusions-Indizes I_{f_4} und I_{f_5} , die die Verteilungen der Varianzen berücksichtigen, heranzuziehen.

5. Auswahl der Lösungsvorschläge und Traitmodelle für die empirische Untersuchung

5.1 Auswahl der Lösungsvorschläge

Im Kapitel 3 wurden die Spezifikationen für die KFA und EFA festgelegt, im Kapitel 4 wurden Vorschläge zur Ermittlung der Repräsentativität von Modellen und der Generalität von Traits entwickelt. Im folgenden wird dargestellt, welche der in Kapitel 4 erarbeiteten Vorschläge anhand der in Kapitel 3 benannten Spezifikationen empirisch untersucht werden sollen.

Sowohl modellspezifische als auch modellübergreifende Faktorenanalysen werden in der psychometrischen Traitforschung immer wieder durchgeführt (z.B. Vassend & Skrandal, 1997; Barbaranelli & Caprara, 1996). Da bei den modellspezifischen Analysen die erwarteten Strukturen bekannt sind, können diese sowohl anhand der EFA als auch anhand der KFA untersucht werden. Die modellspezifischen Analysen haben darüber hinaus den Vorteil, dass die Nachweisbarkeit der Modelle anhand der in Kapitel 3 vorab festgelegten und ausführlich begründeten Spezifikationen für die KFA und EFA untersucht werden kann.

Wie bereits dargelegt, dienen die modellspezifischen Analysen dem Nachweis der jeweils theoretisch erwarteten Struktur, nicht dem Nachweis alternativer Strukturen. Daraus folgt, dass, wenn die Extraktionskriterien der EFA eine andere Faktorenzahl als die theoretisch postulierte nahe legen, diese Modelle nicht weiterverfolgt werden. Aus diesem Grund werden in den Analysen mit dem modellspezifischen Variablenpool die Ladungsmuster von Lösungen mit anderen als den postulierten Faktorenanzahlen nicht dargestellt. In dem Fall, dass die Extraktionskriterien der postulierten Anzahl von Faktoren entsprechen, ist allerdings die Darstellung des Ladungsmusters erforderlich, um sicherzustellen, dass es sich um die intendierten Faktoren und nicht um eine substanziiell andere Struktur mit zufällig gleicher Faktorenzahl handelt. Wenn jedoch die Extraktionskriterien nicht die postulierte Faktorenzahl indizieren, müssen nur die Ergebnisse der Extraktionskriterien dokumentiert werden, da das Ladungsmuster dann ohnehin nicht mehr der intendierten Struktur entsprechen kann. Auch erscheint der Nachweis alternativer Strukturen innerhalb eines modellspezifischen Variablenpools wenig attraktiv. Alternative Strukturen werden lediglich in den Analysen mit einem modellübergreifenden Variablenpool verfolgt.

Die modellübergreifenden Analysen können, wegen ihres exploratorischen Charakters, nur anhand der EFA durchgeführt werden. Mit diesen Analysen kann die Frage beantwortet werden, welche Faktoren welcher Modelle sich modellübergreifend im gegebenen Itempool abbilden

lassen. Derartige Analysen sind der übliche Weg zum Vergleich psychometrischer Traitmodelle (z.B. Barbaranelli & Caprara, 1996) und sollen daher hier nicht ausgespart werden. Es ist auch möglich, dass bei den gemeinsamen Analysen ein neues Traitmodell als Integration der einbezogenen Modelle aufgewiesen wird. Die Erstellung eines neuen Traitmodells ist jedoch nicht primäres Ziel der Arbeit, sondern – wie ausgeführt – die vergleichende Evaluation der bestehenden Modelle auf der Basis der dargestellten Kriterien zur Durchführung multivariater Analysen.

Bezüglich der Evaluation der Generalität ist anzumerken, dass die konventionellen Möglichkeiten diesbezüglich sehr begrenzt sind. Zwar können aus modellübergreifenden, hierarchischen Faktorenanalysen, die ohnehin durchgeführt werden sollen, Ordnungsgrade für Faktoren ermittelt werden. Es bleibt jedoch fraglich, ob die gemeinsamen Analysen eine Struktur ergeben werden, in denen die Ordnungsgrade verschiedener, in unterschiedlichen Modellen postulierter Faktoren direkt verglichen werden können. Die Ordnungsgrade von Faktoren in getrennten, d.h. modellspezifischen Analysen können wiederum nicht direkt verglichen werden, da sie eine Äquivalenz der Generalitäten des Item-Materials der verschiedenen Fragebogen voraussetzen. So ist nicht klar, ob ein Faktor erster Ordnung im EPQ-R dasselbe Generalitätsniveau aufweist wie ein Faktor erster Ordnung im NEO-PI-R. Derartige Generalitätsvergleiche können jedoch anhand der in Kapitel 4.2 entwickelten Inklusions-Indizes durchgeführt werden.

Der Inklusions-Index für Skalen wird hier nicht eingesetzt, da in den bestehenden Modellen zunächst die Faktoren als Indikatoren der Konstrukte gesehen wurden und Fragebogen-Skalen lediglich zur vereinfachten Messung der über Faktoren ermittelten Konstrukte dienen. Daher werden im folgenden nur Inklusions-Indizes für Faktorwerte (s. Kapitel 4.3.2) ermittelt.

Da bisher noch keine Vergleichswerte für die Generalitäts-Indizes vorliegen, haben die hier berichteten Ergebnisse vorläufigen Charakter. Um keine voreiligen Schlüsse zu ziehen, werden die Analysen zum Generalitätsvergleich hier nur für Faktoren durchgeführt, die in verschiedenen Modellen in ähnlicher Weise abgebildet werden bzw. einen ähnlichen Bedeutungshof aufweisen (Extraversion und Neurotizismus). Im folgenden wird eine Kurzübersicht über die durchgeführten Analysen gegeben:

- 1) Modellspezifische Analysen: Überprüfung der Traitmodelle in den für die Modelle spezifischen Fragebogen für die in Kapitel 3 festgelegten Spezifikationen der KFA und EFA (siehe Kapitel 6.2)

- 2) Modellübergreifende Analysen: Untersuchung der Überlappungen zwischen relevanten Traitmodellen in gemeinsamen Analysen der Fragebogen für die in Kapitel 3 festgelegten Spezifikationen der EFA (siehe Kapitel 6.3)
- 3) Vergleichende Ermittlung der Generalität anhand der Inklusions-Indizes für Faktorwerte für ausgewählte Faktoren (siehe Kapitel 6.4)

5.2 Auswahl der Traitmodelle

Die vorliegende Untersuchung sollte „klassische“, konkurrierende Traitmodelle einbeziehen, deren Proponenten seit längerem über die angemessene Anzahl und Generalität von Faktoren diskutieren. Dazu eignen sich besonders die Modelle von Cattell (1994a, b), das FFM in der Variante von Costa und McCrae (1992a,b), sowie das Modell von Eysenck (1991). In den hier benannten Literaturangaben finden sich jeweils Aussagen, die die Überlegenheit des jeweiligen Modells dokumentieren sollen. Gerade die wechselseitigen Überlegenheits- und Integrationsansprüche lassen vergleichende Analysen interessant erscheinen. Vor allem, weil der hier verwendete Methodenstandard (s. Kapitel 3), völlig unabhängig von den Forschungstraditionen aller drei genannten Modelle formuliert und allein anhand einschlägiger Methoden-Arbeiten begründet wurde. Daher erscheint interessant, welches Modell bei Verwendung der hier formulierten Standards am ehesten oder mit den geringsten Problemen nachgewiesen werden kann.

Neben den drei genannten Modellen, die – auch wenn sie durchaus noch aktuell sind – bereits eine forschungshistorische Bedeutung erlangt haben, wurde noch das alternative Fünf-Faktoren-Modell (AFFM) einbezogen (Zuckerman et al., 1993), das eine sehr viel kürzere Entwicklungsperiode hatte und bei dem der Fokus vor allem auf der Abbildung biopsychologisch relevanter Persönlichkeitsdimensionen lag. Wegen dieser Besonderheiten und weil Costa und McCrae (1992b) aufgrund einer Zielrotation eine Umdeutung des AFFM im Sinne des FFM für zweckmäßig hielten, erscheint eine vergleichende Einbeziehung dieses Modells interessant. Es fragt sich, wie gut dieses Modell mit Modellen, denen eine viel breitere Forschungsentwicklung vorausging, konkurrieren kann.

Sicherlich könnten noch weitere Modelle einbezogen werden, etwa die Modelle von Andresen (1995, 2000), Becker (1999), Benet-Martínez und Waller (1997) oder Cloninger et al. (1991).

All diese Modelle haben jedoch nicht die Verbreitung der ersten drei Modelle erreicht. Der Fokus auf die dominierenden Traitmodelle erscheint für diese Arbeit zweckmäßig, da gerade die Diskussion um diese anerkannten Modelle für die weitere Forschung zentral ist.

6. Empirische Anwendung der erarbeiteten Vorschläge auf konkurrierende Traitmodelle

6.1 Beschreibung der Untersuchung

6.1.1 Probandengruppe

Insgesamt wurden 1242 Probanden untersucht. Die Gesamtgruppe wurde in zwei Subgruppen, A und B, geteilt. Einige Fragebogen wurden in der Gesamtgruppe andere in Teilgruppe A oder B eingesetzt (s. Kapitel 6.1.2, Tabelle 6.3). Aus den in Kapitel 3 dargestellten Vorschlägen zur optimalen Durchführung der Faktorenanalysen geht hervor, dass mindestens 500 Probanden untersucht werden sollten. Andererseits sollen insgesamt vier Traitmodelle untersucht werden. Aufgrund der begrenzten Belastbarkeit und wegen des Risikos von Verfälschungstendenzen bei Überlastung von Probanden konnten nicht alle Fragebogen bei der gesamten Probandengruppe eingesetzt werden.

Im folgenden werden die Gesamtgruppe und die Teilgruppen A und B anhand demografischer Merkmale beschrieben. 91,5% der Probanden in Gruppe A und 95,3% der Probanden in Gruppe B lebten zum Zeitpunkt der Erhebung in den neuen Bundesländern.

Häufigkeit

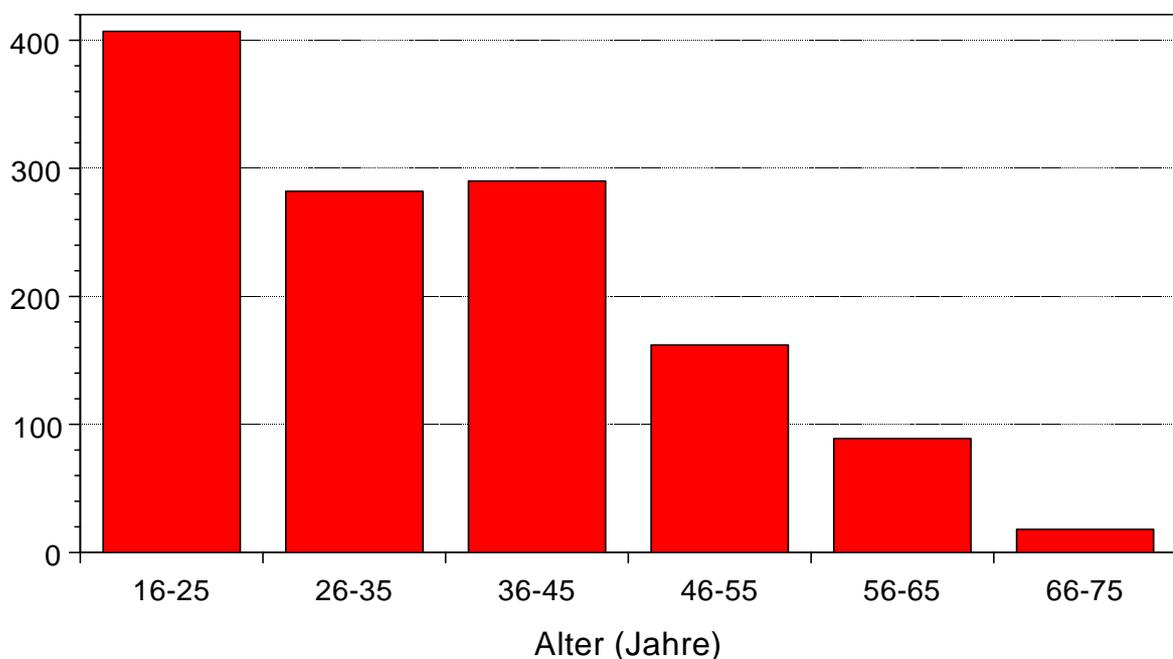


Abbildung 6.1. Häufigkeitsverteilung in der Gesamtgruppe (N=1248)

Die Altersverteilung für die Gesamtgruppe von 1242 Probanden ist in Abbildung 6.1 dargestellt. Die unterschiedlichen Altergruppen waren in den beiden Teilgruppen annähernd gleichhäufig repräsentiert. Die genauen Angaben für die Teilgruppen sind im Anhang (Abbildung A 6.1) dargestellt.

Die Altersverteilung ist sowohl in der Gesamtgruppe, als auch in beiden Teilgruppen deutlich rechtsschief. Damit in den Analysen keine altersbedingten Verzerrungen auftreten, wobei die Altersunterschiede hier Kohortenunterschiede abbilden, erscheint eine Gleichverteilung optimal. Wenn alle Alterskohorten etwa gleichhäufig in der Probandengruppe vertreten sind, können Verzerrungen aufgrund spezifischer Kohorteneffekte minimiert werden. Allerdings kann davon ausgegangen werden, dass sich verschiedene Alterskohorten in Bezug auf Persönlichkeitstraits insgesamt nur wenig unterscheiden (z.B. Goldberg, Sweeney, Merenda & Hughes, 1998). Die Teilgruppe der 66 bis 75jährigen umfasst jedoch nur 18 Probanden und ist damit zu klein um in Gewichtung-Prozeduren oder in Teilgruppenanalysen einzugehen. Aus diesem Grund wurden diese 18 Probanden von den weiteren Analysen ausgeschlossen. Die verbleibende Gesamtgruppe umfasst somit 1224 Probanden im Alter zwischen 16 und 65 Jahren.

Die weiteren demografischen Angaben werden für die verbleibenden 1224 Probanden dargestellt (s. Tabelle 6.1). Es fällt auf, dass an der Untersuchung etwa zwei Drittel Frauen und ein Drittel Männer teilgenommen hat. Die Geschlechtsverteilung und die Altersmittelwerte und Standardabweichungen sind in der Gesamtgruppe und in den beiden Teilgruppen etwa gleich.

Tabelle 6.1

Geschlecht und Alter der Gesamtgruppe und der Teilgruppen A und B

Gruppe	N	Weiblich	Männlich	Alter	
Gesamt	1224	831 (68%)	399 (32%)	M=34.1	S=12.7
Teilgruppe A	590	384 (65%)	206 (35%)	M=33.8	S=12.4
Teilgruppe B	634	442 (70%)	192 (30%)	M=34.4	S=13.0

Anmerkung. Die Prozentwerte in Klammern beziehen sich auf die jeweilige Gruppengröße. M=Mittelwert, S=Standardabweichung

Der jeweils höchste Ausbildungsstand der Probanden ist in Tabelle 6.2 dargestellt. Der insgesamt hohe Anteil an Probanden ohne Abitur wirkt sich vorteilhaft auf die Repräsentativität der Stichprobe aus, dennoch ist die Stichprobe in Richtung auf ein höheres Bildungsniveau verzerrt. Laut Angaben des Statistischen Bundesamtes Deutschland (Pressemitteilungen, November 1999) stieg der Anteil der studienberechtigten Schulabgänger an der

relevanten Altersgruppe der Bevölkerung kontinuierlich von 10.6% im Jahr 1970 auf 36.4% im Jahr 1995. Laut Mikrozensus des Statistischen Bundesamtes (2001) hatten 1999 etwa 30% der Bevölkerung ein Abitur oder eine höhere Ausbildung. In der hier untersuchten Gesamtgruppe haben jedoch 46.7% der Probanden Abitur oder ein abgeschlossenes Hoch- oder Fachhochschulstudium. Insgesamt nahmen nur 110 Psychologiestudenten (56 in Gruppe A, 54 in Gruppe B) und 105 Studenten anderer Fachrichtungen (53 in Gruppe A, 52 in Gruppe B) an der Untersuchung teil.

Tabelle 6.2

Ausbildungsstand der Gesamtgruppe und der Teilgruppen A und B

Ausbildung	Gesamt	Teilgruppe A	Teilgruppe B
keine Ausbildung/Hilfsschule	55 (4.5)	7 (1.2)	47 (7.5)
10. Klasse	71 (5.8)	46 (7.8)	25 (4.0)
Berufs- oder Fachschule	500 (40.7)	302 (51.2)	195 (31.0)
Abitur oder Fachabitur	307 (25.0)	137 (23.2)	170 (27.0)
abgeschl. Hochschul- oder Fachhochschulstudium	291 (23.7)	96 (16.3)	193 (30.6)
keine Angaben	6 (0.5)	2 (0.3)	4 (0.6)

Auch wenn die Stichprobe in Bezug auf die Bildung für die Bevölkerung der Bundesrepublik Deutschland keinesfalls als repräsentativ anzusehen ist, ist sie diesbezüglich ausgewogener als in vielen anderen Untersuchungen im Rahmen psychometrischer Traitforschung, in denen oft nur Studenten untersucht werden (z.B. Benet-Martinez & Waller, 1997; Carver & White, 1994; Cattell & Delhees, 1973; Church & Burke, 1994; Goldberg, 1990). Eine weitere Erhöhung der Repräsentativität der Stichprobe für die Bevölkerung der Bundesrepublik Deutschland könnte durch Probandenauswahl oder Gewichtung (Lynn, Hampson & Lee, 1988; Todt & Fürntratt, 1963) hergestellt werden.

Eine Gewichtung hat gegenüber einer Probandenauswahl den Vorteil, dass auch die Teilstichproben einigermaßen groß bleiben. Aus diesem Grund sollen die Analysen im folgenden sowohl mit den ungewichteten als auch mit nach bestimmten Parametern gewichteten Stichproben durchgeführt werden. Da die Stichprobe in Bezug auf die Bildung ohnehin einigermaßen ausgewogen ist, erscheint eine Gewichtung in Bezug auf Alter und Geschlecht zweckmäßig.

In Bezug auf das Geschlecht bietet eine gleichgewichtete Stichprobe den Vorteil, dass die Ergebnisse nicht wie in den ungewichteten Daten durch die überwiegende Zahl von Frauen bestimmt werden. Dabei sei angemerkt, dass getrennte Analysen für Frauen und Männer zwar

geschlechtsspezifische Effekte aufdecken können, nicht jedoch über die Geschlechtsgruppen maximal generalisierbare Effekte. Gerade wenn sich die Ergebnisse für die Teilgruppen unterscheiden sollten, wäre die Interpretation der Befunde auf Basis der ungewichteten Stichprobe problematisch. Im Falle von Unterschieden stellt eine gleichgewichtete Gruppe eine sinnvolle Möglichkeit der integrativen Betrachtung dar. Insofern stellen geschlechtsspezifische Analysen, gleichwohl ebenfalls interessant, keine Alternative zur Analyse anhand einer gleichgewichteten Stichprobe dar. Geschlechtsspezifische Analysen stellen im Rahmen dieser Arbeit allerdings auch deswegen keine Alternative dar, weil in den hier untersuchten Strukturmodellen, zwar mitunter geschlechtsspezifische Mittelwertsunterschiede für bestimmte Konstrukte, aber keine geschlechtsspezifischen Strukturunterschiede postuliert werden.

Dieselbe Argumentation kann auch für das Alter der Probanden geltend gemacht werden. Die spezifischen Altershäufigkeiten in der Stichprobe führen dazu, dass die Ergebnisse in den ungewichteten Daten stärker von den jüngeren Jahrgängen (vor allem 16-25jährige) bestimmt werden. Da es keinen inhaltlichen Grund gibt, die Altersgruppen zwischen 16 und 65 Jahren ungleich zu gewichten, erscheint auch hier eine Gleichgewichtung angebracht. Die Gleichgewichtung des Alters kann hier anhand der fünf Altersgruppen, die jeweils 10 Jahre umfassen, vorgenommen werden. Die zusätzlich in die Auswertungen einbezogene, gewichtete Stichprobe ist also in Bezug auf Geschlecht und Alter gleichverteilt. Dabei erscheint es angebracht, dass auch innerhalb der fünf Altersgruppen das Geschlecht und innerhalb des Geschlechts die Altersgruppen gleichverteilt sind. Die Geschlechts x Altersverteilung in der ungewichteten Gesamtgruppe und deren Gleichverteilung in den gewichteten Daten kann Tabelle 6.3 entnommen werden.

Tabelle 6.3

Kreuztabelle für die Geschlechts- und Altershäufigkeiten in der Gesamtgruppe

Alter / Jahre	ungewichtet		gewichtet	
	weiblich	männlich	weiblich	männlich
16 – 25	286	120	122	122
26 – 35	180	98	122	122
36 – 45	200	89	122	122
46 – 55	101	61	122	122
56 – 65	59	30	122	122
Summe:	826	398	610*	610*

Anmerkung. * Eigentlich müssten nach der Gewichtung $2 \times 612 = 1224$ Probanden vorliegen. Dies würde jedoch 122,4 Probanden in den einzelnen Teilgruppen erfordern, was natürlich nicht möglich ist. Die Werte sind somit das Ergebnis ganzzahliger Rundung. Die Gewichtung wurde mit der Gewichtungsfunktion von SPSS für Windows über eine Variable, die im Zähler die Gesamthäufigkeit dividiert durch die Zellenanzahl (hier $1224/10$) und im Nenner die jeweilige Zellenhäufigkeit enthält, durchgeführt.

In analoger Weise wird eine Gleichgewichtung der Geschlechts x Altersverteilung für die Subgruppen A und B durchgeführt. Im Falle fehlender Werte bei einzelnen Fragebogen (s. Tabelle 6.4), wurden die Gewichte zur Herstellung einer Gleichverteilung entsprechend angepasst.

6.1.2 Inventare

Es wurden die Inventare für die hier interessierenden Modelle von Cattell (1994a, b), Costa und McCrae (1992a), Eysenck (1991) und Zuckerman et al. (1993) einbezogen. Tabelle 6.4 gibt Aufschluss über die eingesetzten Fragebogen, deren Zuordnung zu den Modellen sowie die Stichprobengröße, wobei aufgrund fehlender Werte für einzelne Fragebogen teilweise etwas weniger Probanden verfügbar sind, als in den in Tabelle 6.1 ausgewiesenen Gesamt- und Teilstichproben. Die modellspezifischen Analysen anhand der einzelnen Fragebogen können anhand der Gesamtgruppe bzw. anhand der Teilgruppe, in denen die Fragebogen eingesetzt wurden, durchgeführt werden.

Tabelle 6.4

Eingesetzte Fragebogen und Stichprobenumfang

Modell	Fragebogen	Gruppe	Stichprobengröße
PEN-Theorie	Eysenck Personality Questionnaire-Revised (EPQ-R; Ruch, 1999)	Gesamt	1217
16-PF	16-Persönlichkeits-Faktoren-Test (16 PF-R; Cattell et al., 1970; Schneewind & Graf, 1998)	A	556
“BIG-Five“	NEO-PI-R (Costa & McCrae, 1992a; Ostendorf & Angleitner, in Druck)	Gesamt	1213
“Alternative Five“	Zuckerman-Kuhlman-Personality Questionnaire (ZKPQ; Zuckerman et al., 1993; Ostendorf & Angleitner, 1994)	B	636
	Sensation Seeking Scales Form V (SSS-V; Zuckerman et al., 1978; Beauducel et al., 1999)	Gesamt	1219

Anmerkung. Die Stichprobengröße ergibt sich aus der Größe der Gesamtgruppe, Teilgruppe A oder B minus der Anzahl fehlender Werte in dem jeweiligen Fragebogen.

Demgegenüber können die modellübergreifenden Analysen aufgrund der Aufteilung der Fragebogen in zwei Gruppen nur für jeweils drei Modelle gleichzeitig durchgeführt werden. Allerdings reduziert sich die Probandengruppe wegen fehlender Werte für die gemeinsamen Analysen in der Gruppe A auf 554 Probanden und in der Gruppe B auf 620 Probanden. Abbildung 6.2 gibt einen Überblick über die in gemeinsamen Analysen zu vergleichenden Modelle.

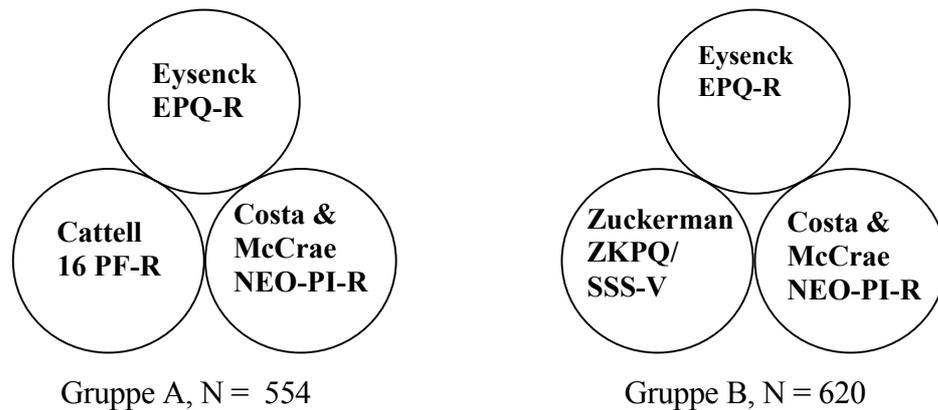


Abbildung 6.2. Vergleich konkurrierender Traitmodelle in gemeinsamen Analysen

6.1.3 Durchführung

Bei 185 Probanden erfolgte die Untersuchung in Räumen der Technischen Universität Dresden, den übrigen 1057 Probanden wurden die Fragebogen zugeschickt. Die Probanden wurden über Aushänge an der Technischen Universität Dresden sowie über Anzeigen in lokalen Zeitungen (z.B. Sächsische Zeitung, Dresdner Neueste Nachrichten, Morgenpost) gewonnen. Die Teilnahme der Probanden war freiwillig. Als Rückmeldung erhielten die Probanden ein auf der Basis des NEO-PI-R ermitteltes Persönlichkeitsprofil. Die Validität der Ergebnisse sollte durch das Interesse der Probanden an der Rückmeldung ihrer Ergebnisse gewährleistet werden.

Man kann zwar trotz des Interesses der Probanden an ihrer persönlichen Rückmeldung die Validität zu Hause ausgefüllter Fragebogen anzweifeln. Dabei sollte man jedoch in Rechnung stellen, dass in vielen großen Fragebogenstudien im Bereich psychometrischer Traitforschung entweder überhaupt keine Angaben darüber zu finden sind, wo die Fragebogen ausgefüllt wurden (z.B. Benet & Waller, 1997; Costa, Busch, Zonderman & McCrae, 1986; Goldberg et al., 1998; McCrae & Costa, 1985), oder aber die Stichprobe auf Studenten bzw. Psychologiestudenten beschränkt ist (z.B. Church & Burke, 1994; Goldberg, 1990; Paunonen, Jackson, Trzebinski, Forsterling, 1992), wobei oft keine Angaben über den Ort des Ausfüllens vorliegen. In einigen Studien wird allerdings auch explizit angegeben, dass die Fragebogen zumindest teilweise zu Hause ausgefüllt wurden (z.B. Eysenck, Eysenck & Barrett, 1985; Heath, Cloninger & Martin, 1994; McCrae & Costa, 1987; Ruch, 1999; Saucier & Ostendorf, 1999; van Heck et al., 1994) und in wenigen Studien wird explizit angegeben, dass die Fragebogen in der Universität ausgefüllt wurden (z.B. Borkenau & Ostendorf, 1991; Draycott & Kline, 1995). Die vorliegende Studie

weicht somit in dieser Hinsicht nicht von dem in der psychometrischen Traitforschung üblichen Vorgehen ab und sollte insofern keine systematischen Verzerrungen gegenüber anderen großangelegten Traitstudien aufweisen. Dies ist auch insofern sinnvoll, als diese Studie auf Aspekte der Variablenauswahl, der statistischen Auswertung und Interpretation in der psychometrischen Traitforschung fokussiert. Eine weitergehende Überprüfung der Validität der Daten wird im Kapitel 6.2.1 anhand der Lügen-Skala des EPQ-R vorgenommen.

Amelang, Schmidt-Rathjens, und Yousfi (2001) legten eine Studie zum Vergleich zwischen den Ergebnissen von per Post verschickten, zu Hause ausgefüllten und in Gruppen unter kontrollierten Bedingungen ausgefüllten Fragebogen vor. Es ergaben sich eine Reihe signifikanter Unterschiede zwischen den Mittelwerten und Standardabweichungen verschiedener Persönlichkeitsskalen für zu Hause und in Gruppen durchgeführte Untersuchungen. Jedoch kommen Amelang et al. (2001) zu dem Ergebnis, dass die Unterschiede aufgrund geringer Effektgrößen ohne praktische Relevanz waren. Die geringen Unterschiede waren aufgrund der großen Stichprobe (5133 Probanden) statistisch bedeutsam. Darüber hinaus wurde zwar festgestellt, dass sich die Varianz-Kovarianzmatrizen der zu Hause und in Gruppen ausgefüllten Fragebogen unterschieden, jedoch waren die Rangfolgen der Koeffizienten extrem hoch korreliert. Somit kommen Amelang et al. (2001) auch in Bezug auf die Korrelationsmatrizen zu dem Schluss, dass die Unterschiede zwischen zu Hause und in Gruppen unter kontrollierten Bedingungen ausgefüllten Fragebogenwerten ohne praktische Relevanz waren. Auch wenn die Ergebnisse von Amelang et al. (2001) nur unter Vorbehalt auf die hier untersuchte Item- und Probandengruppe übertragen werden können, deuten sie darauf hin, dass das Verschicken von Fragebogen nicht zu nennenswerten Validitätseinbußen führen muss.

Die Probanden der Gruppen A und B benötigten für das Ausfüllen der Fragebogen zwischen einer und drei Stunden. Die Folge der eingesetzten Fragebogen für die Gruppen A und B ist in Tabelle 6.5 dargestellt.

Tabelle 6.5

Abfolge der eingesetzten Tests in Gruppe A und B in den Fragebogen-Heften

Gruppe A	Gruppe B
16 PF	SSS-V
SSS-V	ZKPQ
EPQ-R	EPQ-R
NEO-PI-R	NEO-PI-R

6.1.4 Statistische Analyse

Die Spezifikationen der durchgeführten EFA und KFA finden sich in Kapitel 3.2. Die EFA wurden mit SPSS 9.0 für Windows (1999), die KFA mit LISREL 8 (Jöreskog & Sörbom, 1993), die Trasid-Rotation mit einem Programm von Beauducel (1997) durchgeführt. Die in Kapitel 4 dargestellten Indizes wurden mit Hilfe der Matrizen-Syntax von SPSS 9.0 für Windows berechnet.

6.1.5 Interne Konsistenzen der analysierten Fragebogenskalen

Obwohl Maße der internen Konsistenz im Rahmen dieser Arbeit nur bedingt zur Bewertung der Qualität der Daten und der eingesetzten Fragebogen herangezogen werden, da geringe Konsistenzen mit hoher Generalität von Traits einhergehen können (siehe Kapitel 4), sollen zu Vergleichszwecken Cronbach's Alpha und die Split-Half-Reliabilitäten im Folgenden berichtet werden. Tabelle 6.6 enthält Cronbach's Alpha, und Tabelle 6.7 enthält die Split-Half-Reliabilität für die ungewichtete Stichprobe. Die internen Konsistenzen für die gewichtete Stichprobe wichen nur minimal von denen der ungewichteten Stichprobe ab, weshalb diese Werte hier nicht berichtet werden. Zur leichteren Orientierung wurden Werte unter .70, die auf Reliabilitätsprobleme hindeuten können, in Tabelle 6.6 und 6.7 fett gedruckt.

Für den 16 PF-R wurden die Reliabilitäten nur für die 16 Skalen berichtet, da im 16 PF-R die sechs Globalfaktoren (zweiter Ordnung) nur auf der Ebene von Faktorwerten ausgewertet werden. Die Ermittlung der Reliabilitäten von Faktorwerten nach Mosier (1943) setzt Kennwerte für die Reliabilitäten der Einzelitems voraus, die hier nicht vorliegen. Das in Amthauer, Brocke, Liepmann und Beauducel (1999) dargestellte Verfahren könnte zwar im Prinzip verwendet werden, ist jedoch bisher wenig erprobt. Aus diesem Grund wurde die Darstellung der internen Konsistenzen hier auf Skalenwerte beschränkt.

Es zeigt sich, dass im 16 PF-R, im NEO-PI-R und in der SSS-V Skalen mit Cronbach's Alpha und Split-Half-Reliabilitäten unter .70 auftreten. Es fällt auf, dass interne Konsistenzen unter .70 nur bei Skalen auftreten, die zwischen acht und elf Items lang sind, bei den längeren Skalen des EPQ-R und ZKPQ sowie bei den ebenfalls sehr langen Gesamtskalen des NEO-PI-R sind alle Konsistenzen über .70. Dies deutet darauf hin, dass man im Bereich von Persönlichkeitstraiten mit langen Skalen operieren sollte.

Tabelle 6.6
Cronbach's Alpha der eingesetzten Fragebogen

EPQ-R N=1217	16 PF-R N=556	NEO-PI-R N=1213		SSS-V N=1219	ZKPQ N=636
E .87	A .67	N1 .84	V1 .70	TAS .80	Imp .84
N .88	B .74	N2 .73	V2 .66	DIS .69	N-Ä .84
P .71	C .69	N3 .87	V3 .68	ES .60	A-F .76
	E' .74	N4 .75	V4 .60	BS .44	So .85
	F .62	N5 .62	V5 .75	SSS .82	Akt .75
	G .60	N6 .81	V6 .59		
	H .88	E1 .72	G1 .67		
	I .70	E2 .79	G2 .68		
	L .69	E3 .80	G3 .69		
	M .70	E4 .72	G4 .68		
	N' .68	E5 .64	G5 .80		
	O' .78	E6 .81	G6 .79		
	Q1 .70	O1 .75	N .93		
	Q2 .77	O2 .77	E .90		
	Q3 .66	O3 .72	O .87		
	Q4 .74	O4 .66	V .87		
		O5 .77	G .91		
		O6 .49			

Anmerkungen. E = Extraversion; N = Neurotizismus; P = Psychotizismus; O = Offenheit für Erfahrungen; A = Verträglichkeit; A = Wärme; B = Log. Schlußfolgern; C = Emotionale Stabilität; E' = Dominanz; F = Lebhaftigkeit; G = Regelbewußtsein; H = Soziale Kompetenz; I = Empfindsamkeit; L = Wachsamkeit; M = Abgehobenheit; N' = Privatheit; O' = Besorgtheit; Q1 = Offenheit für Veränderung; Q2 = Selbstgenügsamkeit; Q3 = Perfektionismus; Q4 = Anspannung; TAS = Thrill and Adventure Seeking; DIS = Disinhibition; ES = Experience Seeking; BS = Boredom Susceptibility; SSS = Sensation Seeking Scale Gesamtwert; Imp= Impulsivität-Sensation Seeking, N-Ä= Neurotizismus-Ängstlichkeit, A-F= Aggression-Feindseligkeit, So= Soziabilität, Akt= Aktivität; Die Abkürzungen der 30 NEO-PI-R Facetten-Skalen finden sich im Anhang Tabelle A 6.3.

Um die Datenqualität bewerten zu können erscheint ein Vergleich mit den Konsistenzen, die in den Manualen bzw. Referenz-Publikationen zu den Fragebogen angegeben werden, interessant. Für den EPQ-R können Vergleichswerte aus Ruch (1999) herangezogen werden. Für Cronbach's Alpha wurden in Ruch (1999) Werte zwischen .83 und .87 angegeben, während die Werte hier zwischen .71 und .87 liegen. Insbesondere der Wert von .71 für Psychotizismus liegt deutlich unter dem in Ruch (1999) angegebenen Wert von .83. Für die Split-Half-Reliabilität gibt Ruch (1999) Werte zwischen .84 und .88 an, während die Werte hier zwischen .69 und .88 liegen. Die deutlichste Abweichung mit .69 betrifft die Psychotizismus-Skala. Da die Kennwerte für die übrigen Skalen mit denen von Ruch (1999) vergleichbar sind, ist nicht anzunehmen, dass die vorliegende Stichprobe allgemein zu einer geringere Datenqualität geführt hat. Die Unterschiede in der Konsistenz sind somit vermutlich eher auf eine geringere Robustheit der Psychotizismus-Skala zurückzuführen.

Tabelle 6.7
Split-Half-Reliabilitäten der eingesetzten Fragebogen

EPQ-R N=1217	16 PF-R N=556	NEO-PI-R N=1213		SSS-V N=1219	ZKPQ N=636
E .88	A .58	N1 .82	V1 .64	TAS .67	Imp .83
N .87	B .69	N2 .71	V2 .62	DIS .70	N-Ä .85
P .69	C .74	N3 .82	V3 .58	ES .58	A-F .78
	E' .72	N4 .75	V4 .60	BS .40	SY .89
	F .62	N5 .63	V5 .66	SSS .83	Akt .76
	G .61	N6 .83	V6 .53		
	H .88	E1 .69	G1 .58		
	I .64	E2 .82	G2 .66		
	L .66	E3 .82	G3 .74		
	M .68	E4 .68	G4 .68		
	N' .60	E5 .62	G5 .77		
	O' .78	E6 .78	G6 .78		
	Q1 .69	O1 .77	N .93		
	Q2 .76	O2 .75	E .91		
	Q3 .57	O3 .69	O .88		
	Q4 .74	O4 .67	V .81		
		O5 .77	G .85		
		O6 .48			

Anmerkungen. Abkürzungen, siehe Tabelle 6.6. Die Korrelationen zwischen den Testhälften wurden anhand der Spearman-Brown Formel auf die Gesamtestlänge hochkorrigiert.

Für den 16 PF-R liegt Cronbach's Alpha teilweise etwas unter den im Manual angegebenen Werten. Schneewind und Graf (1998) berichten Werte zwischen .66 und .88, während hier für die Skalen F und G Werte unter .66 auftreten (s. Tabelle 6.6). Für die deutsche Fassung des NEO-PI-R liegt noch kein Manual vor (Ostendorf & Angleitner, in Druck), aus dem dem Werte für die interne Konsistenz zu entnehmen wären. Allerdings sind im Katalog der Testzentrale des Hogrefe Verlages für das Publikationsjahr 2000/01 Konsistenzen zwischen .86 und .92 für die fünf Hauptskalen angegeben, wobei aus den Angaben im Katalog nicht entnommen werden kann, ob es sich um Cronbach's Alpha oder Split-Half-Reliabilitäten handelt. Die hier berichteten Werte für die fünf Hauptskalen liegen mit einer Ausnahme (Split-Half Reliabilität von .81 für Verträglichkeit) in diesem Bereich. Für die internen Konsistenzen der Facetten-Skalen werden im Testkatalog Werte zwischen .52 und .85 berichtet, während hier Werte zwischen .48 und .87 auftraten.

In Beauducel et al. (1999) werden für Cronbach's Alpha der SSS-V Subskalen Werte zwischen .46 und .75 und für die Gesamtskala ein Wert von .78 angegeben. In der vorliegenden Stichprobe liegt Cronbach's Alpha für die SSS-V Subskalen zwischen .44 und .80 und für die Gesamtskala bei .82, insgesamt also im Bereich der in Beauducel et al. (1999) angegebenen Werte. Für den

ZKPQ werden in Ostendorf und Angleitner (1994) für Cronbach's Alpha Werte zwischen .71 und .86 angegeben. Die Werte für Cronbach's Alpha liegen hier in derselben Größenordnung, nämlich zwischen .75 und .85.

Insgesamt kann festgehalten werden, dass die internen Konsistenzen etwa in dem Bereich liegen, der auch in entsprechenden Referenz-Publikationen angegeben wird. Lediglich für zwei Skalen des 16 PF-R sowie die Psychotizismus-Skala des EPQ-R traten nennenswert geringere Konsistenzen auf. Da für die internen Konsistenzen jedoch in der vorliegenden Stichprobe teilweise auch größere Werte als in den Referenz-Publikationen festgestellt wurden, sind die genannten geringeren Konsistenzen eher ein Hinweis auf Probleme mit den einzelnen Skalen. Unter dem Gesichtspunkt der internen Konsistenzen kann davon ausgegangen werden, dass insgesamt die Datenqualität der vorliegenden Stichprobe etwa mit der Datenqualität in den Referenz-Publikationen vergleichbar ist.

6.2 Modellspezifische, getrennte Faktorenanalysen für die Fragebogen

6.2.1 Eysenck: EPQ-R

Exploratorische Faktorenanalysen

Der EPQ-R (Ruch, 1999) wurde in der Gesamtgruppe eingesetzt. Aufgrund einiger Probanden mit fehlenden Werten liegen EPQ-R Daten für insgesamt 1217 Probanden vor. Die sogenannte Lügenskala, d.h. die Skala zur Messung der Tendenz zur Beantwortung im Sinne sozialer Erwünschtheit, wurde nicht in die Faktorenanalysen einbezogen, da diese Skala nur bedingt als inhaltlicher Persönlichkeitstrait interpretiert werden kann. Insofern werden hier Lösungen mit jeweils einem Faktor für Extraversion, Neurotizismus und Psychotizismus erwartet.

Der Mittelwert der Lügenskala beträgt für die ungewichtete Stichprobe 7.66 (SD=4.19) und entspricht somit dem in Ruch (1999) publizierten Mittelwert von 7.81 (SD=4.43). Ruch (1999) führte seine psychometrischen Analysen ohne Eliminierung der Probanden mit hohen Lügenwerten durch. Die Beibehaltung der Probanden mit hohen Werten auf der Lügenskala erfolgte auch in Untersuchungen mit der englischen (z.B. Chan & Joseph, 2000) und spanischen Form (Grau & Ortet, 1999) des EPQ-R. Auch Eysenck und Eysenck (1991) empfehlen nicht die pauschale Eliminierung von Probanden mit hohen Lügenwerten und geben aus diesem Grund auch keinen cut-off Wert an. Daher wird im folgenden lediglich eine Überprüfung der Ähnlichkeit der Struktur von Probanden mit hohen und niedrigen Lügenwerten durchgeführt (siehe unten).

Die Ergebnisse der Parallelanalyse auf der Basis der gemittelten Eigenwerte von 100 Zufallslösungen sind in Abbildung 6.3 für die ungewichtete (a) und die gewichtete (b) Gesamtgruppe dargestellt.

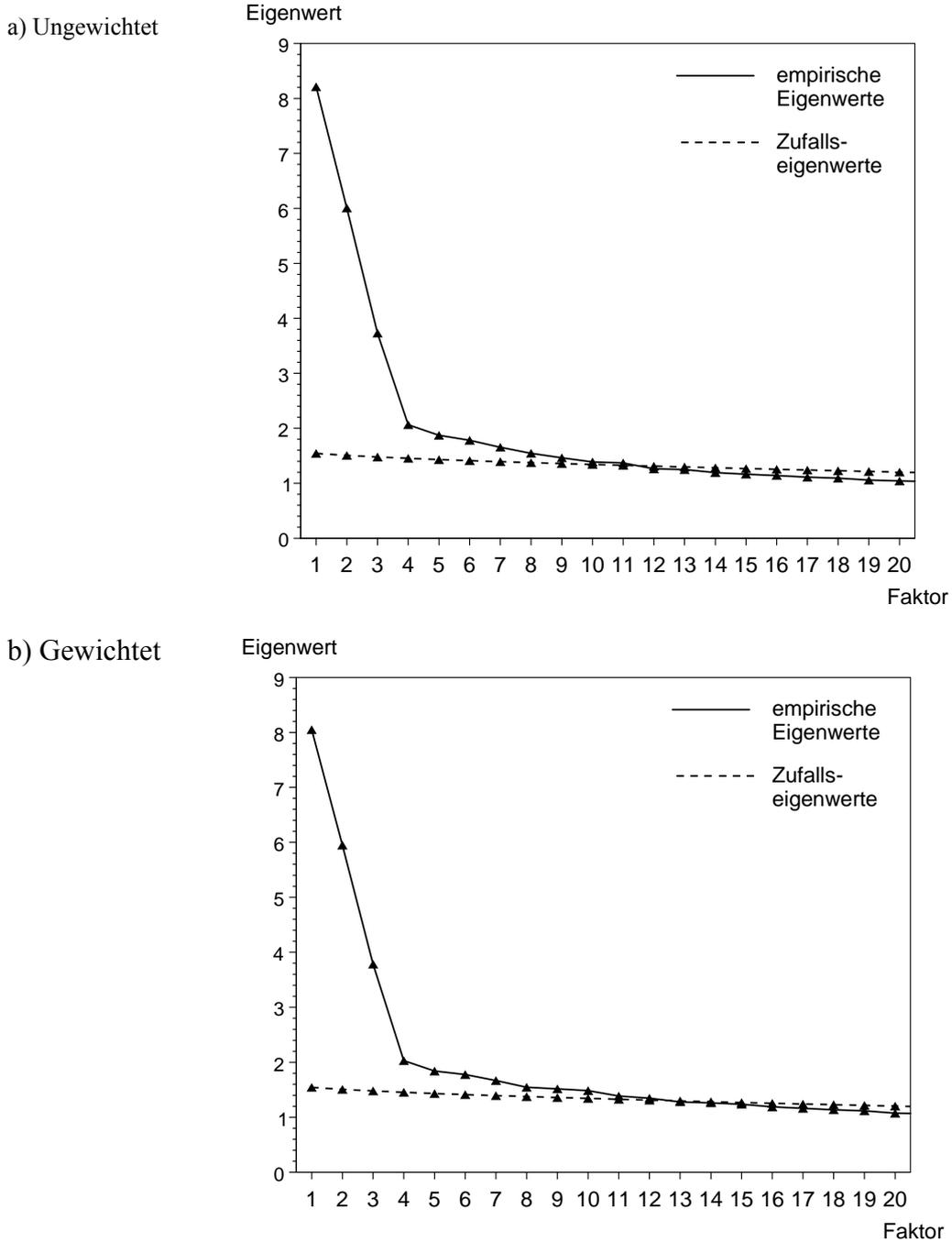


Abbildung 6.3. Parallelanalyse für den EPQ-R (gewichtete und ungewichtete Stichprobe)

Bei der ungewichteten Stichprobe sind 11 empirische Eigenwerte und bei der gewichteten Stichprobe sind 12 empirische Eigenwerte größer als die mittleren Eigenwerte auf der Basis der Zufallslösungen⁴. Die Parallelanalyse spricht somit für die Extraktion von 11 Faktoren in der ungewichteten Stichprobe und für die Extraktion von 12 Faktoren in der gewichteten Stichprobe. Interessant ist, dass der Scree-Test sowohl in der gewichteten als auch in der ungewichteten Stichprobe für die intendierte Faktorenzahl spricht. Die Ergebnisse der Komparabilitäten auf der Basis der Oblimin-rotierten Faktoren sind für die unreduzierte Stichprobe in Tabelle 6.8 dargestellt.

Tabelle 6.8

Faktor-Komparabilitäten für den EPQ-R (N=1217)

		ungewichtete Stichprobe						
	Faktor1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	
2 Faktoren	1.00	.99						
3 Faktoren	1.00	1.00	.99					
4 Faktoren	1.00	1.00	.98	.88				
5 Faktoren	.99	.99	.96	.88	.78			
6 Faktoren	.99	.98	.94	.93	.75	.68		
7 Faktoren	.99	.97	.96	.86	.83	.73	.49	
		gewichtete Stichprobe						
	Faktor1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	
2 Faktoren	.99	.98						
3 Faktoren	.99	.99	.98					
4 Faktoren	.99	.98	.82	.49				
5 Faktoren	.98	.94	.89	.69	.11			
6 Faktoren	.99	.97	.90	.80	.75	.56		
7 Faktoren	.93	.90	.89	.89	.77	.69	.14	

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Es zeigt sich in der ungewichteten Stichprobe, dass die intendierten drei Faktoren zwar äußerst hohe Komparabilitäten aufweisen, allerdings weist der vierte Faktor eine Komparabilität von .88 auf, die über dem Schwellenwert von .85 liegt. In der gewichteten Stichprobe weisen nur in der dreifaktoriellen Lösung alle Faktoren Komparabilitäten > .85 auf, so dass hier die Komparabilitäten die Extraktion von drei Faktoren indizieren.

Da es bei den hier durchgeführten Analysen im modellspezifischen Variablenpool um die Überprüfung der jeweils postulierten Strukturen und nicht um die Entwicklung alternativer Modelle geht, werden in den EFA nur Ladungsmuster dargestellt, wenn die Anzahl der zu extrahierenden Faktoren mit der Anzahl postulierter Faktoren übereinstimmt (siehe Kapitel 5.1). Daher wird im

⁴ Die empirischen Eigenwerte der Konstruktionsstichprobe des EPQ-R sind einer persönlichen Mitteilung von Ruch (24.05.2000) entsprechend den hier dargestellten Eigenwerten sehr ähnlich.

folgenden lediglich das Ladungsmuster für die dreifaktorielle Lösung in der gewichteten Stichprobe dargestellt (siehe Tabelle 6.9).

Tabelle 6.9
Oblimin-Ladungsmuster (delta=0) für den EPQ-R in der gewichteten Stichprobe

Item/Skala	Oblimin (delta=0)			Trasid (Hyperebenendicke=.10)			h ²
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3	
N 027	.60	-.12	.05	.61	-.12	.02	.38
N 031	.60	-.09	-.07	.60	-.09	-.10	.38
N 003	.59	.06	.15	.59	.05	.12	.37
N 062	.57	-.01	.01	.57	.00	-.02	.32
N 022	.56	-.06	-.14	.56	-.05	-.17	.34
N 037	.56	-.05	-.30	.54	-.03	-.34	.41
N 045	.55	-.22	.07	.57	-.22	.04	.36
N 049	.54	-.08	.08	.55	-.08	.05	.31
N 012	.54	.01	-.24	.52	.03	-.27	.35
N 016	.51	.07	-.02	.50	.07	-.04	.26
N 082	.51	-.14	.10	.52	-.15	.07	.29
N 080	.51	-.04	-.05	.51	-.03	-.08	.26
N 098	.48	.02	.05	.48	.02	.02	.23
N 007	.47	-.02	.05	.47	-.02	.02	.22
N 076	.47	-.12	-.25	.46	-.09	-.28	.31
N 034	.46	.07	-.02	.45	.07	-.04	.21
N 085	.46	-.08	-.02	.46	-.08	-.05	.22
N 042	.40	.00	-.27	.38	.02	-.30	.24
N 069	.39	-.03	-.01	.39	-.03	-.03	.16
N 067	.36	-.09	.16	.38	-.10	.14	.16
N 095	.36	.04	-.08	.35	.05	-.10	.14
N 072	.35	-.12	.09	.37	-.12	.07	.15
N 058	.33	.00	-.12	.32	.01	-.14	.12
N 091	.30	.10	.14	.30	.09	.13	.12
N 052	.28	.00	-.16	.27	.01	-.18	.10
P 089	.19	.03	.18	.19	.02	.17	.07
E 010	.04	.69	.00	-.00	.69	.01	.48
E 005	.02	.65	-.08	-.02	.65	-.07	.41
E 074	-.09	.62	-.05	-.13	.63	-.03	.40
E 024	-.09	.62	.09	-.12	.61	.10	.41
E 093	.06	.61	-.02	.02	.61	-.01	.37
E 051	-.10	.59	-.04	-.14	.59	-.02	.36
E 039	-.26	.56	-.03	-.30	.56	-.01	.40
E 088	.03	.55	.00	-.01	.55	.01	.30
E 047	-.06	.54	-.02	-.10	.54	-.01	.30
E 035	-.05	.51	-.15	-.09	.52	-.14	.27
E 015	-.01	.49	.15	-.03	.48	.16	.28
E 029	.07	.49	.07	.04	.48	.08	.26
E 056	.00	.47	-.05	-.04	.48	-.05	.22
E 033	.00	.46	.00	-.03	.46	.01	.22
E 044	-.08	.45	.03	-.11	.45	.04	.22
E 019	.02	.42	.02	.00	.41	.02	.17
E 054	-.02	.40	.02	-.05	.40	.03	.17
E 064	-.11	.33	.14	-.13	.32	.15	.15
E 066	.28	.28	.21	.27	.26	.20	.21
E 060	-.12	.28	-.02	-.14	.28	-.01	.10
E 001	-.02	.28	.01	-.04	.28	.01	.08
E 036	.05	.26	.20	.04	.25	.20	.12

(Fortsetzung und Anmerkungen nächste Seite)

Fortsetzung Tabelle 6.9Oblimin-Ladungsmuster ($\delta=0$) für den EPQ-R in der gewichteten Stichprobe

Item/Skala	Oblimin ($\delta=0$)			Trasid (Hyperebenenendicke=.10)			h ²
	Faktor 1	Faktor 2	Faktor 3	Faktor 1	Faktor 2	Faktor 3	
P 021	-.05	.00	.48	-.02	-.04	.49	.24
P 040	.01	-.03	.47	.04	-.07	.47	.21
P 086	-.03	.06	.41	-.01	.03	.42	.18
P 013	-.07	.06	.40	-.05	.03	.40	.17
P 017	-.00	.03	.39	.02	.00	.39	.15
P 101	-.27	.00	.36	-.25	-.03	.38	.21
P 077	.22	.18	.34	.23	.15	.34	.21
P 046	.10	.09	.33	.11	.07	.33	.14
P 002	.15	.16	.33	.16	.14	.33	.17
P 070	.10	.05	.33	.11	.02	.33	.12
P 006	-.12	.03	.31	-.11	.00	.32	.11
P 071	.00	.00	.27	.01	-.02	.27	.07
P 079	-.04	.06	.27	-.03	.04	.28	.08
P 025	.09	.05	.26	.10	.03	.26	.08
P 099	.00	-.08	.26	.02	-.10	.26	.07
P 061	-.04	.03	.26	-.03	.01	.26	.07
P 057	-.16	-.08	.26	-.14	-.10	.27	.09
P 075	.11	.00	.25	.13	-.02	.25	.07
E 090	-.03	.19	.24	-.03	.17	.24	.11
P 020	-.07	-.07	.22	-.05	-.09	.23	.06
P 094	.01	-.07	.22	.03	-.08	.22	.05
P 041	.11	.15	.21	.12	.13	.21	.09
P 102	-.04	.14	.20	-.04	.12	.21	.07
P 026	-.07	-.16	.20	-.04	-.17	.20	.06
P 009	-.06	.06	.20	-.05	.05	.20	.05
P 100	-.05	-.01	.17	-.04	-.02	.18	.03
P 084	.10	.00	.17	.11	-.01	.16	.04
P 030	.13	.05	.16	.14	.04	.16	.05
P 048	.04	.00	.14	.05	-.02	.14	.02
P 011	-.05	-.03	.14	-.04	-.04	.14	.02
P 065	.12	-.04	.12	.13	-.05	.12	.03
P 097	-.02	-.09	.10	-.01	-.10	.11	.02
Interkorrelationen der Faktoren							
Faktor 2	-.06			.02			%- Varianz:
Faktor 3	-.01	.13		-.01	.19		19.4

Empirische Eigenwerte*: 8.05 5.95 3.78 2.03 1.84 1.78 1.67 1.54 1.52 1.48**Zufalls-Eigenwerte*:** 1.54 1.51 1.48 1.45 1.43 1.41 1.39 1.38 1.36 1.34

Anmerkungen. * Es wurden die (unreduzierten) PCA-Eigenwerte der unrotierten Faktoren dargestellt. Die Kommunalitäten (h²) und die erklärte Varianz wurden jedoch für die unrotierte Hauptachsenlösung nach Kommunalitäteniteration dargestellt. E= Extraversion, N= Neurotizismus, P= Psychotizismus. Ladungen mit Beträgen > .30 wurden fett gedruckt.

Sowohl das Oblimin- als auch das Trasid-rotierte Ladungsmuster entspricht den theoretischen Erwartungen, da jeweils ein Faktor für Neurotizismus, Extraversion und Psychotizismus aufgezeigt werden konnte. Allerdings haben nur 11 der 32 Psychotizismus-Items Hauptladungsbeträge über .30. Dies könnte zu Problemen bei der Replikation des Psychotizismus-Faktors in Datenanalysen anhand eines modellübergreifenden Variablenpools führen. Die Varianzaufklärung der vorliegenden Lösung ist etwas geringer als in Ruch (1999), wo sie bei 25% lag.

Zur Überprüfung des Einflusses der Tendenz zur Beantwortung im Sinne der sozialen Erwünschtheit auf die Ergebnisse, wurden die Faktorenanalysen, wie in Eysenck und Eysenck (1991) empfohlen, für Probanden mit hohen und niedrigen Werten auf der Lügenskala getrennt durchgeführt. Diese Analysen wurden nur für die gewichtete Stichprobe dargestellt, da nur hier die intendierte Struktur aufgezeigt werden konnte⁵. Dazu wurde die Gruppe am Mittelwert der Lügenskala für die gewichtete Stichprobe ($M=8.27$; $SD=4.34$) geteilt. Die dreifaktoriellen Oblimin-Lösungen wurden getrennt für die 640 Probanden mit unterdurchschnittlichen Lügenwerten und die 577 Probanden mit überdurchschnittlichen Lügenwerten ermittelt. Dann wurden die Kongruenzkoeffizienten zur Ermittlung der Ähnlichkeit der Ladungsmuster der drei Faktoren in beiden Stichproben ermittelt. Für Extraversion und Neurotizismus ergab sich jeweils eine Kongruenz von .96, für Psychotizismus ergab sich eine Kongruenz von .91. Vor dem Hintergrund, dass viele Hauptladungen des Psychotizismus-Faktors gering sind, ist auch die Kongruenz des dritten Faktors noch beachtlich. Die Ergebnisse sind ein Hinweis darauf, dass die Faktorenstruktur des EPQ-R nur wenig durch die Tendenz zur Beantwortung im Sinne der sozialen Erwünschtheit beeinflusst wurde. Aus diesem Grund werden auch im folgenden die Analysen für die Gesamtgruppe berichtet.

Konfirmatorische Faktorenanalysen

Gemäß den Spezifikationen in Kapitel 3 wurde sowohl für die ungewichtete als auch für die gewichtete Stichprobe ein Modell mit orthogonalen und ein Modell mit schiefwinkligen Faktoren spezifiziert. Die Hauptladungen wurden freigesetzt und die Nebenladungen auf null fixiert. Als datenkonform gelten Modelle, für die der $RMSEA \leq .06$ und der $SRMR \leq .10$ ist. Sowohl für die ungewichtete als auch für die gewichtete Stichprobe weist nach diesem Kriterium das orthogonale und das oblique Modell eine hinreichende Anpassung an die Daten auf (siehe Tabelle 6.10).

Dass diese Modelle in anderen Kontexten häufig zurückgewiesen würden, zeigt sich an den geringen GFI- und CFI-Werten. Die Korrelationen zwischen den Faktoren sind relativ gering, was auch aus dem geringen Unterschied zwischen der Anpassungsgüte der orthogonalen und obliquen KFA-Modelle hervorgeht. Da jedoch Ruch (1999) oblique Lösungen berichtete, wird auch hier die oblique Lösung dargestellt (siehe Tabelle 6.11).

⁵ Die intendierte Struktur des EPQ-R konnte in der ungewichteten Stichprobe auch dann nicht aufgezeigt werden, wenn die Analysen nur für die Probanden mit unterdurchschnittlichen Lügenwerten durchgeführt wurden.

Tabelle 6.10

Anpassungsgüte an die Daten für die KFA-Modelle des EPQ-R

ungewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	10675.40 (3080)	< .001	.045	.065	.79	.64
oblique	10573.94 (3077)	< .001	.045	.058	.80	.64
gewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	11897.66 (3080)	< .001	.048	.066	.78	.59
oblique	11810.21 (3077)	< .001	.048	.060	.77	.60

Anmerkungen. RMSEA= Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness of Fit Index; CFI = Comparative Fit Index.

Da sich die Lösungen insgesamt kaum unterscheiden, wurden hier nur die Lösungen auf Basis der ungewichteten Stichprobe dargestellt. Die orthogonale Lösung findet sich im Anhang (siehe Tabelle A 6.1). Insgesamt entspricht die Lösung in Tabelle 6.11 den Erwartungen. Allerdings weisen nur 11 der 32 Psychotizismus-Items Ladungen über .30 auf, wobei Ladungen von .10 in dieser Lösung bereits auf dem 1%-Niveau signifikant sind. Offenbar treten nach wie vor Probleme bei der Operationalisierung von Psychotizismus auf (vgl. Ruch, 1999).

Die negative Korrelation zwischen den Faktoren für Extraversion und Neurotizismus und die positive Korrelation zwischen Extraversion und Psychotizismus war bei der vorliegenden Stichprobengröße statistisch bedeutsam. Die Korrelationen sind zwar gering, können aber als Hinweis darauf gelten, dass Eysencks Faktoren nicht unbedingt als orthogonal anzusehen sind.

Diskussion der exploratorischen und konfirmatorischen Faktorenanalysen zum EPQ-R

Nur die Komparabilitäten in der gewichteten Stichprobe sprechen für die Extraktion von drei Faktoren. Die Ladungsmuster der EFA für die gewichtete Stichprobe sowie die KFA für die ungewichtete und die gewichtete Stichprobe sprechen dafür, dass drei modellkonforme Faktoren im EPQ-R nachweisbar sind. Allerdings sind die Beträge der Hauptladungen der meisten Psychotizismus-Items sowohl in der EFA als auch in der KFA kleiner .30. Zusammen mit den Ergebnissen der Komparabilitäten für die ungewichtete Stichprobe und den Ergebnissen der Parallelanalyse, kann dies als Hinweis darauf gelten, dass der Psychotizismus-Faktor bei Analysen mit einem modellübergreifenden Variablenpool sich auf mehrere Faktoren aufteilen könnte. Auch wenn die Ergebnisse der KFA global als Bestätigung der intendierten Struktur gelten können, muss doch in Rechnung gestellt werden, dass die geplanten Analysen in einem modellübergreifenden Variablensatz (siehe Kapitel 6.3) anhand exploratorischer Faktoren-

analysen durchgeführt werden, für die sich keine besonders eindeutige Modellbestätigung ergab.

Tabelle 6.11

Ladungsmuster der KFA mit korrelierten Faktoren für den EPQ-R in der ungewichteten Stichprobe (completely standardized solution)

Item/Skala	P	E	N	Item/Skala	P	E	N
P 025	.28	-	-	E 051	-	.60	-
P 030	.14	-	-	E 054	-	.40	-
P 041	.26	-	-	E 056	-	.46	-
P 046	.32	-	-	E 060	-	.34	-
P 048	.12	-	-	E 064	-	.37	-
P 071	.26	-	-	E 066	-	.27	-
P 079	.29	-	-	E 074	-	.65	-
P 084	.12	-	-	E 088	-	.53	-
P 089	.15	-	-	E 090	-	.28	-
P 102	.21	-	-	E 093	-	.62	-
P 002	.35	-	-	E 024	-	.64	-
P 006	.29	-	-	E 033	-	.43	-
P 009	.19	-	-	E 047	-	.55	-
P 011	.10	-	-	N 003	-	-	.56
P 013	.47	-	-	N 007	-	-	.48
P 017	.39	-	-	N 012	-	-	.58
P 020	.17	-	-	N 016	-	-	.48
P 021	.56	-	-	N 022	-	-	.56
P 026	.11	-	-	N 027	-	-	.61
P 040	.56	-	-	N 031	-	-	.62
P 057	.22	-	-	N 034	-	-	.48
P 061	.33	-	-	N 037	-	-	.62
P 065	.06	-	-	N 042	-	-	.46
P 070	.31	-	-	N 045	-	-	.60
P 075	.22	-	-	N 049	-	-	.52
P 077	.38	-	-	N 052	-	-	.27
P 086	.41	-	-	N 058	-	-	.29
P 094	.20	-	-	N 062	-	-	.59
P 097	.03	-	-	N 067	-	-	.37
P 099	.22	-	-	N 069	-	-	.37
P 100	.17	-	-	N 072	-	-	.37
P 101	.39	-	-	N 076	-	-	.51
E 001	-	.28	-	N 080	-	-	.52
E 005	-	.65	-	N 082	-	-	.54
E 010	-	.69	-	N 085	-	-	.46
E 015	-	.55	-	N 091	-	-	.29
E 019	-	.41	-	N 095	-	-	.34
E 029	-	.50	-	N 098	-	-	.40
E 035	-	.49	-	Korrelation			
E 036	-	.26	-	E	.23		
E 039	-	.60	-	N	-.09	-.21	
E 044	-	.47	-				

Anmerkungen. P= Psychotizismus, E= Extraversion, N= Neurotizismus. Ladungen mit Beträgen > .30 wurden fett gedruckt. Die auf null fixierten Ladungen wurden mit „-“ gedruckt. Zur Modellgüte siehe Tabelle 6.10.

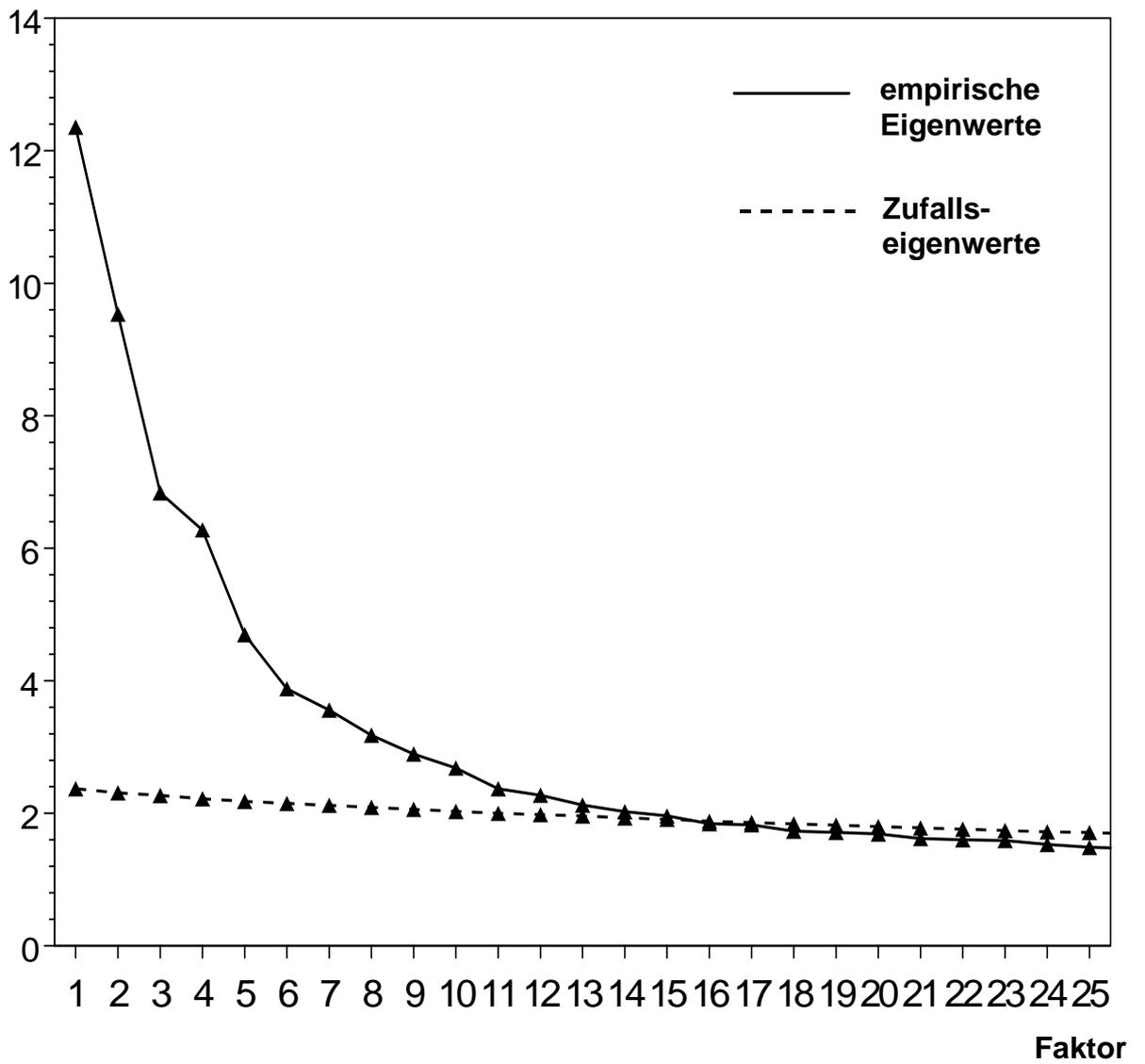
6.2.2 Cattell: 16 PF-R

Der 16-PF-R (Schneewind & Graf, 1998) wurde in Gruppe A (N=590) eingesetzt. Komplette 16-PF-R Daten liegen jedoch nur für 556 Probanden vor. Die Analysen wurden hier, wie bei allen übrigen Fragebogen, anhand der Einzelitems durchgeführt. Der 16-PF-R umfasst 174 Items, die sich den 16 Primärfaktoren zuordnen lassen. Darüber hinaus enthält der 16-PF-R eine 10 Items umfassende Skala zum „Impression Management“ (IM), mit der die Tendenz sozial erwünschte Antworten zu geben, erfasst werden soll. Wie beim EPQ-R empfehlen die Autoren auch bezüglich dieser Skala eine flexible Handhabung von Cut-off-Werten in Abhängigkeit von der Fragestellung. Allerdings sind Schneewind und Graf (1998) der Auffassung, dass Cut-Off-Werte, die bei 5% bzw. 95% liegen, üblich seien. Der 95. Perzentil liegt nach den Angaben im Handbuch bei 26 Punkten, der Mittelwert der IM-Skala bei 19,3 (SD= 4.2). Demgegenüber liegt der Mittelwert der IM-Skala für die hier untersuchte Stichprobe bei 17,9 (SD=4.0). Da die Tendenz sozial erwünscht zu antworten in der vorliegenden Stichprobe etwas geringer ausgefallen ist als in der Eichstichprobe liegen nur sieben Probanden (1,3%) über dem von Schneewind und Graf (1998) angegebenen 95. Perzentil. Da sieben Probanden bei einer Gesamtgruppe von 556 Probanden die Korrelationen zwischen den Items nicht wesentlich verzerren können, wurde davon abgesehen, diese Probanden aus den Analysen auszuschließen.

Die Ergebnisse der Parallelanalyse auf der Basis der gemittelten Eigenwerte von 100 Zufallslösungen sind in Abbildung 6.4 für die ungewichtete (a) und die gewichtete (b) Gesamtgruppe dargestellt. Bei der Parallelanalyse für die ungewichtete Stichprobe sind 15 empirische Eigenwerte größer als die Zufallseigenwerte, während bei der Parallelanalyse für die gewichtete Stichprobe 21 empirische Eigenwerte größer sind als die Zufallseigenwerte. Die Parallelanalyse indiziert also in keinem Fall die Extraktion von 16 Faktoren, obgleich eingeräumt werden muss, dass eine Ungenauigkeit von nur einem Faktor in der ungewichteten Stichprobe bei insgesamt 16 Faktoren eher gering erscheint.

a) Ungewichtet

Eigenwert



b) Gewichtet

Eigenwert

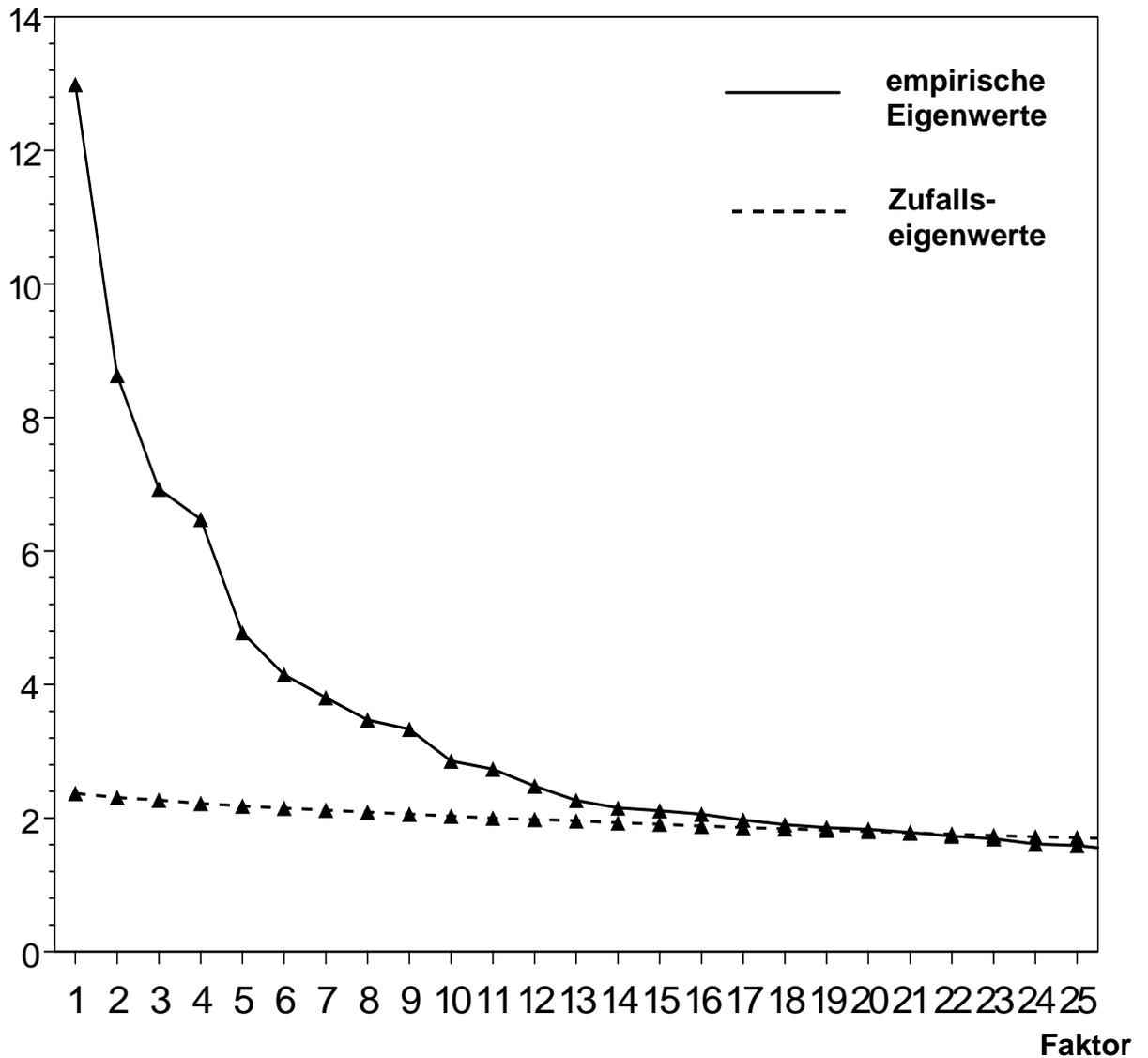


Abbildung 6.4. Parallelanalyse für den 16-PF-R (gewichtete und ungewichtete Stichprobe; N=556)

Die Komparabilitäten für die Lösungen mit zwei bis 16 Faktoren sind für die ungewichtete Stichprobe in Tabelle 6.12, für die gewichtete Stichprobe in Tabelle 6.13 aufgeführt. Die Komparabilitäten der ungewichteten Stichprobe indizieren die Extraktion von fünf Faktoren, die der gewichteten Stichprobe indizieren die Extraktion von zwei Faktoren. Somit indiziert keine der Extraktionsmethoden die Anzahl von 16 Faktoren.

Tabelle 6.12

Faktor-Komparabilitäten für 2-16 Faktoren im 16-PF-R in der ungewichteten Stichprobe (N=556)

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	Faktor 9	Faktor 10	Faktor 11	Faktor 12	Faktor 13	Faktor 14	Faktor 15	Faktor 16
2 Faktoren	.95	.95														
3 Faktoren	.87	.76	.01													
4 Faktoren	.96	.94	.93	.86												
5 Faktoren	.93	.90	.89	.87	.86											
6 Faktoren	.88	.87	.86	.86	.83	.49										
7 Faktoren	.92	.91	.87	.87	.81	.66	.59									
8 Faktoren	.93	.92	.92	.87	.84	.76	.57	.56								
9 Faktoren	.92	.88	.87	.84	.79	.75	.64	.55	.37							
10 Faktoren	.90	.89	.84	.79	.74	.71	.68	.64	.42	.37						
11 Faktoren	.91	.88	.86	.85	.84	.79	.69	.69	.65	.39	.25					
12 Faktoren	.89	.89	.88	.88	.68	.62	.62	.58	.60	.41	.35	.28				
13 Faktoren	.88	.88	.87	.87	.71	.70	.58	.56	.55	.44	.40	.28	.13			
14 Faktoren	.90	.89	.88	.84	.82	.74	.73	.61	.58	.54	.54	.48	.30	.12		
15 Faktoren	.91	.87	.74	.72	.72	.71	.70	.68	.59	.50	.47	.45	.26	.26	.18	
16 Faktoren	.92	.89	.73	.73	.71	.69	.68	.68	.58	.56	.54	.46	.31	.31	.28	.12

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Tabelle 6.13

Faktor-Komparabilitäten für 2-16 Faktoren im 16-PF-R in der gewichteten Stichprobe (N=556)

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	Faktor 9	Faktor 10	Faktor 11	Faktor 12	Faktor 13	Faktor 14	Faktor 15	Faktor 16
2 Faktoren	.88	.87														
3 Faktoren	.87	.87	.65													
4 Faktoren	.91	.91	.87	.85												
5 Faktoren	.84	.79	.73	.71	.30											
6 Faktoren	.80	.79	.77	.76	.57	.48										
7 Faktoren	.83	.82	.81	.79	.72	.60	.08									
8 Faktoren	.85	.85	.74	.73	.66	.61	.57	.10								
9 Faktoren	.85	.83	.82	.69	.62	.58	.51	.50	.40							
10 Faktoren	.84	.83	.81	.76	.74	.71	.70	.62	.43	.36						
11 Faktoren	.84	.81	.79	.78	.73	.69	.63	.60	.42	.20	.02					
12 Faktoren	.85	.85	.82	.79	.76	.75	.63	.56	.49	.31	.22	.03				
13 Faktoren	.87	.87	.85	.80	.78	.77	.74	.70	.40	.37	.25	.23	.09			
14 Faktoren	.88	.86	.84	.79	.72	.69	.65	.64	.60	.53	.51	.48	.26	.21		
15 Faktoren	.87	.85	.79	.73	.66	.60	.57	.55	.51	.44	.36	.31	.27	.23	.14	
16 Faktoren	.88	.86	.85	.82	.75	.75	.62	.62	.59	.56	.48	.46	.44	.32	.23	.11

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Konfirmatorische Faktorenanalysen zum 16 PF-R

Bei der Überprüfung der Struktur mit 16 Faktoren sind nur Lösungen mit korrelierten Faktoren mit Cattells bzw. Schneewind und Grafs (1998) Modellvorstellungen vereinbar, da die 16 Faktoren sich auf einer höheren Ebene zu sechs Faktoren gruppieren sollen. Die orthogonalen Lösungen werden daher lediglich zu Vergleichszwecken aufgeführt. Die Hauptladungen wurden freigesetzt und die Nebenladungen auf null fixiert. Es wurden auch hierarchische Modelle spezifiziert, indem statt der Interkorrelationen zwischen den Faktoren, Ladungen der jeweils sechs zugehörigen Faktoren erster Ordnung auf sechs orthogonalen Faktoren zweiter Ordnung freigesetzt wurden⁶.

Im Rahmen von Cattells Modellvorstellungen sollten jedoch auch die Faktoren zweiter Ordnung eher oblique sein (z.B. Cattell, 1994a) weshalb auch hierarchische Modelle berechnet wurden, bei denen Interkorrelationen zwischen den sechs Faktoren zugelassen wurden (hierarchisch-oblique in Tabelle 6.14).

Tabelle 6.14

Anpassungsgüte an die Daten für die KFA-Modelle des 16 PF-R

ungewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	29786.41 (14877)	p < .001	.042	.092	.57	.47
oblique	26994.62 (14757)	p < .001	.039	.066	.61	.56
hierarchisch-orthogonal	*					
hierarchisch-oblique	393387.85 (14843)	p < .001	.21	.086	.54	.12
gewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	36029.61 (14877)	p < .001	.051	.096	.53	.38
oblique	33195.85 (14757)	p < .001	.047	.071	.56	.46
hierarchisch-orthogonal	*					
hierarchisch-oblique	*					

Anmerkungen. RMSEA= Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness of Fit Index; CFI = Comparative Fit Index. * Mindestens ein Parameter konnte nicht identifiziert werden, so dass keine verlässlichen Fit-Indizes ermittelt werden konnten.

⁶ Es ist zu beachten, dass bei den konventionellen, nicht-hierarchischen konfirmatorischen Faktorenanalysen nur latente exogene Variablen als Faktoren spezifiziert wurden, während bei der hierarchischen Analyse die 16 Faktoren erster Ordnung als latente endogene Variablen spezifiziert werden und die sechs Faktoren zweiter Ordnung als latente exogene Variablen.

Der RMSEA beträgt für die nicht-hierarchischen Modelle $\leq .06$ und der SRMR $\leq .10$. Die hierarchischen Modelle wiesen entweder eine unzureichende Anpassung an die Daten auf, oder aber einzelne Parameter der Modelle konnten nicht identifiziert werden, so dass keine verlässliche Schätzung der Modell-Anpassung an die Daten vorgenommen werden konnte. Die Anpassung an die Daten war für die Modelle mit korrelierten Faktoren, sowohl in der ungewichteten Stichprobe ($df_{\text{diff}} = 120$; $\chi^2_{\text{diff}} = 2791,79$; $p < .001$) als auch in der gewichteten Stichprobe ($df_{\text{diff}} = 120$; $\chi^2_{\text{diff}} = 2833,76$; $p < .001$), höher als für die Modelle mit unkorrelierten Faktoren. Zwar decken die Modelle mit 16 korrelierten Faktoren erster Ordnung nicht das gesamte mit dem 16PF-R intendierte Modell ab, sie bestätigen jedoch einen bedeutsamen Aspekt dieses Modells, weshalb im folgenden das Ladungsmuster der korrelierten Faktoren erster Ordnung für die ungewichtete Stichprobe dargestellt wurde (s. Tabelle 6.15).

Da die Nebenladungen alle auf null fixiert wurden und das Ladungsmuster für die Faktoren erster Ordnung sehr umfassend ist, wurden die Null-Nebenladungen in Tabelle 6.15 nicht dargestellt. Die Interkorrelationen der in Tabelle 6.15 dargestellten 16 Faktoren sind im Anhang aufgeführt (s. Tabelle A 6.2). Die meisten Hauptladungen der Lösung liegen über .30, was als Hinweis auf die Angemessenheit der Skalenbildung im 16 PF-R gelten kann. Die KFA anhand der gewichteten Stichprobe erbrachte nahezu dieselbe Lösung, weshalb diese Lösung hier nicht dargestellt wurde.

Diskussion der exploratorischen und konfirmatorischen Faktorenanalysen zum 16 PF-R

Zwar wies die Parallelanalyse in der ungewichteten Stichprobe auf die Extraktion von 15 Faktoren hin, alle übrigen Berechnungen deuteten jedoch auf deutlich andere Faktorenzahlen als die intendierten 16 Faktoren. Somit konnte die Struktur des 16 PF-R anhand der hier vorgenommenen EFA nicht repliziert werden. Dieses Ergebnis ist insofern nicht erstaunlich, als auch Schneewind und Graf (1998) die intendierte Struktur mit 16 Faktoren lediglich anhand von Analysen mit voraggregierten Aufgaben nachweisen konnten. Für jede der 16 Skalen wurden durch Zufallsaufteilung zwei Teilskalen gebildet, die dann beide auf dem jeweils intendierten Faktor luden. Die von Schneewind und Graf (1998) verwendete Methode der gezielten Voraggregation würde vermutlich bei den meisten Fragebogen den Nachweis der intendierten Strukturen begünstigen. Die Methode der Voraggregation wurde bereits von Bartussek (1988) beim 16 PF und von Rodenhausen (1999) beim 16 PF-R kritisiert, der im übrigen auch kritisch anmerkt, dass die Eigenwerte der 16 Faktoren im Manual nicht mitgeteilt werden. Die hier vorgenommenen EFA anhand der einzelnen Items begünstigen die intendierte Struktur weniger. Im vorliegenden Fall erschien die konservativere Strategie angemessener,

zumal die Struktur des EPQ-R ebenfalls ohne Voraggregation untersucht wurde. Eine Analyse anhand voraggrierter Items könnte qualitative Unterschiede zwischen den Fragebogen verdecken.

Tabelle 6.15

Ladungsmuster für die 16 korrelierten Faktoren erster Ordnung des 16 PF-R in der ungewichteten Stichprobe (completely standardized solution)

	A	B	C	E	F	G	H	I
1	.15	172 .73	22 .30	2 .46	3 .42	4 .38	7 .55	8 .27
21	.27	173 .52	25 .29	27 .49	5 .36	6 .40	29 .65	30 .47
23	.39	174 .28	44 .38	46 .44	26 .62	28 .57	52 .66	31 .55
73	.33	175 .76	47 .39	74 .34	48 .69	49 .52	77 .67	53 .47
94	.33	176 -.02	95 .49	97 .48	50 .32	51 .32	78 .54	55 .45
121	.45	177 .65	98 .60	99 .48	75 .38	76 .26	101 .74	79 .50
129	.61	178 .69	120 .47	123 .19	132 .46	100 .51	103 .56	81 .53
142	.39	179 .66	122 .58	124 .58	158 .39	102 .19	126 .45	104 .21
149	.69	180 .35	130 .43	131 .47	159 .19	125 .30	127 .63	128 .40
		181 .34	150 .49	151 .45		167 .24	143 .65	133 .61
		182 .27	165 .44	169 .58		170 .27	152 .63	160 .35
		183 .60					166 .61	
		184 .64						
	L	M	N	O	Q1	Q2	Q3	Q4
10	.36	9 .47	12 .59	15 .77	36 .47	18 .58	19 .47	43 .53
54	.36	11 .18	14 .62	16 .62	37 .45	39 .41	20 .50	69 .37
56	.30	13 .34	34 .73	35 .53	38 .42	41 .50	42 .37	72 .34
80	.54	32 .59	58 .21	60 .54	61 .46	67 .36	68 .37	92 .24
83	.49	33 .55	62 .41	65 .21	64 .62	70 .48	71 .52	105 .57
106	.31	45 .42	86 .69	85 .49	66 .33	89 .62	90 .27	116 .55
135	.39	57 .27	108 .43	88 .38	87 .32	91 .38	93 .41	119 .32
145	.49	59 .64	112 .17	96 .64	111 .34	117 .42	115 .25	141 .62
154	.43	82 .44	136 .28	110 .46	113 .29	139 .37	118 .43	148 .56
171	.62	107 .40	155 .26	137 .66	138 .51	146 .36	140 .48	157 .58
		109 .28	162 .33	156 .57	163 .49	168 .72	147 .51	164 .41

Anmerkungen. Die Itemnummern sind fett gedruckt. A = Wärme, B = Log. Schlussfolgern, C = Em. Stabilität, E = Dominanz, F = Lebhaftigkeit, G = Regelbewußtsein, H = Soziale Kompetenz, I = Empfindsamkeit, L = Wachsamkeit, M = Abgehobenheit, N = Privatheit, O = Besorgtheit, Q1 = Offenheit für Verändg., Q2 = Selbstgenügsamkeit, Q3 = Perfektionismus, Q4 = Anspannung. Die ausformulierten Items werden hier – zur Wahrung des Copyrights – nicht mitgeteilt, sie finden sich in Schneewind und Graf (1998). Der Modellfit ist in Tabelle 6.14 dargestellt.

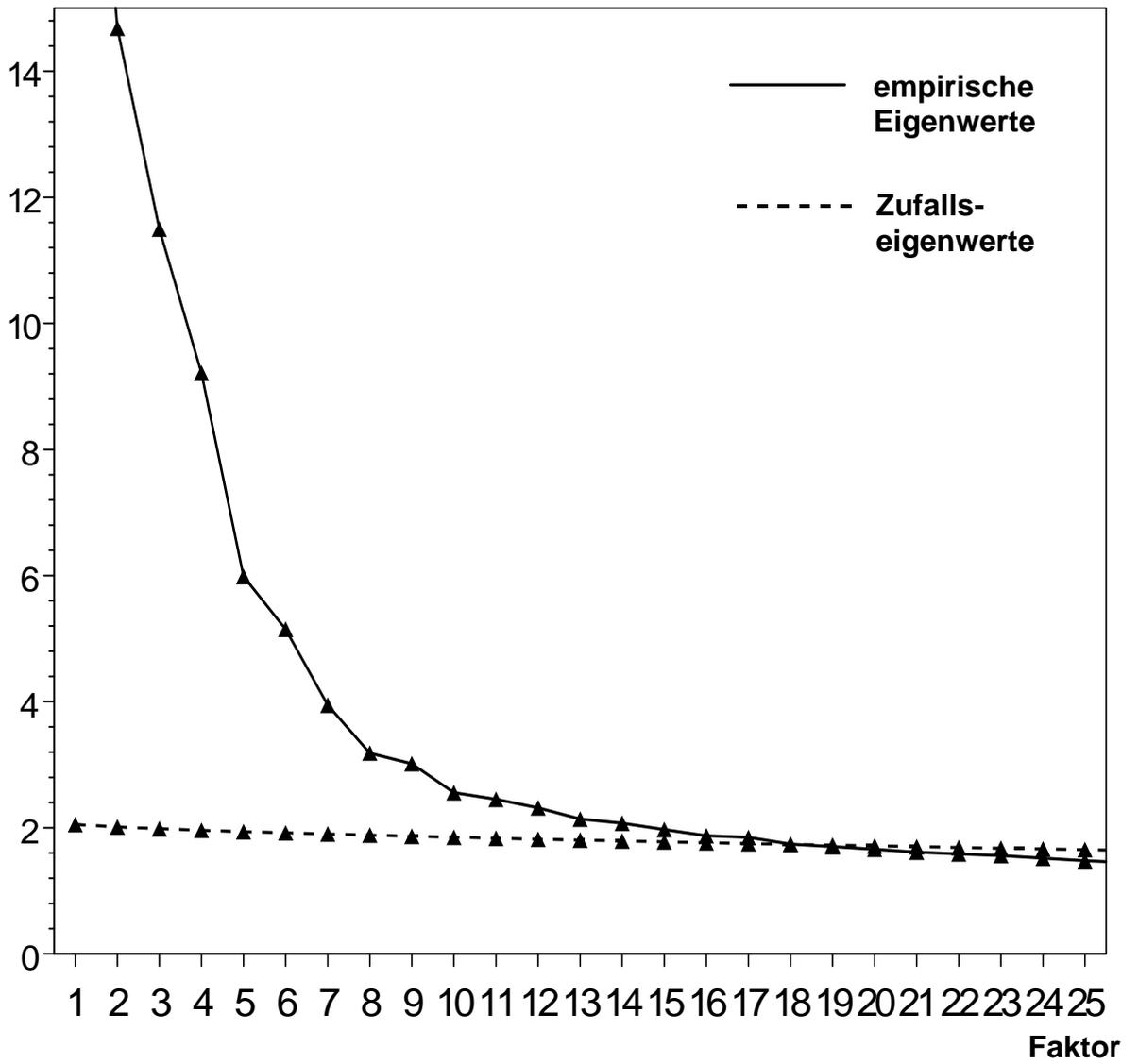
Die KFA erbrachte eine hinreichende Anpassung an die Daten für Modelle, in denen nur die 16 Primärfaktoren spezifiziert wurden. Die hierarchischen Modelle, in denen zusätzlich sechs Faktoren zweiter Ordnung spezifiziert wurden, konnten entweder nicht geschätzt werden oder wiesen eine unzureichende Anpassung an die Daten auf. Da die weiterführenden, modellübergreifenden Analysen anhand der EFA erfolgen sollen, erscheinen die Voraussetzungen für den Nachweis des mit dem 16 PF-R intendierten Modells insgesamt ungünstig.

6.2.3 Costa & McCrae: NEO-PI-R

Der NEO-PI-R (Costa & McCrae, 1992a; Ostendorf & Angleitner, in Druck) wurde in der Gesamtgruppe eingesetzt. Aufgrund einiger Probanden mit fehlenden Werten liegen für den NEO-PI-R Daten von insgesamt 1213 Probanden vor. Die Analysen wurden hier, wie bei allen übrigen Fragebogen, anhand der Einzelitems durchgeführt (s. auch Costa et al., 1991). Der NEO-PI-R umfasst 240 Items, die sich 30 Facetten-Skalen zuordnen lassen, so daß 30 von den 8 Items der jeweiligen Facetten-Skalen markierte Faktoren erwartet werden. Analysen anhand der 30 Facetten-Skalen zur Überprüfung der Fünf-Faktoren-Struktur wären auch denkbar gewesen. Allerdings müsste bei derartigen Analysen die Geltung der 30 Facetten-Skalen vorausgesetzt werden. Eine vollständigere Überprüfung der intendierten Struktur des NEO-PI-R erfordert jedoch den Nachweis der Angemessenheit der 30 Facetten-Skalen, weshalb hier die Item-basierten Analysen berichtet werden. Hierarchische Analysen machen in diesem Kontext nur Sinn, wenn auf der Ebene der Primärfaktoren die 30 den Facetten-Skalen entsprechenden Faktoren nachgewiesen werden können. Die Ergebnisse der Parallelanalyse für die 240 Items des NEO-PI-R sind in Abbildung 6.5 für die ungewichtete (a) und die gewichtete (b) Gesamtgruppe dargestellt.

a) Ungewichtet

Eigenwert



b) Gewichtet

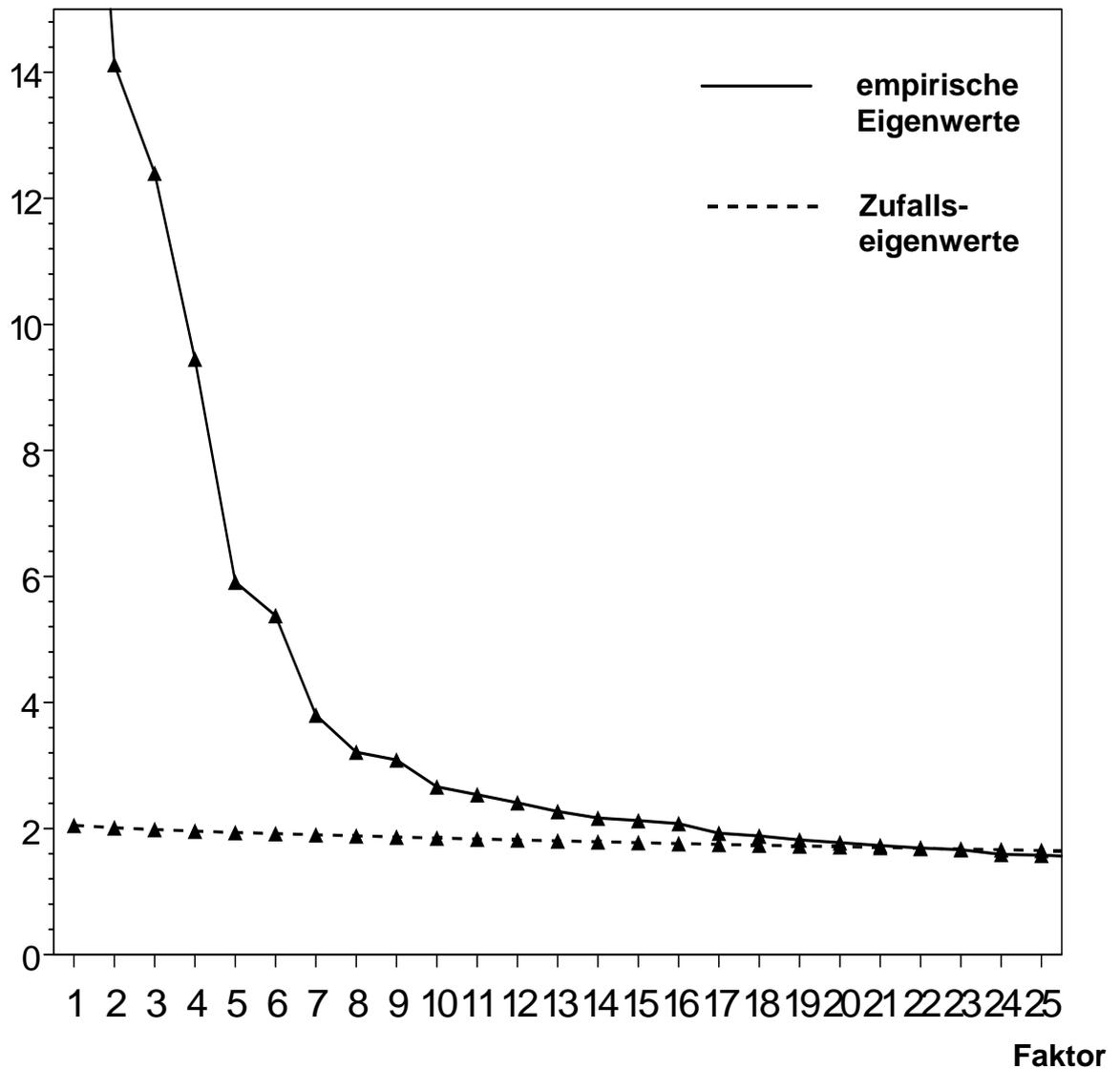
Eigenwert

Abbildung 6.5. Parallelanalyse für den NEO-PI-R (gewichtete und ungewichtete Stichprobe; N=1213)

In der ungewichteten Stichprobe sind 18 empirische Eigenwerte größer als die entsprechenden gemittelten Eigenwerte der Zufallslösungen. In der gewichteten Stichprobe sind 22 empirische Eigenwerte größer als die entsprechenden Zufallseigenwerte. Die Ergebnisse der Parallelanalyse sprechen somit weder für die Extraktion der 30 Faktoren, was der Anzahl an Facetten-Skalen entsprechen würde, noch für die Extraktion von fünf Faktoren. Die Ergebnisse der Komparabilitäten sind in Tabelle 6.16 dargestellt.

Tabelle 6.16

Faktor-Komparabilitäten für den NEO-PI-R (N=1213)
 ungewichtete Stichprobe

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	Faktor 9	Faktor 10
2 Faktoren	.98	.94								
3 Faktoren	.99	.97	.96							
4 Faktoren	1.00	.97	.96	.31						
5 Faktoren	.99	.99	.98	.97	.96					
6 Faktoren	.99	.99	.99	.99	.99	.98				
7 Faktoren	.99	.99	.96	.96	.96	.96	.87			
8 Faktoren	.97	.96	.94	.93	.84	.80	.76	.54		
9 Faktoren	.99	.98	.93	.90	.89	.85	.84	.78	.53	
10 Faktoren	.98	.98	.93	.88	.86	.83	.83	.82	.73	.50

gewichtete Stichprobe

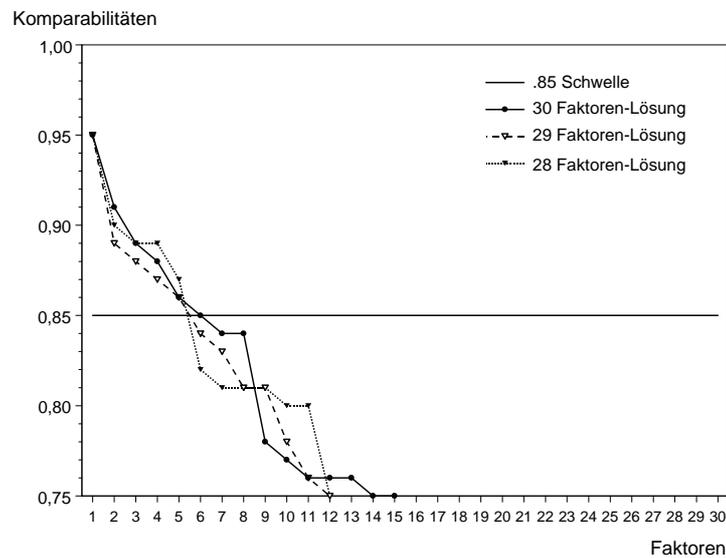
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	Faktor 9	Faktor 10
2 Faktoren	.99	.96								
3 Faktoren	.99	.92	.92							
4 Faktoren	.99	.98	.97	.97						
5 Faktoren	.99	.98	.98	.98	.97					
6 Faktoren	.99	.98	.98	.96	.94	.94				
7 Faktoren	.99	.98	.97	.97	.96	.96	.92			
8 Faktoren	.97	.96	.92	.91	.90	.82	.34	.11		
9 Faktoren	.98	.96	.90	.88	.86	.80	.77	.73	.13	
10 Faktoren	.98	.96	.93	.79	.79	.79	.79	.73	.53	.35

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Da sowohl in der gewichteten als auch in der ungewichteten Stichprobe in der siebenfaktoriellen Lösung alle sieben Faktoren Komparabilitäten > .85 aufweisen, wurden die Komparabilitäten hier zunächst für bis zu zehn Faktoren berechnet, damit die Komparabilitäten – wie bei den übrigen Fragebogen – bis zu einer Faktorenzahl dargestellt werden, bei der mindestens ein Faktor eine Komparabilität unterhalb des Schwellenwertes aufweist. Es wird somit die maximale Anzahl von Faktoren mit Komparabilitäten oberhalb des Schwellenwertes ermittelt. Bei den Lösungen mit acht bis zehn Faktoren treten stets auch Komparabilitäten unterhalb des Schwellenwertes auf, was bedeutet, dass hier maximal sieben Faktoren angemessene Komparabilitäten aufweisen. Auf dieser Basis indizieren die Komparabilitäten die Extraktion von sieben Faktoren.

Auf der Basis der in Tabelle 6.16 dargestellten Ergebnisse kann jedoch nicht ausgeschlossen werden, dass die Komparabilitäten die Extraktion von 30 Faktoren indizieren, die den 30 Facetten-Skalen entsprechen könnten. In diesem Fall müssten 30 Faktoren Komparabilitäten $> .85$ aufweisen. Aus diesem Grund wurden die Komparabilitäten auch für Lösungen mit 30 Faktoren ermittelt. Dabei ist zu bedenken, dass ein oder zwei der 30 Faktoren unzureichende Komparabilitäten aufweisen könnten, während die übrigen Faktoren der intendierten Struktur entsprechen würden. Aus diesem Grund wurden auch die Komparabilitäten für Lösungen mit 28 und 29 Faktoren berechnet. Aufgrund der Menge der numerischen Information wurden die Komparabilitäten für die Lösungen für 28 bis 30 Faktoren in Abbildung 6.6 veranschaulicht.

a) Ungewichtet



b) Gewichtet

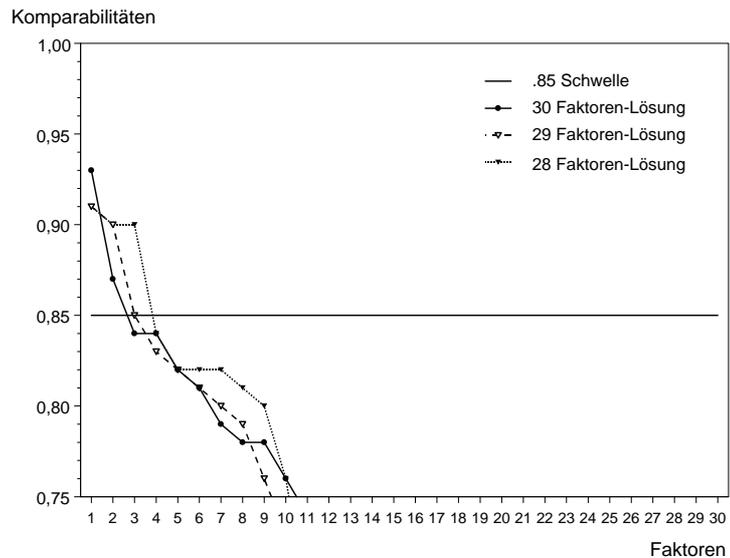


Abbildung 6.6. Komparabilitäten für die NEO-PI-R Lösungen mit 28-30 Faktoren, gewichtete (a) und ungewichtete (b) Stichprobe; N=1213

Es zeigt sich, dass in den Lösungen mit 28 bis 30 Faktoren nur zwei bis fünf Faktoren Komparabilitäten über .85 aufweisen. Insofern indizieren die Komparabilitäten nicht die Extraktion von 28, 29 oder 30 Faktoren. Damit sprechen die Komparabilitäten – wie die Parallelanalyse – weder für die Extraktion von fünf noch für die Extraktion von 30 Faktoren. Da hier nur die Überprüfung der intendierten Struktur vorgenommen wird, werden Lösungen mit sieben oder 18 Faktoren, die durch Komparabilitäten bzw. Parallelanalyse indiziert werden, hier nicht dargestellt.

Konfirmatorische Faktorenanalysen zum NEO-PI-R

Bei der Überprüfung der Struktur mit 30 den Facetten-Skalen entsprechenden Faktoren sind nur Lösungen mit korrelierten Faktoren mit dem Fünf-Faktoren-Modell vereinbar, da die 30 Faktoren sich auf einer höheren Ebene zu fünf Faktoren gruppieren sollen. Die orthogonalen Lösungen werden daher lediglich zu Vergleichszwecken aufgeführt. Die Hauptladungen wurden freigesetzt und die Nebenladungen auf null fixiert. Es wurden auch hierarchische Modelle spezifiziert, indem statt der Interkorrelationen zwischen den Faktoren, Ladungen der jeweils sechs zugehörigen Faktoren erster Ordnung auf fünf orthogonalen Faktoren zweiter Ordnung freigesetzt wurden⁷. In diesen hierarchischen Modellen sind alle Annahmen, die in den NEO-PI-R eingehen, spezifiziert. Die Orthogonalität der fünf Faktoren wurde jedoch mitunter in Zweifel gezogen (z.B. Digman, 1997), weshalb auch hierarchische Modelle berechnet wurden, bei denen Interkorrelationen zwischen den fünf Faktoren zugelassen wurden (hierarchisch-oblique in Tabelle 6.17).

Als Datenkonform gelten Modelle, für die der RMSEA \leq .06 und der SRMR \leq .10 ist. Sowohl für die ungewichtete als auch für die gewichtete Stichprobe weist nach diesem Kriterium das oblique nicht-hierarchische Modell nicht aber das orthogonale nicht-hierarchische Modell eine hinreichende Anpassung an die Daten auf (siehe Tabelle 6.17). Dass diese Modelle, in anderen Kontexten allerdings meist zurückgewiesen würden, zeigt sich an den geringen GFI- und CFI-Werten.

⁷ Es ist zu beachten, dass bei den konventionellen, nicht-hierarchischen konfirmatorischen Faktorenanalysen nur latente exogene Variablen als Faktoren spezifiziert wurden, während bei der hierarchischen Analyse die 30 Faktoren erster Ordnung als latente endogene Variablen spezifiziert werden und die fünf Faktoren zweiter Ordnung als latente exogene Variablen.

Tabelle 6.17

Anpassungsgüte an die Daten für die KFA-Modelle des NEO-PI-R

ungewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	90466.51 (28440)	< .001	.042	.120	.44	.43
oblique	70481.87 (28005)	< .001	.035	.071	.57	.61
hierarchisch-orthogonal	17040.55 (28410)	> .99			perfekter fit	
hierarchisch-oblique	20282.84 (28400)	> .99			perfekter fit	
gewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	104632.79 (28440)	< .001	.047	.120	.41	.37
oblique	84469.70 (28005)	< .001	.041	.077	.52	.54
hierarchisch-orthogonal	36667.59 (28410)	< .001	.015	.11	.44	.83
hierarchisch-oblique	40927.13 (28400)	< .001	.019	.096	.44	.77

Anmerkungen. RMSEA= Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness of Fit Index; CFI = Comparative Fit Index.

Für die hierarchisch-orthogonalen sowie die hierarchisch-obliquen Modelle wurde mit LISREL 8 in der ungewichteten Stichprobe einen perfekten Modellfit ermittelt. Da das hierarchische Modell mit orthogonalen Faktoren zweiter Ordnung den Annahmen von Costa und McCrae (1992a) direkt entspricht, wird das Ladungsmuster dieses Modells für die ungewichtete Stichprobe im folgenden dargestellt (s. Tabelle 6.18 und 6.19). Da die Nebenladungen alle auf null fixiert wurden und das Ladungsmuster für die Faktoren erster Ordnung sehr umfassend ist, wurden die Null-Nebenladungen in Tabelle 6.18 nicht dargestellt.

Für das hierarchisch-orthogonale Modell wurde allerdings in der gewichteten Stichprobe kein ausreichender Modellfit ermittelt, da der SRMR dort .11 betrug (s. Tabelle 6.17). Demgegenüber wies das hierarchisch-oblique Modell auch in der gewichteten Stichprobe einen akzeptablen Modellfit auf. Um die Interkorrelationen zwischen den fünf Faktoren zweiter Ordnung mit den Ergebnissen von Digman (1997) vergleichen zu können, wurde die Lösung für korrelierte Faktoren zweiter Ordnung in der ungewichteten Stichprobe im Anhang dargestellt (s. Tabelle A 6.4, A 6.5) sowie die Inter-Faktorkorrelationen für die gewichtete Stichprobe in Tabelle A 6.6.

Tabelle 6.18

Ladungsmuster der Faktoren erster Ordnung in der KFA mit orthogonalen Faktoren zweiter Ordnung für den NEO-PI-R in der ungewichteten Stichprobe (completely standardized solution)

N1	N2	N3	N4	N5	N6	E1	E2	E3	E4	E5	E6	O1	O2	O3
091 .64	036 .47	041 .76	046 .62	051 .49	056 .62	002 .40	007 .57	012 .70	047 .49	022 .34	027 .49	003 .48	008 .54	013 .51
061 .74	006 .53	101 .60	016 .64	021 .20	026 .70	032 .54	037 .83	042 .51	017 .46	082 .59	057 .40	033 .67	038 .48	043 .57
031 .58	096 .48	131 .55	106 .53	111 .39	116 .46	062 .49	067 .46	102 .63	107 .61	142 .55	087 .69	063 .70	068 .46	073 .54
001 .58	066 .35	161 .58	076 .54	081 .35	086 .56	092 .37	097 .52	072 .68	077 .45	172 .44	117 .54	093 .65	158 .50	133 .64
151 .74	156 .56	191 .77	136 .68	171 .47	176 .61	122 .73	157 .34	162 .53	167 .29	052 .40	147 .80	123 .72	128 .66	103 .33
211 .57	126 .24	011 .63	166 .40	231 .46	146 .56	152 .57	127 .42	132 .38	197 .33	202 .44	177 .83	153 .40	098 .67	163 .55
121 .57	186 .47	221 .74	196 .44	201 .53	206 .61	182 .47	217 .75	192 .42	227 .66	112 .29	207 .35	183 .20	188 .65	193 .37
181 .70	216 .73	071 .75	226 .31	141 .30	236 .63	212 .37	187 .62	222 .76	137 .57	232 .32	237 .57	213 .46	218 .45	223 .50
O4	O5	O6	V1	V2	V3	V4	V5	V6	G1	G2	G3	G4	G5	G6
018 .52	023 .58	028 .39	034 .60	009 .49	044 .49	019 .28	054 .37	029 .40	005 .51	040 .61	015 .60	050 .72	025 .58	060 .77
048 .42	083 .48	058 .25	094 .56	069 .51	104 .57	079 .38	114 .36	089 .59	065 .56	010 .22	075 .46	110 .79	085 .60	120 .79
078 .34	143 .41	088 .22	154 .60	129 .26	014 .60	139 .32	174 .55	149 .37	125 .49	100 .39	135 .63	170 .57	145 .56	180 .74
108 .44	053 .58	118 .38	184 .39	039 .72	074 .61	049 .43	204 .30	179 .43	185 .52	160 .31	165 .35	200 .39	235 .64	210 .32
138 .58	113 .54	148 .33	214 .56	099 .23	164 .35	109 .51	024 .56	209 .66	035 .13	070 .60	045 .57	230 .49	055 .59	240 .52
198 .44	173 .57	178 .47	004 .43	159 .51	194 .31	169 .54	084 .67	059 .10	095 .38	130 .64	195 .62	020 .22	115 .61	030 .24
228 .51	203 .52	208 .28	064 .31	189 .63	224 .38	199 .48	144 .62	119 .50	155 .49	190 .31	225 .47	080 .25	175 .50	090 .65
168 .35	233 .64	238 .24	124 .34	219 .14	134 .27	229 .32	234 .77	239 .27	215 .73	220 .46	105 .19	140 .39	205 .49	150 .52

Anmerkungen. Die Itemnummern sind fett gedruckt. Faktoren N1-6 repräsentieren Neurotizismus, E1-6 Extraversion, O1-6 Offenheit für Erfahrungen, V1-6 Verträglichkeit, G1-6 Gewissenhaftigkeit. Die Zuordnung der Facetten-Nummern zu den Facetten-Bezeichnungen findet sich im Anhang (Tabelle A 6.3). Eine Zuordnungstabelle der Itemnummern zu den ausformulierten Items des NEO-PI-R wird hier nicht präsentiert, da die ausformulierten Items hier nicht mitgeteilt werden können, sie finden sich in Ostendorf und Angleitner (in Druck). Der Modellfit ist in Tabelle 6.17 dargestellt und die Ladungen der Faktoren erster Ordnung auf den Faktoren zweiter Ordnung in Tabelle 6.17.

Hervorzuheben sind die sowohl in der gewichteten als auch in der ungewichteten Stichprobe auftretenden deutlich negativen Korrelationen von Neurotizismus mit Extraversion und Gewissenhaftigkeit (ca. $-.50$) sowie die deutlich positive Korrelation (ca. $.50$) von Extraversion und Offenheit für Erfahrungen.

Tabelle 6.19

Ladungsmuster der orthogonalen Faktoren zweiter Ordnung in der KFA für den NEO-PI-R in der ungewichteten Stichprobe (completely standardized solution)

Faktor	N	E	O	A	G
1. Ordnung					
N1	.88	-	-	-	-
N2	.86	-	-	-	-
N3	.93	-	-	-	-
N4	.85	-	-	-	-
N5	.51	-	-	-	-
N6	.93	-	-	-	-
E1	-	.85	-	-	-
E2	-	.81	-	-	-
E3	-	.62	-	-	-
E4	-	.72	-	-	-
E5	-	.63	-	-	-
E6	-	.77	-	-	-
O1	-	-	.65	-	-
O2	-	-	.81	-	-
O3	-	-	.81	-	-
O4	-	-	.53	-	-
O5	-	-	.62	-	-
O6	-	-	.68	-	-
A1	-	-	-	.52	-
A2	-	-	-	.78	-
A3	-	-	-	.86	-
A4	-	-	-	.78	-
A5	-	-	-	.65	-
A6	-	-	-	.58	-
G1	-	-	-	-	.84
G2	-	-	-	-	.84
G3	-	-	-	-	.80
G4	-	-	-	-	.79
G5	-	-	-	-	.97
G6	-	-	-	-	.52

Anmerkungen. N = Neurotizismus, E = Extraversion, O = Offenheit für Erfahrungen, V = Verträglichkeit, G = Gewissenhaftigkeit. Die Zuordnung der Facetten-Nummern zu den Facetten-Bezeichnungen findet sich im Anhang (Tabelle A 6.3). Der Modellfit ist in Tabelle 6.17 dargestellt und das Ladungsmuster der Faktoren erster Ordnung in Tabelle 6.16.

Diskussion der exploratorischen und konfirmatorischen Faktorenanalysen zum NEO-PI-R

Die Parallelanalyse indiziert in der ungewichteten Stichprobe die Extraktion von 18, in der gewichteten Stichprobe die Extraktion von 22 Faktoren. Die Komparabilitäten der ungewichteten und der gewichteten Stichprobe sprechen für die Extraktion von sieben Faktoren. Für ein hierarchisches Modell mit 30 Faktoren erster Ordnung und fünf orthogonalen Faktoren zweiter Ordnung wurde in der KFA mit der ungewichteten Stichprobe ein perfekter Modellfit erreicht, während in der gewichteten Stichprobe für dieses Modell kein akzeptabler Modellfit erreicht wurde. Demgegenüber erreichte ein hierarchisches Modell mit 30 Faktoren erster Ordnung und fünf korrelierten Faktoren zweiter Ordnung sowohl in der ungewichteten als auch in der gewichteten Stichprobe einen akzeptablen Modellfit.

Insgesamt fällt hier die geringe Konvergenz der Ergebnisse der verschiedenen Analyse-Verfahren auf. Da jedoch sowohl die Parallelanalyse als auch die Komparabilitäten weniger als 30 Faktoren indizieren, können die 30 Facettenskalen nicht als äquivalent zu 30 Faktoren erster Ordnung im Rahmen hierarchischer Faktorenanalysen aufgefasst werden. Dennoch weisen diese Skalen ein Mindestmaß an Qualität auf, da mit Hilfe der KFA mehrere Modelle mit akzeptablem Fit auf der Basis von 30, den Facetten-Skalen entsprechenden Faktoren, formuliert werden konnten. Bezüglich der Orthogonalität der Faktoren zweiter Ordnung konnte hier keine sichere Entscheidung getroffen werden, da das Modell mit orthogonalen Faktoren zweiter Ordnung in der ungewichteten Stichprobe zwar einen perfekten Modellfit aufwies, während es in der gewichteten Stichprobe keinen akzeptablen Modellfit hatte. Da das Modell mit korrelierten Faktoren zweiter Ordnung sowohl in der ungewichteten als auch in der gewichteten Stichprobe einen akzeptablen Modellfit aufwies und eine Reihe von substantziellen Interkorrelationen zwischen den Faktoren auftraten, liegt die Vermutung nahe, dass das Fünf-Faktoren-Modell um Faktoren dritter Ordnung erweitert werden sollte (vgl. Digman, 1997).

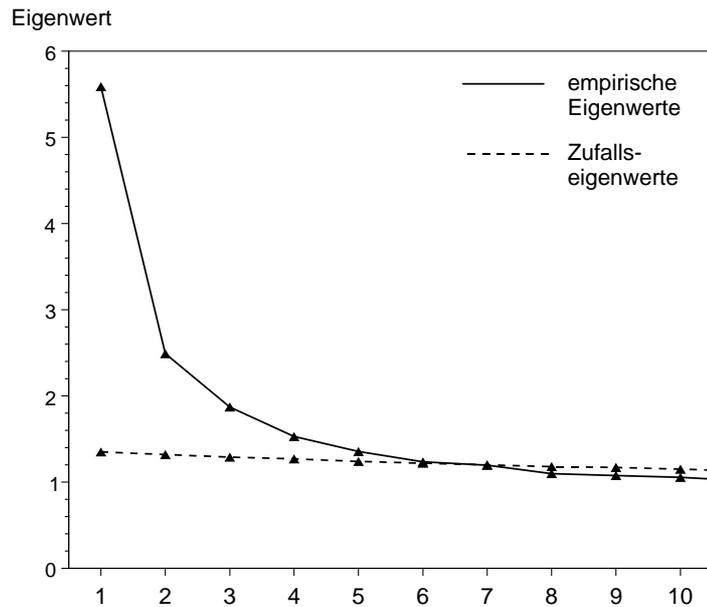
Insgesamt kann zwar anhand der KFA davon ausgegangen werden, dass – bei geeigneter Auswertung – mit dem NEO-PI-R die Faktoren des Fünf-Faktoren-Modells erfasst werden können, allerdings konnten anhand der EFA weder die 30 Facettenskalen noch die fünf Faktoren des Modells aufgezeigt werden. Von daher sind die Faktoren des Fünf-Faktorenmodells in den geplanten exploratorischen Analysen mit einem modellübergreifenden Itempool (siehe Kapitel 6.3) anhand der NEO-PI-R Items nicht optimal repräsentiert.

6.2.4 Zuckerman: SSS-V und ZKPQ

Exploratorische Faktorenanalysen zur SSS-V

Wegen fehlender Werte umfasst die Gesamtgruppe für die Analyse der SSS-V nicht 1224 sondern lediglich 1213 Probanden. Die Ergebnisse der Parallelanalyse sind in Abbildung 6.7 für die ungewichtete (a) und die gewichtete (b) Gesamtgruppe dargestellt.

a) Ungewichtet



b) Gewichtet

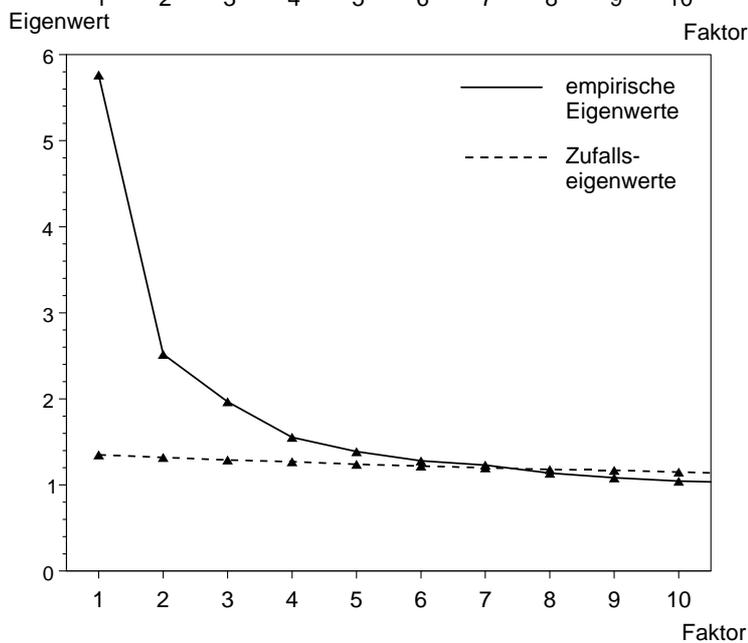


Abbildung 6.7. Parallelanalyse für die SSS-V (gewichtete und ungewichtete Stichprobe; N=1213)

Sowohl bei der gewichteten als auch bei der ungewichteten Stichprobe sind fünf empirische Eigenwerte eindeutig größer als die korrespondierenden Eigenwerte auf der Basis von 100 gemittelten Zufallslösungen. Numerisch sind jedoch in der ungewichteten Stichprobe (s. Tabelle 6.21) sechs und in der gewichteten Stichprobe sieben empirische Eigenwerte größer als die korrespondierenden Eigenwerte auf Basis der Zufallsdaten. Die Parallelanalyse indiziert somit die Extraktion von sechs bzw. sieben Faktoren. Die Ergebnisse für die Komparabilitäten sind in Tabelle 6.20 dargestellt.

Die Komparabilitäten auf Basis der ungewichteten Stichprobe sprechen für eine dreifaktorielle Struktur, wobei die Komparabilität für den vierten Faktor genau auf dem Schwellenwert liegt. Demgegenüber sprechen die Komparabilitäten der gewichteten Stichprobe eindeutig für eine dreifaktorielle Struktur.

Tabelle 6.20

Faktor-Komparabilitäten für die SSS-V (N=1213)

		Ungewichtete Stichprobe						
	Faktor1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	
2 Faktoren	.99	.99						
3 Faktoren	.99	.94	.90					
4 Faktoren	.99	.98	.96	.85				
5 Faktoren	.99	.91	.85	.80	.40			
6 Faktoren	.97	.96	.92	.90	.82	.11		
7 Faktoren	.98	.95	.91	.89	.85	.77	.64	
		Gewichtete Stichprobe						
	Faktor1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	
2 Faktoren	.96	.92						
3 Faktoren	.99	.97	.95					
4 Faktoren	.99	.96	.82	.36				
5 Faktoren	.99	.88	.80	.79	.59			
6 Faktoren	.99	.83	.80	.78	.63	.08		
7 Faktoren	.92	.89	.87	.83	.53	.45	.44	

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Die mit dem Fragebogen intendierte vierfaktorielle Struktur wird demnach anhand der Komparabilitäten nicht bestätigt. Da allerdings der Wert von .85 für den vierten Faktor in der ungewichteten Stichprobe gerade erreicht wird, wurde die vierfaktorielle Struktur auf Basis der ungewichteten Stichprobe in Tabelle 6.21 dargestellt. Das korrespondierende Trasid-Ladungsmuster wurde im Anhang (Tabelle A 6.7) dargestellt, da es kaum von der Oblimin-Lösung abweicht. Allerdings ist die Korrelation zwischen den Faktoren drei und vier in der Trasid-Lösung etwas höher.

Tabelle 6.21

Oblimin-Ladungsmuster (delta=0) für die SSS-V in der ungewichteten Stichprobe

Item/Skala	Faktor 1	Faktor 2	Faktor 3	Faktor 4	h ²
16 TAS	.71	-.01	-.13	-.06	.45
17 TAS	.71	-.02	-.11	.01	.45
23 TAS	.59	.05	.01	-.04	.38
21 TAS	.55	.09	-.03	-.01	.32
3 TAS	.50	-.04	.07	.00	.27
20 TAS	.47	.07	-.05	-.00	.23
40 TAS	.46	-.03	.06	.06	.24
38 TAS	.42	-.07	.20	.02	.27
28 TAS	.41	.13	.02	-.04	.21
14 ES	.14	.03	.13	-.02	.06
1 DIS	.05	.52	.13	-.09	.32
29 DIS	.06	.51	-.14	-.02	.26
30 DIS	-.04	.48	-.07	-.04	.21
13 DIS	.12	.43	.06	.07	.27
32 DIS	-.07	.40	.03	.04	.16
35 DIS	.07	.39	-.00	-.08	.16
12 DIS	.03	.38	.15	-.11	.19
36 DIS	.01	.34	-.05	.07	.13
33 DIS	.08	.30	.04	.06	.13
34 BS	.01	.29	-.01	.01	.09
10 ES	.14	.19	.12	.14	.14
19 ES	.02	.04	.53	.15	.33
26 ES	.03	-.02	.43	.01	.20
24 BS	.07	.09	.42	-.04	.23
18 ES	.13	-.02	.38	.11	.21
11 TAS	.27	.17	.34	.03	.35
25 DIS	.10	.26	.33	.06	.28
7 BS	-.10	.19	.32	.16	.18
37 ES	.07	-.12	.32	-.18	.16
31 BS	-.09	.22	.32	-.00	.15
22 ES	-.02	-.09	.32	-.20	.14
4 ES	.07	-.03	.28	-.15	.12
6 ES	.22	-.07	.26	.07	.16
9 ES	.16	.24	.24	.12	.24
8 BS	.06	-.13	.10	.42	.18
5 BS	.05	.10	.16	.33	.18
15 BS	-.04	-.04	-.03	.33	.11
2 BS	-.03	-.01	-.06	.26	.07
39 BS	-.04	.13	-.13	.26	.11
27 BS	.05	.13	.15	.16	.09
Interkorrelationen der Faktoren					%-erklärte Varianz: 21.0
Faktor 2	.24				
Faktor 3	.41	.20			
Faktor 4	-.00	.24	.01		
empirische Eigenwerte*: 5.59 2.49 1.87 1.53 1.35 1.23 1.20 1.10 1.08 1.06					
Zufalls-Eigenwerte*: 1.35 1.32 1.29 1.27 1.24 1.22 1.20 1.18 1.17 1.15					

Anmerkungen. * Es wurden die (unreduzierten) PCA-Eigenwerte der unrotierten Faktoren dargestellt. Die Kommunalitäten (h²) und die erklärte Varianz wurden jedoch für die unrotierte Hauptachsenlösung nach Kommunalitäteniteration dargestellt. TAS= Thrill and Adventure Seeking, DIS= Disinhibition, ES= Experience Seeking, BS= Boredom Susceptibility. Ladungen mit Beträgen > .30 wurden fett gedruckt.

Konfirmatorische Faktorenanalysen zur SSS-V

Gemäß den Spezifikationen in Kapitel 3 wurde sowohl für die ungewichtete als auch für die gewichtete Stichprobe ein Modell mit orthogonalen und ein Modell mit schiefwinkligen Faktoren spezifiziert. Die Hauptladungen wurden freigesetzt und die Nebenladungen auf null fixiert. Anhand der Indizes, die hier nach Hu und Bentler (1999) zur Bewertung der Modelle herangezogen werden, konnten alle Modelle, wenn auch einige nur sehr knapp, bestätigt werden, d.h. der RMSEA ist bei allen Modellen $\leq .06$ und der SRMR ist bei allen Modellen $\leq .10$ (s. Tabelle 6.22). Dass diese Modelle in anderen Kontexten häufig zurückgewiesen würden, zeigt sich an den geringen GFI- und CFI-Werten. Andererseits liegt die Stichprobengröße für die hier untersuchten Modellen mit 1213 Probanden deutlich über der von Hu und Bentler (1999) für die hier verwendete Strategie empfohlenen Mindestgröße von 500 Probanden, was die Verwendung dieser Strategie legitimiert.

Da die Hauptladungen frei geschätzt wurden, bleibt die Frage offen, ob überhaupt in ausreichendem Maße nennenswerte Hauptladungen auftreten. Eine Interpretation der Modelle setzt für die überwiegende Zahl der Variablen bedeutsame Hauptladungen voraus. Diese Information findet sich für die oblique Lösung der ungewichteten Stichprobe in Tabelle 6.23. Die Lösung mit korrelierten Faktoren ist hier den orthogonalen Lösungen vorzuziehen, da die vier Subskalen einen Sensation Seeking Gesamtwert abbilden sollen.

Tabelle 6.22

Anpassungsgüte an die Daten für die KFA-Modelle der SSS-V

ungewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	3387.38 (740)	< .001	.054	.095	.86	.61
oblique	2598.33 (734)	< .001	.046	.053	.90	.73
gewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	3810.23 (740)	< .001	.058	.099	.85	.60
oblique	3009.11 (734)	< .001	.050	.057	.88	.70

Anmerkungen. RMSEA= Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness of Fit Index; CFI = Comparative Fit Index.

Tabelle 6.23

Ladungsmuster der KFA mit korrelierten Faktoren für die SSS-V in der ungewichteten Stichprobe (completely standardized solution)

Item/Skala	DIS	BS	ES	TAS
1 DIS	.56	-	-	-
12 DIS	.43	-	-	-
13 DIS	.50	-	-	-
25 DIS	.47	-	-	-
29 DIS	.44	-	-	-
30 DIS	.39	-	-	-
32 DIS	.36	-	-	-
33 DIS	.36	-	-	-
35 DIS	.39	-	-	-
36 DIS	.32	-	-	-
2 BS	-	.01	-	-
5 BS	-	.35	-	-
7 BS	-	.45	-	-
8 BS	-	.16	-	-
15 BS	-	.03	-	-
24 BS	-	.45	-	-
27 BS	-	.31	-	-
31 BS	-	.40	-	-
34 BS	-	.23	-	-
39 BS	-	.06	-	-
4 ES	-	-	.27	-
6 ES	-	-	.38	-
9 ES	-	-	.51	-
10 ES	-	-	.35	-
14 ES	-	-	.24	-
18 ES	-	-	.45	-
19 ES	-	-	.55	-
22 ES	-	-	.20	-
26 ES	-	-	.40	-
37 ES	-	-	.22	-
3 TAS	-	-	-	.52
11 TAS	-	-	-	.49
16 TAS	-	-	-	.62
17 TAS	-	-	-	.63
20 TAS	-	-	-	.47
21 TAS	-	-	-	.56
23 TAS	-	-	-	.62
28 TAS	-	-	-	.46
38 TAS	-	-	-	.49
40 TAS	-	-	-	.49
Interkorrelationen der Faktoren				
BS	.68			
ES	.54	.72		
TAS	.45	.42	.62	

Anmerkungen. TAS= Thrill and Adventure Seeking, DIS= Disinhibition, ES= Experience Seeking, BS= Boredom Susceptibility. Ladungen mit Beträgen > .30 wurden fett gedruckt. Die auf null fixierten Ladungen wurden mit „-“ gedruckt. Zur Modellgüte siehe Tabelle 6.22.

Es zeigt sich, dass alle Faktoren mindestens fünf Ladungen über .30 aufweisen. Es sei darauf hingewiesen, dass selbst Ladungen um .20 bei der vorliegenden Stichprobe bereits signifikant

sind. Insofern kann selbst der BS-Faktor in dieser Analyse als bestätigt angesehen werden, wenn auch fünf BS-Items unzureichende Ladungen aufweisen. Die Interkorrelationen zwischen den Faktoren sind sehr hoch, was im Einklang mit der Bildung eines Sensation Seeking Gesamtwertes ist. Die Ergebnisse für die gewichtete Stichprobe ergeben dasselbe Bild (siehe Anhang, Tabelle A 6.8).

Diskussion der exploratorischen und konfirmatorischen Faktorenanalysen zur SSS-V

Die Ergebnisse der EFA bringen keine Bestätigung der intendierten vierfaktoriellen Struktur. Dies ergibt sich daraus, dass die Extraktionskriterien weder in der gewichteten noch in der ungewichteten Stichprobe auf die vierfaktorielle Lösung konvergieren. Allerdings findet sich in den Komparabilitäten der ungewichteten Stichprobe ein schwacher Anhaltspunkt für eine vierfaktorielle Lösung.

Wie in Beauducel et al. (1999) zeigt sich in der vierfaktoriellen Lösung für die ungewichtete Stichprobe, dass die Faktoren Thrill and Adventure Seeking (TAS) und Disinhibition (DIS) in der EFA gut repliziert werden konnten, während die Faktoren für Experience Seeking (ES) und vor allem Boredom Susceptibility (BS) nur unzureichend durch die zugehörigen Items markiert werden. Auch wenn die vierfaktorielle Lösung noch als schwacher Hinweis auf die intendierte Struktur gewertet werden kann, ist der Nachweis keinesfalls ausreichend um als Bestätigung der intendierten Struktur zu gelten.

In der Oblimin-Lösung sind zwei Korrelationen und in der Trasid-Lösung (s. Anhang; Tabelle A 6.7) ist eine Korrelation zwischen den Faktoren null, was nicht im Einklang mit der Bildung einer Sensation Seeking Gesamtskala steht. Es ist somit nicht davon auszugehen, dass die Items der SSS-V in Analysen in einem modellübergreifenden Variablenpool auf einem Faktor laden werden.

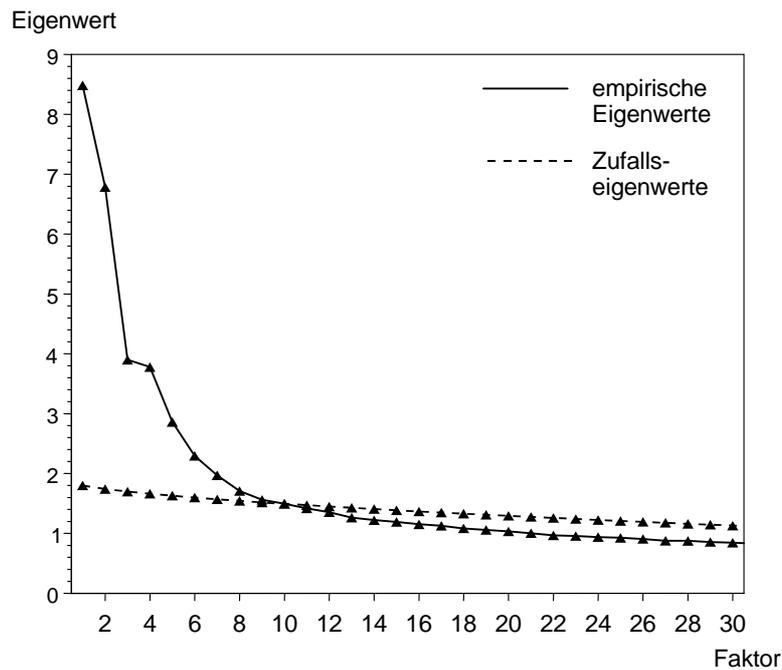
Die Ergebnisse der KFA sind demgegenüber günstiger zu bewerten: Die Modellgüte ist nach den für die Traitforschung spezifizierten Kriterien (s. Kapitel 3) ausreichend und die Faktoren weisen eine genügende Anzahl an Hauptladungen auf. Die modellübergreifenden Analysen werden jedoch anhand der EFA durchgeführt. Da das Ladungsmuster in der EFA deutlich weniger prägnant ist und in der EFA nicht die Möglichkeit besteht unerwünschte Nebenladungen zu unterdrücken, ist nicht davon auszugehen, dass die intendierte Struktur in exploratorischen Analysen mit einem modellübergreifenden Variablenpool ohne weiteres aufgezeigt werden kann. Die Ergebnisse der KFA können auch dahingehend gedeutet werden, dass die SSS-V zwar keine optimalen, aber leidlich verwendbare Skalen enthält, während die

Ergebnisse der EFA darauf hindeuten, dass der Itempool für das Aufzeigen der vier Subskalen ohne Vorwissen über die zu erwartende Struktur nicht geeignet ist. Somit ist der Variablenpool der SSS-V als theoretische Basis für den Nachweis der Konstrukte als unbefriedigend anzusehen, während der Test, wenn auch als suboptimales Verfahren, in Forschungsuntersuchungen eingesetzt werden kann (z.B. Brocke et al., 2000).

Exploratorische Faktorenanalysen zum ZKPQ

Der ZKPQ wurde in einer gegenüber der von Ostendorf und Angleitner (1994) eingesetzten Form um 11 Items gekürzten Form in Teilgruppe B (N=636) eingesetzt, für die Beauducel et al. (1999) zufriedenstellende Ergebnisse berichteten. Die Infrequency-Skala, mit der Antworttendenzen erfasst werden sollen, wurde nicht in die Faktorenanalysen einbezogen. Die Ergebnisse der Parallelanalyse sind in Abbildung 6.8 für die ungewichtete (a) und die gewichtete (b) Teilgruppe B dargestellt.

a) Ungewichtet



b) Gewichtet

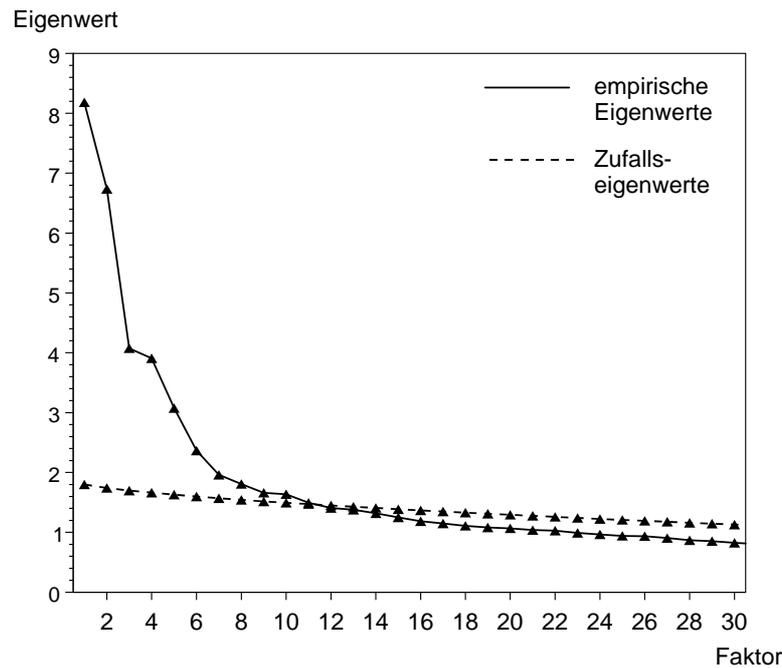


Abbildung 6.8. Parallelanalyse für den ZKPQ (gewichtete und ungewichtete Stichprobe; N=636)

Die Parallelanalyse auf Basis der gemittelten Eigenwerte von 100 Zufallslösungen spricht in der ungewichteten Stichprobe für die Extraktion von neun (siehe auch Tabelle 6.25), und in der gewichteten Stichprobe für die Extraktion von 11 Faktoren. Somit sprechen beide Parallelanalysen nicht für die intendierte Anzahl von fünf Faktoren. Die Komparabilitäten sind in Tabelle 6.24 dargestellt.

Die Komparabilitäten legen sowohl für die ungewichtete als auch für die gewichtete Stichprobe die Extraktion von fünf Faktoren nahe, was mit den Annahmen von Zuckerman et al. (1993) kompatibel ist. Im folgenden ist die Fünf-Faktorenlösung für die ungewichtete Stichprobe exemplarisch dargestellt (s. Tabelle 6.25).

Tabelle 6.24

Faktor-Komparabilitäten für die ZKPQ Items (N=636)

Ungewichtete Stichprobe							
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7
2 Faktoren	.99	.99					
3 Faktoren	.98	.84	.18				
4 Faktoren	.99	.99	.97	.99			
5 Faktoren	.97	.99	.97	.99	.93		
6 Faktoren	.92	.99	.96	.99	.96	.54	
7 Faktoren	.91	.99	.95	.92	.96	.44	.51
Gewichtete Stichprobe							
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7
2 Faktoren	.98	.97					
3 Faktoren	.95	.82	.80				
4 Faktoren	.88	.96	.96	.80			
5 Faktoren	.97	.98	.98	.96	.96		
6 Faktoren	.98	.97	.95	.93	.93	.50	
7 Faktoren	.97	.97	.94	.93	.93	.48	.32

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Das Ladungsmuster der fünf Faktoren entspricht bis auf minimale Abweichungen der intendierten Struktur, d.h. die Faktoren Impulsives Sensation Seeking, Neurotizismus-Ängstlichkeit, Soziabilität, Aktivität und Aggression-Feindseligkeit konnten problemlos identifiziert werden. Zu Vergleichszwecken ist im Anhang eine Trasid-Lösung für die ungewichtete Stichprobe dargestellt (s. Tabelle A 6.9), die allerdings nicht wesentlich von der hier dargestellten Lösung abweicht und somit ebenfalls der intendierten Struktur entspricht. Da sowohl das Ladungsmuster als auch die Interfaktor-Korrelationen für die gewichtete Stichprobe ebenfalls nur unwesentlich von der hier dargestellten Lösung abweichen, wurde auf eine Darstellung der Lösungen auf Basis der gewichteten Stichprobe hier verzichtet.

Tabelle 6.25Oblimin-Ladungsmuster ($\delta=0$) für den ZKPQ in der ungewichteten Stichprobe

Item/Skala	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	h ²
Imp 39	.60	-.16	.00	.05	.03	.43
Imp 34	.56	-.04	-.07	-.02	-.15	.30
Imp 45	.55	.02	.17	.01	.04	.38
Imp 24	.55	.03	.00	.12	.01	.33
Imp 70	.54	.02	.16	.03	.05	.37
Imp 55	.51	-.03	-.08	.07	.16	.32
Imp 65	.50	.02	.10	.15	-.03	.31
Imp 84	.45	.06	.17	.08	.13	.33
Imp 29	.43	.07	.01	-.17	.12	.24
Imp 79	.42	-.02	.15	.03	.12	.27
Imp 75	.40	-.12	-.15	.01	-.06	.19
Imp 60	.38	-.01	.04	.13	-.00	.18
Imp 95	.35	-.01	.24	.03	.15	.27
Imp 19	.34	.01	-.04	-.08	.05	.12
Imp 01	.31	.12	.01	-.06	.16	.16
Akt 28	.28	-.23	.05	.15	-.01	.20
Imp 50	.28	.21	.05	-.07	.14	.16
N-Ä 90	.11	.75	-.12	.00	-.08	.55
N-Ä 61	-.03	.70	-.02	-.13	-.13	.52
N-Ä 25	.08	.68	-.08	.06	-.01	.45
N-Ä 15	.12	.60	-.01	.13	.08	.39
N-Ä 51	.05	.55	-.05	-.05	.01	.31
N-Ä 80	.02	.54	-.03	.03	.08	.32
N-Ä 07	.01	.53	.00	-.14	-.10	.30
N-Ä 76	.03	.52	-.08	.08	-.02	.27
N-Ä 71	.03	.49	.02	-.12	-.06	.26
N-Ä 35	-.03	.49	.01	.01	.11	.28
N-Ä 46	-.11	.48	-.03	.04	.09	.28
Akt 99	-.09	.46	-.08	.16	-.01	.24
N-Ä 20	.01	.44	-.06	.03	.04	.20
N-Ä 41	-.08	.44	.04	.03	.02	.21
N-Ä 56	-.01	.42	.05	.02	.01	.18
N-Ä 96	-.01	.41	-.01	.04	.06	.18
N-Ä 66	.01	.40	.09	.10	.03	.18
N-Ä 02	-.04	.31	.11	-.08	-.02	.12
Akt 88	.11	-.21	-.03	.16	.02	.10
So 87	.01	-.01	.63	-.12	-.04	.39
So 58	-.06	.00	.59	-.05	.03	.35
So 78	.05	-.14	.59	.15	.09	.45
So 22	.07	-.04	.58	.06	-.05	.36
So 52X	.15	-.08	.57	.05	.01	.39
So 98	.12	-.17	.54	.00	-.02	.36
So 68	-.06	-.10	.53	.16	.01	.33
So 48	.27	-.12	.52	.02	.00	.42
So 63	-.29	.08	.50	.08	.07	.32
So 92	.17	.01	.50	-.07	-.10	.29
So 17	-.24	.12	.49	.03	.09	.30
So 12	-.13	.00	.48	-.12	.00	.23
So 27	.21	-.09	.48	.06	.04	.34
So 09	.26	-.15	.33	.01	.00	.24
So 37	.17	.07	.29	.02	.10	.17
So 82	-.01	.16	.28	.02	-.05	.10

(Fortsetzung und Anmerkungen nächste Seite)

Fortsetzung Tabelle 6.25

Oblimin-Ladungsmuster (delta=0) für den ZKPQ in der ungewichteten Stichprobe

Item/Skala	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	h ²				
Akt 83	.08	.05	.15	.67	-.01	.50				
Akt 05	-.16	.07	-.01	.65	.00	.42				
Akt 44	.03	.09	.00	.62	.02	.38				
Akt 33	.18	.00	.12	.56	-.02	.40				
Akt 74	.01	.02	.06	.53	-.04	.29				
Akt 38	-.18	.17	.00	.47	-.07	.26				
Akt 64	.19	.00	-.05	.41	-.02	.22				
Akt 54	-.05	.15	.05	.39	.09	.18				
Akt 18	.12	-.11	-.12	.35	-.05	.18				
Akt 94	.05	-.11	.09	.33	-.02	.15				
Akt 13	.29	-.11	.05	.33	.03	.26				
Akt 59	.09	-.12	-.06	.28	-.13	.14				
Akt 49	-.16	-.03	-.10	.22	.02	.07				
A-F 72	-.03	.03	.10	.10	.60	.41				
A-F 77	.05	-.03	-.14	.01	.54	.29				
A-F 97	-.03	.02	.03	-.06	.51	.27				
A-F 08	.01	.10	-.05	-.08	.50	.29				
A-F 86	-.14	-.08	-.04	-.14	.49	.24				
Imp 89	.14	.00	.13	.20	.47	.36				
A-F 11	.03	.02	-.01	-.09	.46	.23				
A-F 57	.05	-.03	-.02	-.14	.44	.22				
A-F 91	.10	.09	-.01	-.02	.41	.21				
Imp 14	.25	.03	.07	.11	.41	.32				
A-F 42	.01	-.21	-.01	.22	.41	.23				
A-F 21	.03	.11	-.07	.01	.36	.15				
A-F 16	.14	-.01	.00	.04	.32	.15				
A-F 47	-.06	-.14	.12	.05	.30	.12				
A-F 67	.03	.04	.16	.05	.30	.15				
N-Ä 85	-.12	.15	.02	.01	.24	.11				
Imp 06	.20	.14	.06	-.12	.22	.16				
A-F 03	.10	.14	-.02	-.08	.19	.09				
Interkorrelationen der Faktoren						%-erklärte				
Faktor 2	-.13					Varianz:				
Faktor 3	.17	.01				26.9				
Faktor 4	.11	-.13	.09							
Faktor 5	.20	.20	.19	-.01						
<hr/>										
Empirische Eigenwerte*:	8.48	6.78	3.90	3.78	2.86	2.29	1.97	1.71	1.56	1.49
Zufalls-Eigenwerte*:	1.80	1.74	1.70	1.66	1.63	1.60	1.57	1.54	1.52	1.49

Anmerkungen. Die Kommunalitäten (h²) und die erklärte Varianz wurden für die unrotierte Hauptachsenlösung nach Kommunalitäteniteration dargestellt. * Es wurden jedoch die (unreduzierten) PCA-Eigenwerte der unrotierten Faktoren dargestellt, da diese für die Extraktionsentscheidung relevant sind. Imp= Impulsivität-Sensation Seeking, N-Ä= Neurotizismus-Ängstlichkeit, So= Soziabilität, Akt= Aktivität, A-F= Aggression-Feindseligkeit. Ladungen mit Beträgen > .30 wurden fett gedruckt.

Konfirmatorische Faktorenanalysen zum ZKPQ

Gemäß den Spezifikationen in Kapitel 3 wurde sowohl für die ungewichtete als auch für die gewichtete Stichprobe ein Modell mit orthogonalen und ein Modell mit schiefwinkligen Faktoren spezifiziert. Die Hauptladungen wurden freigesetzt und die Nebenladungen auf null fixiert. Auch für den ZKPQ konnten alle Modelle, wenn auch einige nur sehr knapp, bestätigt

werden, d.h. RMSEA ist bei allen Modellen $\leq .06$ und SRMR ist bei allen Modellen $\leq .10$ (s. Tabelle 6.26). Wie bei der SSS-V sind jedoch auch hier die GFI- und CFI-Werte sehr gering.

Tabelle 6.26

Anpassungsgüte an die Daten für die KFA-Modelle zum ZKPQ

ungewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	8532.06 (3320)	< .001	.050	.090	.72	.62
oblique	8260.12 (3310)	< .001	.048	.074	.73	.64
gewichtete Stichprobe						
	χ^2 (df)	p	RMSEA	SRMR	GFI	CFI
orthogonal	10020.65 (3320)	< .001	.056	.091	.69	.56
oblique	9803.52 (3310)	< .001	.055	.078	.69	.57

Anmerkungen. RMSEA= Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness of Fit Index; CFI = Comparative Fit Index.

Da Zuckerman et al. (1988) Zusammenhänge zwischen den fünf Faktoren annehmen, wird hier die KFA-Lösung mit korrelierten Faktoren dargestellt. Tabelle 6.27 zeigt die Lösung für die ungewichtete Stichprobe. Die KFA-Lösung für die gewichtete Stichprobe findet sich im Anhang (Tabelle A 6.10). Beide Lösungen sind mit den Annahmen von Zuckerman et al. (1993) kompatibel. Die meisten Hauptladungen liegen über .30 (wobei hier Ladungen von .30 bereits auf dem 1%-Niveau statistisch bedeutsam sind). Bemerkenswert sind die substanziellen Korrelationen von Impulsivem Sensation Seeking (Imp) mit Aggression-Feindseligkeit (A-F), Soziabilität (So) und Aktivität (Akt), die einen Hinweis auf einen Faktor höher Ordnung darstellen.

Tabelle 6.27

Ladungsmuster der KFA mit korrelierten Faktoren für den ZKPQ in der ungewichteten Stichprobe (completely standardized solution)

Item/Skala	Imp	N-Ä	A-F	So	Akt
Imp 01	.36	-	-	-	-
Imp 06	.29	-	-	-	-
Imp 14	.49	-	-	-	-
Imp 19	.33	-	-	-	-
Imp 24	.56	-	-	-	-
Imp 29	.44	-	-	-	-
Imp 34	.41	-	-	-	-
Imp 39	.60	-	-	-	-
Imp 45	.61	-	-	-	-
Imp 50	.31	-	-	-	-
Imp 55	.54	-	-	-	-
Imp 60	.40	-	-	-	-
Imp 65	.50	-	-	-	-
Imp 70	.62	-	-	-	-
Imp 75	.27	-	-	-	-
Imp 79	.52	-	-	-	-
Imp 84	.59	-	-	-	-
Imp 89	.45	-	-	-	-
Imp 95	.48	-	-	-	-
N-Ä 02	-	.30	-	-	-
N-Ä 07	-	.51	-	-	-
N-Ä 15	-	.61	-	-	-
N-Ä 20	-	.45	-	-	-
N-Ä 25	-	.69	-	-	-
N-Ä 35	-	.51	-	-	-
N-Ä 41	-	.44	-	-	-
N-Ä 46	-	.53	-	-	-
N-Ä 51	-	.55	-	-	-
N-Ä 56	-	.42	-	-	-
N-Ä 61	-	.69	-	-	-
N-Ä 66	-	.40	-	-	-
N-Ä 71	-	.46	-	-	-
N-Ä 76	-	.51	-	-	-
N-Ä 80	-	.58	-	-	-
N-Ä 85	-	.21	-	-	-
N-Ä 90	-	.74	-	-	-
N-Ä 96	-	.41	-	-	-
A-F 03	-	-	.32	-	-
A-F 08	-	-	.57	-	-
A-F 11	-	-	.52	-	-
A-F 16	-	-	.37	-	-
A-F 21	-	-	.39	-	-
A-F 42	-	-	.31	-	-
A-F 47	-	-	.27	-	-
A-F 57	-	-	.51	-	-
A-F 67	-	-	.34	-	-
A-F 72	-	-	.54	-	-
A-F 77	-	-	.52	-	-
A-F 86	-	-	.41	-	-
A-F 91	-	-	.46	-	-
A-F 97	-	-	.53	-	-

(Fortsetzung und Anmerkungen nächste Seite)

Fortsetzung Tabelle 6.27

Ladungsmuster der KFA mit korrelierten Faktoren für den ZKPQ in der ungewichteten Stichprobe (completely standardized solution)

Item/Skala	Imp	N-Ä	A-F	So	Akt
So 09	-	-	-	.44	-
So 12	-	-	-	.39	-
So 17	-	-	-	.35	-
So 22	-	-	-	.61	-
So 58	-	-	-	.53	-
So 37	-	-	-	.35	-
So 9	-	-	-	.59	-
So 48	-	-	-	.64	-
So 52x	-	-	-	.64	-
So 63	-	-	-	.36	-
So 68	-	-	-	.53	-
So 78	-	-	-	.67	-
So 82	-	-	-	.23	-
So 87	-	-	-	.57	-
So 92	-	-	-	.51	-
So 98	-	-	-	.60	-
Akt 05	-	-	-	-	.56
Akt 13	-	-	-	-	.43
Akt 18	-	-	-	-	.34
Akt 28	-	-	-	-	.24
Akt 33	-	-	-	-	.64
Akt 38	-	-	-	-	.40
Akt 44	-	-	-	-	.61
Akt 49	-	-	-	-	.13
Akt 54	-	-	-	-	.36
Akt 59	-	-	-	-	.30
Akt 64	-	-	-	-	.44
Akt 74	-	-	-	-	.55
Akt 83	-	-	-	-	.73
Akt 88	-	-	-	-	.20
Akt 94	-	-	-	-	.36
Akt 99	-	-	-	-	.05
Interkorrelationen der Faktoren					
N-Ä	-.05				
A-F	.39	.23			
So	.45	-.14	.19		
Akt	.30	-.12	-.01	.27	

Anmerkungen. Imp= Impulsivität-Sensation Seeking, N-Ä= Neurotizismus-Ängstlichkeit, So= Soziabilität, Akt= Aktivität, A-F= Aggression-Feindseligkeit. Ladungen mit Beträgen > .30 wurden fett gedruckt. Ladungen mit Beträgen > .30 wurden fett gedruckt. Die auf null fixierten Ladungen wurden mit „-“ gedruckt. Zur Modellgüte siehe Tabelle 6.26.

Diskussion der exploratorischen und konfirmatorischen Faktorenanalysen zum ZKPQ

Wenn man von den Ergebnissen der Parallelanalyse absieht, erbringen die Ergebnisse der EFA eine Bestätigung der intendierten fünffaktoriellen Struktur. Auch die Ergebnisse der KFA bestätigen die intendierte fünffaktorielle Struktur. Vor dem Hintergrund, dass der ZKPQ ein vergleichsweise neues Erhebungsinstrument ist und hier in einer um 11 Items verkürzten Form eingesetzt wurde, erscheint die vorliegende Bestätigung der intendierten Struktur besonders

erfreulich. Aufgrund der Deutlichkeit der von Zuckerman et al. (1993) postulierten Struktur in den dargestellten Analysen erscheint der ZKPQ für weiterführende Analysen in einem größeren Variablenpool als besonders geeignet.

6.2.5 Zusammenfassende Diskussion der Ergebnisse zu den modellspezifischen Analysen

Die Ergebnisse der modellspezifischen Analysen wurden in Tabelle 6.28 zusammenfassend dargestellt. Modellkonforme Ergebnisse wurden in der Tabelle mit „+“, diskonforme Ergebnisse mit „-“ und einige uneindeutige Fälle mit „+/-“ gekennzeichnet. Die uneindeutigen Fälle treten nur beim 16 PF-R und beim NEO-PI-R auf, da diese Fragebogen eine hierarchische Struktur abbilden sollen und demzufolge die Modellbestätigung auf verschiedenen Hierarchieebenen bewertet werden muss.

Tabelle 6.28

Zusammenfassung der Ergebnisse der modellspezifischen Analysen

Stichprobe/ Fragebogen	Parallelanalyse		Komparabilität		KFA-Modellfit	
	ungewichtet -tet	gewichtet	ungewichtet -tet	gewichtet	ungewichtet -tet	gewichtet
EPQ-R	-	-	-	+	+	+
16 PF-R	-	-	-	-	+/-	+/-
NEO-PI-R	-	-	-	-	+	+/-
SSS-V	-	-	-	-	+	+
ZKPQ	-	-	+	+	+	+

Anmerkung. Beim 16 PF-R bezieht sich die „+/-“ Bewertung darauf, dass nur Modelle mit Faktoren erster Ordnung bestätigt werden konnten. Beim NEO-PI-R bezieht sich das „+/-“ darauf, dass in der gewichteten Stichprobe die intendierte Struktur mit fünf orthogonalen Faktoren zweiter Ordnung einen unzureichenden Modellfit aufweist, während ein Modell mit obliquen Faktoren zweiter Ordnung einen angemessenen Modellfit aufweist.

Es fällt auf, dass die Parallelanalyse in keinem der Fälle die intendierte Faktorenanzahl indizierte. Die eindeutigste Bestätigung der intendierten Struktur ergab sich für den ZKPQ, da beim ZKPQ sowohl die Komparabilitäten der EFA als auch die KFA-Modelle modellkonforme Ergebnisse erbrachten. Den „zweitbesten“ Rang der Modellbestätigung weist hier der EPQ-R auf, bei dem der Grad der Bestätigung offenbar von der Verwendung der gewichteten oder der ungewichteten Stichprobe abhängt. Die Modellbestätigung des NEO-PI-R und der SSS-V konnte nur anhand der KFA erfolgen, weshalb diese in einer Rangfolge der Bestätigung den

nächsten Platz belegen. Der geringste Grad an Modellbestätigung wurde für den 16 PF-R ermittelt, dessen intendierte Struktur auch anhand der KFA nur in Bezug auf die Faktoren erster Ordnung bestätigt werden konnte.

Die vergleichsweise eindeutige Replikation der „Alternative Five“ mit dem ZKPQ ist insofern erstaunlich als dieser Fragebogen noch relativ neu ist und – im Vergleich zu den meisten übrigen Fragebogen – ein weniger etabliertes Modell repräsentiert. Das günstige Abschneiden des EPQ-R im Vergleich zum 16 PF-R und zum NEO-PI-R kann auf die deutlich geringere Komplexität der intendierten Strukturen zurückgeführt werden. Vermutlich bereitet die Abbildung komplexer hierarchischer Strukturen mit Hilfe aufsteigender EFA besondere Schwierigkeiten. Insgesamt fällt auf, dass die Fragebogen zur Abbildung einfacherer Strukturen, also der EPQ-R und der ZKPQ (nicht die SSS-V) im Vergleich zum 16 PF-R und zum NEO-PI-R eine größere Anzahl von Items pro Faktor erster Ordnung aufweisen. Vermutlich sollten die Ergebnisse der Simulationsstudie von Velicer und Fava (1998), die darauf hindeuten, dass pro Faktor in vielen Fällen eine extrem hohe Anzahl von Variablen erforderlich ist, bei der Fragebogen-Konstruktion im Trait-Bereich stärkere Berücksichtigung finden.

Durch eine gezielte Voraggregation von 32 Halb-Skalen im Falle des 16 PF-R und von 30 Facetten-Skalen im Falle des NEO-PI-R könnten bestimmte Aspekte der intendierten Strukturen wahrscheinlich auch mit der EFA aufgezeigt werden. Dabei muss jedoch bedacht werden, dass die Voraggregation bereits die Gültigkeit der intendierten Struktur auf der Ebene der Voraggregate oder Facetten voraussetzt. Insofern sind derartige Analysen nicht mehr eindeutig explorativ, sondern stellen ein nicht immer leicht durchdringbares Gemisch konfirmatorischer Setzungen und exploratorischer Prüfungen voraus. Im Gegensatz zu den Arbeiten zum Berliner Intelligenzstrukturmodell, bei dem der deduktive Charakter von Bündellungen explizit diskutiert und mit den theoretischen Modellvorstellungen in Einklang gebracht wird (z.B. Jäger, 1982; Jäger et al., 1997) ist der theoretische Status der Voraggregationen bei den genannten Fragebogen weniger eindeutig.

Die vorliegenden Analysen sollten die Gültigkeit der intendierten Faktoren im modell-spezifischen Kontext für den Fall eines eindeutig exploratorischen Vorgehens auf Basis einzelner Items evaluieren, da nur itembasierte exploratorische Analysen die Grundlage für modellübergreifende Analysen sind, bei denen überlappende Strukturen oder eine wechselseitige Integration der Faktoren ermittelt werden kann. Wenn Skalen oder Faktoren erster Ordnung als Grundlage für die modellübergreifenden Analysen verwendet würden, würde ja die Gültigkeit der Modelle im modellübergreifenden Variablenpool bereits partiell voraus-

gesetzt werden. Es ist beispielsweise möglich, dass der Psychotizismus-Faktor, der in den modellspezifischen Analysen auftritt, in modellübergreifenden Analysen nicht nachgewiesen werden könnte. Durch Voraggregation einer derartigen Skala könnte das Auftreten eines Psychotizismus-Faktors im modellübergreifenden Kontext natürlich nicht geprüft werden.

Auch wenn bei den modellspezifischen Analysen die einfacheren Strukturen des EPQ-R und ZKPQ besser nachgewiesen werden konnten als die komplexeren Strukturen des 16 PF-R und des NEO-PI-R so muss dies nicht zwangsläufig dazu führen, dass diese Strukturen in den modellübergreifenden Analysen dominieren. Die komplexeren Fragebogen zeichnen sich insgesamt durch eine besonders große Anzahl von Items aus, auch wenn die Zahl der Items pro Primärfaktor vermutlich teilweise zu gering ist. Wegen der absolut großen Itemzahl erscheint es nicht ausgeschlossen, dass auch die intendierten Strukturen des 16 PF-R und des NEO-PI-R in den Analysen mit einem modellübergreifenden Itempool eine große Rolle spielen.

Schließlich stellt sich die Frage, wie die Diskrepanzen zwischen EFA und KFA zu erklären sind, da im Traitbereich eher eine Modellbestätigung mit der EFA als anhand der KFA erwartet worden wäre (s. Costa et al., 1996). Zum einen ist dieser Unterschied auf die neue Bewertungsstrategie nach Hu und Bentler (1999) bzw. Raykov (1998) zurückzuführen, denn GFI und CFI deuten weiterhin auf einen ungenügenden Modellfit hin. Zum anderen könnte ein Grund für die günstigen Ergebnisse mit der KFA darin liegen, dass die Hauptladungen der meisten Items eher gering sind, dass aber die Zuordnung der Items zu den Faktoren häufig einigermaßen modellkonform ist. Mit der KFA besteht somit die Möglichkeit, eine in den Daten moderat aufzeigbare Struktur durch die systematische Unterdrückung der Nebenladungen zu optimieren.

6.3 Modellübergreifende, gemeinsame Faktorenanalysen für die Fragebogen

In den Analysen anhand eines modellübergreifenden Itempools können jeweils nur drei Modelle verglichen werden (s. Kapitel 6.1.2). In der Teilgruppe A können die Modelle von Cattell, Eysenck und Costa und McCrae verglichen werden, während in der Teilgruppe B die Modelle von Zuckerman, Eysenck und Costa und McCrae verglichen werden können.

6.3.1 Modellübergreifende Analysen für die Modelle von Cattell, Eysenck sowie Costa und McCrae

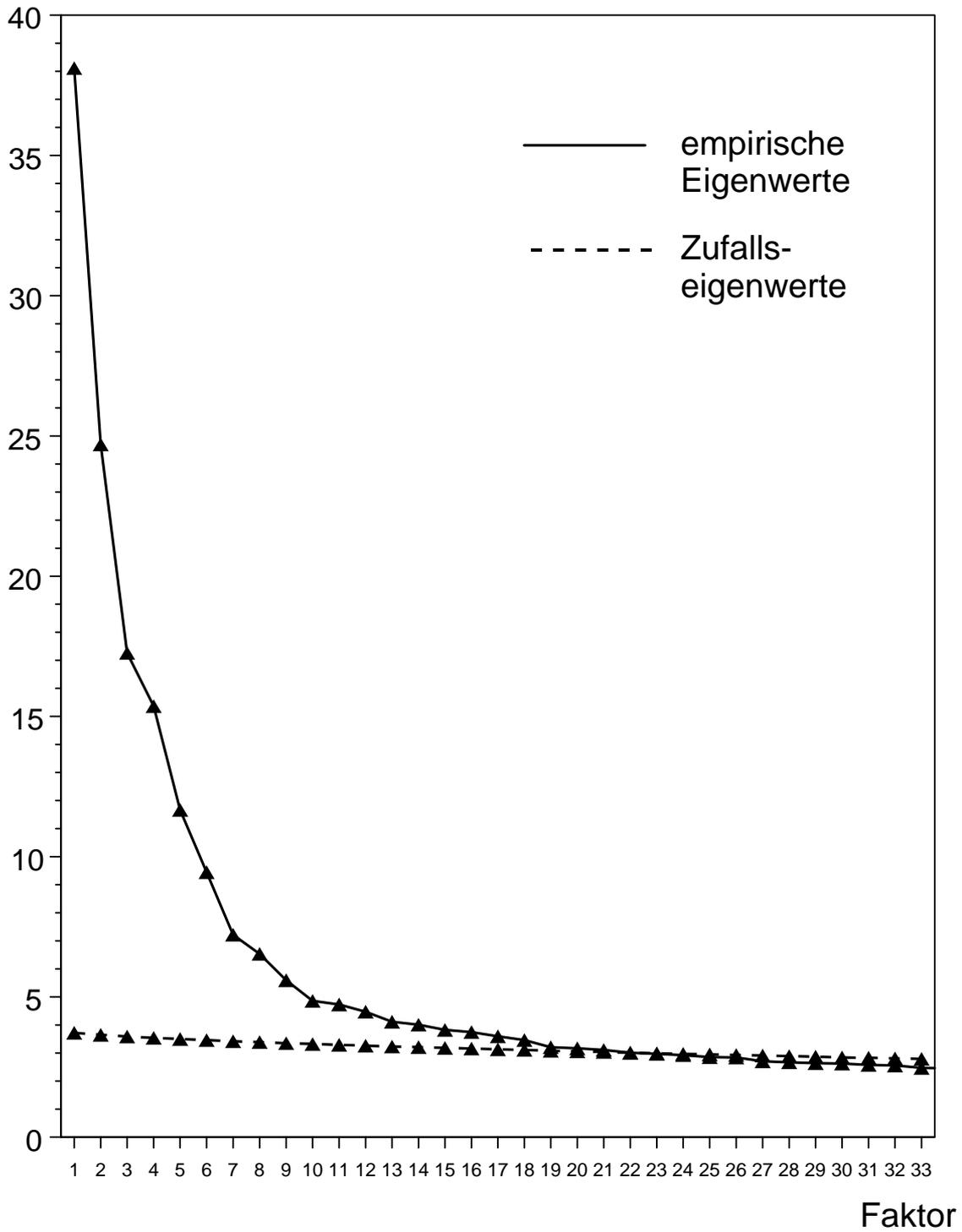
Für die Analysen anhand eines die Modelle von Cattell, Eysenck sowie Costa und McCrae übergreifenden Itempools sind die Datensätze von 554 Probanden aus Teilgruppe A verfügbar. Vor der gemeinsamen EFA wurde untersucht, in welchem Ausmaß der gemeinsame Itempool extrem ähnliche Items enthält. Die Anzahl extrem ähnlicher Items ist insgesamt sehr klein, was teilweise auf das unterschiedliche Item-Format des EPQ-R, 16 PF-R und NEO-PI-R zurückgeführt werden kann. So enthalten zwei der drei Antwortalternativen im 16 PF-R häufig nicht nur eine einfache „stimmt/stimmt nicht“ Wertung sondern komplette Aussagen. Im EPQ-R werden Fragen formuliert, auf die mit „Ja“ oder „Nein“ zu antworten ist. Im NEO-PI-R werden – wie im 16 PF-R – Aussagen formuliert, die auf einer 5-stufigen Rating-Skala eingestuft werden. Insgesamt wurden lediglich sieben einander sehr ähnliche Item-Paare identifiziert, die im Anhang aufgeführt sind (siehe Tabelle A 6.11), wobei zwei Items dieser Paare nicht in die Analysen eingehen, weil sie Items einer Antwortstil-Skala (Impression Management) des 16 PF-R betreffen. Somit sind nur fünf der insgesamt 497 Items paarweise im gemeinsamen Itempool enthalten, was als vernachlässigbar angesehen wird. Außerdem ist die Eliminierung eines der beiden Items eines Paares ein subjektiver Prozess und es erscheint angemessener, den Wert der geplanten Analysen nicht durch die Subjektivität einer derartigen Item-Auswahl einzuschränken.

6.3.1.1 Faktorenanalysen erster Ordnung

Zunächst wurde eine Parallelanalyse anhand von 100 Zufallsläufen für die gewichtete und die ungewichtete Stichprobe durchgeführt (siehe Abbildung 6.9). In der ungewichteten Stichprobe sind 21 empirische Eigenwerte größer als die entsprechenden Zufallseigenwerte (siehe auch im Anhang, Tabelle A 6.13, unten) und in der gewichteten Stichprobe sind 32 empirische Eigenwerte größer als die entsprechenden Zufallseigenwerte (siehe auch Tabelle A 6.16, unten).

a) Ungewichtet

Eigenwert



b) Gewichtet

Eigenwert

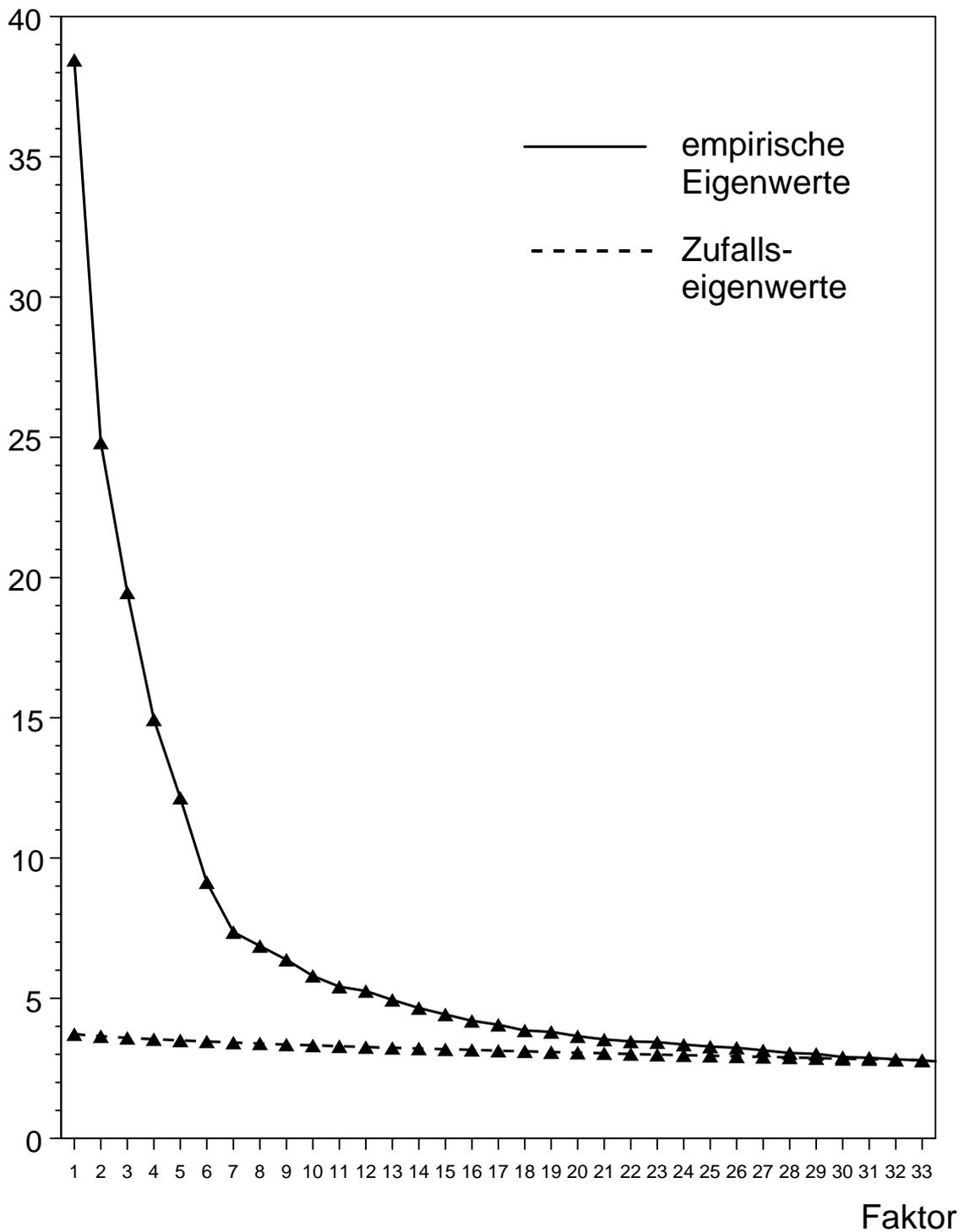


Abbildung 6.9. Parallelanalyse für den gemeinsamen Itempool des EPQ-R, 16 PF-R und NEO-PI-R (gewichtete und ungewichtete Stichprobe; N=554)

Die Anzahl der Probanden war für die Berechnung der Faktor-Komparabilitäten im modellübergreifenden Variablenpool zu gering, da die Berechnung der Komparabilitäten eine Halbierung der Probandengruppe erfordert. Bei einer Halbierung der Probandengruppe gehen weniger Probanden (277) als Variablen (494) in die Analysen ein. Aus diesem Grund wurden die Ergebnisse der Komparabilitäten nur aus Gründen der Vollständigkeit aufgeführt, jedoch nicht für die Festlegung der Anzahl zu extrahierender Faktoren herangezogen. Außerdem führte das ungünstige Verhältnis der Probandenanzahl zur Variablenanzahl dazu, dass die Komparabilitäten hier nicht auf der Basis von Hauptachsenanalysen sondern auf der Basis von Hauptkomponentenanalysen durchgeführt werden mussten (bei Hauptachsenanalysen traten nicht invertierbare Matrizen auf). Auch wenn es bisher keine Arbeiten gibt, die darauf hindeuten, dass Hauptkomponenten- oder Hauptachsenanalyse bei der Ermittlung der Komparabilitäten grundsätzlich zu präferieren sind, so wären Hauptachsenanalysen hier vorzuziehen, weil nach der Entscheidung für eine bestimmte Faktorenanzahl, gemäß Kapitel 3 in jedem Fall eine Hauptachsenanalyse durchgeführt wird. Dies schränkt die Bedeutung der Ergebnisse der Komparabilitäten hier noch weiter ein.

Die Komparabilitäten sind für die ungewichtete Stichprobe in Tabelle 6.29 für Lösungen mit zwei bis 21 Faktoren dargestellt, da die Parallelanalyse die Extraktion von 21 Faktoren indiziert. Die geringen Komparabilitäten können hier durch die bei geteilter Stichprobe zu kleine Anzahl an Probanden im Verhältnis zur Variablenanzahl bedingt sein. Es bestehen allerdings auch noch keine Erfahrungswerte für Komparabilitäten bei extrem großen Faktorenzahlen, so dass die Ursachen der geringen Komparabilitäten nicht eindeutig zu bestimmen sind. Hier sind weiterführende Methoden-Studien erforderlich. Die Komparabilitäten für die gewichtete Stichprobe sind im Anhang aufgeführt (Tabelle A 6.12). Da die Entscheidung über die Anzahl zu extrahierter Faktoren hier allein anhand der Parallelanalyse erfolgte, wurde im folgenden für die ungewichtete Stichprobe eine Hauptachsen-Lösung mit 21 Faktoren berechnet.

Faktoren erster Ordnung in der ungewichteten Stichprobe

Die komplette Ladungsmatrix der Oblimin-Lösung für 21 Faktoren und 494 Variablen ist im Anhang dargestellt (s. Tabelle A 6.13)¹. Ein grober Überblick über die Lösung kann auch anhand Tabelle 6.30 gewonnen werden, in der allerdings nur die drei Variablen mit den größten Hauptladungen für jeden der 21 Faktoren dargestellt sind.

¹ Die aufwendige Darstellung der kompletten Ladungsmuster im Anhang ermöglicht dem Leser die Überprüfung der hier vorgenommenen Interpretationen der Faktoren.

Tabelle 6.29

Faktor-Komparabilitäten für 2-21 Faktoren in Analysen des gemeinsamen Itempools des EPQ-R, 16 PF-R und NEO-PI-R in der ungewichteten Stichprobe (N=554)

	Faktor																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2 Faktoren	.77	.77																			
3 Faktoren	.74	.71	.71																		
4 Faktoren	.75	.73	.72	.70																	
5 Faktoren	.75	.74	.72	.70	.58																
6 Faktoren	.74	.74	.71	.67	.60	.51															
7 Faktoren	.74	.73	.66	.61	.59	.48	.42														
8 Faktoren	.66	.61	.60	.60	.54	.45	.44	.44													
9 Faktoren	.64	.61	.58	.55	.47	.45	.45	.41	.41												
10 Faktoren	.58	.55	.49	.48	.46	.46	.40	.33	.29	.18											
11 Faktoren	.98	.54	.52	.52	.52	.48	.47	.46	.35	.24	.00										
12 Faktoren	.56	.55	.53	.50	.50	.47	.43	.38	.34	.27	.19	.13									
13 Faktoren	.59	.58	.53	.50	.49	.49	.48	.47	.34	.28	.19	.14	.13								
14 Faktoren	.61	.57	.56	.53	.51	.45	.44	.41	.40	.36	.28	.26	.17	.14							
15 Faktoren	.60	.55	.54	.53	.46	.46	.45	.43	.41	.38	.34	.32	.26	.12	.05						
16 Faktoren	.58	.53	.52	.48	.47	.46	.45	.44	.40	.39	.30	.28	.26	.22	.21	.01					
17 Faktoren	.53	.52	.48	.48	.47	.46	.44	.44	.43	.43	.39	.32	.28	.25	.24	.15	.06				
18 Faktoren	.53	.51	.47	.44	.43	.43	.43	.40	.39	.38	.37	.27	.26	.26	.25	.22	.22	.16			
19 Faktoren	.58	.52	.51	.47	.44	.42	.41	.41	.40	.39	.37	.32	.28	.25	.25	.21	.18	.17	.01		
20 Faktoren	.53	.49	.46	.45	.44	.43	.43	.42	.39	.38	.37	.37	.28	.27	.25	.24	.23	.21	.19	.11	
21 Faktoren	.52	.47	.43	.43	.41	.38	.38	.37	.35	.32	.31	.30	.29	.28	.28	.25	.24	.21	.12	.11	.01

Anmerkung. Komparabilitäten > .85 wurden fett gedruckt.

Bei der Interpretation der Lösung ist zu berücksichtigen, dass die Items in der Polung in die Analysen eingingen, in der sie auch in die intendierten Skalen der jeweiligen Fragebogen eingingen. Das kann dazu führen, dass Items, die im EPQ-R zur Erfassung von Psychotizismus gepolt sind, im NEO-PI-R eine umgekehrte Polung zur Erfassung von Verträglichkeit aufweisen. Die wechselnden Polungen wurden bei der Benennung der Faktoren berücksichtigt, d.h. die Bezeichnungen wurden in der Regel im Sinne der Items, die positive Hauptladungen auf den Faktoren haben, gewählt. Wenn sich die Bezeichnung an den Items mit negativen Hauptladungen orientiert, wurde sie mit einem Minuszeichen gekennzeichnet.

Auf das Trasid-Ladungsmuster und dessen Ähnlichkeit zum Oblimin-Ladungsmuster wird weiter unten eingegangen. Wenn ein Faktor im wesentlichen von Variablen einer Subskala eines Fragebogens markiert wird, wurde die Zuordnung dieses Faktors zu dem betreffenden Fragebogen hinter der Faktorenbezeichnung in Tabelle 6.30 markiert und die Bezeichnung des Faktors entspricht dann der Skalenbezeichnung im betreffenden Fragebogen. Wenn ein Faktor zwar primär von Items eines Fragebogens markiert wird, diese Items jedoch unterschiedlichen Skalen des Fragebogens entstammen, orientiert sich die Faktorenbezeichnung natürlich nicht an dem betreffenden Fragebogen. Da die modellübergreifenden Analysen mehr als die modell-spezifischen Analysen exploratorischen Charakter haben, wurden hier die Bezeichnungen der Faktoren, die Ergebnis dieser Analysen sind, in Anführungszeichen gesetzt, um die Subjektivität der Interpretation zu kennzeichnen.

Zunächst muss festgehalten werden, dass die Interpretation des Faktors „selbstbestimmtes Handeln“ in Tabelle 6.30 unsicher ist, da dieser Faktor nur durch vier Items mit einem sehr heterogenen Inhalt geladen wird. Auch spricht das Auftreten eines Faktors mit nur vier Items dafür, dass die Anzahl extrahierter Faktoren im vorliegenden Material eher hoch ist, d.h., dass eine Unterextraktion auf der Basis der Parallelanalyse, wie sie in Turner (1998) und Beauducel (2001b) nachgewiesen wurde, hier wahrscheinlich nicht aufgetreten ist.

Die Faktoren „Offenheit für Ästhetik“, „Offenheit für Phantasie“ und „Besonnenheit (minus)“ entsprechen den gleichnamigen Facetten-Skalen im NEO-PI-R. Die Faktoren „Intelligenz“, „soziale Kompetenz“, „Empfindsamkeit“, „Abgehobenheit“ und „Privatheit“ werden vor allem durch Items der gleichnamigen Skalen des 16 PF-R markiert. Die beiden Faktoren „Perfektionismus“ und „Vertrauen“ werden durch je eine Subskala des 16 PF-R und je eine Facetten-Skala des NEO-PI-R markiert.

Tabelle 6.30

Oblimin-Hauptladungsmuster für die drei Items mit den höchsten Ladungen auf den 21 Faktoren der ungewichteten Stichprobe
(gemeinsamer Itempool: EPQ-R, 16 PF-R und NEO-PI-R)

Item-Kodierung/ Item in Stichworten / Ladungen	
1. Depressivität	
N3 NEO 191 zieml. düster & hoffnungslos	-.65
N3 NEO 071 bin selten traurig oder deprimiert	-.60
N EPQ 082 Fühlen Sie sich oft einsam?	-.59
2. Geselligkeit	
E2 NEO 037 habe gerne viele Leute um mich	.63
E2 NEO 217 mag Parties mit vielen Leuten	.63
16PF F26 verbringe d. Abend lieber auf Party	.59
3. Kompetenz	
C1 NEO 125 stolz auf gesundes Urteilsvermögen	.55
C1 NEO 185 bin in vielem kompetente Person	.48
A2 NEO 219 Geschick i. Umgang mit Mensch.	-.48
4. Berechnend/kalt	
A3 NEO 074 Leute halten mich f. kalt/berechn.	-.50
A2 NEO 189 Leuten zusetzen od. schmeicheln	-.48
A2 NEO 039 Menschen notfalls manipulieren	-.44
5. Offenheit für Ästhetik / NEO	
O2 NEO 098 begeistern Motive i. Kunst & Natur	.60
O2 NEO 188 wenn Lit./Kunstwerk, begeistert	.57
O2 NEO 128 Poesie beeindruckt mich wenig	.56
6. Intelligenz / 16 PF	
16PF B175 Intelligenz/schlußfolg. Denken	.82
16PF B172 Intelligenz/schlußfolg. Denken	.67
16PF B177 Intelligenz/schlußfolg. Denken	.67
7. Reizbarkeit, Ungeduld	
N2 NEO 066 aufbrausend/temperamentvoll	.57
N EPQ 016 Sind Sie leicht reizbar?	.52
N2 NEO 096 nicht reizbar/leicht erregbar	.52
8. Offenheit für Phantasie / NEO	
O1 NEO 033 Ausflüge ins Reich der Phantasie.	.36
O1 NEO 093 Zeit nicht mit Tagträumereien	.33
O3 NEO 103 wenig Aufmerksam. für Gefühle	.29
9. Konservatismus	
16PF Q1163 Bewährtes oder neue Ideen	.43
16PF F158 vertraute od. aufregende Situationen.	.42
P EPQ 017 immer das Gesetz befolgen?	.40
10. soziale Kompetenz / 16 PF	
16PF H52 Gespräch mit Fremden anfangen	.63
16PF H77 gegen Schüchternheit ankämpfen	.63
16PF H101 kontaktfreudig & selbstsicher	.62
11. Empfindsamkeit / 16 PF	
16PF I133 Interesse für technische Dinge	-.57
16PF I81 Interesse f. mech Dinge & Reparatur	-.55
16PF I31 Musiker interessanter als Mechaniker	-.53
12. Besonnenheit (-) /NEO	
C6 NEO 120 vor Handeln Konsequ. überdenk.	-.70
C6 NEO 090 erst handeln, dann denken	-.67
C6 NEO 060 gründl. nachdenk. vor Entscheid.	-.62
13. Benehmen vs. Unbefangenheit	
N4 NEO 226 Unfug v. Bekannten mir peinlich	-.33
P EPQ 013 Abneigung gegen nicht benehmen	.33
E EPQ 036 Lieben Sie handfeste Streiche?	.32
14. Sorge, Ängstlichkeit	
N1 NEO 151 beunr. über Dinge, die schiefgeh.	.56
16PF O15 zu viele Sorgen, über etwas getanes	.53
N EPQ 012 Sorg. üb. Dinge, d. n. hätte tun sollen	.52
15. Vertrauen /NEO,16PF	
A1 NEO 094 die meisten Menschen sind ehrlich	.53
A1 NEO 034 d. meisten Mensch. gute Absichten	.52
A1 NEO 214 habe Vertrauen in menschl. Natur	.48
16. Perfektionismus /16PF,NEO	
16PF Q371 Ich bin etwas perfektionistisch	.58
C4 NEO 200 strebe nach Perfektion bei allem	.48
C2 NEO 190 bin beim Putzen nicht pingelig	.44
17. hedonistische Kontrolle	
N5 NEO 111 esse meist zu viel v. Lieblingssp.	.47
N5 NEO 171 esse bis mir schlecht wird	.42
C4 NEO 080 gesund. Lebensw. nur paar Tage	-.38
18. selbstbestimmtes Handeln (?)	
16PF I104 lese lieber Action-Erzählungen	.81
16PF G102 Regeln umgehen, wenn Grund	.73
16PF Q3115 Plan, um zw. Aufg. keine Zeit	.73
19. Abgehobenheit /16 PF	
16PF M57 viele Ideen, weiß nicht wie umsetzen	.40
16PF M9 zerstreut, nicht immer praktisch	.39
16PF M32 immer praktische Sachen machen	.39
20. Altruismus	
A6 NEO 209 Mitgef. f. M., denen weniger gut	.38
A6 NEO 089 für arme, ältere M. nie genug tun	.36
A3 NEO 224 unterbreche. Tät., um and. zu helf.	.33
21. Privatheit/ 16 PF	
16PF N34 behalte meine Probleme für mich	-.38
16PF B180 Intelligenz: Schlußfolg. Denken	.37
16PF N14 meine Probleme für mich behalten	-.34

Anmerkungen. Die Item-Kodierungen enthalten die Test- und Skalenzugehörigkeit sowie die Nummern der Items in den jeweiligen Tests. Wenn Faktoren überwiegend durch Items einer Skala eines Tests markiert werden, wurde die Zugehörigkeit zu dem betreffenden Test hinter der Benennung des Faktors markiert. Unsichere Interpretationen bzw. Benennungen von Faktoren wurden mit „(?)“ gekennzeichnet. EPQ = EPQ-R; NEO = NEO-PI-R; 16PF = 16 PF-R.

Gemäß den in Kapitel 4.2 getroffenen Unterscheidungen sind die Faktoren, die eindeutig durch die Items einer Subskala eines Tests markiert werden, modellspezifische Faktoren. Modellspezifische Faktoren sprechen dafür, dass die im Modell postulierten Faktoren auf einem Itempool beruhen, der vom Itempool der übrigen Modelle abweicht und nicht ohne weiteres in die jeweils anderen Modelle integriert werden kann. Insofern spricht die große Zahl modellspezifischer Faktoren, drei für den NEO-PI-R und fünf für den 16 PF-R gegen einen allumfassenden Generalitätsanspruch der durch die beiden Fragebogen repräsentierten Modelle. Denn offenbar enthalten beide Modelle Items, die sich nicht ohne weiteres mit den übrigen Items im Gesamtpool zu allgemeineren Faktoren kombinieren. Selbst wenn man den „Intelligenz“-Faktor der Persönlichkeitssphäre nicht zurechnet, verbleiben noch einige Faktoren des 16 PF-R, die sich kaum mit den NEO-PI-R Items kombinieren, und es verbleiben auch einige NEO-PI-R Items, die sich nicht mit den 16 PF-R Items kombinieren. Dabei muss festgehalten werden, dass eine große Zahl modellspezifischer Faktoren nicht in dieser gemeinsamen Analyse nachgewiesen werden konnten.

Insbesondere die Items des EPQ-R bilden hier keine einheitlichen Faktoren, bzw. verteilen sich auf eine größere Anzahl von Faktoren. Viele Extraversions-Items des EPQ-R laden auf den Faktoren „Geselligkeit“ und „soziale Kompetenz“, viele Neurotizismus-Items laden auf den Faktoren „Depressivität“ und „Sorge, Ängstlichkeit“ und einige Psychotizismus-Items laden auf den Faktoren „Berechnend/kalt“ und „Benehmen vs. Unbefangenheit“. Dies kann als Hinweis darauf gelten, dass die drei Skalen des EPQ-R eine höhere Generalität aufweisen als die hier nachgewiesenen Primärfaktoren, da sie sich zu keinem einzelnen Primärfaktor kombinieren. Es bleibt jedoch zu prüfen, ob sich die genannten Faktorenpaare „Geselligkeit“ und „Kompetenz“, „Depressivität“ und „Sorge, Ängstlichkeit“, sowie „Berechnend/kalt“ und „Benehmen vs. Unbefangenheit“ bei den Faktorenanalysen zu Faktoren höherer Ordnung kombinieren, die im Sinne der Persönlichkeitsdimensionen des EPQ-R interpretiert werden können.

Als nächstes wird die Trasid-rotierte Lösung dargestellt. Da die Faktorladungsmatrizen mit 494 Variablen und 21 Faktoren extrem umfassend sind, wird die komplette Ladungsmatrix im Anhang dargestellt (s. Tabelle A 6.14), während in Tabelle 6.31 ein grober Überblick über die Lösung anhand der drei Items mit den jeweils höchsten Ladungen pro Faktor gegeben wird.

Es zeigt sich, dass insgesamt 13 der 21 Faktoren sehr ähnlich wie in der Oblimin-Lösung interpretiert werden können, während sich bei acht Faktoren deutliche Abweichungen zeigen. Deutliche Abweichungen zeigen sich zunächst im ersten Faktor, der in der Oblimin-Lösung als Depressivitäts- und in der Trasid-Lösung als Neurotizismus-Faktor interpretiert werden muss, da neben den Depressivitäts-Items auch die Items zu Ängstlichkeit und Sorge in der Trasid-Lösung auf dem ersten Faktor laden. Die Tatsache, dass somit deutliche Generalitätsunter-

schiede von Faktoren in Abhängigkeit von der Rotationsmethode auftreten, erscheint dabei bemerkenswert.

Tabelle 6.31

Trasid-Hauptladungsmuster für die drei Items mit den höchsten Ladungen auf den 21 Faktoren der ungewichteten Stichprobe
(gemeinsamer Itempool: EPQ-R, 16 PF-R und NEO-PI-R)

Item-Kodierung/ Item in Stichworten / Ladungen	
1. Neurotizismus	
N1 NEO 151 beunr. über Dinge, die schiefg.	.78
N3 NEO 191 zieml. düster & hoffnungslos	.71
N1 NEO 091 oft angespannt und nervös	.70
2. Geselligkeit	
16PF H101 kontaktfreudig & selbstsicher	.63
16PF H52 Gespräch mit Fremden anfangen	.62
16PF H166 gesprächig oder zurückhaltend	.61
3. Leistungsstreben	
C4 NEO 110 arbeite hart, um Ziele zu err.	.61
C1 NEO 215 arbeite zielstrebig & effektiv	.61
C5 NEO 055 verträdele Zeit, bevor Arbeit	.59
4. Soziale Kompetenz / 16 PF	
E2 NEO 037 gerne viele Leute um mich	.66
E2 NEO 217 mag Parties mit vielen Leuten	.66
16PF Q218 am liebsten Menschen um mich	-.61
5. Offenheit für Ästhetik / NEO	
O2 NEO 098 begeistern Motive in Kunst	.65
O2 NEO 128 Poesie beeindruckt mich wenig	.63
O2 NEO 188 wenn Lit./Kunstwerk begeistert	.62
6. Bescheidenheit (-)	
A2 NEO 039 Menschen notfalls manipulier.	.55
A5 NEO 174 anderen nicht überlegen föhl.	.52
A5 NEO 234 glaube, daß anderen überlegen	.49
7. Intelligenz / 16 PF	
16PF B175 Intelligenz/schlußf. Denken	.85
16PF B172 Intelligenz/schlußf. Denken	.68
16PF B177 Intelligenz/schlußf. Denken	.67
8. Besonnenheit (-) / NEO	
C6 NEO 120 vor Handeln Konsequenz. Überd.	-.70
C6 NEO 090 erst handeln, dann denken	-.67
C6 NEO 060 gründl. Nachdenk. vor Entsch.	-.61
9. Konservatismus/Benehmen (-)	
16PF G100 Freiheit wichtiger als Manier.	-.46
P EPQ 086 Regeln der Ges. od. eig. Weg	.45
16PF G28 gute Manieren respektieren	-.44
10. Reizbarkeit, Ungeduld	
N2 NEO 066 Aufbrausend/temperamentvoll	.52
A4 NEO 109 zeige offen wenn M. nicht mag	-.47
N2 NEO 096 nicht reizbar/leicht erregbar	.45
11. Vertrauen	
A1 NEO 094 die meisten Menschen ehrlich	.50
16PF L171 zuviel erzählt, nützen and. aus	-.48
A1 NEO 064 von meisten Leuten ausgenutzt	.47
12. Empfindsamkeit / 16 PF	
16PF I81 Interesse für mechan. Dinge	.55
16PF I133 Interesse für technische Dinge	.55
16PF I31 Musiker interess. als Mechan.	.48
13. Kompetenz	
C1 NEO 125 stolz auf gesundes Urteilsvm.	.48
A2 NEO 219 Geschick im Umgang mit Mensch.	-.46
A3 NEO 164 meisten Menschen mögen mich	.41
14. Berechnend/kalt	
E1 NEO 092 halten mich für kühl/distanz.	.46
A3 NEO 074 halten mich für kalt/berechn.	.44
16PF M11 Ideen verfolge, ohne besprech.	-.36
15. Selbstunsicherheit	
N4 NEO 106 nicht verl., wenn andere spott.	.31
N EPQ 082 Fühlen Sie sich oft einsam?	-.30
16PF Q4105 leicht ärgerlich & ungeduldig	.28
16. Privatheit / 16 PF	
16PF N86 über meine Gefühle spreche ich	.37
16PF N14 meine Probleme für mich beh.	.37
16PF N34 behalte meine Prob. für mich	.36
17. Dominanz und Unbefangenheit	
E3 NEO 162 Lieber eigene Wege als Gruppe	.29
E3 NEO 102 bei Vers. Reden and. Überlass.	.27
E3 NEO 222 fällt schwer führende Rolle	.27
18. Unstrukturiertheit (?)	
C2 NEO 220 nach Dingen suche, die verlegt	-.31
N EPQ 034 einen nervösen Menschen?	.30
16PF G76 wenn Konto nicht bel., Hinweis	.28
19. Selbstbestimmtes Handeln (?)	
16PF I104 lieber Action-Erzählungen	.75
16PF G102 Regeln umgehen, wenn Grund	.68
16PF Q3115 Plan, um zw. Aufg. keine Zeit	.67
20. Intelligenz, Selbstkontrolle	
16PF B180 Intelligenz: Schlußf. Denken	.34
N5 NEO 111 esse zu viel von Lieblingssp.	.31
16PF B182 Intelligenz: Schlußf. Denken	.31
21. Abgehobenheit	
16PF M32 immer prakt. Sachen machen	.34
16PF E124 nicht leicht was gefall. lass.	-.33
16PF I160 empfind./gef. od. rob./sachl.	.33

Anmerkungen. Die Item-Kodierungen enthalten die Test- und Skalenzugehörigkeit sowie die Nummern der Items in den jeweiligen Tests. Wenn Faktoren überwiegend durch Items einer Skala eines Tests markiert werden, wurde die Zugehörigkeit zu dem betreffenden Test hinter der Benennung des Faktors markiert. Unsichere Interpretationen bzw. Benennungen von Faktoren wurden mit „(?)“ gekennzeichnet. EPQ = EPQ-R; NEO = NEO-PI-R; 16PF = 16 PF-R.

Weitere Abweichungen zeigten sich für den achten Faktor, der in der Oblimin-Lösung die „Offenheit für Phantasie“ abbildet, während hier ein Faktor für die „Unstrukturiertheit“ des

eigenen Verhaltens auftritt. Der neunte Faktor, der in der Oblimin-Lösung „Konservatismus“ repräsentiert, beinhaltet in der Trasid-Lösung zusätzlich gutes Benehmen/Manieren und wird daher hier als „Konservatismus/Benehmen (minus)“ bezeichnet. Der 13. Faktor, der in der Oblimin-Lösung „Benehmen versus Unbefangenheit“ abbildet, entspricht in der Trasid-Lösung eher der Kombination „Dominanz und Unbefangenheit“. Für „Ängstlichkeit/Sorge“, den 14. Faktor der Oblimin-Lösung, gibt es in der Trasid-Lösung keine Entsprechung. Auch für den 16. Faktor der Oblimin-Lösung, „Perfektionismus“, kann in der Trasid-Lösung kein entsprechender Faktor gefunden werden. Der 17. Oblimin-Faktor, „hedonistische Kontrolle“ verändert seine Bedeutung in der Trasid-Lösung und wird nun „Intelligenz und Selbstkontrolle“ bezeichnet. Der 20. Faktor, der in der Oblimin-Lösung eher durch „Altruismus“ gekennzeichnet ist, wird in der Trasid-Lösung angemessener durch „Bescheidenheit“ gekennzeichnet, da er vor allem von Items der NEO-PI-R Facetten-Skala „Bescheidenheit“ markiert wird, obgleich auch einige Items der NEO-PI-R Facettenskala „Freimütigkeit“ auf diesem Faktor laden.

Die Unterschiede zwischen der Trasid- und Oblimin-rotierten Lösung sind beachtlich. Dies mag mit der großen Anzahl rotierter Faktoren und Variablen zusammenhängen. Eine große Ladungsmatrix bietet insgesamt mehr Möglichkeiten für unterschiedliche Optimierungsprozesse. In Bezug auf die Unterschiedlichkeit der Lösungen ist festzuhalten, dass in der Trasid-Lösung zwar nur 3% mehr Ladungsbeträge $\leq .05$ sind als in der Oblimin-Lösung, dass dies jedoch einer Absoluten Anzahl von 307 Ladungen entspricht, die in der Trasid-Lösung eindeutig als Nullladungen gelten können, während sie in der Oblimin-Lösung geringe aber statistisch bedeutsame Anteile repräsentieren. Vor dem Hintergrund der unterschiedlichen Optimierungsprozesse erscheint die Unterschiedlichkeit bei der Interpretation von acht Faktoren erklärbar.

Als nächstes werden die Interkorrelationsmatrizen der beiden schiefwinkligen Lösungen dargestellt. Die Interkorrelationen zwischen den Oblimin-Faktoren erster Ordnung, die die Grundlage für eine Faktorenanalyse zweiter Ordnung darstellen, sind eher gering (s. Tabelle 6.32). Nur fünf der insgesamt 210 Korrelationen (2.4%) haben Beträge über .20. Die Interkorrelationen zwischen den Trasid-Faktoren finden sich im Anhang (siehe Tabelle A 6.15). Auch hier haben nur fünf der 210 Korrelationen Beträge über .20.

Tabelle 6.32

Interkorrelationen zwischen den 21 Oblimin-rotierten Faktoren erster Ordnung in der Analyse des gemeinsamen Itempools des EPQ-R, 16 PF-R und NEO-PI-R in der ungewichteten Stichprobe (N=554)

	Faktor																			
	Depr.	Gesell.	Komp.	Be- rechn./ kalt	Off. für Ästhet	Intell.	Reizb., Unge- duld	Offen. für Phan.	Kons.	sozial. Komp.	Empf.	Beson- nenh.	Be- nehm. vs. Unbef.	Sorge, Ängst.	Ver- trauen	Per- fekt.	hedon. Kontr.	selb. Handl.	Abge- hob.	Altru- ismus.
Gesell.	.06																			
Komp.	.17	.11																		
Berechn./kalt	-.02	-.05	.03																	
Off. für Ästhet.	-.02	.06	.10	-.05																
Intell.	.02	.04	.04	.03	.09															
Reizb., Ungeduld	-.13	.01	-.01	.17	-.03	-.00														
Offen. für Phan.	.06	.15	-.03	.03	.10	.05	-.02													
Kons.	.04	.04	.04	.09	.13	.06	.04	.07												
sozial. Komp.	.17	.24	.25	.01	.08	-.02	.10	.03	.08											
Empf.	.17	-.03	.15	.09	-.04	-.08	-.03	-.04	.03	.05										
Besonnenh. (-)	-.11	.16	-.11	.07	.00	-.07	.13	.07	.13	.10	-.09									
Benehm. vs. Unbef.	.01	.11	.04	.05	.08	-.04	.01	.13	.13	.10	.01	.14								
Sorge, Ängst.	-.30	.02	-.11	-.03	.03	.07	.07	-.04	-.15	-.17	-.20	.02	-.10							
Vertrauen	.12	.10	-.04	-.15	.11	-.00	-.17	.11	.01	.09	-.04	.06	.04	-.07						
Perfekt.	.06	-.03	.22	-.03	.02	-.04	.05	-.12	-.12	.05	.10	-.19	-.13	.05	-.13					
hedon. Kontr.	-.15	.08	-.05	.14	.03	-.05	.05	.06	-.01	-.00	-.01	.14	.10	.09	-.04	-.08				
selb. Handl.	.09	.15	.19	-.10	.16	.03	.00	-.02	.01	.25	.02	-.00	.04	-.06	.10	.09	-.09			
Abgehob.	-.14	.03	.02	.09	.19	-.00	.01	.01	.08	.04	.03	.09	.11	.08	.03	-.06	.16	.00		
Altruismus.	-.02	.11	.02	-.17	.08	.03	-.08	-.07	-.10	.04	-.03	-.00	-.05	.12	.02	.01	.00	.12	.03	
Privatheit	.13	.03	.02	.07	.02	.05	.03	.20	.07	.08	.02	.01	.06	-.12	.11	-.07	-.06	-.00	-.02	-.09

Anmerkung. Korrelationen mit Beträgen > .20 wurden fett gedruckt. Die Vorzeichen der Korrelationen hängen mit der Polung der Faktoren bzw. Items zusammen. Die Abkürzungen der Faktorenbezeichnungen sind in Tabelle 6.30 erläutert.

Die insgesamt geringen Korrelationen zwischen den Faktoren erster Ordnung relativieren die Bedeutung von Faktoren höherer Ordnung. Dennoch korrelieren die beiden Faktoren der Oblimin-Lösung „Depressivität“ und „Sorge, Ängstlichkeit“, die als Aspekte von Neurotizismus gesehen werden können, sowie die Faktoren „Geselligkeit“ und „Soziale Kompetenz“, die als Aspekte von Extraversion gelten können, substantiell. Auch die Korrelation zwischen den Oblimin-rotierten Faktoren „Kompetenz“ und „Perfektionismus“ könnte als Hinweis auf einen Faktor „Gewissenhaftigkeit (minus)“ gelten.

Faktoren erster Ordnung in der gewichteten Stichprobe

Die komplette Ladungsmatrix der Oblimin-Lösung für die 32 Faktoren der gewichteten Stichprobe ist im Anhang dargestellt (s. Tabelle A 6.16). Ein grober Überblick über die Lösung kann auch anhand Tabelle 6.33 gewonnen werden, in der die drei Variablen mit den größten Hauptladungen für jeden der 32 Faktoren dargestellt sind. Bei der Interpretation der Lösung ist zu berücksichtigen, dass die Items in der Polung in die Analysen eingingen, in der sie auch in die intendierten Skalen der jeweiligen Fragebogen eingingen.

Die Oblimin-Lösung mit 32 Faktoren unterscheidet sich natürlich deutlich von der Oblimin-Lösung mit 21 Faktoren für die ungewichtete Stichprobe. Dabei ist zu beachten, dass einige der Interpretationen und Benennungen der 32 Faktoren mit deutlicher Unsicherheit behaftet sind (insbesondere für die Faktoren „Selbstbestimmtes Handeln“ und „Extremes Verhalten, Erleben meiden“).

In Bezug auf die Trasid-Lösung ist festzustellen, dass in der Varimax-Lösung, die hier als Ausgangslösung für die Trasid-Rotation verwendet wird, zwei Faktoren mit nur einer Hauptladung, ein Faktor mit nur zwei Hauptladungen und vier Faktoren mit nur drei Faktoren auftraten. Aufgrund der geringen Anzahl von Hauptladungen bei insgesamt sieben der 32 Faktoren erhält man den Eindruck, dass insgesamt eine zu große Anzahl von Faktoren extrahiert wurde. Andererseits trat bei der Oblimin-rotierten Lösung, die auf derselben unrotierten Lösung basiert, nur ein Faktor mit nur drei Hauptladungen auf (Faktor 18, „Selbstbestimmtes Handeln“), während alle übrigen Faktoren deutlich mehr Hauptladungen aufwiesen. Dies deutet darauf hin, dass die Anzahl der Hauptladungen pro Faktor in hohem Maße von der Rotationsmethode abhängt und somit für die Bewertung der Anzahl zu extrahierender Faktoren nur bedingt geeignet ist. In jedem Fall stellt die Varimax-Lösung eine ungünstige Startversion für die Trasid-Rotation dar, da in der Regel die Zahl der Hauptladungen pro Faktor durch die Trasid-Rotation nicht erhöht wird.

Tabelle 6.33

Oblimin-Hauptladungsmuster für die drei Items mit den höchsten Ladungen auf den 32 Faktoren der gewichteten Stichprobe (Itempool: EPQ-R, 16 PF-R und NEO-PI-R)

Item-Kodierung/ Item in Stichworten, rechts daneben: Ladungen	
1. Denressivität (-)	
N EPO 082 Fühlen Sie sich oft einsam?	-.63
N3 NEO 011 fühle mich selten einsam oder traurig	-.62
N3 NEO 071 bin selten traurig oder deprimiert	-.57
2. Besonnenheit (-)	
C6 NEO 090 erst handeln. dann denken	-.65
C6 NEO 120 vor Handeln Konsequ. überdenken	-.65
P EPO 077 Grundsatz: „erst Denken. dann handeln“	.62
3. Altruismus	
A6 NEO 209 Mitgeföhl. für M..denen weniger gut	.63
A6 NEO 089 für arme.ältere M. nie genug tun	.46
P EPO 097 würde Tier in der Falle leid tun?	-.43
4. Berechnend/kalt	
E1 NEO 092 Leute halten mich für kühl/distanziert	-.53
E1 NEO 062 als herzl./freundl. Person bekannt	-.51
A3 NEO 074 Leute halten mich für kalt/berechnend	-.50
5. Offenheit für Ästhetik / NEO	
O2 NEO 098 begeistern Motive in Kunst & Natur	.64
O2 NEO 188 wenn Literatur/Kunstwerk begeistert.	.55
O2 NEO 008 Kunst bedeutet mir nicht viel	.45
6. Pünktlichkeit/Verlässlichkeit	
P EPO 079 Zug in letzter Minute erreichen	-.60
C5 NEO 055 verträdele Zeit. bevor Arbeit beginne	.52
C5 NEO 025 kann meine Zeit gut einteilen	.45
7. Kompetenz (-)	
A5 NEO 234 glaube. daß ich anderen überlegen bin	.55
A5 NEO 084 bin besser als die meisten Menschen	.55
C1 NEO 125 bin stolz auf gesundes Urteilsvmögen	-.51
8. Intelligenz / 16PF	
16PF B175 Intelligenz/schlußfolg. Denken	-.76
16PF B177 Intelligenz/schlußfolg. Denken	-.67
16PF B172 Intelligenz/schlußfolg. Denken	-.66
9. Konservatismus	
16PF O164 Bewährtes oder neue Wege	.49
16PF A121 Maschinenpark od. Mitarbeitergespräch	.40
16PF O1138 bewährte oder neue Methoden	.37
10. Sorge. Ängstlichkeit	
16PF O342 halte meine Sachen in Ordnung	-.58
N1 NEO 151 beunr. über Dinge.die schiefg.	.39
16PF O187 Wege ausdenken. wie Welt verbesserbar	.34
11. Berufliches Vertrauen	
A2 NEO 099 abs. Ehrlichk. Bringt i. Geschäft nichts	.51
16PF L80 wenn offen. dann nützen andere aus	-.48
A1 NEO 064 von meisten Leuten ausgenutzt	.45
12. Geselligkeit (-)	
E2 NEO 037 habe gerne viele Leute um mich	-.58
16PF O218 am liebsten Menschen um mich	.56
E2 NEO 217 mag Parties mit vielen Leuten	-.55
13. Soziale Kompetenz / 16PF	
16PF H77 gegen Schüchternheit ankämpfen	-.60
16PF H52 Gespräch mit Fremden anfangen	-.59
16PF H101 kontaktfreudig & selbstsicher	-.59
14. Privatheit	
16PF N108 v. Antw. Drücken. wenn Fragen zu pers.	-.49
16PF N86 über meine Geföhle spreche ich	-.48
16PF N12 fällt leicht. über persönl. Dinge zu reden	-.43
15. Persönliches Vertrauen	
16PF N112 sehr vorsichtig.wem ich mich anvertraue	-.53
16PF B180 Intelligenz: Schlußfolg. Denken	.36
A1 NEO 094 die meisten Menschen sind ehrlich	.35
16. Freimütigkeit	
A4 NEO 109 zeige offen wenn ich M. nicht mag	.37
16PF E46 wenn andere falsch.sage was ich denke	-.37
E6 NEO 057 manchmal intensive Freude/Ekstase	-.36
17. Dominanz/Durchsetzungsfähigkeit	
P EPO 041 oft gegen Willen der Eltern entschieden	.56
E3 NEO 012 bin dominant. selbstsicher. durchsetzfäh.	.42
16PF E151 Herausforderung od. Wünsche anderer	.41
18. Robustheit	
16PF N14 meine Probleme für mich behalten	.43
16PF I79 schönes Gedicht od. Fußballstrategie	-.39
A4 NEO 019 lieber zusammenarb. als wetteifern.	-.36
19. Regelbewußtsein	
16PF G76 wenn Konto f. etw. nicht belastet. Hinweis	.37
16PF O138 solide Bürger oder Reformier	-.31
16PF G6 Fähigkeit od. Verantwortungsbewußtsein.	.30
20. Extremes Verhalten. Erleben meiden	
A2 NEO 189 Leuten zusetzen oder schmeicheln.	.44
N5 NEO 051 Schwier.. meinen Begierd. widerstehen	-.43
O1 NEO 183 als Kind selten Spaß am Schauspieln	-.32
21. Empfindsamkeit / 16PF	
16PF I31 Musiker interessanter als Mechaniker	-.57
16PF I81 Interesse für mechan. Dinge & Reparatur	-.55
16PF I133 Interesse für technische Dinge	-.52
22. Benehmen/Anpassungsbereitschaft (-)	
P EPO 013 Abneigung gegen nicht benehm.	.48
P EPO 021 Sind gute Manieren wichtig?	.48
16PF G28 Regeln & gute Manieren respektieren	-.47
23. Gemütlichkeit. zeitl./mental.	
E4 NEO 077 arbeite langsam aber stetig	.54
E4 NEO 017 arbeite und spiele gemächlich	.39
A6 NEO 239 lieber als gütig denn als gerecht bekannt	-.36
24. Reizbarkeit/Ungeduld	
N2 NEO 066 aufbrausend/temperamentvoll	.55
N EPO 016 Sind Sie leicht reizbar?	.53
N2 NEO 096 nicht reizbar/leicht erregbar	.53
25. Selbstbestimmtes Handeln	
16PF I104 lese lieber Action-Erzählungen	-.64
16PF O3115 Plan.um zw. Aufg. keine Zeit	-.56
16PF G102 Regeln umgehen. wenn Grund	-.50
26. Offenheit für Phantasie (-)	
O3 NEO 103 wenig Aufmerksamkeit für Geföhle	-.53
O1 NEO 033 Ausflüge ins Reich der Phantasie	-.38
E6 NEO 027 nie vor Freude in die Luft gesprungen	-.35
27. Schuldgefühle/Scham	
16PF O15 zu viele Sorg..über etwas getanes	.51
16PF O137 Schuldgefühle wegen belangloser Dinge	.44
N EPO 095 in manchen Dingen bes. empfindlich	.43
28. Leistungsstreben, Energie (-)	
C4 NEO 170 strebe. alles mir mögliche erreichen	-.56
C4 NEO 110 arbeite hart. um Ziele zu erreichen	-.51
16PF M57 viele Ideen.weiß nicht wie umsetzen	-.42
29. Perfektionismus	
16PF O320 in ungeordneten Umständen wohl fühlen	.51
16PF O371 Ich bin etwas perfektionistisch.	.51
C4 NEO 200 strebe nach Perfektion bei allem	.48
30. Verantwortungsbewußtsein/Eigensinnigkeit.	
E2 NEO 127 bevorzuge allein zu erledigend. Arbeiten	.30
16PF O65 zu sehr verantwortlich für Dinge fühlen	.29
E3 NEO 162 lieber eigene Wege als Gruppe anführen	.29
31. Prinzipientreue	
O6 NEO 148 Treue zu eigenen Idealen wichtig	-.30
O4 NEO 168 Zuhause etwas Neues ausprobieren	-.27
16PF G51 jede kl. Sicherheitsvorschrift einhalten	.27
32. Hedonistische Kontrolle	
N5 NEO 171 esse bis mir schlecht wird	-.57
N5 NEO 111 esse meist zu viel von Lieblingssp.	-.56
C2 NEO 220 Zeit f. Suche nach Dingen die verlegt	.32

Anmerkung. Die Item-Kodierungen enthalten die Test- und Skalenzugehörigkeit sowie die Nummern der Items in den jeweiligen Tests. Die Interkorrelationen der Faktoren finden sich in Tabelle A 6.17. Wenn Faktoren überwiegend durch Items einer Skala eines Tests markiert werden, wurde die Zugehörigkeit zu dem betreffenden Test hinter der Benennung des Faktors markiert. EPQ = EPQ-R; NEO = NEO-PI-R; 16PF = 16 PF-R.

Dementsprechend können aufgrund ungünstiger Ladungsverteilungen acht der 32 Trasid-rotierten Faktoren nur unter Vorbehalt interpretiert werden (s. Anhang, Tabelle A 6.18). Aufgrund der deutlichen Unsicherheiten bei der Interpretation der 32 Trasid-rotierten Faktoren wird die im Anhang dargestellte Lösung nicht in den hierarchischen Analysen weiterverfolgt.

Die Interkorrelationen der 32 Oblimin-rotierten Faktoren finden sich im Anhang in Tabelle A 6.17 (für die Trasid-rotierten Faktoren in Tabelle A 6.19). Die Interkorrelationen der 32 Oblimin-rotierten Faktoren fallen insgesamt noch etwas geringer aus als die Interkorrelationen der 21 Oblimin- und Trasid-rotierten Primärfaktoren, da nur eine der insgesamt 496 Interkorrelationen über .20 liegt (auch die Interfaktor-Korrelationen der Trasid-Lösung fallen äußerst gering aus). Aus o.g. Gründen wurde nur für die 32 Primärfaktoren der Oblimin-Lösung eine Faktorenanalyse zweiter Ordnung durchgeführt.

6.3.1.2 Faktorenanalysen zweiter Ordnung

Ungewichtete Stichprobe

Das Ergebnis der Faktorenanalyse zweiter Ordnung auf der Basis der 21 Oblimin-rotierten Faktoren der ungewichteten Stichprobe ist in Tabelle 6.34 dargestellt.

Faktor-Komparabilitäten, die die Berechnung von Faktorwerten erfordern, können auf der Basis einer Korrelationsmatrix nicht berechnet werden, so dass die Entscheidung für die Anzahl der zu extrahierenden Faktoren auf der Ebene der Faktoren zweiter Ordnung lediglich anhand der Parallelanalyse erfolgen konnte. Die Parallelanalyse für die Oblimin-rotierten Faktoren erster Ordnung deutete auf die Extraktion von sechs Faktoren hin (s. Tabelle 6.34). Mit 23.6% ist die Varianzaufklärung der Lösung eher gering, was auf die geringen Interkorrelationen zwischen den Oblimin-rotierten Faktoren erster Ordnung zurückgeführt werden kann.

Der erste Oblimin-rotierte Faktor zweiter Ordnung wird primär durch die Faktoren „soziale Kompetenz“, „Geselligkeit“ und „eigenständiges Handeln“ geladen, wobei auch eine Ladung des Faktors „Kompetenz“ auffällt. Diese Kombination von Faktoren erster Ordnung zu einem Faktor zweiter Ordnung kann vermutlich am besten mit dem Konzept der Extraversion in Verbindung gebracht werden, weshalb dieser Faktor auch so bezeichnet wurde. Der zweite Faktor in Tabelle 6.33 wird durch „Besonnenheit (minus)“, „Perfektionismus“, „Kompetenz“ und „Benehmen“ markiert und wurde hier mit „Gewissenhaftigkeit (minus)“ bezeichnet. Man könnte diesen Faktor aber auch mit dem Faktor „Selbstkontrolle“ des 16 PF-R (Schneewind & Graf, 1998) in Verbindung bringen. Der dritte Faktor, der durch die Primärfaktoren „Sorge,

„Ängstlichkeit“, „Empfindsamkeit“ und „Depressivität“ markiert wird, wurde mit „Neurotizismus“ bezeichnet.

Tabelle 6.34

Oblimin-rotiertes Ladungsmuster der sechs Faktoren zweiter Ordnung auf Basis der 21 Primärfaktoren (ungewichtete Stichprobe)

Faktor	1	2	3	4	5	6	
	E	G-	N	V-/P	unk. Annäh.	In-sich-gek.	h²
Soziale Kompetenz	.59	.04	-.16	.11	-.02	.02	.39
Geselligkeit	.43	.16	.09	-.02	.01	.06	.22
Selbstbestimmtes Handeln	.43	-.13	.00	-.10	.03	-.02	.23
Besonnenheit (-)	.19	.48	.08	.13	.06	-.03	.31
Perfektionismus	.14	-.43	-.02	.14	-.06	-.16	.27
Kompetenz	.34	-.35	-.22	.09	.10	-.01	.31
Benehmen	.10	.21	-.13	.01	.20	.08	.15
Sorge, Ängstlichkeit	-.05	-.08	.58	.08	.09	-.09	.38
Empfindsamkeit	-.01	-.13	-.41	.03	.07	-.12	.19
Depressivität	.15	-.11	-.41	-.16	-.20	.16	.33
Reizbarkeit, Ungeduld	.14	.03	.14	.48	-.06	.04	.24
Berechnend/kalt	-.11	.01	-.14	.43	.13	.11	.27
Vertrauen	.10	.16	-.04	-.38	.02	.15	.22
Altruismus	.19	-.04	.18	-.21	.04	-.19	.16
Abgehobenheit	-.03	.02	.00	.00	.51	-.07	.26
Offenheit für Ästhetik	.12	-.18	.09	-.16	.42	.17	.29
Hedonistische Kontrolle	-.02	.20	.06	.13	.25	-.11	.15
Privatheit	.02	.03	-.09	.04	-.08	.40	.18
Offenheit für Phantasie	.03	.13	.02	-.04	.05	.38	.18
Intelligenz	.00	-.16	.16	.01	.04	.26	.10
Konservatismus	.02	.09	-.15	.06	.17	.20	.13
Interkorrelationen der Faktoren							
G-	-.04						%-erkl.
N	-.07	.01					Varianz:
V-/P	-.12	.08	-.07				23.6
unk. Annäh.	.14	.16	.04	.09			
In-sich-gek.	.10	.11	-.12	-.06	.10		
Empirische Eigenwerte*:	2.07	1.88	1.61	1.51	1.21	1.19	1.05
	.96	.91	.85	.83	.78	.77	
Zufalls-Eigenwerte*:	1.36	1.30	1.25	1.21	1.17	1.14	1.11
	1.08	1.05	1.02	.99	.96	.94	

Anmerkungen. Ladungen mit Beträgen > .20 wurden fett gedruckt. Die Bezeichnungen der Faktoren zweiter Ordnung sind wie folgt: E = Extraversion; G- = Gewissenhaftigkeit (minus); N = Neurotizismus; V-/P= Verträglichkeit (minus)/Psychotizismus, unk. Annäh.= unkontrollierte Annäherung; In-sich-gek.= In-sich-gekehrt sein.

Der vierte Faktor wird durch Primärfaktoren repräsentiert, die mit „Verträglichkeit (minus)“ zusammenhängen. Die Kombination der Primärfaktoren „Reizbarkeit, Ungeduld“, „Berechnend/kalt“ und mit negativem Vorzeichen „Vertrauen“ und „Altruismus“ geben diesem Faktor auch eine große Nähe zum Psychotizismus-Konzept, vor allem, wenn man die Interpretation von Psychotizismus wie Zuckerman (1989) eher auf die antisozialen Persönlichkeitsmerkmale ausrichtet. Der fünfte Faktor wird durch die Primärfaktoren „Abgehobenheit“, „Offenheit für Ästhetik“ und „hedonistische Kontrolle“ markiert. Bei diesen Faktoren stehen Offenheit, geringe Selbstkontrolle und Hedonismus im Vordergrund, weshalb dieser Faktor hier versuchsweise als „unkontrollierte Annäherung“ interpretiert wurde.

Der sechste Faktor, der durch die Primärfaktoren „Privatheit“, „Offenheit für Phantasie“ und „Intelligenz“ markiert wird, wurde hier versuchsweise als „In-sich-gekehrt-sein“ interpretiert, wobei es sich hier nicht um Introversion im herkömmlichen, primär sozialen Sinne handelt, sondern eher um die Tendenz sich in die eigenen Kognitionen zurückzuziehen. Dieser Faktor entspricht vielleicht eher dem von Jung (1960) ursprünglich mit dem Begriff der Introversion intendierten Konzept der Beschäftigung mit der „inneren Realität“.

Es bleibt noch festzuhalten, dass die beiden letzten Faktoren, die man inhaltlich auch als zwei Aspekte der Offenheit für Erfahrungen interpretieren könnte, in der schiefwinkligen Lösung nur mit .10 interkorrelieren (s. Tabelle 6.34). Somit können diese beiden Faktoren nicht einen Faktor höherer Ordnung bilden. Die geringen Interkorrelationen der Oblimin-rotierten Faktoren zweiter Ordnung sprechen insgesamt dafür, hier keine Faktoren dritter Ordnung zu ermitteln.

Die Faktoren zweiter Ordnung auf Basis der Trasid-rotierten Primärfaktoren sind in Tabelle 6.35 dargestellt. Die Parallelanalyse sprach auch hier für die Extraktion von sechs Faktoren. Die Varianzaufklärung der sechsfaktoriellen Lösung ist mit 24.4% gering, was auf die geringen Interkorrelationen zwischen den Trasid-rotierten Primärfaktoren zurückzuführen ist. Mit den sechs Faktoren zweiter Ordnung wurde wiederum eine Trasid-Rotation durchgeführt.

Die in Tabelle 6.35 dargestellte Trasid-Lösung zweiter Ordnung auf Basis der Trasid-rotierten Primärfaktoren weicht deutlich von der auf Oblimin-Rotation basierenden Faktorenlösung zweiter Ordnung ab. Auf dem ersten Faktor lädt fast ausschließlich der Faktor „Bescheidenheit“, der zwar als Primärfaktor auch einige weitere Aspekte aber keineswegs alle Aspekte der Verträglichkeit einschließt, wie sie im NEO-PI-R konzeptualisiert ist. Dieser Faktor weist somit zwar eine große Nähe zu einem Verträglichkeitsfaktor i.S. des Fünf-Faktorenmodells auf, kann aber nicht damit gleichgesetzt werden. Auch der zweite Faktor wird nur durch einen einzigen Primärfaktor markiert. Dieser Primärfaktor war durch äußerst viele verschiedene

Items aus dem Neurotizismus-Bereich markiert und wurde deswegen auch in diesem Sinne interpretiert. Insofern als dieser Primärfaktor bereits einem vermutlich sehr generellen Faktor entspricht, verwundert es nicht, dass er auf der Ebene der Faktoren zweiter Ordnung im Prinzip unverändert bleibt.

Tabelle 6.35

Trasid-rotiertes Ladungsmuster der sechs Faktoren zweiter Ordnung auf Basis der 21 Primärfaktoren (ungewichtete Stichprobe)

Faktor	1	2	3	4	5	6							
	Be- scheid.	N	E	G	V-/P	Off. für Ästhet.	h ²						
Bescheidenheit	.98	-.02	.04	.01	.01	.01	.94						
Empfindsamkeit	.16	.15	.04	.13	.12	-.04	.08						
Neurotizismus	-.03	.94	-.01	.04	.03	-.04	.87						
Abgehobenheit	-.03	.16	-.01	.18	.08	.07	.08						
Soziale Kompetenz	-.02	-.04	.69	-.02	.02	.04	.51						
Geselligkeit	.05	.04	.41	.01	.11	.17	.23						
Reizbarkeit, Ungeduld	-.18	.05	.22	.16	-.19	-.21	.16						
Leistungsstreben	-.03	-.03	.15	-.59	-.05	.01	.36						
Konservatismus/Benehmen (-)	-.07	-.13	-.02	.41	-.14	.12	.23						
Besonnenheit (-)	-.05	.14	.19	.44	-.03	-.15	.23						
Selbstbestimmtes Handeln	.09	.05	.03	-.13	.04	.08	.02						
Intelligenz	-.08	.05	-.00	-.10	.02	.03	.03						
Berechnend/kalt	-.11	-.02	.03	-.01	.58	.04	.34						
Vertrauen	.05	-.21	-.05	.05	.32	.19	.20						
Privatheit	-.01	-.03	-.21	.01	-.23	-.02	.09						
Unstrukturiertheit	.02	-.02	-.11	.01	-.21	.03	.05						
Selbstunsicherheit	-.05	.00	.03	-.12	.15	.00	.03						
Offenheit für Ästhetik	.04	.02	.02	.02	-.05	.61	.38						
Kompetenz	-.04	-.13	.14	-.30	-.15	.26	.21						
Intelligenz Kontrolle	-.05	-.12	-.08	-.05	-.02	.17	.02						
Dominanz/Unbefangenheit	-.09	-.05	.04	-.04	-.09	.11	.04						
Interkorrelationen der Faktoren													
	N	.15					%-erkl.						
	E	-.10	-.25				Varianz:						
	G	-.11	-.02	.10			24.4						
	V-/P	.09	-.10	-.03	.17								
	Offenheit für Ästhet.	-.10	.07	.20	.23	.04							
Empirische Eigenwerte*:	1.83	1.71	1.53	1.29	1.21	1.16	1.08	1.06	1.00	.97	.95	.87	.86
Zufalls-Eigenwerte*:	1.36	1.30	1.25	1.21	1.17	1.14	1.11	1.08	1.05	1.02	.99	.96	.94

Anmerkungen. Ladungen mit Beträgen > .20 wurden fett gedruckt. Die Bezeichnungen der Faktoren zweiter Ordnung sind wie folgt: Bescheid.= Bescheidenheit; N = Neurotizismus; E = Extraversion; G = Gewissenhaftigkeit (minus); V-/P = Verträglichkeit (minus)/Psychotizismus, Off. für Ästhet. = Offenheit für Ästhetik.

Der dritte Faktor zweiter Ordnung wird vor allem durch die Faktoren „soziale Kompetenz“ und „Geselligkeit“ markiert, so dass dieser Faktor als Extraversionsfaktor interpretiert werden kann. Der vierte Faktor ist durch Leistungsstreben, Konservatismus und Besonnenheit (minus) markiert und kann als Faktor „Gewissenhaftigkeit (minus)“ i.S. des Fünf-Faktoren-Modells oder „Selbstkontrolle“ i.S. des 16 PF-R interpretiert werden. Der fünfte Faktor, der vor allem durch die Primärfaktoren „Berechnend/kalt“ und „Vertrauen“ markiert wird, wurde hier als Faktor „Verträglichkeit (minus)/Psychotizismus“ interpretiert, wobei die positive Ladung des Primärfaktors „Vertrauen“ auf diesem Faktor ungewöhnlich erscheint. Der sechste Faktor wird vor allem durch den Primärfaktor „Offenheit für Ästhetik“ markiert, weshalb dieser Faktor zweiter Ordnung auch in diesem Sinne interpretiert wird.

Insgesamt fällt auf, dass die beiden Faktoren „Bescheidenheit“ und „Verträglichkeit (minus)“, die beide als Aspekte eines allgemeineren Faktors für „Verträglichkeit“ gelten könnten, nur zu .09 miteinander korrelieren, weshalb sie empirisch keinen Faktor höherer Ordnung bilden können. Die Interkorrelationen zwischen den Trasid-rotierten Faktoren zweiter Ordnung sind zwar etwas höher als die zwischen den Oblimin-rotierten Faktoren zweiter Ordnung, jedoch sind sie nicht so hoch, dass eine Faktorenanalyse dritter Ordnung zur weiteren Reduktion der Daten angemessen erscheint.

Gewichtete Stichprobe

Das Ergebnis der Faktorenanalyse zweiter Ordnung auf der Basis der 32 Oblimin-rotierten Primärfaktoren der gewichteten Stichprobe ist in Tabelle 6.36 dargestellt. Bei einer Parallelanalyse waren die Eigenwerte der ersten fünf empirischen Faktoren größer als der Mittelwert der Eigenwerte von Zufalls-Faktoren (siehe Tabelle 6.36). Aus diesem Grund wurden fünf Faktoren zweiter Ordnung extrahiert, die insgesamt jedoch nur 16.1% der Varianz der Primärfaktoren erklären.

Die fünf Faktoren lassen sich im Sinne des Fünf-Faktoren-Modells der Persönlichkeit interpretieren, wobei die Faktoren „Extraversion“ und „Gewissenhaftigkeit“ negativ gepolt sind. In den Faktor „Verträglichkeit“ gehen hier Faktoren wie „Benehmen/Anpassungsbereitschaft (minus)“, „Regelbewusstsein“, „Perfektionismus“ und „Gemütlichkeit zeitl./mental“ ein, was darauf hindeutet, dass dieser Faktor, wie in den Analysen für die ungewichtete Stichprobe, einen deutlichen (negativen) Bezug zu Psychotizismus hat. Insbesondere auf dem Faktor „Benehmen/Anpassungsbereitschaft (minus)“ laden sehr viele Psychotizismus-Items des EPQ-R (siehe Tabelle A 6.16). Die Interkorrelationen zwischen den Faktoren zweiter Ordnung sind insgesamt gering, weshalb Faktorenanalysen dritter Ordnung hier nicht sinnvoll erscheinen.

Tabelle 6.36

Oblimin-rotiertes Ladungsmuster der fünf Faktoren zweiter Ordnung auf Basis der 32 Primärfaktoren (gewichtete Stichprobe)

Faktor	1	2	3	4	5										
	N	V	E-	G-	O	h²									
Depressivität (-)	-.47	-.13	-.06	-.07	-.18	.30									
Reizbarkeit/Ungeduld	.39	-.03	-.03	.05	-.07	.16									
Hedonistische Kontrolle	-.38	.12	.09	-.02	.01	.17									
Schuldgefühle/Scham	.38	.22	.06	.13	.12	.25									
Berufliches Vertrauen	-.32	-.05	-.12	.18	.08	.15									
Persönlich/privates Vertrauen	-.27	.01	-.08	.02	.16	.11									
Extremes Verhalten/Erleben meiden	-.23	.08	.15	.13	-.05	.10									
Sorge/Ängstlichkeit	.18	.18	.05	.09	.17	.11									
Altruismus	-.02	.45	-.21	-.01	.06	.24									
Selbstbestimmtes Handeln	.03	-.37	.08	.16	-.08	.17									
Berechnend/kalt	.14	-.32	.26	-.07	-.09	.19									
Benehmen/Anpassungsbereitschaft (-)	-.04	-.30	-.09	-.06	.12	.13									
Regelbewußtsein	-.04	.28	.01	.03	-.12	.10									
Perfektionismus	.06	.27	.12	-.26	-.13	.18									
Gemütlichkeit zeitl./mental	-.14	-.21	-.09	-.15	-.04	.11									
Geselligkeit (-)	-.07	-.09	.47	.07	-.01	.23									
Soziale Kompetenz (-)	.04	.00	.42	.25	.08	.26									
Besonnenheit (-)	.25	-.17	-.30	.12	.02	.20									
Privatheit (-)	-.01	.05	-.29	.01	.00	.09									
Prinzipientreue	.07	.22	.24	-.01	-.08	.14									
Verantwortungsbewusstsein/Eigensinnigkeit	-.16	-.09	-.23	.01	-.02	.10									
Kompetenz (-)	.06	.04	-.02	.47	.03	.23									
Leistungsstreben (-)	.01	-.16	.03	.43	-.11	.21									
Dominanz/Durchsetzungsfähigkeit	.04	-.17	-.04	-.36	.10	.17									
Freimütigkeit	.01	-.07	.19	.28	.00	.12									
Empfindsamkeit (-)	.01	-.05	.10	-.23	-.22	.12									
Offen für Ästhetik	-.08	.18	.02	-.22	.50	.31									
Konservatismus	.06	-.13	-.04	-.02	.29	.11									
Offenheit für Phantasie (-)	.02	.22	.13	.01	-.25	.15									
Pünktlichkeit/Verlässlichkeit	-.20	.14	.01	-.08	-.25	.14									
Robustheit	.00	-.03	.02	-.14	-.23	.08									
Intelligenz	.03	.03	-.10	.04	-.13	.02									
Interkorrelationen der Faktoren															
	V	.00				%-erkl.									
	E-	.10	.11			Varianz:									
	G-	.13	.02	.07		16.1									
	O	.03	-.06	-.16	.07										
Empir. Eigenwerte*:	2.35	2.06	1.82	1.63	1.37	1.25	1.14	1.06	1.03	1.01	.98	.95	.91	.88	.86
Zufalls-Eigenwerte*:	1.47	1.42	1.37	1.33	1.29	1.26	1.23	1.20	1.17	1.14	1.12	1.09	1.06	1.04	1.02

Anmerkungen. Ladungen mit Beträgen > .20 wurden fett gedruckt. Die Bezeichnungen der Faktoren zweiter Ordnung sind wie folgt: N = Neurotizismus; V = Verträglichkeit; E- = Extraversion (minus); G- = Gewissenhaftigkeit (minus); O = Offenheit für Erfahrungen.

Allerdings hängt die insgesamt geringe erklärte Varianz der Lösung damit zusammen, dass bereits die Primärfaktoren sehr gering interkorrelieren. In diesem Zusammenhang fällt auf, dass sechs Faktoren eine Kommunalität von .10 oder geringer haben, was bedeutet, dass die Lösung einige Primärfaktoren kaum repräsentiert.

6.3.2 Modellübergreifende Analysen für die Modelle von Zuckerman, Eysenck und Costa und McCrae

Für die Analysen anhand eines die Modelle von Zuckerman, Eysenck sowie Costa und McCrae übergreifenden Itempools sind die Datensätze von 620 Probanden aus Teilgruppe B verfügbar. Vor der gemeinsamen EFA wurde untersucht, in welchem Ausmaß der gemeinsame Itempool extrem ähnliche Items enthält. Da der EPQ-R und der NEO-PI-R bereits bei der vorherigen Analyse (siehe Kapitel 6.3.1) in Bezug auf ähnliche Items abgeglichen wurden, wurden hier nur der ZKPQ und die SSS-V miteinander und mit dem EPQ-R und dem NEO-PI-R verglichen. Es wurden nur vier extrem ähnliche Items von ZKPQ und NEO-PI-R festgestellt (siehe Tabelle A 6.20). Für die SSS-V ergab sich nur eine Überlappung mit drei Items des ZKPQ. Die geringe Anzahl ähnlicher Items kann vermutlich auch auf das unterschiedliche Item-Format des EPQ-R, NEO-PI-R, SSS-V und ZKPQ zurückgeführt werden. Dies gilt vor allem für die SSS-V, bei der – im Gegensatz zu allen übrigen hier eingesetzten Fragebogen – sehr handlungsbezogene Items im „forced-choice“-Format (d.h. mit zwei Aussagen, zwischen denen man sich entscheiden muss) dargeboten werden. Da die Anzahl sehr ähnlicher Items gering war, wurde auch hier vermieden, eine subjektive Itemauswahl durchzuführen. Der Itempool umfasst somit insgesamt alle Items des EPQ-R, NEO-PI-R, SSS-V und des ZKPQ, allerdings ohne die Lügen-Items.

6.3.2.1 Faktorenanalysen erster Ordnung

Zunächst wurde für die insgesamt 443 Variablen und 620 Personen eine Parallelanalyse anhand von 100 Zufallsläufen für die gewichtete und die ungewichtete Stichprobe durchgeführt (s. Abbildung 6.10). In der ungewichteten Stichprobe sind 21 empirische Eigenwerte größer als die entsprechenden Zufallseigenwerte (siehe Anhang, Tabelle A 6.23, unten) und in der gewichteten Stichprobe sind 30 empirische Eigenwerte größer als die entsprechenden Zufallseigenwerte (siehe Anhang, Tabelle A 6.27, unten).

Wie in den Analysen zu Teilgruppe A gehen bei einer Halbierung der Probandengruppe weniger Probanden (310) als Variablen (443) in die Analysen ein. Aus diesem Grund wurden auch

hier die Ergebnisse der Komparabilitäten nur der Vollständigkeit halber berichtet und nicht für die Extraktionsentscheidung herangezogen. Außerdem führte das ungünstige Verhältnis der Probandenanzahl zur Variablenanzahl dazu, dass die Komparabilitäten hier nicht auf der Basis von Hauptachsenanalysen sondern auf der Basis von Hauptkomponentenanalysen durchgeführt werden mussten (bei Hauptachsenanalysen traten nicht invertierbare Matrizen auf), was den Wert der Komparabilitäten weiter verringert. Wegen ihrer geringen Bedeutung für die Extraktionsentscheidung wurden die Komparabilitäten im Anhang aufgeführt (siehe Tabellen A 6.21, A 6.22). Da die Entscheidung über die Anzahl zu extrahierender Faktoren anhand der Parallelanalyse erfolgte, wurden für die ungewichtete Stichprobe 21 und für die gewichtete Stichprobe 30 Faktoren extrahiert.

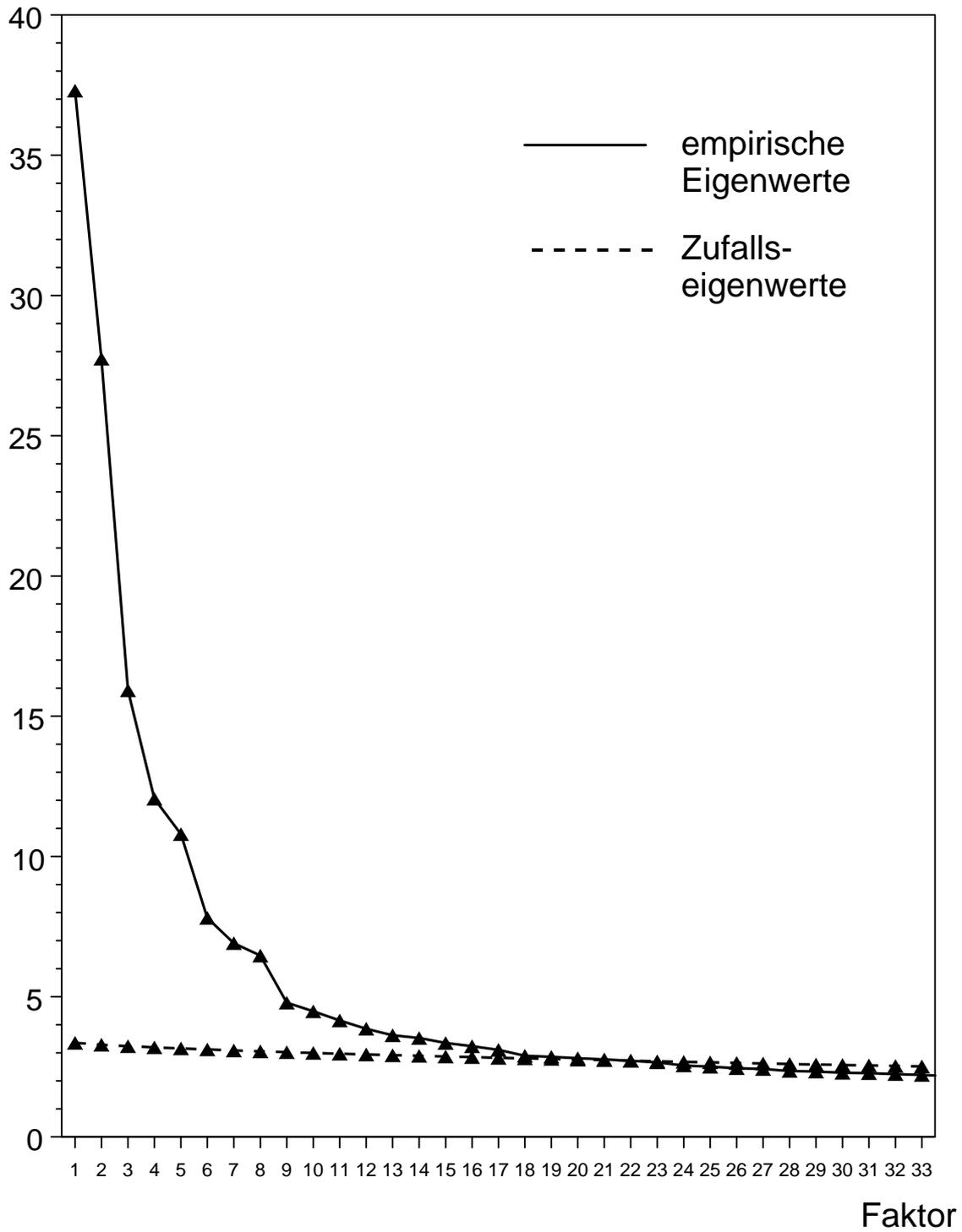
Faktoren erster Ordnung in der ungewichteten Stichprobe

Das komplette Oblimin-rotierte Ladungsmuster für die 21 Primärfaktoren der ungewichteten Stichprobe findet sich im Anhang (s. Tabelle A 6.23). Tabelle 6.37 enthält einen Kurz-Überblick über die Lösung anhand der drei Items mit den höchsten Ladungen. Bei der Interpretation der Lösung ist zu berücksichtigen, dass die Items in der Polung in die Analysen eingingen, in der sie auch in die intendierten Skalen der jeweiligen Fragebogen eingingen.

Obwohl einige Primärfaktoren in Tabelle 6.37 denen der Analysen mit dem gemeinsamen Itempool von 16 PF-R, EPQ-R und NEO-PI-R sehr ähnlich sind, bestehen auch erhebliche Abweichungen. Faktoren, die im anderen Itempool zumindest in einer der Analysen in ähnlicher Weise aufgetreten sind, sind „Neurotizismus“, „Offenheit für Ästhetik“, „Reizbarkeit/Aggressivität“ und „Vertrauen“. Faktoren, für die eine moderate Ähnlichkeit mit den Faktoren der vorherigen Analysen besteht, sind „Bescheidenheit (plus)/Kompetenz (minus)“, „Benehmen“, „Soziabilität“ und „Robustheit“. Die Faktoren „Aktivität“, „Gefahr- und Abenteuer-Suche“, „Gefühlsbetontheit“, „Planen (minus)“, „Lebhaftigkeit“, „Erlebnishunger und Geradlinigkeit“, „Selbstdisziplin“, „Bereitschaft Drogen zu nehmen“, „Spontaneität (minus), sich absichern“, „Bewegtes Leben/Langeweile meiden (minus)“, „Aufregende Erfahrungen suchen“, „Selbstkontrolle“ und „Offenheit für Ideen (minus)“ wurden in den vorherigen Analysen nicht festgestellt. Es erscheint naheliegend, dass gerade die Faktoren zum Reizsuche-Verhalten in dem Variablensatz ohne SSS-V und ZKPQ nicht nachgewiesen werden konnten. Entsprechend treten die Faktoren „Intelligenz“, „Empfindsamkeit“ und „Abgehobenheit“, die in den vorherigen Analysen primär durch Items des 16 PF-R markiert wurden, in der folgenden Analyse nicht auf.

a) Ungewichtet

Eigenwert



b) Gewichtet

Eigenwert

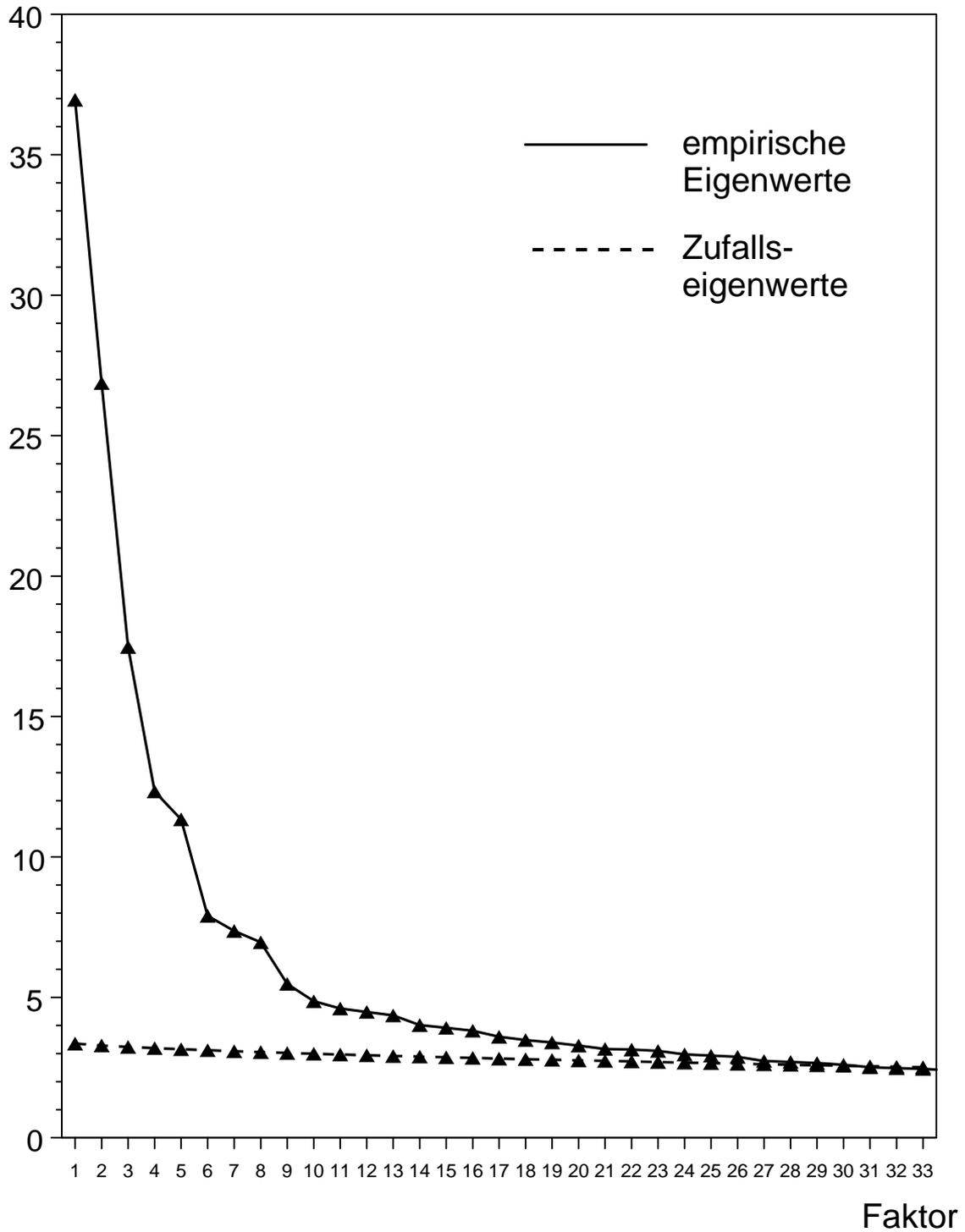


Abbildung 6.10. Parallelanalyse für den gemeinsamen Itempool von EPQ-R, NEO-PI-R, SSS-V und ZKPQ (gewichtete und ungewichtete Stichprobe; N=620)

Tabelle 6.37

Oblimin-Hauptladungsmuster für die drei Items mit den höchsten Ladungen auf den 21 Faktoren der ungewichteten Stichprobe (Itempool: EPQ-R, NEO-PI-R, SSS-V und ZKPQ)

Item-Kodierung/ Item in Stichworten / Ladungen	
1. Neurotizismus (-)	
N EPQ 062 ohne Grund matt und erschöpft	-.70
AN ZKPQ 90 oft ohne besonderen Grund unwohl	-.69
AN ZKPQ 25 oft ohne ersichtlichen Grund angesp.	-.67
2. Planen (-)	
C6 NEO 060 gründl. nachdenk. vor Entsch.	-.70
C6 NEO 120 vor Handeln Konsequenz. überd.	-.69
P EPQ 077 „erst Denken, dann handeln“	.69
3. Vertrauen / NEO	
A1 NEO 154 1. Reaktion, M. zu vertrauen	.57
A1 NEO 094 die meisten Menschen ehrlich	.46
A1 NEO 184 von anderen das beste anneh.	.46
4. Aktivität / ZKPQ	
AC ZKPQ 05 nur ungern herumsitzen und ausruhen	.59
AC ZKPQ 83 bin gerne ununterbrochen beschäftigt	.59
AC ZKPQ 44 nicht das Bedürfnis, mich ständig	.58
5. Soziabilität / ZKPQ	
E2 NEO 067 Dinge lieber allein tun	-.65
SO ZKPQ 87 bevorzuge es, Dinge alleine zu erled.	-.62
SO ZKPQ 63 irgendwo ohne soziale Kontakte leben	-.58
6. Gefahr- und Abenteuer-Suche / SSS-V	
SSSV 16TAS würde gern lernen, Wasserski zu l.	.68
SSSV 17TAS würde gern versuchen zu surfen	.67
SSSV 23TAS gern versuchen, Fallschirm zu spr.	.58
7. Bescheidenheit (+)/Kompetenz (-)	
A5 NEO 234 glaube, dass anderen überlegen	.62
A5 NEO 084 bin besser als die meisten M.	.56
A5 NEO 144 habe hohe Meinung von mir	.52
8. Gefühlsbetontheit	
E6 NEO 117 manch. sprudel vor Glück über	.46
O3 NEO 043 selten starke Empfindungen	.41
E6 NEO 027 nie vor Freude in die Luft	.39
9. Lebhaftigkeit	
E EPQ 074 Party in Schwung bringen?	.60
E EPQ 051 leicht Leben in Party bringen?	.56
E EPQ 010 Sind Sie ziemlich lebhaft?	.49
10. Erlebnishunger und Geradlinigkeit (?)	
SSSV 39BS mit dummen & langw. M. keine Ged.	.31
A2 NEO 129 schlimm für einen Heuchler	.26
O4 NEO 228 stets eine bewährte Route	.25
11. Selbstdisziplin / NEO	
C5 NEO 025 kann meine Zeit gut einteilen	-.63
C5 NEO 055 vertrödele Zeit, bevor Arbeit	-.57
C3 NEO 045 manchmal nicht so verlässlich	-.52
12. Reizbarkeit/Aggressivität	
N2 NEO 066 Aufbrausend/temperamentvoll	.59
AG ZKPQ 72 ich habe ein aufbrausendes Temper.	.57
AG ZKPQ 86 immer geduldig mit anderen	.52
13. Bereitschaft Drogen zu nehmen	
IM ZKPQ 95 Ich mag „wilde“,hemmungsl. Parties.	.45
SSSV 13DIS ich bin gern im Rauschzustand	.41
SSSV 10ES gern Drogen prob.,die Halluzinationen	.40
14. Spontaneität (-), sich absichern	
N1 NEO 121 selten beunruhigt über Zuk.	.33
P EPQ 048 die Ehe altmodische Sache?	-.29
P EPQ 071 um Zukunft kümmern,mit Spar.	-.29
15. Benehmen (-)	
P EPQ 040 gutes Ben. & Sauberk. wichtig?	.52
P EPQ 021 Sind gute Manieren wichtig?	.48
P EPQ 013 Abneigung gegen nicht benehm.	.41
16. Bewegtes Leben/Langeweile meiden (-)	
SSSV 33DIS Leben eines „Jet-Setters“ Spaß	-.37
SSSV 5BS langweilt mich, ständig dieselben Ges.	-.35
N EPQ 067 das Leben langweilig?	-.34
17. Offenheit für Ästhetik (-)	
O2 NEO 128 Poesie beeindruckt mich wenig	-.53
O2 NEO 068 Ballet/modern. Tanz langweilig	-.50
O2 NEO 188 wenn Lit./Kunstwerk begeistert.	-.49
18. Aufregende Erfahrungen suchen (-)	
IM ZKPQ 24 gerne neue und aufregende Erfahr.	-.36
SSSV 25DIS liebe neue & aufreg. Erf. auch wenn	-.36
A2 NEO 069 könnte niemanden betrügen	.34
19. Robustheit	
N4 NEO 106 nicht verl.,wenn andere spott.	.31
N EPQ 085 gekränkt,wenn Leute Sie krit.	.31
SO ZKPQ 82 sehr aktiver Teil einer Gruppe sein	.30
20. Selbstkontrolle	
N5 NEO 111 esse zu viel von Lieblingssp.	-.40
C3 NEO 105 schwindeln,wenn alleine spiel.	.34
E6 NEO 237 leicht zum Lachen zu bringen	-.25
21. Offenheit für Ideen (-) / NEO	
O5 NEO 083 Löse gern Probleme od. Aufg.	-.53
O5 NEO 113 verl. Int., wenn abstrakt	-.50
O5 NEO 023 Spaß mit Theorien/abst. Ideen	-.50

Anmerkungen. Die Item-Kodierungen enthalten die Test- und Skalenzugehörigkeit sowie die Nummern der Items in den jeweiligen Tests. Wenn Faktoren überwiegend durch Items einer Skala eines Tests markiert werden, wurde die Zugehörigkeit zu dem betreffenden Test hinter der Benennung des Faktors markiert. EPQ = EPQ-R; NEO = NEO-PI-R.

Die Interkorrelationen der Oblimin-rotierten Primärfaktoren sind auch in diesem Datensatz gering (siehe Anhang, Tabelle A 6.24). Nur sieben der insgesamt 210 Korrelationen (3.3%) haben einen Betrag über .20.

Eine Zusammenfassung der wichtigsten Markiertvariablen der Trasid-rotierten Lösung mit 21 Faktoren ist in Tabelle 6.38 dargestellt (Gesamtlösung, siehe Anhang, Tabelle A 6.25). Die meisten der 21 Trasid-rotierten Primärfaktoren entsprechen weitgehend denen der Oblimin-Lösung. Faktoren, die in der Oblimin-Lösung nicht auftreten, sind: „Empfänglichkeit für Langeweile“, „Fröhlich/Warmherzig“, „Soziale Einstellung“, „Freunde/Freizeit“, „Erlebnishunger“ und „Selbstsicherheit“. Dies sind überwiegend Faktoren, bei denen die Interpretation nur unter Vorbehalt erfolgen konnte. Insgesamt konnten in dieser Lösung vier Faktoren nur unter Vorbehalt interpretiert werden, diese sind in Tabelle 6.38 markiert. Die Trasid-rotierten Primärfaktoren weisen – wie die Oblimin-rotierten Faktoren – insgesamt nur niedrige Interkorrelationen auf (siehe Tabelle A 6.26): Nur acht der insgesamt 210 Korrelationen (3.8 %) haben einen Betrag über .20, was die Bedeutung von Faktoren höherer Ordnung relativiert.

Faktoren erster Ordnung in der gewichteten Stichprobe

Die komplette Ladungsmatrix der Oblimin-Lösung für die 30 Faktoren der gewichteten Stichprobe ist im Anhang dargestellt (s. Tabelle A 6.27). Ein grober Überblick über die Lösung kann auch anhand Tabelle 6.39 gewonnen werden, in der die drei Variablen mit den größten Hauptladungen für jeden der 30 Faktoren dargestellt sind.

Faktoren der 30 Faktoren-Lösung, die in der Lösung mit 21 Faktoren keine Entsprechung finden, sind: „Soziale Einstellung“, „Ärger offen zeigen“, „Allein sein können“, „Depressivität (minus)“, „soziale Anpasstheit“, „Dominanz“, „Sorge und Ängstlichkeit“, „Negative soziale Einstellung“, „Befangenheit“ und „Impulsivität/Anpassung“.

Wie in den Analysen mit dem anderen Variablensatz können hier die Faktoren „Neurotizismus“, „Offenheit für Ästhetik“, „Reizbarkeit/ Aggressivität“ und „Vertrauen“ nachgewiesen werden. Auch in der vorliegenden Analyse treten einige Faktoren zum Reizsuche-Verhalten auf, die in den Analysen mit dem Variablensatz ohne die SSS-V und den ZKPQ nicht nachgewiesen werden konnten („Gefahr- und Abenteuer-Suche“, „Aufregende Erfahrungen suchen (minus)“, „Bereitschaft Drogen zu nehmen“). Insgesamt besteht bei sechs der 30 Faktoren eine Unsicherheit bezüglich der Interpretation, was bei den Faktorenanalysen zweiter Ordnung berücksichtigt werden muss.

Tabelle 6.38

Trasid-Hauptladungsmuster für die drei Items mit den höchsten Ladungen auf den 21 Faktoren der ungewichteten Stichprobe (Itempool: EPQ-R, NEO-PI-R, SSS-V und ZKPQ)

Item-Kodierung/ Item in Stichworten, rechts daneben: Ladungen	
1. Neurotizismus	
N3 NEO 071 selten traurig oder deprim.	.75
N3 NEO 041 manchmal völlig wertlos	.73
N3 NEO 191 zieml. Duster & hoffnungslos	.72
2. Geselligkeit	
E2 NEO 037 gerne viele Leute um mich	.71
SO ZKPQ 78 Ich bin ein sehr geselliger Mensch.	.62
E2 NEO 217 mag Parties mit vielen Leuten	.63
3. Gefühlsbetontheit	
E6 NEO 117 manch. Sprudel vor Glück über	.54
O2 NEO 188 wenn Lit./Kunstwerk begeistert.	.56
O3 NEO 133 erl. Weit. Bereich von Gefühl.	.52
4. Selbstdisziplin/Gewissenhaftigkeit	
C5 NEO 025 kann meine Zeit gut einteilen	.63
C5 NEO 055 verträdelte Zeit, bevor Arbeit	.58
C1 NEO 215 arbeite zielstrebig & effektiv	.55
5. Gefahr und Abenteuersuche / SSS-V	
SSSV 17TAS würde gern versuchen zu surfen	.66
SSSV 16TAS würde gern lernen, Wasserski zu l.	.66
SSSV 23TAS gern versuchen, Fallschirm zu spr.	.60
6. Reizbarkeit/Aggressivität	
N2 NEO 066 Aufbrausend/temperamentvoll	.58
AG ZKPQ 72 ich habe ein aufbrausendes Temper.	.55
A4 NEO 079 zögere meinen Ärger auszudr.	-.51
7. Planen (-)	
C6 NEO 060 gründl. Nachdenk. Vor Entsch.	-.69
C6 NEO 120 vor Handeln Konsequenz. Überd.	-.68
P EPQ 077 „erst Denken, dann handeln“	.68
8. Vertrauen/Freundlichkeit	
A1 NEO 154 1. Reaktion, M. zu vertrauen	.56
E1 NEO 062 als herzl./freundl. Pers. Bek.	.43
A1 NEO 094 die meisten Menschen ehrlich	.48
9. Bescheidenheit (+) / Kompetenz (-)	
A5 NEO 234 glaube, dass anderen überlegen	-.64
A5 NEO 084 bin besser als die meisten M.	-.58
A5 NEO 144 habe hohe Meinung von mir	-.53
10. Aktivität / ZKPQ	
AC ZKPQ 83 bin gerne ununterbrochen beschäftigt	.57
AC ZKPQ 54 Meistens habe ich es eilig.	.55
AC ZKPQ 05 nur ungern herumsitzen und ausruhen	.54
11. Enthemmtheit / SSS-V	
IM ZKPQ 95 Ich mag „wilde“,hemmungsl. Parties.	.45
SSSV 1DIS Ich mag wilde und zügellose Parties.	.41
SSSV 9ES schon einmal Marihuana/Haschisch ger.	.36
12. Lebhaftigkeit	
E EPQ 074 Party in Schwung bringen?	.48
E EPQ 051 leicht Leben in Party bringen?	.46
E EPQ 010 Sind Sie ziemlich lebhaft?	.40
13. Empfänglichkeit für Langeweile (?)	
O4 NEO 138 vertraute Umgebung vorziehen	.33
P EPQ 026 Bewundern Sie, was Eltern tun	.30
C6 NEO 210 Reisen sorgfältig planen	-.27
14. Vertrauen (-)	
A1 NEO 124 mißtr., wenn jmd. Gefallen tut	-.42
A1 NEO 064 von meisten Leuten ausgen.	-.41
P EPQ 048 die Ehe altmodische Sache?	.33
15. Offenheit für Ideen / NEO	
O5 NEO 113 verl. Int., wenn abstrakt	.48
O5 NEO 083 Löse gern Probleme od. Aufg.	.46
O5 NEO 023 Spaß mit Theorien/abst. Ideen	.43
16. Benehmen (-)	
P EPQ 040 gutes Ben. & Sauberk. Wichtig?	.49
P EPQ 021 Sind gute Manieren wichtig?	.43
P EPQ 013 Abneigung gegen nicht benehm.	.39
17. Fröhlich/Warmherzig	
E6 NEO 147 nicht besonders fröhlich	.32
E1 NEO 092 halten mich für kühl/distanz.	.31
E6 NEO 237 leicht zum Lachen zu bringen	.24
18. Soziale Einstellung (?)	
A6 NEO 029 Polit. Mehr um menschl. Seite	-.29
N5 NEO 231 stets Gefühle unter Kontrolle	.26
A6 NEO 149 Menschl. Bed. vor wirt. Über.	-.26
19. Freunde/Freizeit	
SO ZKPQ 98 habe mehr Freunde als die meisten	.33
E EPQ 035 Haben Sie viele Freunde?	.30
AC ZKPQ 49 Arbeit, die mir ein Max. an Freizeit	-.28
20. Erlebnishunger (?)	
E EPQ 019 gerne neue Bekanntschaften?	.31
N5 NEO 111 esse zu viel von Liebesssp.	-.26
IM ZKPQ 45 viele Dinge nur wg. des Nervenkitzels	.24
21. Selbstsicherheit (?)	
SSSV 6ES ich erkunde gern eine fremde Stadt	.27
N EPQ 062 ohne Grund matt und erschöpft	-.26
IM ZKPQ 75 erkunde gerne fremde Stadt auf eigen.	.26

Anmerkungen. Die Item-Kodierungen enthalten die Test- und Skalenzugehörigkeit sowie die Nummern der Items in den jeweiligen Tests. Wenn Faktoren überwiegend durch Items einer Skala eines Tests markiert werden, wurde die Zugehörigkeit zu dem betreffenden Test hinter der Benennung des Faktors markiert. EPQ = EPQ-R; NEO = NEO-PI-R.

Auch die Interkorrelationen der 30 Oblimin-rotierten Primärfaktoren sind insgesamt eher gering (siehe Anhang, Tabelle A 6.28). Von den insgesamt 435 Korrelation hatten nur vier Korrelationen (0.9 %) einen Betrag größer als .20. Dieser Befund relativiert die Bedeutung von Faktoren höherer Ordnung.

Für die Trasid-rotierte Lösung mit 30 Faktoren ergaben sich – wie bei der Lösung mit 32 Faktoren im Variablenpool von EPQ-R, 16 PF-R und NEO-PI-R – Probleme mit der Varimax-Vorrotation. Bei der Varimax-Vorrotation der 30 Faktoren traten zwei Faktoren mit nur einer Hauptladung, vier Faktoren mit zwei Hauptladungen und sechs Faktoren mit drei Hauptladungen auf. Die große Anzahl von Faktoren mit wenigen Hauptladungen führte in der Trasid-rotierten Lösung dazu, dass vier Faktoren nicht interpretiert werden konnten und acht Faktoren nur unter Vorbehalt interpretiert werden konnten (siehe Tabelle A 6.30). Da somit bei 12 Faktoren Interpretationsprobleme auftraten, wird die Trasid-Lösung in den hierarchischen Analysen nicht weiterverfolgt. Die Interkorrelationen der 30 Trasid-rotierten Faktoren sind – wie die der Oblimin-rotierten Faktoren – insgesamt gering: Nur sieben der 435 Korrelationen (1.6 %) weisen Beträge über .20 auf.

Tabelle 6.39

Oblimin-Hauptladungsmuster für die drei Items mit den höchsten Ladungen auf den 30 Faktoren der ungewichteten Stichprobe (Itempool: EPQ-R, NEO-PI-R, SSS-V und ZKPQ)

Item-Kodierung/ Item in Stichworten, rechts daneben: Ladungen	
1. Neurotizismus (-)	
AN ZKPO 25 oft ohne ersichtlichen Grund angesp.	-.67
N EPO 062 ohne Grund matt und erschöpft	-.64
N1 NEO 091 oft angespannt und nervös	-.54
2. Planen (-)	
P EPO 077 „erst Denken, dann handeln“	.69
P EPO 002 überdenken, vor handeln?	.68
C6 NEO 060 gründl. nachdenk. vor	-.66
3. Soziale Einstellung	
A6 NEO 149 Menschl. Bed. vor wirtschaftl. Über.	.47
A2 NEO 189 Leuten zusetzen od. schmeicheln	.37
A6 NEO 029 Politik mehr um menschl. Seite	.31
4. Ärger offen zeigen (-)	
AG ZKPO 47 wenn zornig, nicht verbergen	-.61
A4 NEO 109 zeige offen wenn M. nicht mag	.59
AG ZKPO 42 wenn and. mich ärg. zögere ich nicht	-.58
5. Allein sein können	
SO ZKPO 63 irgendwo ohne soziale Kontakte leben	-.62
SO ZKPO 17 an einem Ort mehrere Tage allein	-.62
SO ZKPO 68 im allgemeinen bin ich gern alleine	-.59
6. Gefahr- und Abenteuersuche / SSS-V	
SSSV 23TAS gern versuchen, Fallschirm zu springen.	.64
SSSV 17TAS würde gern versuchen zu surfen	.62
SSSV 16TAS würde gern lernen, Wasserski zu laufen.	.58
7. Selbstdisziplin	
C5 NEO 055 verträdele Zeit, bevor Arbeit	.62
C5 NEO 025 kann meine Zeit gut einteilen	.62
C5 NEO 115 Schwier. zu tun was ich sollte	.54
8. Aktivität / ZKPO	
AC ZKPO 83 bin gerne ununterbrochen beschäftigt	.64
AC ZKPO 44 nicht das Bedürfnis, mich ständig	.60
AC ZKPO 05 nur ungern herumsitzen und ausruhen	.58
9. Offenheit für Ideen (-) / NEO	
O5 NEO 113 verl. Int., wenn abstrakt	-.62
O5 NEO 173 Int. an Natur des Universums	-.57
O5 NEO 023 Spaß mit Theorien/abstrakten Ideen	-.50
10. Depressivität (minus)	
N EPO 082 Fühlen Sie sich oft einsam?	-.46
N3 NEO 011 selten einsam oder traurig	-.45
N EPO 049 ab und zu Selbstmitleid?	-.39
11. Robustheit (-) (?)	
N6 NEO 116 In Notsituationen kühlen Kopf	.38
E3 NEO 132 and. erwart. von mir Entscheidungen	-.36
SSSV 15BS Erinnerungsfotos and. langweilen mich	-.28
12. Soziale Angewandtheit	
O6 NEO 208 wenn 25j. keine feste Meinung	.40
P EPO 057 Stört, wenn Fehler in eig. Arbeit	-.37
P EPO 065 Ihre Mutter eine gute Frau?	-.34
13. Dominanz (?)	
E4 NEO 077 arbeite langsam aber stetig	.46
E3 NEO 222 fällt schwer führende Rolle	.38
AC ZKPO 88 kann Gef. an Routinetätigk. Finden	.34
14. Geselligkeit	
SO ZKPO 92 mehr Zeit mit Freunden als ich sollte	.55
SO ZKPO 09 auf Party Unterhaltung beginnen	.51
IM ZKPO 45 viele Dinge nur wg. des Nervenkitzels	.43
15. Sorge und Ängstlichkeit (-)	
N1 NEO 151 beunruhigt über Dinge, die schiefe gehen	-.50
C3 NEO 105 schwindeln, wenn alleine spielen	.43
N1 NEO 121 selten beunruhigt über Zukunft	-.40
16. Erlebnishunger / Geradlinigkeit (?)	
O6 NEO 148 Treue zu eigenen Idealen wichtig.	.41
SSSV 5BS langweilt mich, ständig dieselben Gesicht	.34
SSSV 19ES mit ausgefall. Leuten Freundschaft	.34
17. Negative soziale Einstellung (?)	
A1 NEO 064 von meisten Leuten ausgen.	-.39
N2 NEO 006 ärg. Mich, wie and. mich behandeln	.36
A1 NEO 124 mißtr., wenn imd. Gefallen tut	-.31
18. Erlebnishunger / Enthemmtheit	
E2 NEO 217 mag Parties mit vielen Leuten	.48
SSSV 33DIS Leben eines „Jet-Setters“ Spaß	.44
IM ZKPO 95 Ich mag „wilde“, hemmungsl. Parties.	.39
19. Hedonistische Kontrolle	
N5 NEO 171 esse bis mir schlecht wird	.49
N5 NEO 111 esse zu viel von Lieblings Speisen	.44
SSSV 36DIS fühle mich nach einigen Gläs. Alkohol	.31
20. Erfahrungssuche	
SSSV 6ES ich erkunde gern eine fremde Stadt	.52
IM ZKPO 75 erkunde gerne fremde Stadt	.51
IM ZKPO 34 gern spont. verreisen, ohne die Reiser.	.35
21. Benehmen	
P EPO 040 gutes Benehmen & Sauberk. wichtig?	.59
P EPO 021 Sind gute Manieren wichtig?	.58
P EPO 013 Abneigung gegen nicht benehmen	.45
22. Offenheit für Ästhetik / NEO	
O2 NEO 128 Poesie beeindruckt mich wenig	.61
O2 NEO 098 begeistern Motive in Kunst	.56
O2 NEO 008 Kunst bedeutet mir nicht viel	.53
23. Befangenheit	
SSSV 34BS mag witzige Menschen, auch wenn sie	-.41
SSSV 12DIS gern in Gesell. unverklemmter Leute	-.36
P EPO 102 Glauben Sie, Streiten verbindet?	-.30
24. Gefühlsbetontheit (minus)	
E6 NEO 057 manchm. intens. Freude/ Ekstase	-.55
O3 NEO 133 erl. weit. Bereich von Gefühl.	-.48
E6 NEO 117 manchmal sprudel vor Glück über	-.47
25. Planen/ Absicherung (?)	
C2 NEO 010 Mögk. Offen statt alles planen	.35
C6 NEO 210 Reisen sorgfältig planen	.34
P EPO 020 gut. Versicherung abzuschliessen?	-.28
26. Impulsivität/ Anpassung (-) (?)	
E5 NEO 082 etwas wg. Nervenkitzel getan	-.38
IM ZKPO 84 von neuen Sachen und Ideen gebannt	-.36
SSSV 30DIS gefüllte Gläser garant. gelung. Fest	-.35
27. Reizbarkeit/ Aggressivität	
AG ZKPO 86 immer geduldig mit anderen	.52
N2 NEO 096 nicht reizbar/ leicht erregbar	.47
AG ZKPO 08 wenn verärg. sage ich häßl. Dinge	.47
28. Lebhaftigkeit	
E EPO 051 leicht Leben in Party bringen?	.69
E EPO 074 Party in Schwung bringen?	.68
E EPO 005 Sind sie sehr gesprächig?	.52
29. Bescheidenheit (+)/ Kompetenz (-)	
A5 NEO 234 glaube, dass anderen überlegen	.60
A5 NEO 084 bin besser als die meisten Menschen	.48
A5 NEO 144 habe hohe Meinung von mir	.47
30. Vertrauen (-)	
A1 NEO 154 1. Reaktion, Menschen zu vertrauen	-.55
A1 NEO 214 Vertrauen in menschliche Natur	-.47
E1 NEO 002 die meisten Mensch. wirkl.	-.46

Anmerkung. Die Item-Kodierungen enthalten die Test- und Skalenzugehörigkeit sowie die Nummern der Items in den jeweiligen Tests. Wenn Faktoren überwiegend durch Items einer Skala eines Tests markiert werden, wurde die Zugehörigkeit zu dem betreffenden Test hinter der Benennung des Faktors markiert. Unsichere Interpretationen bzw. Benennungen von Faktoren wurden mit „(?)“ gekennzeichnet. Die Interkorrelationen der Faktoren finden sich in Tabelle A 6.28. EPQ = EPQ-R; NEO = NEO-PI-R.

6.3.2.2 Faktorenanalysen zweiter Ordnung

Ungewichtete Stichprobe

Das Ergebnis der Faktorenanalyse zweiter Ordnung auf der Basis der 21 Oblimin-rotierten Faktoren der ungewichteten Stichprobe ist in Tabelle 6.40 dargestellt. Die Parallelanalyse für die Oblimin-rotierten Faktoren erster Ordnung deutete auf die Extraktion von fünf Faktoren hin (s. Tabelle 6.40). Mit 23.0% ist die Varianzaufklärung der Lösung gering, was auf die niedrigen Interkorrelationen zwischen den Oblimin-rotierten Faktoren erster Ordnung zurückgeführt werden kann.

Die Einfachstruktur der rotierten Lösung ist eher moderat. Einige Primärfaktoren weisen Ladungen auf mehreren Sekundärfaktoren auf. So lädt der Primärfaktor „Bescheidenheit (plus)/Kompetenz (minus)“, der Kompetenz und Bescheidenheit in konfundierter Form repräsentiert, gemäß seiner komplexen Zusammensetzung auf mehreren Sekundärfaktoren. Dennoch kann die Lösung relativ eindeutig interpretiert werden: Der erste Oblimin-rotierte Faktor zweiter Ordnung wird durch die Primärfaktoren „Offenheit für Ideen“, „Spontaneität“ und „Gefahr- und Abenteuersuche“ markiert. Wenn man die Nebenladungen der Primärfaktoren „Aufregende Erfahrungen suchen“ und „Offenheit für Ästhetik“ bei der Interpretation einbezieht, zeichnet sich eine Interpretation des Faktors i.S. von „Offenheit für Erfahrungen/Sensation Seeking“ ab. Der zweite Faktor wird vor allem durch die Primärfaktoren „Reizbarkeit/Aggressivität“, „Bereitschaft Drogen zu nehmen“ und „Vertrauen“ markiert. Somit laden einige der Primärfaktoren, die in einen Faktor zweiter Ordnung für „Verträglichkeit“ eingehen könnten, auf diesem Faktor. Andererseits können die Ladungen der Primärfaktoren „Bereitschaft Drogen zu nehmen“, „Aufregende Erfahrungen suchen (minus)“, „Bewegtes Leben/Langeweile meiden (minus)“ und „Planen (minus)“ nicht mit einer Interpretation i.S. von Verträglichkeit in Einklang gebracht werden.

Die Kombination sozial (negativer) Orientierungen mit bestimmten Aspekten des Reizsuche-Verhaltens sowie der geringen Tendenz zu Planen und geringer Selbstdisziplin ist dagegen gut mit Eysencks Psychotizismus-Konzept in Verbindung zu bringen. Daher wird dieser Faktor als „Psychotizismus“-Faktor interpretiert.

Der dritte Faktor zweiter Ordnung kann eindeutig als „Neurotizismus (minus)“-Faktor, der vierte Faktor als „Gewissenhaftigkeits (minus)“-Faktor und der fünfte Faktor als „Extraversion“-Faktor interpretiert werden. Die Lösung kann als eine Variante des Fünf-Faktoren-Modells der Persönlichkeit aufgefasst werden, wobei im vorliegenden Material statt eines Faktors für Verträglichkeit ein Psychotizismus-Faktor auftrat und der Faktor, der Offenheit für Erfahrungen repräsentiert in deutlichem Umfang Aspekte von Sensation Seeking umfasst. Die

Interkorrelationen der Sekundärfaktoren sind mit Ausnahme der Korrelation zwischen „Extraversion“ und „Offenheit für Erfahrungen/Sensation Seeking“ insgesamt so gering, dass eine Faktorenanalyse dritter Ordnung nicht interessant erscheint.

Tabelle 6.40

Oblimin-rotiertes Ladungsmuster der fünf Faktoren zweiter Ordnung auf Basis der 21 Primärfaktoren (ungewichtete Stichprobe)

Faktor	1	2	3	4	5	
	O/SS	P	N-	C-	E	h²
Offenheit für Ideen (-)	-.51	-.01	.04	.15	.05	.26
Spontaneität (-)	-.34	.01	-.04	.00	.01	.12
Gefahr- und Abenteuersuche	.32	.13	.09	.11	.15	.20
Erlebnishunger und Geradlinigkeit (?)	.17	-.07	.09	.16	-.07	.07
Reizbarkeit/Aggressivität	-.06	.44	-.10	-.13	.04	.22
Bereitschaft Drogen zu nehmen	.00	.39	-.02	.06	.01	.16
Vertrauen	.09	-.38	-.13	-.18	.22	.25
Aufregende Erfahrungen suchen (-)	-.30	-.32	.09	-.04	-.13	.26
Bewegtes Leben/Langeweile meiden (-)	-.21	-.32	.07	-.01	-.08	.18
Bescheidenheit (+)/Kompetenz (-)	-.21	-.31	-.20	.25	-.03	.24
Planen (-)	.08	.28	-.09	.19	.21	.21
Neurotizismus (-)	.10	-.17	.57	-.01	.27	.43
Robustheit	-.16	.05	-.39	-.02	-.01	.19
Selbstkontrolle	-.08	.01	.38	.10	-.05	.16
Offenheit für Ästhetik (-)	-.26	.29	.30	-.10	-.10	.25
Aktivität	.14	-.01	-.03	-.42	.03	.20
Benehmen (-)	.18	.06	.15	.39	.11	.23
Selbstdisziplin (-)	-.07	.28	-.22	.35	-.08	.28
Soziabilität (-)	.23	-.02	-.04	-.01	-.67	.41
Lebhaftigkeit	.14	.04	.15	-.21	.47	.34
Gefühlsbetontheit	.09	-.02	-.22	.14	.27	.17

Interkorrelationen der Faktoren						
P	.04					%-erkl.
N	.10	-.05				Varianz:
C-	.03	.08	-.02			23.0
E	.30	.10	-.10	-.04		

Empirische Eigenwerte*: 2.30 1.97 1.61 1.43 1.25 1.07 1.04 .98 .92 .83 .82 .79

Zufalls-Eigenwerte*: 1.35 1.29 1.24 1.20 1.17 1.14 1.10 1.08 1.04 1.02 .99 .96

Anmerkungen. Ladungen mit Beträgen > .20 wurden fett gedruckt. Die Bezeichnungen der Faktoren zweiter Ordnung sind wie folgt: O/SS = Offenheit für Erfahrungen/Sensation Seeking; P = Psychotizismus; N- = Neurotizismus (minus); G- = Gewissenhaftigkeit (minus); E = Extraversion.

Die Trasid-rotierten Faktoren zweiter Ordnung auf Basis der ebenfalls Trasid-rotierten Primärfaktoren sind in Tabelle 6.41 dargestellt. Die Parallelanalyse sprach auch hier für die

Extraktion von fünf Faktoren, obgleich der fünfte empirische Eigenwert mit 1.169 nur minimal über dem mittleren Eigenwert von 1.168 für die 100 zufallsbasierten Lösungen lag. Die Varianzaufklärung der sechsfaktoriellen Lösung ist mit 19.6% gering, was auf die geringen Interkorrelationen zwischen den Trasid-rotierten Primärfaktoren zurückzuführen ist.

Der erste Faktor ist vor allem durch positive Ladungen der Primärfaktoren „Planen (minus)“, „Reizbarkeit/Aggressivität“, „Neurotizismus“, „Enthemmtheit“ und negative Ladungen des Faktors „Selbstdisziplin/Gewissenhaftigkeit (minus)“ gekennzeichnet. Diese Kombination kann am ehesten als „Psychotizismus“-Faktor interpretiert werden. Vor allem die positive Ladung des Faktors „Planen (minus)“ und die negative Ladung des Primärfaktors „Selbstdisziplin/Gewissenhaftigkeit“ sind mit einer Interpretation dieses Faktors i.S. von Verträglichkeit (minus) oder Aggressivität inkompatibel. Der zweite Faktor verbindet Aspekte der Lebhaftigkeit und Geselligkeit mit Aktivität und Bescheidenheit. Dieser Faktor kann insgesamt als „Extraversions“-Faktor interpretiert werden. Der dritte Faktor wird im wesentlichen durch Primärfaktoren aus dem Bereich Sensation Seeking und durch den Faktor „Offenheit für Ideen“ markiert. Er kann somit als „Sensation Seeking/Offenheits“-Faktor interpretiert werden. Der vierte Faktor repräsentiert Aspekte der Verträglichkeit und auf dem fünften Faktor laden die Primärfaktoren „Neurotizismus“ und „Vertrauen (minus)“ positiv, während die Primärfaktoren „Lebhaftigkeit“, „Geselligkeit“ und „Soziale Einstellung“ negativ auf diesem Faktor laden. Dieser Faktor kann somit nur bedingt als „Neurotizismus“-Faktor interpretiert werden. Die positive Ladung von „Vertrauen (minus)“ und die negative Ladung des Primärfaktors „Soziale Einstellung“ weist auf eine Interpretation dieses Faktors i.S. von „Neurotizismus/Negative soziale Einstellung“ hin.

Im Rahmen der in dieser Arbeit dargestellten Analysen treten hier zum ersten mal substantielle Interkorrelationen zwischen den Sekundärfaktoren auf, die eine Faktorenanalyse dritter Ordnung interessant erscheinen lassen. Das Ergebnis einer entsprechenden Analyse führt zu einer Lösung mit zwei Faktoren (siehe Anhang, Tabelle A 6.31). Auf dem ersten Faktor dritter Ordnung lädt der Sekundärfaktor „Neurotizismus“ negativ, während „Sensation Seeking/Offenheit“ und „Psychotizismus“ positiv auf diesem Faktor laden. Dieser Faktor wurde vorläufig als „Robustheits“-Faktor interpretiert, während der zweite Faktor dritter Ordnung, auf dem „Extraversion“ und „Verträglichkeit“ laden, als Faktor für die „Soziale Ausrichtung“ interpretiert wird. Bei Digman (1997) bildeten Verträglichkeit, Gewissenhaftigkeit und Emotionale Stabilität einen Faktor dritter Ordnung, während Extraversion und Intellekt einen weiteren Faktor dritter Ordnung bildeten. Die hier ermittelten Faktoren dritter Ordnung weichen somit von denen ab, die Digman (1997) in seinen Reanalysen aufzeigen konnte. Die Interpretationen der Tertiärfaktoren sind jedoch als sehr vorläufig anzusehen, vor

allem weil aufgrund der geringen Interkorrelationen der Sekundärfaktoren in den bisherigen Analysen keine Faktoren dritter Ordnung sinnvoll berechnet werden konnten.

Tabelle 6.41

Trasid-rotiertes Ladungsmuster der fünf Faktoren zweiter Ordnung auf Basis der 21 Primärfaktoren (ungewichtete Stichprobe)

Faktor	1	2	3	4	5	
	P	E	SS/O	V	N/Neg	h²
Planen (-)	.55	-.04	-.03	.21	-.04	.34
Reizbarkeit/Aggressivität	.58	.20	-.19	-.04	.02	.28
Neurotizismus	.41	-.19	-.19	.00	.47	.34
Enthemmtheit	.32	.20	.19	-.04	-.02	.21
Selbstdisziplin/Gewissenhaftigkeit	-.37	.28	-.02	-.01	-.02	.25
Lebhaftigkeit	.09	.40	.04	.17	-.26	.34
Bescheidenheit (+)/Kompetenz (-)	.02	.46	.20	.00	-.05	.23
Aktivität	-.04	.42	-.01	.02	.04	.19
Geselligkeit	.19	.24	-.01	.29	-.29	.33
Gefahr- und Abenteuersuche	.03	.17	.47	.04	-.04	.24
Benehmen (-)	.03	-.14	.34	.02	.02	.16
Offenheit für Ideen	-.13	.26	.44	.05	.12	.17
Empfänglichkeit für Langeweile	-.07	-.03	.27	-.04	.00	.07
Erlebnishunger	.00	.13	.25	.01	-.01	.07
Nervosität (-)	-.10	-.01	.22	-.05	.05	.04
Vertrauen/Freundlichkeit	-.32	-.04	.00	.54	-.09	.39
Gefühlsbetontheit	.13	.05	.19	.40	.04	.20
Fröhlich/Warmherzig	.04	.04	-.14	.19	-.19	.09
Vertrauen (-)	.24	.24	.03	-.00	.31	.11
Soziale Einstellung	.03	-.13	-.13	-.10	-.23	.06
Freunde/Freizeit	-.04	-.03	-.05	.02	-.15	.02
Interkorrelationen der Faktoren						
E	-.15					%-erkl.
SS/O	.47	-.22				Varianz:
V	-.06	.20	-.13			19.6
N/Neg	-.39	-.13	-.37	-.02		

Empirische Eigenwerte*: 2.29 1.84 1.44 1.25 1.17 1.07 1.06 1.02 .98 .96 .87 .87

Zufalls-Eigenwerte*: 1.35 1.29 1.24 1.20 1.17 1.14 1.10 1.08 1.04 1.02 .99 .96

Anmerkungen. Ladungen mit Beträgen > .20 wurden fett gedruckt. Die Bezeichnungen der Faktoren zweiter Ordnung sind wie folgt: P = Psychotizismus; E = Extraversion; SS/O = Sensation Seeking/Offenheit; V = Verträglichkeit; N/Neg = Neurotizismus/Negative soziale Einstellung.

Gewichtete Stichprobe

Das Ergebnis der Faktorenanalyse zweiter Ordnung auf der Basis der 30 Oblimin-rotierten Primärfaktoren der gewichteten Stichprobe (siehe Tabelle 6.39) ist in Tabelle 6.42 dargestellt. Bei einer Parallelanalyse waren die Eigenwerte der ersten fünf empirischen Faktoren größer als der Mittelwert der Eigenwerte von Zufalls-Faktoren (siehe Tabelle 6.42). Aus diesem Grund

wurden fünf Faktoren zweiter Ordnung extrahiert, die insgesamt jedoch nur 17.7% der Varianz der Primärfaktoren erklären.

Der erste Faktor zweiter Ordnung wird durch positive Ladungen des Primärfaktors „Benehmen (minus)“ sowie von drei Primärfaktoren aus dem Bereich Sensation Seeking gekennzeichnet. Hinzu kommen negative Ladungen der Primärfaktoren „Planung/Absicherung“ sowie „Offenheit für Ideen (minus)“. Die Kombination von intensivem Reizsucheverhalten, geringer Planung/Absicherung und geringer Wertschätzung sozialer Umgangsformen (Benehmen) verbunden mit einer Offenheit für Ideen spricht für die Interpretation dieses Faktors i.S. von Eysencks „Psychotizismus“-Dimension. Der zweite Sekundärfaktor, der vor allem durch positive Ladungen der Primärfaktoren „Neurotizismus (minus)“, „Depressivität (minus)“ sowie „Sorge und Ängstlichkeit (minus)“ gekennzeichnet ist, kann eindeutig als „Neurotizismus (minus)“ interpretiert werden. Der dritte Sekundärfaktor, mit negativen Ladungen der Primärfaktoren „Offenheit für Ästhetik“, „Ärger offen zeigen (minus)“, „Befangenheit“ und „Bescheidenheit (plus)/Kompetenz (minus)“, sowie einer positiven Ladung von „Vertrauen (minus)“, ist weniger eindeutig interpretierbar.

Dieser Faktor, der durch geringe Befangenheit, verbunden mit der Tendenz, Ärger offen zu zeigen und Misstrauen bei gleichzeitig geringer Offenheit für Ästhetik gekennzeichnet ist, wird hier versuchsweise als Faktor für „Negative Grundeinstellungen“ interpretiert. Diese Interpretation muss als sehr vorläufig gelten, da ein derartiger Faktor im Kontext von Persönlichkeitstraits bisher nicht etabliert wurde. Der vierte Sekundärfaktor, der vor allem durch eine positive Ladung des Primärfaktors „Robustheit (minus)“ und durch negative Ladungen der Primärfaktoren „Aktivität“ und „Negative soziale Einstellung“ gekennzeichnet ist, wird hier versuchsweise als „Robustheit (minus)“ interpretiert. Die Unsicherheiten bei der Interpretation sind jedoch auch deswegen beachtlich, weil bereits sechs der Primärfaktoren nicht ganz eindeutig interpretiert werden konnten (siehe Tabelle 6.39 und im Anhang, Tabelle A 6.27). Der fünfte Sekundärfaktor kann eindeutig als „Extraversions“-Faktor interpretiert werden. Abschließend kann festgehalten werden, dass die Interkorrelationen der Sekundärfaktoren so gering sind, dass eine Faktorenanalyse dritter Ordnung hier nicht geboten erscheint.

Tabelle 6.42

Oblimin-rotiertes Ladungsmuster der fünf Faktoren zweiter Ordnung auf Basis

der 30 Primärfaktoren (gewichtete Stichprobe)

Faktor	1	2	3	4	5							
	P	N-	Neg	Rob-	E	h²						
			(?)	(?)								
Benehmen (-)	.44	-.01	.06	.09	-.08	.21						
Erlebnishunger / Geradlinigkeit (?)	.37	-.04	-.01	.10	-.02	.14						
Gefahr- und Abenteuersuche	.35	.00	.08	-.02	.05	.15						
Erfahrungssuche	.33	.05	-.01	.01	.05	.12						
Planen/Absicherung (?)	-.28	-.16	.03	.06	-.04	.12						
Offenheit für Ideen (-)	-.27	-.08	.08	.25	.00	.15						
Neurotizismus (-)	.07	.53	.10	.05	.09	.30						
Depressivität (-)	.04	.43	.14	-.06	.09	.23						
Sorge und Ängstlichkeit (-)	.06	.41	.13	.00	-.11	.19						
Selbstdisziplin	-.19	.36	-.04	-.15	.14	.21						
Hedonistische Kontrolle	-.09	-.32	.18	-.14	.04	.16						
Impulsivität/Anpassung (-) (?)	-.09	.31	-.19	.07	.03	.15						
Reizbarkeit/Aggressivität	-.10	-.30	.17	.08	.17	.16						
Negative soziale Einstellung (?)	.03	-.28	.00	-.26	.01	.13						
Planen (-)	.24	-.26	.06	.16	.15	.19						
Dominanz (?)	.11	.23	.14	-.12	.13	.14						
Offenheit für Ästhetik	.17	-.08	-.45	-.13	.11	.24						
Ärger offen zeigen (-)	-.02	.13	-.32	.18	-.21	.23						
Befangenheit	-.21	-.05	-.32	.01	-.01	.17						
Vertrauen (-)	.01	-.09	.31	.21	-.26	.23						
Bescheidenheit (+)/Kompetenz (-)	-.07	-.04	-.30	.19	-.03	.15						
Soziale Einstellung	-.12	.05	-.26	-.19	-.07	.14						
Soziale Angepasstheit	-.14	-.07	-.19	.06	-.05	.08						
Robustheit (-) (?)	.06	.04	.03	.47	.12	.22						
Aktivität	-.11	.06	.04	-.27	.13	.11						
Lebhaftigkeit	-.05	.17	.06	-.04	.51	.30						
Allein sein können	.06	-.01	.06	-.22	-.50	.25						
Geselligkeit	.14	-.04	-.01	.01	.35	.16						
Gefühlsbetontheit (-)	-.23	.09	.21	.12	-.26	.20						
Erlebnishunger / Enthemmtheit	.07	-.07	.08	-.11	.20	.09						
Interkorrelationen der Faktoren												
N-	.03					%-erkl.						
Neg (?)	.16	-.06				Varianz:						
Rob- (?)	-.04	-.08	.00			17.7						
E	.22	.01	.09	-.19								
Empirische Eigenwerte*:	2.44	2.19	1.78	1.55	1.40	1.15	1.11	1.06	1.00	.98	.95	.93
Zufalls-Eigenwerte*:	1.43	1.37	1.33	1.29	1.26	1.23	1.19	1.17	1.14	1.12	1.09	1.07

Anmerkungen. Ladungen mit Beträgen > .20 wurden fett gedruckt. Die Bezeichnungen der Faktoren zweiter Ordnung sind wie folgt: P = Psychotizismus; N- = Neurotizismus (minus); Neg (?) = Negative Grundeinstellungen; Rob- (?) = Robustheit (minus); E = Extraversion. Als unsicher eingestufte Interpretationen sind mit „(?)“ gekennzeichnet.

6.3.3 Diskussion der Ergebnisse der modellübergreifenden Analysen

Im folgenden werden die Ergebnisse der modellübergreifenden Analysen spezifisch für die beiden Variablengruppen (Eysenck, Cattell sowie Costa und McCrae bzw. Eysenck, Zuckerman sowie Costa und McCrae) diskutiert. Die über beide Variablengruppen generalisierbaren Ergebnisse werden in Kapitel 7 abschließend diskutiert.

Überlappungen der Modelle von Eysenck, Cattell sowie Costa und McCrae

Zunächst fällt auf, dass die Faktorenlösungen erster Ordnung mit keinem der drei untersuchten Modelle vollständig in Einklang zu bringen sind. Dies deutet darauf hin, dass die untersuchten Inventare (EPQ-R, 16 PF-R und NEO-PI-R) Varianzquellen erfassen, die wechselseitig in den jeweils anderen Inventaren nur partiell enthalten sind. Natürlich hängt die Abweichung der Primärfaktoren-Lösungen von den Primärfaktoren, wie sie sich gemäß der eingesetzten Inventare ergeben sollten, auch damit zusammen, dass bereits in den getrennten exploratorischen Analysen die intendierten Primärfaktoren für den 16 PF-R und den NEO-PI-R nur sehr partiell nachgewiesen werden konnten (siehe Kapitel 6.2).

Modellspezifische Faktoren

In Kapitel 4.2 wurde die Unterscheidung zwischen modellübergreifenden und modellspezifischen Faktoren getroffen. Einige modellspezifische Faktoren lassen sich in den modellübergreifenden Analysen aufzeigen: So tritt in allen Analysen ein Primärfaktor „Intelligenz“ auf, der fast ausschließlich von den entsprechenden Items des 16 PF-R markiert wird. Allerdings muss eingeräumt werden, dass Intelligenz im Gegensatz zu den anderen Faktoren des 16 PF-R anhand von Leistungsvariablen erfasst wird, so dass vermutlich eine gemeinsame Methodenvarianz vorliegt und diesem Faktor im Rahmen der Analysen somit eine Sonderrolle zukommt. Auch der mit dem 16 PF-R erfasste Primärfaktor „Empfindsamkeit“ tritt als modellspezifischer Faktor in den modellübergreifenden Analysen auf. Allerdings kann auch für diesen Faktor, der im wesentlichen Interessen erfasst, eine gemeinsame Methodenvarianz vermutet werden, wobei der Übergang zwischen Interessen, die eher dem Bereich der Motivation zuzuordnen sind, zu Neigungen, die im klassischen Sinne traitrelevant sind, vermutlich fließend ist. Ein modellspezifischer Faktor des 16 PF-R, der ebenfalls in allen Analysen auftritt und der keine methodenspezifische Varianz binden dürfte, ist der Faktor „Privatheit“. Dieser Faktor stellt somit einen modellspezifischen Faktor dar, der in den übrigen Inventaren nicht repräsentiert wird. Ein Faktor des NEO-PI-R, der nur in der Trasid-rotierten Lösung für die gewichtete Stichprobe nicht ganz rein auftritt, ist der Faktor „Offenheit für Ästhetik“. Da die Trasid-Lösung für die 32 Faktoren der ungewichteten Stichprobe viele kaum interpretierbare Faktoren aufweist und daher in den hierarchischen Analysen nicht weiter verfolgt wurde, wird der Umstand, dass der Faktor „Offenheit für Ästhetik“ in der Trasid-Lösung nicht aufgezeigt werden konnte, als vernachlässigbar angesehen. Der Faktor „Offenheit für Ästhetik“

repräsentiert somit einen modellspezifischen Faktor des NEO-PI-R. Auch ein Faktor „Kompetenz“, der mitunter negative Ladungen vieler Items der Facetten-Skala „Bescheidenheit“ aufweist, tritt in allen Lösungen mit Ausnahme der Trasid-rotierten Lösung für die gewichtete Stichprobe auf. Dieser Faktor wird ausschließlich durch Items des NEO-PI-R konstituiert. Er entspricht zwar nicht der intendierten Struktur, da im NEO-PI-R Kompetenz und Bescheidenheit zwei verschiedene Facetten-Skalen darstellen, aber er repräsentiert offenbar Varianzquellen, die nicht mit dem 16 PF-R und dem EPQ-R erfasst werden. Es treten keine modellspezifischen Faktoren, d.h. nur durch Items des EPQ-R markierte Faktoren, auf. Dies hängt vermutlich damit zusammen, dass die drei mit dem EPQ-R erfassten Faktoren Neurotizismus, Extraversion und Psychotizismus einen so hohen Generalitätsgrad aufweisen, dass sie stets auch von vielen Items anderer Inventare markiert werden.

Modellübergreifende Faktoren

Die Primärfaktoren „Geselligkeit“, „Vertrauen“ und „Reizbarkeit, Ungeduld“ treten mit minimalen Variationen in allen Analysen auf, jedoch werden diese Faktoren nicht von den Items eines einzelnen Traitinventars markiert, sondern durch Items mehrerer Traitinventare und stellen insofern modellübergreifende Primärfaktoren dar. Auch wenn der Faktor „Geselligkeit“ wie im NEO-PI-R bezeichnet wurde, kann er jedoch nicht in der Weise interpretiert werden, dass der NEO-PI-R hier Aspekte anderer Modelle integriert, da im 16 PF-R die Skala „soziale Kompetenz“ ähnlich gelagert ist. Auch für den modellübergreifenden Faktor „Vertrauen“, der ebenfalls eine Bezeichnung trägt, die im NEO-PI-R verwendet wird, findet sich in der Skala „Wachsamkeit“ des 16 PF-R eine Parallele. Aspekte des Faktors „Reizbarkeit, Ungeduld“ treten sowohl in der Facetten-Skala Reizbarkeit als auch in der Skala Anspannung des 16 PF-R auf. Insofern sind diese modellübergreifenden Faktoren bereits in mehreren Modellen explizit repräsentiert, so dass diese Faktoren nicht für eine überlegene Strukturierung eines der Modelle sprechen. Die weitere Untersuchung der modellübergreifenden Faktoren erfordert die Einbeziehung der Analysen zweiter Ordnung.

Bei einem Vergleich der Oblmin- und Trasid-rotierten Faktorenlösungen zweiter Ordnung für die ungewichtete Stichprobe fällt auf, dass beide Lösungen Faktoren für Neurotizismus, Extraversion, Gewissenhaftigkeit (minus) und Verträglichkeit (minus)/Psychotizismus aufweisen. Diese Faktoren sind insofern modellübergreifend als sie von Primärfaktoren, die Variablen der drei eingesetzten Inventare repräsentieren, markiert werden.

Es bleibt festzuhalten, dass alle modellspezifischen Primärfaktoren in die modellübergreifenden Faktoren eingehen. Selbst die Faktoren „Intelligenz“ und „Empfindsamkeit“, die keine typischen Traitfaktoren darstellen und die in nennenswertem Umfang spezifische Methodenvarianz repräsentieren können, weisen mitunter nennenswerte Ladungen auf den

Faktoren zweiter Ordnung auf. Dies kann als Hinweis auf die Angemessenheit der Einbeziehung derartiger Faktoren in Traitsysteme gelten.

In der Oblimin-Lösung zweiter Ordnung treten zwei sehr gering korrelierte Faktoren auf, die Aspekte des Faktors Offenheit für Erfahrungen beinhalten (siehe Tabelle 6.34). Diese beiden Faktoren konnten in der Trasid-Lösung jedoch nicht nachgewiesen werden. Dafür tritt in der Trasid-Lösung ein Faktor „Offenheit für Ästhetik“ auf, der Ähnlichkeiten mit dem Oblimin-Faktor „unkontrollierte Annäherung“ aufweist. Neben dem Faktor „Verträglichkeit (minus)/Psychotizismus“ tritt in der Trasid-Lösung ein Faktor „Bescheidenheit“ auf, der Aspekte der Verträglichkeit repräsentiert, die nicht in den Psychotizismus-nahen „Verträglichkeit (minus)“-Faktor eingehen (siehe Tabelle 6.35).

Die drei Faktoren „Extraversion“, „Neurotizismus“ und „Verträglichkeit (minus)/Psychotizismus“ können als Hinweis auf eine modellübergreifende Bedeutung von Eysencks Persönlichkeitsfaktoren angesehen werden, wobei eine stärkere Akzentuierung des Psychotizismus-Faktors in Bezug auf Aspekte der Verträglichkeit (minus) beachtet werden sollte. Ebenso ergibt sich jedoch aus den Analysen, dass „Gewissenhaftigkeit (minus)“ nicht unter dem Faktor „Verträglichkeit (minus)/Psychotizismus“ subsumiert werden kann. Dieser Faktor ist auch im Einklang mit den Analysen zum 16 PF-R (Schneewind & Graf, 1998), in denen die mit Gewissenhaftigkeit (minus) hoch korrelierte „Selbstkontrolle“ einen Faktor zweiter Ordnung darstellt. Die Tatsache, dass der Faktor „Gewissenhaftigkeit (minus)“ nicht mit dem „Faktor Verträglichkeit (minus)/Psychotizismus“ zu einem Faktor höherer Ordnung verschmelzen kann, ergibt sich auch aus den geringen Korrelationen dieser beiden Faktoren zweiter Ordnung (.08 bei Oblimin-Rotation; .17 bei Trasid-Rotation). Dass der Faktor für „Gewissenhaftigkeit (minus)“ hier nicht unter dem Faktor „Verträglichkeit (minus)/Psychotizismus“ subsumiert werden kann, ist im Einklang mit dem Fünf-Faktoren-Modell und mit Cattells Modell der Persönlichkeit und deutet auf eine Unvollständigkeit in Eysencks Persönlichkeitssystem. Andererseits tritt in diesen Analysen kein Faktor „Offenheit für Erfahrungen“ auf, wie er nach dem Fünf-Faktoren-Modell zu erwarten wäre. Wohl aber traten Faktoren auf, die mit „Offenheit für Erfahrungen“ inhaltlich zusammenhängen. In der Trasid-Lösung für die ungewichtete Stichprobe tritt der Faktor „Offenheit für Ästhetik“ als Faktor zweiter Ordnung auf, in der Oblimin-Lösung treten zwei kaum zusammenhängende Faktoren auf, die verschiedene Aspekte der Offenheit repräsentieren. Beide Ergebnisse sprechen dafür, dass der Bereich „Offenheit für Erfahrungen“, in der ungewichteten Stichprobe, so heterogen strukturiert ist, dass er nicht durch einen einheitlichen Faktor abgebildet werden kann. Andererseits wird an den vorliegenden Daten auch deutlich, dass im Bereich von „Offenheit für Erfahrungen“ relevante Varianzquellen sind, die mit dem dreifaktoriellen Modell von Eysenck nicht angemessen beschrieben werden können.

Aus den Analysen zweiter Ordnung für die gewichtete Stichprobe ergibt sich für die Oblimin-rotierte Lösung eine Variante des Fünf-Faktoren-Modells (siehe Tabelle 6.36), wobei der Faktor „Verträglichkeit“ auch Aspekte von Psychotizismus beinhaltet. Bei der Varimax-Vorrotation, die als Grundlage für die Trasid-Rotation verwendet wurde, hatten sieben Faktoren nur ein bis drei Hauptladungen, was eine zu geringe Sättigung der Faktoren darstellt (siehe Velicer & Fava, 1998). Da demzufolge in der Trasid-Lösung acht Faktoren nur unter Vorbehalt interpretiert werden konnten, wurde die Trasid-Lösung in der Analyse zweiter Ordnung nicht weiterverfolgt.

Auch die Replikation des Fünf-Faktorenmodells auf der Ebene der Faktoren zweiter Ordnung deutet darauf hin, dass dieses Modell – zumindest im Bereich „Offenheit für Erfahrungen“ – Varianz repräsentiert, die weder in Cattells noch in Eysencks Modell umfassend repräsentiert ist.

Im übrigen muss festgehalten werden, dass sich auf der Ebene der Primärfaktoren teilweise völlig neue Strukturierungen ergaben, während in den Faktoren zweiter Ordnung im wesentlichen bekannte Strukturierungen auftraten. Die Faktoren „Extraversion“, „Neurotizismus“ und „Gewissenhaftigkeit“ konnten in allen Analysen zweiter Ordnung anhand des vorliegenden Variablensatzes bestätigt werden. Für den Bereich „Offenheit für Erfahrungen“ ergaben sich teilweise unterschiedliche Strukturierungen. Der Bereich „Verträglichkeit“ bzw. „Psychotizismus“ war in allen Analysen mit unterschiedlichen Akzentuierungen vertreten. Insofern sind die hier aufgezeigten Faktoren zweiter Ordnung denen der drei untersuchten Modelle ähnlich, wobei die Struktur des Fünf-Faktoren-Modells dominant erscheint, allerdings mit der Einschränkung, dass der Faktor für „Verträglichkeit“ in deutlichem Umfang auch Aspekte von Psychotizismus einschließt (z.B. Unstrukturiertheit und Regelbewusstsein), die mit den bisherigen Formulierungen dieses Faktors im Rahmen des Fünf-Faktoren-Modells nur bedingt kompatibel sind.

Überlappungen der Modelle von Eysenck, Costa und McCrae sowie Zuckerman

Modellspezifische Faktoren

Der Primärfaktor „Gefahr- und Abenteuersuche“ ist spezifisch für Zuckermans Modell und der Primärfaktor „Offenheit für Ideen“ spezifisch für das Fünf-Faktorenmodell. Die übrigen Primärfaktoren konnten weniger eindeutig einem einzigen Modell zugewiesen werden, auch wenn der Faktor „Aktivität“ sehr stark von den entsprechenden Items des ZKPQ dominiert wird. Somit ist die Anzahl der modellspezifischen Primärfaktoren in diesen Analysen eher gering.

Modellübergreifende Faktoren

In allen Analysen erster Ordnung in diesem Variablensatz treten Faktoren die „Neurotizismus“, „Planen“, „Benehmen“, „Geselligkeit/Soziabilität“, „Bescheidenheit/Kompetenz (minus)“, „Gefühlsbetontheit“ und „Reizbarkeit/Aggressivität“ auf. Diese Primärfaktoren werden durch die Items mehrerer Fragebogen repräsentiert, obwohl sie nicht die Struktur eines der Fragebogen widerspiegeln. So konstituierten beispielsweise nur die SSS-V Items für Gefahr- und Abenteuersuche einen modellspezifischen Faktor (s.o.), während sich die Items der übrigen drei Subskalen der SSS-V jeweils anderen Faktoren zuordneten. Genauso entsprach nur der Primärfaktor „Offenheit für Ideen“ direkt einer der 30 Facetten-Skalen des NEO-PI-R.

In allen Analysen zweiter Ordnung ergab sich eine fünffaktorielle Struktur, die allerdings nicht der des Fünf-Faktoren-Modells entsprach. Neben den modellübergreifenden Faktoren „Extraversion“ und „Neurotizismus“ konnte vor allem ein modellübergreifender „Psychotizismus“-Faktor in allen Analysen zweiter Ordnung nachgewiesen werden (siehe Tabellen 6.40, 6.41, 6.42). Der „Psychotizismus“-Faktor, wie er hier auftrat, integrierte auch Varianz von „Verträglichkeit (minus)“ und „Gewissenhaftigkeit (minus)“. In den Analysen mit der ungewichteten Stichprobe trat ein Faktor „Offenheit für Erfahrungen/Sensation Seeking“ auf, während die übrigen hier nicht genannten Sekundärfaktoren stets nur in einer Analyse auftraten.

Auch bei den Analysen der Gruppe B treten teilweise Primärfaktoren auf, die in dieser Form bisher nicht in Traitmodellen aufgezeigt werden konnten. Allerdings bedürfen die Befunde auf dieser Ebene zunächst der Replikation, vor allem da ein beachtlicher Einfluss der Stichprobe und Rotationsmethode auf die Ergebnisse festgestellt wurde. Eine neue Strukturierung auf der Ebene der Faktoren zweiter Ordnung ergibt sich aus der Kombination von „Sensation Seeking“ mit „Offenheit für Erfahrungen“, sowie aus den Faktoren „Negative Grundeinstellungen“ und „Robustheit (minus)“. Diese Faktoren bedürfen vor einer weitergehenden Interpretation in besonderem Maße der Replikation. Insgesamt dominieren jedoch im Bereich der Faktoren zweiter Ordnung die bereits bekannten Faktoren.

6.4 Anwendung der Inklusions-Indizes zur Ermittlung der Generalität

Da die Traitdimensionen im Rahmen dieser Arbeit anhand faktorenanalytischer Analysetechniken untersucht wurden, kommen hier auch lediglich die Inklusions-Indizes zum Einsatz, die sich auf Faktorwerte anwenden lassen. Von den in Kapitel 4.3.2 dargestellten Inklusionsindizes für Faktorwerte sollen hier die in Formel 10, 13, 15 und 16 dargestellten Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} für Faktoren auf der Basis von Korrelationen von Items und Faktorwerten zur

Anwendung kommen. Die Indizes I_{f2} und I_{f3} haben den Vorteil, dass sie die Differenz der wechselseitig durch die zu vergleichenden Faktoren erklärten Varianzen widerspiegeln. Die Indizes I_{f2} und I_{f3} unterscheiden sich voneinander dadurch, dass in I_{f2} die Faktoren auch wechselseitig Varianz an den Items erklären können, die die jeweils zugehörigen Faktoren nicht an den Items erklären. In I_{f3} dürfen jeweils nur die Varianzen an den Items erklärt werden, die auch von den zugehörigen Faktoren erklärt wird. Insofern erlaubt der Vergleich der Ergebnisse zu beiden Indizes die Beurteilung der Relevanz dieses Unterschiedes. Die Indizes I_{f4} und I_{f5} sollen zusätzlich die Verteilungen der erklärten Varianzen auf die Itemgruppen als Aspekt der Generalität erfassen, weshalb diese Indizes nicht im Sinne wechselseitig erklärter Varianz interpretiert werden können. Index I_{f4} entspricht dem Index I_{f2} , nur dass die Verteilungen der Korrelationen berücksichtigt werden (Formel 15, siehe Kapitel 4.3.2) und I_{f5} entspricht I_{f3} mit zusätzlicher Berücksichtigung der Verteilungen der Korrelationen (Formel 16). Vor der Darstellung der inhaltlichen Ergebnisse zum Generalitätsvergleich werden zunächst die Ergebnisse zur Validität der Inklusions-Indizes anhand des Vergleichs zwischen Faktoren erster und zweiter Ordnung einer hierarchischen Faktorenanalyse präsentiert.

6.4.1 Zur Validität der Generalitäts-Indizes anhand des Vergleichs zwischen Faktoren erster und zweiter Ordnung

Die Grundidee bei den hier dargestellten Analysen ist, dass in hierarchischen Analysen die Faktoren zweiter Ordnung in der Regel eine höhere Generalität aufweisen sollten, als die Faktoren erster Ordnung, die die Faktoren zweiter Ordnung markieren. Wenn sich bei einem Vergleich der Generalitäten anhand der Inklusions-Indizes die zu erwartenden Generalitätsunterschiede nachweisen lassen, wäre dies ein Hinweis auf die Angemessenheit der Inklusions-Indizes.

Es wird im Rahmen dieser Arbeit zwar nicht davon ausgegangen, dass Sekundärfaktoren notwendiger Weise eine höhere Generalität aufweisen sollten als Primärfaktoren (siehe Kapitel 4.2). Allerdings handelt es sich hier um Primärfaktoren, die mit nennenswerten Varianzanteilen in die Sekundärfaktoren eingehen. Auch für derartige Primärfaktoren kann nur unter der Bedingung einer ähnlichen Generalität davon ausgegangen werden, dass sie eine geringere Generalität aufweisen als die Sekundärfaktoren. Für die Annahme ähnlicher Generalität der Primärfaktoren spricht jedoch, dass sie sich aus sehr ähnlichem Itemmaterial zusammensetzen. Natürlich hätte man auf der Basis der Inklusions-Indizes die Ähnlichkeit der Generalitäten der Primärfaktoren untersuchen können. Diese Prüfung wurde hier jedoch aus folgendem Grund nicht durchgeführt: Eine derartige Prüfung wäre zirkulär, da ja die Prüfung der Ähnlichkeit der

Primärfaktoren bereits die Validität der Indizes voraussetzen würde. Dabei sollen aus dem Vergleich der Inklusions-Indizes für die Faktoren erster und zweiter Ordnung erste Hinweise auf die Validität der Inklusions-Indizes gewonnen werden. Es erscheint nicht sinnvoll, das Ergebnis der Prüfung bei der Prüfung vorauszusetzen. Insofern muss die – aufgrund des Itemmaterials nicht unbegründete – Voraussetzung der Ähnlichkeit der Generalitäten der Primärfaktoren hier getroffen werden. Ein Hinweis auf die Angemessenheit dieser Voraussetzung ergibt sich aus der faktoriellen Komplexität der Primärfaktoren in den Faktorenlösungen zweiter Ordnung. Die vergleichsweise eindeutige Ladung aller Primärfaktoren auf nur einem Faktor kann als Hinweis auf die Ähnlichkeit der Generalität der Primärfaktoren gewertet werden. Da die Voraussetzung der Ähnlichkeit der Generalitäten der Primärfaktoren hier nicht eingehender geprüft werden kann, muss dies bei der Interpretation der Ergebnisse einschränkend berücksichtigt werden.

Die Generalitätsvergleiche wurden hier für Faktoren erster und zweiter Ordnung aus den Bereichen Neurotizismus und Extraversion vorgenommen. Da den Ergebnissen für die gewichtete Stichprobe wegen der vermutlich höheren Repräsentativität eine größere Bedeutung zukommt, wurden diese Analysen aus Aufwandsgründen nur für die gewichtete Stichprobe durchgeführt.

Da die Berechnung der Inklusions-Indizes Faktorwerte erfordert, mussten die Faktorenanalysen zweiter Ordnung für die in Kapitel 6.3 dargestellten Faktoren zweiter Ordnung in einer Weise durchgeführt werden, dass Faktorwerte für die Faktoren zweiter Ordnung geschätzt werden konnten. Dazu waren hier gesonderte Berechnungen erforderlich, da die Analysen zweiter Ordnung, wie sie in Kapitel 6.3 durchgeführt wurden, direkt anhand der Korrelationen der Faktoren erster Ordnung durchgeführt wurden. Dies ist zwar das konventionelle Vorgehen bei einer aufsteigenden Faktorenanalyse, führt allerdings im Falle einer Hauptachsenanalyse dazu, dass die Faktorwerte für die Faktoren zweiter Ordnung nicht direkt geschätzt werden können. Dies hängt damit zusammen, dass bereits die Interkorrelationen der geschätzten Faktorwerte erster Ordnung bei einer Hauptachsenanalyse nicht direkt den Korrelationen zwischen den Faktoren erster Ordnung entsprechen (s. Gorsuch, 1983). Insofern ergeben sich bei einer Hauptachsenanalyse kleine Verschiebungen wenn die Faktoren zweiter Ordnung nicht auf der Basis der Interkorrelationen der Faktoren erster Ordnung, sondern auf der Basis der Interkorrelationen der Faktorwerte der Faktoren erster Ordnung ermittelt werden. Die Faktorwerte wurden hier nach der Methode von Bartlett (1937) ermittelt.

Die Ähnlichkeit zwischen den in Kapitel 6.3 dargestellten konventionellen Lösungen zweiter Ordnung und den anhand der Faktorwerte ermittelten Lösungen zweiter Ordnung wurde anhand von Kongruenzkoeffizienten geprüft. Die Kongruenzen der anhand der Interkorre-

lationen der Faktorwerte ermittelten Oblimin-rotierten Lösungen mit den anhand der Interkorrelationen der Faktoren ermittelten Oblimin-Lösungen für die gewichteten Stichproben der Gruppen A (s. Tabelle 6.36) und B (s. Tabelle 6.42) sind in Tabelle 6.43 dargestellt.

Insgesamt sind die Kongruenz-Koeffizienten äußerst hoch, was bedeutet, dass durch die Interkorrelationen der Faktorwerte der Faktoren erster Ordnung die Interkorrelationen der Faktoren erster Ordnung zufriedenstellend abgebildet werden können. Im übrigen sind die Kongruenzen so hoch, dass eine nochmalige Darstellung der Lösungen, die mit denen in Tabelle 6.36 bzw. Tabelle 6.42 nahezu identisch sind, nicht erforderlich war. Besonders relevant sind hier die Kongruenzen der Faktoren Neurotizismus und Extraversion, da für diese Faktoren der Vergleich der Generalität mit den entsprechenden Faktoren des EPQ-R und NEO-PI-R durchgeführt wurde.

Tabelle 6.43

Kongruenzen der Faktoren der anhand der Faktorwerte ermittelten Lösungen mit den anhand der Korrelationen ermittelten Lösungen

Lösung in Tabelle 6.36, Gruppe A	N	V	E-	G-	O
Kongruenz	.97	.96	.99	1.00	.98
Lösung in Tabelle 6.42, Gruppe B	P	N-	Neg (?)	Rob- (?)	E
Kongruenz	1.00	1.00	.99	.99	1.00

Anmerkung. N = Neurotizismus; V = Verträglichkeit; E- = Extraversion (minus); G- = Gewissenhaftigkeit (minus); O = Offenheit für Erfahrungen; P = Psychotizismus; N- = Neurotizismus (minus); Neg (?) = Negative Grundeinstellungen; Rob- (?) = Robustheit (minus); E = Extraversion.

Zunächst wurde der Generalitätsvergleich für die Faktoren erster und zweiter Ordnung der gewichteten Stichproben A (N=554) und B (N=620) durchgeführt. Dazu wurden die vier Primärfaktoren mit den höchsten Hauptladungen auf dem Faktor „Neurotizismus“ der Oblimin-Lösung zweiter Ordnung (siehe Tabelle 6.36 für Gruppe A und Tabelle 6.42 für Gruppe B) mit diesem „Neurotizismus“-Faktor in Bezug auf die Inklusion verglichen. Da diese vier Primärfaktoren in den „Neurotizismus“-Faktor eingehen, wurde erwartet, dass sie kleinere Inklusions-Indizes aufweisen, als der „Neurotizismus“-Faktor zweiter Ordnung.

In der Oblimin-Lösung zweiter Ordnung der Gruppe A weisen die vier Primärfaktoren „Depressivität (minus)“, „Reizbarkeit/Ungeduld“, „Hedonistische Kontrolle“ und „Schuldgefühle/Scham“ die höchsten Ladungsbeträge auf dem „Neurotizismus“-Faktor auf. Für diese Primärfaktoren wurde der Vergleich der Inklusions-Indizes mit dem „Neurotizismus“-Faktor durchgeführt. In der Oblimin-Lösung zweiter Ordnung der Gruppe B weisen die vier

Primärfaktoren „Neurotizismus (minus)“, „Depressivität (minus)“, „Sorge und Ängstlichkeit (minus)“ sowie „Selbstdisziplin“ die höchsten Ladungsbeträge auf. Daher wurde der Generalitätsvergleich mit dem Neurotizismusfaktor zweiter Ordnung in der Gruppe B für diese vier Faktoren durchgeführt. Die Ergebnisse zu den Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} sind in Tabelle 6.44 dargestellt.

Die Inklusions-Indizes I_{f2} , I_{f3} , und I_{f4} sind für alle Faktoren in beiden Gruppen positiv, was auf eine höhere Generalität des Faktors zweiter Ordnung hindeutet und zugleich darauf hinweist, dass die Inklusions-Indizes die erwarteten Generalitätsunterschiede angemessen abbilden. Für den Inklusions-Index I_{f5} ergaben sich für die beiden Faktoren „Depressivität (minus)“ und „Sorge und Ängstlichkeit (minus)“ negative Werte, was auf eine höhere Generalität dieser beiden Primärfaktoren im Vergleich zum Faktor zweiter Ordnung oder auf Probleme bei der Erfassung der Generalität mit Hilfe des I_{f5} -Indexes hindeutet.

Tabelle 6.44

Vergleich der Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} für die „Neurotizismus“-Faktoren zweiter Ordnung mit den jeweils vier am höchsten auf diesem Faktor ladenden Primärfaktoren (gewichtete Stichprobe)

	I_{f2}	I_{f3}	I_{f4}	I_{f5}
N, 2. Ordnung* – Depressivität (-)				
N (Gruppe A)	.064	.034	8.453	6.910
N, 2. Ordnung* – Reizbarkeit/Ungeduld				
N (Gruppe A)	.100	.062	6.118	4.381
N, 2. Ordnung* – Hedonistische Kontrolle				
N (Gruppe A)	.159	.101	3.423	.962
N, 2. Ordnung* – Schuldgefühle/Scham				
N (Gruppe A)	.094	.058	3.819	2.243
N, 2. Ordnung* – Neurotizismus (minus)				
N (Gruppe B)	.097	.052	6.286	4.786
N, 2. Ordnung* – Depressivität (minus)				
N (Gruppe B)	.199	.113	.155	-2.401
N, 2. Ordnung* – Sorge und Ängstlichkeit (minus)				
N (Gruppe B)	.234	.234	.530	-2.282
N, 2. Ordnung* – Selbstdisziplin				
N (Gruppe B)	.121	.080	4.192	2.910

Anmerkungen. N = Neurotizismus; Gruppe A: N = 554; Gruppe B: N = 620. * Ein positiver Inklusions-Index deutet auf eine höhere Generalität der hier zuerst genannten Faktoren hin.

Von einer höheren Generalität der Primärfaktoren im Vergleich zu einem Faktor zweiter Ordnung, in den diese eingehen, kann vor allem dann ausgegangen werden, wenn die Primärfaktoren substanzielle Nebenladungen auf anderen Faktoren zweiter Ordnung aufweisen. In diesem Fall gehen die Primärfaktoren nicht unbedingt mit dem Hauptteil ihrer gemeinsamen Varianz in den Faktor zweiter Ordnung ein, weshalb letzterer eine weniger hohe Generalität aufweisen kann. Die Primärfaktoren „Depressivität (minus)“ und „Sorge und Ängstlichkeit (minus)“ weisen jedoch keine besonders hohen Nebenladungen auf weiteren Faktoren zweiter Ordnung auf (s. Tabelle 6.42).

Des Weiteren sind die Kommunalitäten der Primärfaktoren in den Lösungen zweiter Ordnung insgesamt eher gering, weshalb eine höhere Generalität der Primärfaktoren aufgrund von Varianz, die nicht in die Lösungen zweiter Ordnung eingeht, denkbar ist. Da jedoch weder die Kommunalitäten der beiden Faktoren „Depressivität (minus)“ und „Sorge und Ängstlichkeit (minus)“ geringer sind als die der übrigen Primärfaktoren (Tabelle 6.42) noch die Nebenladungen dieser Primärfaktoren auf anderen Faktoren zweiter Ordnung besonders hoch sind und alle übrigen Inklusions-Indizes auf eine höhere Generalität des Faktors zweiter Ordnung hindeuten, wird die Abweichung von der erwarteten Generalität für die beiden Primärfaktoren hier eher auf Probleme des I_{15} -Indexes zurückgeführt.

Der Generalitätsvergleich zwischen dem Extraversionsfaktor zweiter Ordnung und den vier Primärfaktoren mit den höchsten Ladungen auf diesem Faktor wurde in der Gruppe A für die Primärfaktoren „Geselligkeit (minus)“, „Soziale Kompetenz (minus)“, „Besonnenheit (minus)“ und „Privatheit (minus)“ durchgeführt. In der Gruppe B wurde der Generalitätsvergleich für die Faktoren „Lebhaftigkeit“, „Allein sein können“, „Geselligkeit“ und „Gefühlsbetontheit (minus)“ durchgeführt (s. Tabelle 6.45).

Insgesamt ergaben sich für alle Inklusions-Indizes positive Werte, was – wie erwartet – auf eine höhere Generalität des Faktors zweiter Ordnung hindeutet. Da die untersuchten Primärfaktoren mit substanziellen Varianzanteilen in den jeweiligen „Extraversions“-Faktor zweiter Ordnung eingingen, wird dieses Ergebnis als Hinweis auf die Validität der Inklusions-Indizes gedeutet. Insgesamt ergaben sich nur für den I_{15} -Index Hinweise auf eine möglicherweise suboptimale vergleichende Erfassung der Generalität.

Tabelle 6.45

Vergleich der Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} für die „Extraversions“-Faktoren zweiter Ordnung mit den jeweils vier am höchsten auf diesem Faktor ladenden Primärfaktoren (gewichtete Stichprobe)

	I_{f2}	I_{f3}	I_{f4}	I_{f5}
N, 2. Ordnung* – Geselligkeit (minus)				
E (Gruppe A)	.031	.017	8.027	7.063
N, 2. Ordnung* – Soziale Kompetenz (minus)				
E (Gruppe A)	-.004	-.006	6.077	5.464
N, 2. Ordnung* – Besonnenheit (minus)				
E (Gruppe A)	.011	.005	5.462	4.899
N, 2. Ordnung* – Privatheit (minus)				
E (Gruppe A)	.138	.092	3.153	1.306
N, 2. Ordnung* – Lebhaftigkeit				
E (Gruppe B)	.077	.043	4.972	3.783
N, 2. Ordnung* – Allein sein können				
E (Gruppe B)	.111	.058	5.960	4.145
N, 2. Ordnung* – Geselligkeit				
E (Gruppe B)	.161	.098	1.792	3.180
N, 2. Ordnung* – Gefühlsbetontheit (minus)				
E (Gruppe B)	.114	.086	1.798	.990

Anmerkungen. E = Extraversion; Gruppe A: N= 554; Gruppe B: N = 620. * Ein positiver Inklusions-Index deutet auf eine höhere Generalität der hier zuerst genannten Faktoren hin.

6.4.2 Zur Generalität der Neurotizismus-Faktoren des EPQ-R, NEO-PI-R und ZKPQ

Die Anwendung der Inklusions-Indizes erscheint zunächst besonders in den Fällen interessant, in denen dasselbe Konstrukt mit Hilfe unterschiedlicher Faktoren erfasst werden soll. So wird sowohl mit dem EPQ-R, dem NEO-PI-R und dem ZKPQ Neurotizismus erfasst. Im NEO-PI-R ist für Neurotizismus eine Binnenstruktur vorgesehen, die allerdings in exploratorischen Faktorenanalysen nicht aufgezeigt werden konnte (siehe Kapitel 6.2.3). Insofern ist die intendierte Binnenstruktur keine Garantie für eine höhere Generalität des Neurotizismus-Faktors im NEO-PI-R gegenüber dem Neurotizismus-Faktor im EPQ-R. Im folgenden soll die Generalität des anhand

des NEO-PI-R ermittelten Neurotizismus-Faktors mit der Generalität des anhand des EPQ-R ermittelten Neurotizismus-Faktors verglichen werden.

Es geht hier nicht wie in Kapitel 6.3 um den Nachweis der Faktoren, sondern um die vergleichende Generalitätsbewertung der Faktoren, wie sie von den Modellproponenten vorgeschlagen werden. Daher werden die Faktoren hier auf der Variablenbasis und weitgehend mit den Methoden gewonnen wie in Ruch (1999) für den EPQ-R und in McCrae et al. (1996) für den NEO-PI-R und in Zuckerman et al. (1993) für den ZKPQ vorgeschlagen wurde. D.h., der Neurotizismus-Faktor des EPQ-R wird anhand einer Hauptkomponentenanalyse der Items und der Neurotizismus-Faktor des NEO-PI-R anhand einer Hauptkomponentenanalyse der Facetten-Skalen ermittelt. Die Hauptkomponentenanalyse wird hier verwendet, weil Ruch (1999), McCrae et al. (1996) und Zuckerman et al. (1993) diese Methode wählten, obwohl sie vermutlich ein Verfahren ist, das zu weniger robusten Ergebnissen führt (Beauducel, 2001b; Widaman, 1993). Die Verwendung der Facetten-Skalen statt der Items bedeutet, dass eine theoriekonforme Aggregation dem Nachweis der Faktoren vorangestellt wird, was hier als suboptimales Vorgehen angesehen wird. Diese Vorgehensweisen werden hier nur gewählt, um die von den Autoren intendierten Faktoren möglichst gut abzubilden, um auf dieser Basis Generalitätsvergleiche exemplarisch und mit Bezug zu aktuellen Inventaren durchführen zu können. Die Faktoren wurden hier Varimax-rotiert, weil diese Methode im Kontext des Fünf-Faktorenmodells am häufigsten eingesetzt wird und weil die verglichenen Faktoren auf derselben Rotationsmethode beruhen sollten.

Die für den Generalitätsvergleich durchgeführten Faktorenanalysen können natürlich nicht im Sinne der Bestätigung der jeweiligen Modelle interpretiert werden, da die Anzahl der zu extrahierenden Faktoren den Modellannahmen entsprechend vorgegeben wurde. Die für den Generalitätsvergleich durchgeführten Faktorenanalysen demonstrieren vielmehr die Möglichkeit, intendierte Modellstrukturen auf einem methodisch suboptimalen Niveau darzustellen. Für die Bewertung der Modell-Gültigkeit sei auf die für die Modelle getrennt durchgeführten Faktorenanalysen in Kapitel 6.2 verwiesen.

Die Analysen wurden für den EPQ-R und den NEO-PI-R getrennt für die gemeinsame ungewichtete Stichprobe der Gruppe A (N=554) und der Gruppe B (N=620) durchgeführt, um die Robustheit der Ergebnisse zu prüfen. Ebenso wurden die Berechnungen auch für die gewichteten Stichproben der Gruppen A und B durchgeführt, um den Einfluss der Stichprobenzusammensetzung auf die Generalitätsschätzungen zu untersuchen. Die entsprechenden Analysen wurden für den ZKPQ nur in der ungewichteten und gewichteten Gruppe B durchgeführt, da er nur in dieser Gruppe dargeboten wurde.

Der Neurotizismus-Faktor konnte in allen drei Inventaren in den gewichteten und ungewichteten Stichproben auf Basis der Gruppen A bzw. B anhand der beschriebenen Hauptkomponentenanalysen ermittelt werden. Die Ladungsmuster der entsprechenden Lösungen finden sich im Anhang (siehe Tabellen A 6.32 - A 6.37).

Der Inklusions-Index I_{f2} liegt für den Vergleich von EPQ-R und NEO-PI-R sowohl in den gewichteten als auch in den ungewichteten Stichproben A und B im negativen Bereich (siehe Tabelle 6.46). Die Werte wurden so in Formel 10 eingesetzt, dass negative Werte auf eine höhere Generalität des NEO-PI-R Faktors und positive Werte auf eine höhere Generalität des EPQ-R Faktors hindeuten. Der Inklusions-Index I_{f2} kann somit dahingehend interpretiert werden, dass der Neurotizismus-Faktor des NEO-PI-R rund 5% mehr Varianz am Neurotizismus-Faktor des EPQ-R erklärt als der Neurotizismus-Faktor des EPQ-R an dem des NEO-PI-R. Für den Inklusions-Index I_{f3} ergibt sich, dass der Neurotizismus-Faktor des NEO-PI-R rund 4% mehr Varianz am Neurotizismus-Faktor des EPQ-R erklärt als der Neurotizismus-Faktor des EPQ-R an dem des NEO-PI-R (siehe Tabelle 6.46). Die Inklusions-Indizes I_{f4} und I_{f5} deuten darauf hin, dass sich die erklärte Varianz des Neurotizismus-Faktors des NEO-PI-R auf eine größere Itemmenge verteilt als die erklärte Varianz des Neurotizismus-Faktors des EPQ-R. Diese Ergebnisse werden als Hinweis darauf angesehen, dass der Neurotizismus-Faktor des NEO-PI-R eine höhere Generalität aufweist als der des EPQ-R.

Tabelle 6.46

Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} für Neurotizismus im EPQ-R, NEO-PI-R und ZKPQ

	I_{f2}	I_{f3}	I_{f4}	I_{f5}
EPQ-R* – NEO-PI-R				
N (unge., Gruppe A)	-0.049	-0.049	-1.721	-1.697
N (unge., Gruppe B)	-0.031	-0.031	-2.660	-2.638
N (ge., Gruppe A)	-0.037	-0.039	-2.004	-2.008
N (ge., Gruppe B)	-0.032	-0.033	-2.707	-2.689
EPQ-R* – ZKPQ				
N (unge., Gruppe B)	-0.012	-0.012	2.522	2.520
N (ge., Gruppe B)	-0.021	-0.022	3.573	3.504
ZKPQ* – NEO-PI-R				
N (unge., Gruppe B)	-0.027	-0.028	-3.118	-3.106
N (ge., Gruppe B)	-0.016	-0.018	-4.258	-4.242

Anmerkungen. N = Neurotizismus; unge. = ungewichtet; ge. = gewichtet; Gruppe A: N = 554; Gruppe B: N = 620.
* Ein positiver Inklusions-Index deutet auf eine höhere Generalität des hier zuerst genannten Fragebogens hin. Der Neurotizismus-Faktor des NEO-PI-R basiert hier auf Analysen der Facetten-Skalen.

Beim Vergleich des Neurotizismus-Faktors des EPQ-R mit dem des ZKPQ ergaben sich für die Inklusions-Indizes I_{f2} und I_{f3} sehr geringe Unterschiede, die auf eine etwas höhere Generalität des Neurotizismus-Faktors im ZKPQ hindeuten. Andererseits ergab sich für I_{f4} und I_{f5} eine etwas breitere Verteilung der erklärten Varianzen des Neurotizismus-Faktors des EPQ-R im Vergleich zum Neurotizismus-Faktor des ZKPQ. Insgesamt sollte man somit davon ausgehen, dass die Generalitätsunterschiede zwischen den Neurotizismus-Faktoren des EPQ-R und ZKPQ gering sind.

In Einklang damit ist, dass die Generalitäts-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} auf eine höhere Generalität des Neurotizismus-Faktors des NEO-PI-R im Vergleich zum Neurotizismus-Faktor des ZKPQ hindeuten, auch wenn der Unterschied für I_{f3} nur gering ist. Der Neurotizismus-Faktor des NEO-PI-R hat gemäß der hier ermittelten Indizes die höchste Generalität im Vergleich zu den Neurotizismus-Faktoren des EPQ-R und ZKPQ, die eine ähnliche Generalität aufweisen.

6.4.3 Zur Generalität der Extraversions-Faktoren des EPQ-R und NEO-PI-R sowie der Soziabilitäts- und Aktivitäts-Faktoren im ZKPQ

Der Extraversions-Faktor konnte im EPQ-R und NEO-PI-R in den gewichteten und ungewichteten Stichproben auf Basis der Gruppen A bzw. B anhand der beschriebenen Faktorenanalysen bestätigt werden (siehe Anhang, Tabellen A 6.32 - A 6.35). Im ZKPQ erfassen die Faktoren Aktivität und Soziabilität Varianzen, die auch einem Extraversions-Faktor zugeordnet werden können (s. Zuckerman et al., 1988). Die Einbeziehung dieser Faktoren erscheint insofern interessant als diese vermutlich geringere Generalität aufweisen sollten als die Extraversions-Faktoren des EPQ-R und NEO-PI-R. Auch die Faktoren Aktivität und Soziabilität konnten in entsprechend der o.g. Spezifikationen durchgeführten Hauptkomponentenanalysen nachgewiesen werden (siehe Anhang, Tabellen A 6.36 und A 6.37).

Für den Vergleich der Extraversions-Faktoren des EPQ-R und NEO-PI-R ergab sich für den Inklusions-Index I_{f2} in allen Stichproben ein sehr knappes Ergebnis. Der Extraversions-Faktor des EPQ-R erklärte demnach etwa 1.5% mehr Varianz am Extraversions-Faktor des NEO-PI-R als der Extraversions-Faktor des NEO-PI-R an dem des EPQ-R (siehe Tabelle 6.47).

Für den Inklusions-Index I_{f3} lagen die Unterschiede in der wechselseitig erklärten Varianz, die auf eine etwas höhere Generalität des Extraversions-Faktors des EPQ-R hindeuten, bei etwa 3.5%. Für die Inklusions-Indizes I_{f4} und I_{f5} ergaben sich uneinheitliche Ergebnisse, die jedoch insgesamt eher auf eine etwas höhere Generalität des Extraversions-Faktors im NEO-PI-R

hinweisen. Insgesamt deuten die hier ermittelten Indizes auf geringe Generalitäts-Unterschiede zwischen den Extraversions-Faktoren des EPQ-R und des NEO-PI-R hin.

Für den Vergleich der Extraversions-Faktoren des EPQ-R und NEO-PI-R mit dem Sozialisierbarkeit-Faktor des ZKPQ ergaben sich erwartungsgemäß für alle ermittelten Inklusions-Indizes sowohl in der ungewichteten als auch in der gewichteten Stichprobe etwas höhere Generalitäten für die Extraversions-Faktoren. Der Vergleich der Extraversions-Faktoren des EPQ-R und NEO-PI-R mit dem Aktivitäts-Faktor ergab für die Indizes I_{f2} und I_{f3} für die gewichtete und ungewichtete Stichprobe etwas höhere Generalitäten der Extraversions-Faktoren.

Tabelle 6.47

Inklusions-Index I_{f2} , I_{f3} , I_{f4} und I_{f5} für Extraversion im EPQ-R, NEO-PI-R und ZKPQ

	I_{f2}	I_{f3}	I_{f4}	I_{f5}
EPQ-R* – NEO-PI-R				
E (unge., Gruppe A)	.047	.043	-1.368	-1.516
E (unge., Gruppe B)	.036	.030	.536	.350
E (ge., Gruppe A)	.044	.044	-1.109	-1.123
E (ge., Gruppe B)	.032	.028	-.929	-1.067
EPQ-R* - ZKPQ/So				
E/So (unge., Gruppe B)	.075	.062	.684	.186
E/So (ge., Gruppe B)	.064	.054	.547	.191
ZKPQ/So* - NEO-PI-R				
E/So (unge., Gruppe B)	-.049	-.042	-8.418	-7.986
E/So (ge., Gruppe B)	-.037	-.031	-10.248	-9.842
EPQ-R* - ZKPQ/Akt				
E/Akt (unge., Gruppe B)	.064	.052	-8.392	-8.842
E/Akt (ge., Gruppe B)	.065	.054	-7.799	-8.176
ZKPQ/Akt* - NEO-PI-R				
E/Akt (unge., Gruppe B)	-.057	-.050	-1.802	-1.340
E/Akt (ge., Gruppe B)	-.050	-.046	-2.236	-1.846

Anmerkungen. E = Extraversion; unge. = ungewichtet; ge. = gewichtet; Gruppe A: N= 554; Gruppe B: N = 620. * Ein positiver Inklusions-Index deutet auf eine höhere Generalität des hier zuerst genannten Fragebogens hin. Der Extraversions-Faktor des NEO-PI-R basiert hier auf Analysen der Facetten-Skalen.

Für die Indizes I_{f4} und I_{f5} ergab sich eine etwas höhere Generalität des Extraversions-Faktors des NEO-PI-R aber eine deutlich geringere Generalität des Extraversions-Faktors des EPQ-R im Vergleich zum Aktivitäts-Faktor des ZKPQ. Dieses unerwartete Ergebnis kommt vor allem durch die geringe Varianz der Korrelationen des Aktivitäts-Faktors mit den Extraversions-

Items zustande. Die geringe Varianz der Korrelationen könnte damit zusammenhängen, dass insgesamt viele Items des EPQ-R verhaltensnahe Fragen nach bestimmten Aktivitäten darstellen (z.B. Item 79: „Wenn Sie einen Zug erreichen wollen, kommen sie dann oft in der letzten Minute?“). Da die Fragen nach Aktivitäten sich im EPQ-R nicht auf die Extraversions-Items beschränken, sondern auch auf die übrigen Items, ist es denkbar, dass Aktivität im EPQ-R vergleichsweise breit im Hintergrund erfasst wird. Dieser Effekt könnte im NEO-PI-R dadurch kompensiert werden, dass dort eine spezifische Facetten-Skala „Aktivität“ erfasst wird, die dazu beiträgt, die Varianz der Korrelationen zu erhöhen.

Die hier durchgeführten Analysen deuten insgesamt darauf hin, dass der Soziabilitäts-Faktor des ZKPQ die geringste Generalität aufweist. Für die übrigen Faktoren sind die Ergebnisse weniger einheitlich. In Bezug auf die anhand von I_{f2} und I_{f3} erfassten, wechselseitig erklärten Varianzen zeichnet sich eine etwas höhere Generalität für den Extraversions-Faktor des EPQ-R gegenüber dem Extraversions-Faktor des NEO-PI-R ab. Dieses Ergebnis wird jedoch durch die Ergebnisse von I_{f4} und I_{f5} relativiert, die auf eine etwas höhere Generalität des NEO-PI-R Extraversions-Faktors hindeuten. Auch der Vergleich zwischen dem Aktivitäts-Faktor des ZKPQ und dem Extraversions-Faktor des EPQ-R anhand von I_{f4} und I_{f5} deutet auf eine geringere Generalität des Extraversions-Faktors des EPQ-R hin. Da die Ergebnisse für die verschiedenen Faktoren und Indizes etwas uneinheitlich sind, wurden sie im folgenden grafisch veranschaulicht (siehe Abbildung 6.11).

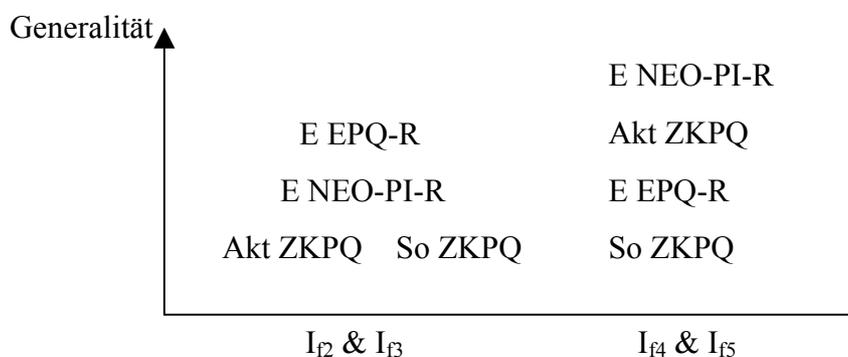


Abbildung 6.11. Zusammenfassung der Generalitätsvergleiche für Extraversion (E), Soziabilität (So) und Aktivität (Akt)

Die größten Unterschiede zwischen den Indizes I_{f2} und I_{f3} im Vergleich zu I_{f4} und I_{f5} treten für den Extraversions-Faktor des EPQ-R und den Aktivitäts-Faktor des ZKPQ auf. Die Unterschiede in den Indizes deuten darauf hin, dass die Generalität von Faktoren möglicherweise kein einheitliches Konzept ist, sondern dass Faktoren unter verschiedenen Gewichtspunkten eine unterschiedliche Generalität aufweisen können.

Andererseits ergaben sich für den Neurotizismus-Faktor vergleichsweise einheitliche Ergebnisse, weshalb die Unterschiede nicht auf eine grundsätzliche Gegenläufigkeit der in den Inklusions-Indizes operationalisierten Generalitätsaspekte schließen lassen. Um die Einheitlichkeit der mit den Indizes erfassten Generalitätsaspekte weiter zu untersuchen, bietet sich ein Generalitätsvergleich zwischen modellspezifischen Faktoren und den im Kapitel 6.3 ermittelten Faktoren zweiter Ordnung an.

Insgesamt ist bemerkenswert, dass für die Extraversions- und Neurotizismus-Faktoren des EPQ-R und NEO-PI-R unterschiedliche Ergebnisse erzielt wurden. Obwohl die Aggregationsniveaus der Variablen, die in die Faktorenanalysen gingen, für die beiden Inventare unterschiedlich waren (beim EPQ-R Items, beim NEO-PI-R Facetten Skalen) führten diese Unterschiede nicht zu einer einheitlich höheren Generalitätsschätzung für den NEO-PI-R. Dies spricht dafür, dass der Index vermutlich nur in geringerem Umfang Methodenvarianz widerspiegelt. Dabei muss jedoch in Rechnung gestellt werden, dass die Facetten-Skalen in den exploratorischen Analysen zum NEO-PI-R nicht bestätigt werden konnten, so dass der Nutzen der Aggregation bei den Facetten-Skalen vermutlich nicht voll zum tragen kam.

6.4.4 Zum Generalitätsvergleich modellspezifischer Faktoren mit den in Kapitel 6.3 ermittelten Extraversions- und Neurotizismus-Faktoren zweiter Ordnung

Der Generalitätsvergleich wurde auch hier auf die Faktoren Extraversion und Neurotizismus begrenzt, da diese Faktoren sowohl in den modellspezifischen als auch in den modellübergreifenden Analysen aufgezeigt werden konnten. Insofern können die modellspezifischen und modellübergreifenden Extraversions- und Neurotizismus-Faktoren hier direkt aufeinander bezogen werden. Auch wurden die Vergleiche hier nur für die Faktoren der gewichteten Stichprobe durchgeführt, da den Ergebnissen der gewichteten Stichprobe aufgrund der verringerten Stichprobenspezifität eine höhere Bedeutung beizumessen ist. Es wurden die in Kapitel 6.3 dargestellten Faktoren zweiter Ordnung mit den Faktoren verglichen, die sich bei modellspezifischen Analysen ergaben. Die modellspezifischen Analysen wurden in Bezug auf die Anzahl der zu extrahierenden Faktoren den spezifischen Modellannahmen angepasst (drei Faktoren beim EPQ-R, fünf beim NEO-PI-R), allerdings wurden die Analysen hier auf der Basis der Hauptachsenanalyse mit Oblimin-Rotation durchgeführt, da die in Kapitel 6.3 dargestellten Faktoren zweiter Ordnung auf diese Weise gewonnen wurden und es nicht sinnvoll erscheint, den Generalitätsvergleich für Faktoren durchzuführen, die auf der Basis unterschiedlicher Extraktions- und Rotationsmethoden entstanden sind. Auch wurden die Analysen zum NEO-PI-R sowohl auf der Basis der Facetten-Skalen (siehe Anhang, Tabelle A 6.39) als auch auf der Basis der Items durchgeführt (siehe Anhang, Tabelle A 6.40 und A 6.41), um den Einfluss des

Aggregationsniveaus der Ausgangsvariablen zu untersuchen. Auch die Hauptachsenlösungen mit anschließender Oblimin-Rotation für den EPQ-R finden sich im Anhang (siehe Tabellen A 6.38).

Für den Neurotizismus-Faktor sind die Vergleiche zwischen den Faktoren zweiter Ordnung in den modellspezifischen Faktoren in Tabelle 6.48 aufgeführt. Die Inklusions-Indizes I_{f2} und I_{f3} deuten auf eine höhere Generalität der modellspezifischen Faktoren hin, während die Inklusions-Indizes I_{f4} und I_{f5} auf eine höhere Generalität des Faktors zweiter Ordnung hindeuten. Die auf Itembasis und auf Basis der Facetten-Skalen ermittelten Neurotizismus-Faktoren des NEO-PI-R erbringen im Generalitätsvergleich mit dem Neurotizismus-Faktor zweiter Ordnung nahezu dieselben Ergebnisse.

Tabelle 6.48

Vergleich der Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} für die Neurotizismus-Faktoren zweiter Ordnung mit den modellspezifischen Neurotizismus-Faktoren des EPQ-R und NEO-PI-R

	I_{f2}	I_{f3}	I_{f4}	I_{f5}
2. Ordnung* – EPQ-R				
N (ge., Gruppe A)	-.095	-.071	5.104	5.712
N (ge., Gruppe B)	-.043	-.064	3.852	3.128
2. Ordnung* – NEO-PI-R/Facetten-Skalen				
N (ge., Gruppe A)	-.190	-.153	4.858	5.775
N (ge., Gruppe B)	-.072	-.064	2.785	3.128
2. Ordnung* – NEO-PI-R/Item				
N (ge., Gruppe A)	-.119	-.077	4.199	5.143
N (ge., Gruppe B)	-.045	-.029	1.888	2.263

Anmerkungen. N = Neurotizismus; ge. = gewichtet; Gruppe A: N= 554; Gruppe B: N = 620. * Ein positiver Inklusions-Index deutet auf eine höhere Generalität der hier zuerst genannten Faktoren hin.

Die geringere Generalität der Faktoren zweiter Ordnung in den Indizes I_{f2} und I_{f3} , die ausschließlich die wechselseitig erklärte Varianz widerspiegeln, kann damit zusammenhängen, dass die Faktoren zweiter Ordnung aufgrund der äußerst geringen Interkorrelationen der Primärfaktoren nur wenig Varianz der Primärfaktoren und damit auch wenig Varianz der Items erklären. Dadurch, dass die modellspezifischen Faktoren hier als Faktoren erster Ordnung ermittelt wurden, erklären diese Faktoren mehr Varianz. Zugleich verteilt sich die erklärte Varianz bei dem Neurotizismus-Faktor zweiter Ordnung offenbar gleichmäßiger auf die Items, weshalb die Inklusions-Indizes I_{f4} und I_{f5} auf eine höhere Generalität des Faktors zweiter Ordnung hindeuten. Wie in den Analysen in Kapitel 6.4.3 unterscheiden sich die Faktoren somit hinsichtlich der verschiedenen Generalitätsaspekte.

In Bezug auf die Extraversions-Faktoren deuten sowohl die Inklusions-Indizes I_{f2} und I_{f3} als auch die Indizes I_{f4} und I_{f5} auf eine höhere Generalität der modellspezifischen Faktoren hin (s. Tabelle 6.49). Ausnahmen ergaben sich lediglich für I_{f4} und I_{f5} beim Extraversions-Faktor des EPQ-R und bei I_{f2} und I_{f3} beim Item-basierten Extraversions-Faktor des NEO-PI-R in der Gruppe B. Die insgesamt höhere Generalität der modellspezifischen Faktoren, ist ein Hinweis darauf, dass der Generalitätsgrad von Faktoren nicht durchgängig proportional zum Ordnungsgrad der Faktoren ist. Faktoren erster Ordnung können durchaus genereller sein, als Faktoren höherer Ordnung.

Allerdings ist im Hinblick auf die hier untersuchten modellspezifischen Faktoren anzumerken, dass zwar deren Generalität hoch, jedoch vor allem im Falle des NEO-PI-R ihre Validität eingeschränkt ist, da bei den modellspezifischen exploratorischen Faktorenanalysen (siehe Kapitel 6.2) die Parallelanalysen und Komparabilitäten nicht für die Extraktion von fünf Faktoren sprachen. Auch beim EPQ-R konnten die drei Faktoren nicht durchgängig nachgewiesen werden. Es ist durchaus denkbar, dass eine verminderte konvergente und diskriminante Validität der Markieritems von Faktoren und die damit zusammenhängende faktorielle Komplexität von Items, die einen Faktor konstituieren, zu einer erhöhten Generalitätsschätzung anhand der Inklusions-Indizes beigetragen hat. Erhöhte Generalitätsschätzungen von Faktoren können somit durchaus mit einer empirisch unscharfen Konzeption zusammenhängen. Insofern muss für eine sichere Generalitätsschätzung die strukturelle Validität der Faktoren in vollem Umfang gegeben sein, was die Aussagekraft der vorliegenden Ergebnisse relativiert.

Tabelle 6.49

Vergleich der Inklusions-Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} für die Extraversions-Faktoren zweiter Ordnung mit den modellspezifischen Extraversions-Faktoren des EPQ-R und NEO-PI-R (gewichtete Stichprobe)

	I_{f2}	I_{f3}	I_{f4}	I_{f5}
2. Ordnung* – EPQ-R				
E (Gruppe A)	-.124	-.097	6.031	6.675
E (Gruppe B)	-.072	-.035	2.367	-2.895
2. Ordnung* – NEO-PI-R/Facetten-Skalen				
E (Gruppe A)	-.068	-.063	-2.345	-1.831
E (Gruppe B)	-.046	-.035	-3.480	-2.895
2. Ordnung* – NEO-PI-R/Item				
E (Gruppe A)	-.006	-.004	-.087	-.049
E (Gruppe B)	.033	.023	-6.026	-6.309

Anmerkungen. E = Extraversion; Gruppe A: N = 554; Gruppe B: N = 620. * Ein positiver Inklusions-Index deutet auf eine höhere Generalität der hier zuerst genannten Faktoren hin.

7. Zusammenfassende Diskussion und Ausblick

In dieser Arbeit wurden Beiträge zur Lösung wesentlicher Probleme der psychometrischen Traitforschung aufgezeigt. Dabei wurde von der Einschätzung ausgegangen, dass ein wesentliches Problem in einer *theoretisch-interpretativen und empirisch-methodischen Unverbindlichkeit* der psychometrischen Traitforschung besteht (siehe Kapitel 1). Die empirisch-methodische Unverbindlichkeit ist vor allem auf einen suboptimalen bzw. wenig disziplinierten Einsatz der Faktorenanalyse zurückzuführen, die theoretische Unverbindlichkeit auf eine eher willkürliche Einschätzung der Generalität von Faktoren. In der vorliegenden Arbeit wurden zu beiden Hauptproblemen der psychometrischen Traitforschung Lösungsvorschläge unterbreitet. Demzufolge wurden in dieser Arbeit Vorschläge unterbreitet, die dazu beitragen sollen, die *empirisch-methodische Unverbindlichkeit* der psychometrischen Traitforschung durch einen methodisch anspruchsvollen, weniger willkürlichen Einsatz der faktorenanalytischen Verfahren zu reduzieren. Zum anderen wurden Beiträge zur Quantifizierung der Generalität von Faktoren als Weg zur Reduktion der *theoretisch-interpretativen Unverbindlichkeit* psychometrischer Traitmodelle geleistet. Im folgenden soll zusammenfassend diskutiert werden, inwieweit diese Ziele – zumindest im Rahmen dieser Arbeit – erreicht werden konnten. Natürlich bedarf eine wirkungsvolle Reduktion der Unverbindlichkeiten der psychometrischen Traitforschung nicht nur der Aufarbeitung bestehender und der Neuformulierung methodischer Standards, wie dies im Rahmen dieser Arbeit erfolgte, sondern auch der sozialen Durchsetzung dieser Standards (siehe z.B. Westmeyer, 1995). Insofern soll mit dieser Arbeit auch dafür geworben werden, in der Reduktion der Freiheitsgrade bei der Durchführung von Faktorenanalysen und bei der Interpretation von Faktoren eine Möglichkeit zu erkennen, den empirischen Gehalt psychometrischer Traitmodelle zu erhöhen.

Entsprechend diesem Ziel wurden in der Arbeit zunächst theoretische und methodische Überlegungen zur methodisch anspruchsvollen Durchführung von Faktorenanalysen (Kapitel 3) sowie zur Bewertung der Generalität von Faktoren dargestellt (Kapitel 4). In einem zweiten Schritt wurden diese Überlegungen exemplarisch auf einen Datensatz, der in zwei Gruppen geteilt wurde, angewendet (Kapitel 6). Dementsprechend werden in Kapitel 7.1 die Ergebnisse der theoretischen und methodischen Überlegungen, in Kapitel 7.2 die faktorenanalytischen Ergebnisse der empirischen Untersuchungen und in Kapitel 7.3 die Ergebnisse des empirischen Generalitätsvergleichs zusammengefasst und diskutiert. In Kapitel 7.4 werden die wesentlichen Einschränkungen und Grenzen der vorliegenden Arbeit zusammenfassend diskutiert. Schließlich wird aus den wesentlichen Grenzen der vorliegenden Arbeit das verbleibende Arbeitsprogramm zur Realisierung einer verbindlicheren psychometrischen Traitforschung abgeleitet.

7.1 Ergebnisse der theoretischen und methodischen Überlegungen

Mit Bezug auf die im Rahmen dieser Arbeit unterbreiteten Vorschläge zur Durchführung von Faktorenanalysen kann als Fazit folgendes festgehalten werden:

- 1) McCrae et al. (1996) schlugen Procrustes-Verfahren als Ersatz für die KFA vor, da sie die KFA im Kontext der psychometrischen Traitforschung für wenig geeignet hielten. Raykov (1998) schlug demgegenüber vor, in der psychometrischen Traitforschung den RMSEA-Index zu verwenden. Da die Überlegungen von Raykov (1998) nicht ausreichen, um eine Auswertungsstrategie für die psychometrische Traitforschung zu fundieren und im übrigen die Orientierung an einem einzelnen Fit-Index das Risiko einer sehr einseitigen Bewertung der Datenanpassung birgt, wurde der Vorschlag von Raykov (1998) hier mit den Ergebnissen der Simulationsstudie von Hu und Bentler (1999) verknüpft (siehe Kapitel 3.2.1). Da Hu und Bentler (1999) eine Kombinationsstrategie aus RMSEA und SRMR vorschlugen, ergab sich mit dem Rückgriff auf diese Strategie für die psychometrische Traitforschung die Möglichkeit, einerseits den Argumenten von Raykov (1998) mit einer deutlichen Betonung des RMSEA zu folgen und andererseits auf eine in Simulationen fundierte Strategie zurückzugreifen. Auf dieser Basis erscheinen die Argumente von McCrae et al. (1996) weitgehend entkräftet.
- 2) Die Willkür bei der Durchführung exploratorischer Faktorenanalysen kann anhand der verfügbaren Literatur in Bezug auf die Extraktionsmethoden, Extraktionskriterien und Rotationsmethoden erheblich reduziert werden. Es ist eine methodisch begründete Auswahl leistungsfähiger Verfahren für die einzelnen Schritte möglich (Kapitel 3.2.2). Auch wenn die Auswahl nicht in allen Fällen auf eine einzige Variante konvergiert, können bestimmte Methoden, wie beispielsweise die Eigenwerte > 1 Regel (das Guttman-Kaiser Kriterium), anhand der verfügbaren Literatur eindeutig als defizitär eingestuft werden, während andere Verfahren, wie beispielsweise die Parallelanalyse, sich als besonders leistungsfähig erweisen (z.B. Velicer et al. 2000). Im Rahmen dieser Arbeit wird eine Kombination der Hauptachsenanalyse als Extraktionsmethode, der Parallelanalyse und der Komparabilitäten als Extraktionskriterien und der direkten Oblimin-Rotation (mit $\Delta=0$) bzw. der Trasid-Rotation empfohlen. Allerdings ergeben sich für die Varimax-Vorrotation im Rahmen der Trasid-Rotation aus den empirischen Analysen deutliche Einschränkungen, auf die bei der Darstellung der Befunde und Konsequenzen aus den empirischen Analysen näher eingegangen wird.

In Bezug auf die Generalitätseinschätzung von Traitdimensionen ergeben sich folgende Schlussfolgerungen:

- 2) Die Generalität von Faktoren oder Skalen hängt auch von der Generalität der Ausgangsvariablen (Items) ab (Kapitel 4.2). Daher lässt sich die Generalität von Faktoren nicht automatisch aus dem Ordnungsgrad der Faktoren und das Generalitätsniveau von Skalen nicht automatisch aus ihrem Aggregationsniveau bestimmen.
- 3) Auch Heterogenitäts- bzw. Homogenitätsmaße können nicht direkt als Generalitätsindikatoren verwendet werden. Denn eine derartige Gleichsetzung würde dazu führen, dass extrem geringe Korrelationen zwischen den Items einer Skala, die auch einen Hinweis auf Reliabilitäts- und/oder Validitätsprobleme einer Skala darstellen können, Voraussetzung für eine maximale Generalität wären (Kapitel 4.2).
- 4) Bislang lagen keine direkten Quantifizierungen der Generalität vor. Da jedoch in vergleichenden Diskussionen von Persönlichkeitsfaktoren immer wieder der Aspekt der Generalität angesprochen wurde, erschien eine Quantifizierung dieses Aspektes weiterführend. Daher wurden hier Generalitätsmaße entwickelt, die weder vom Ordnungsgrad der Faktoren bzw. vom Aggregationsniveau der Skalen noch von der Homogenität direkt abhängen. Diese Generalitätsmaße dienen zur relativen Generalitätsermittlung von Skalen und Faktoren. Sie beruhen im wesentlichen auf dem Prinzip der wechselseitigen Vorhersage bzw. Inklusion der Varianzen (Kapitel 4.3).

Die Ergebnisse der theoretischen und methodischen Analysen können teilweise bereits auf einer argumentativen Ebene als Beiträge zur Reduktion der Unverbindlichkeit der psychometrischen Traitforschung dienen. Wichtig ist jedoch, welche Aussagen sich bei einer Anwendung der o.g. methodischen Kriterien und Indizes auf einen empirischen Datensatz ergeben. Dazu wurden in der Probandengruppe A (N=554) die Fragebogen der Traitmodelle von Cattell, Eysenck sowie Costa und McCrae und in der Probandengruppe B (N=620) die Fragebogen der Traitmodelle von Eysenck, Zuckerman sowie Costa und McCrae untersucht.

7.2 Faktorenanalytische Ergebnisse der empirischen Untersuchungen

Im folgenden werden zunächst die wesentlichen Ergebnisse der modellspezifischen Analysen aufgeführt:

- 6) Zunächst konnte erfreulicherweise festgestellt werden, dass alle Faktoren erster Ordnung der hier untersuchten Inventare (EPQ-R, 16 PF-R, NEO-PI-R, SSS-V, ZKPQ) in den konfirmatorischen Analysen auf Basis des von Hu und Bentler (1999) vorgeschlagenen kombinierten Kriteriums von $SRMR \leq .10$ und $RMSEA \leq .06$ bestätigt wurden. Allerdings ergaben sich für den CFI durchgängig Werte unter $.90$. Nicht bestätigt werden konnten lediglich die Faktoren zweiter Ordnung des 16 PF-R und die Orthogonalität der fünf Faktoren zweiter Ordnung in der gewichteten Stichprobe für den NEO-PI-R. Wegen der durchgängigen Modellbestätigungen sollte in weiterführenden Studien jedoch geprüft werden, ob die Modellprüfung auf der Basis des SRMR und des RMSEA eine ausreichend strenge Prüfstrategie darstellt.

- 7) In den exploratorischen Analysen konnte die Anzahl der zu extrahierenden Faktoren nur für den ZKPQ und in der gewichteten Stichprobe für den EPQ-R anhand der Komparabilitäten bestätigt werden. Die Parallelanalyse indizierte in keinem Fall die in den Fragebogen intendierte Faktorenanzahl. Dieses Ergebnis ist insofern bemerkenswert als alle genannten Fragebogen anhand exploratorischer Faktorenanalysen entwickelt wurden und die Parallelanalyse als eins der leistungsfähigsten Extraktionskriterien gilt (Hubbard & Allen, 1987; Velicer et al., 2000; Zwick & Velicer, 1986). In Beauducel (2001b) wurde zwar eine Tendenz der Parallelanalyse zur Unterschätzung der Faktorenanzahl bei hoch korrelierenden Faktoren festgestellt. Diese Tendenz kann jedoch für die hier festgestellten Abweichungen kaum verantwortlich sein, da die Parallelanalyse in den Analysen des EPQ-R, SSS-V, ZKPQ und teilweise des 16 PF-R die Extraktion einer größeren Anzahl von Faktoren indizierte, als gemäß der intendierten Strukturen zu erwarten war. Insofern ergeben sich zwei mögliche Alternativen: Entweder weist die Parallelanalyse weitere über die in Turner (1998) und Beauducel (2001b) berichteten Probleme hinausgehende Defizite auf, oder aber die untersuchten Fragebogen enthalten tatsächlich in nennenswertem Umfang nicht intendierte Varianz, die zur Extraktion zusätzlicher Faktoren führen könnte. Letzteres könnte darauf zurückzuführen sein, dass bei der Entwicklung der untersuchten Inventare häufig die Anzahl der zu extrahierenden Faktoren gemäß den Erwartungen der Autoren vorgegeben wurde (z.B. Ruch, 1999; Schneewind & Graf, 1998; Zuckerman et al., 1978). Auch ein Vorgehen, bei dem die Entscheidung über die Anzahl der extrahierten Faktoren nicht direkt von den Extraktionskriterien abhängig gemacht wurde (z.B. Zuckerman et al., 1993), oder aber problematische Extraktionskriterien wie die Eigenwerte > 1 Regel verwendet wurden (z.B. Costa et al., 1991), könnte zu den Schwierigkeiten des Nachweises der intendierten Faktorenanzahl mit der Parallelanalyse geführt haben.

Folgende Befunde der modellübergreifenden Analysen sind festzuhalten und zu diskutieren:

- 8) Weder in den gemeinsamen Analysen der Modelle von Eysenck, Cattell sowie Costa und McCrae noch in den gemeinsamen Analysen der Modelle von Eysenck, Zuckerman sowie Costa und McCrae entsprachen die 21 bis 32 aufgrund der Parallelanalysen extrahierten Primärfaktoren durchgängig den Faktoren eines der untersuchten Modelle. Dabei war die Varianzaufklärung der Lösungen erster Ordnung mit 34.2% bis 42.5% für den Traitkontext insgesamt zufriedenstellend.

Es zeigten sich in beiden Variablengruppen mehrere modellspezifische Primärfaktoren, was darauf hindeutet, dass keines der eingesetzten Inventare die Persönlichkeitssphäre umfassend beschreibt. Es traten die Faktoren „Intelligenz“, „Empfindsamkeit“, „Privatheit“, „Offenheit für Ästhetik“, „Offenheit für Ideen“, sowie „Gefahr- und Abenteuersuche“ als modellspezifische Primärfaktoren auf. Teilweise repräsentieren diese Faktoren vermutlich auch Methodenvarianz (z.B. Intelligenz). Auch modellübergreifende Primärfaktoren wurden festgestellt: „Geselligkeit“ bzw. „Geselligkeit/Soziabilität“, „Vertrauen“, „Reizbarkeit, Ungeduld“ bzw. „Reizbarkeit/Aggressivität“, „Neurotizismus“, „Planen“, „Benehmen“, „Bescheidenheit/Kompetenz (minus)“ und „Gefühlsbetontheit“. Diese Faktoren integrieren Items verschiedener Inventare. Obgleich die Anzahl der modellübergreifenden Primärfaktoren etwas größer ist als die Anzahl der modellspezifischen Primärfaktoren, deutet dieses Ergebnis darauf hin, dass die Fragebogen der verschiedenen Modelle nur teilweise überlappende Persönlichkeitstrait erfassen.

- 9) Das vielleicht bedeutendste Ergebnis dieser Analysen besteht in den geringen Interkorrelationen der Primärfaktoren. In keiner der Oblimin- und Trasid-rotierten Faktorenlösungen erster Ordnung waren mehr als 4% der Korrelationsbeträge über .20. Somit konnten die Faktoren zweiter Ordnung in allen Analysen auch nur geringe Varianzanteile der Primärfaktoren repräsentieren. Dies kann anhand der erklärten Varianzen für die Variablengruppe aus EPQ-R, 16 PF-R und NEO-PI-R beispielhaft verdeutlicht werden: Die 21 Primärfaktoren der ungewichteten Stichprobe erklärten insgesamt 34.2% an der Gesamtvarianz. 23.6% der 34.2% wurden wiederum von den sechs Faktoren zweiter Ordnung anhand der Oblimin-rotierten Lösung erklärt. Der prozentuale Anteil der von den Faktoren zweiter Ordnung erklärten Varianz an der Gesamtvarianz ergibt sich aus $23.6 \cdot 34.2 / 100 = 8.1\%$. Der prozentuale Anteil der auf der Basis der Trasid-rotierten 21 Primärfaktoren ermittelten Faktoren zweiter Ordnung an der Gesamtvarianz betrug $24.4 \cdot 34.2 / 100 = 8.3\%$. In der gewichteten Stichprobe erklärten die 32 Primärfaktoren 42.4% der Gesamtvarianz. Von den Oblimin-rotierten 32 Primärfaktoren erklärten die fünf Faktoren zweiter Ordnung 16.1% der Varianz. Somit erklärten die fünf Faktoren zweiter Ordnung insgesamt $16.1 \cdot 42.4 / 100 = 6.8\%$ der Gesamtvarianz. Für den Variablensatz aus EPQ-R, NEO-PI-R, SSS-V und ZKPQ

lagen die entsprechenden Varianzanteile der Faktoren zweiter Ordnung zwischen 7.0% und 8.4%.

Somit war für beide Rotationsmethoden, sowohl in der gewichteten als auch in der ungewichteten Stichprobe, in beiden Variablengruppen der Anteil erklärter Varianz an der Gesamtvarianz äußerst gering. Nun mag man einwenden, dass die geringe Reliabilität der Items auch zu geringen erklärten Varianzen führen muss. Andererseits reflektieren die 34.2% bzw. 35.5% der durch die Primärfaktoren in den ungewichteten Stichproben erklärten Varianzen und die 42.4% bzw. 42.5% der durch die Primärfaktoren in den gewichteten Stichproben erklärten Varianzen mit hoher Wahrscheinlichkeit einen reliablen Varianzanteil. Wenn von diesem reliablen Varianzanteil wiederum nur 16-24% der Varianz durch die Faktoren zweiter Ordnung erklärt wird, spricht dies allein bereits für eine relativ geringe Relevanz von Faktoren zweiter Ordnung. An dieser Stelle sei darauf hingewiesen, dass die über die interne Konsistenz der in den Fragebogen intendierten Skalen geschätzten Reliabilitäten in diesem Datensatz akzeptabel sind (s. Kapitel 6.1.5), so dass Besonderheiten bezüglich der Reliabilität der Daten als Erklärungsgrund für die geringen Interkorrelationen der Primärfaktoren und die damit einhergehende geringe Varianzaufklärung der Sekundärfaktoren ausscheiden.

Die geringen Korrelationen zwischen den Primärfaktoren erscheinen unter anderem auch deswegen bemerkenswert, weil die meisten hier analysierten Fragebogen ähnliche Traitmodelle repräsentieren. Der einzige hier eingesetzte Fragebogen, der vermutlich speziellere Varianzen abbildet, war die SSS-V. Somit konnte von einem hohen Maß an Redundanz ausgegangen werden. Offenbar ist jedoch die Redundanz der Fragebogen und damit vermutlich auch die Redundanz der Traitsysteme von Cattell, Costa und McCrae, Eysenck und Zuckerman geringer als erwartet.

Darüber hinaus relativiert die geringe Varianzaufklärung der Faktoren zweiter Ordnung die Bedeutung derjenigen Modelle, die besonders auf diese Faktorebene fokussieren. Möglicherweise ist die geringe Varianzaufklärung der Faktoren zweiter Ordnung der Grund dafür, dass so selten hierarchische, aufsteigende Faktorenanalysen durchgeführt werden und dass regelmäßig versucht wird, Faktoren, die unter dem Gesichtspunkt der Generalität als Faktoren zweiter Ordnung gelten sollen, als Faktoren erster Ordnung zu erfassen (Costa & McCrae, 1992a; Eysenck & Eysenck, 1991; Zuckerman et al., 1993). Allein Cattell (1972, 1994a) hielt strikt an der Durchführung hierarchischer, aufsteigender Faktorenanalysen fest, wobei er leider häufig spezifische Aggregationsmethoden verwendete, die die Nachvollziehbarkeit seiner Lösungen erschweren (siehe z.B. Cattell, 1994b). Dass vor allem Cattell (1972, 1994a) die vergleichsweise größere Relevanz von Primärfaktoren

betonte, mag damit zusammenhängen, dass er zu den wenigen Persönlichkeitsforschern zählte, die konsequent aufsteigende Faktorenanalysen durchgeführt haben. Möglicherweise konnten Eysenck, Costa und McCrae sowie Zuckerman die vergleichsweise geringere Relevanz der Faktoren zweiter Ordnung nicht feststellen, da sie keine aufsteigenden Faktorenanalysen durchführten. In Bezug auf Zuckerman (1994) kann festgehalten werden, dass er in seiner experimentellen und biopsychologischen Forschung besonders häufig Befunde für die vier Subskalen der SSS-V berichtet, worin eine Betonung der Primärfaktoren deutlich wird. Die Bedeutung der spezielleren Sensation-Seeking Subskalen zeigt sich auch in Brocke et al. (1999, 2000). Aber auch in biopsychologischen Untersuchungen zu anderen Merkmalen konnten stärkere Effekte für die spezifischeren Dimensionen nachgewiesen werden: So stellte Debener (2001) in seinen Untersuchungen zur Alpha-Asymmetrie im EEG stärkere Effekte für eine spezifischere Depressivitätsskala im Vergleich zu den Effekten generellerer Emotionsfaktoren (mit Traitcharakter) fest.

Möglicherweise kann auch die Eingangs problematisierte Perpetuierung unterschiedlicher Traitmodelle mit der geringen Varianzaufklärung der Sekundärfaktoren in Verbindung gebracht werden: Wenn die Faktoren zweiter Ordnung nur rund ein Viertel der Varianz der Primärfaktoren erklären, dann können leicht geringfügige Verschiebungen in den erklärten bzw. nicht-erklärten Varianzanteilen auftreten. Anders ausgedrückt: Die Anzahl der substantiell korrelierenden Primärfaktoren scheint sehr begrenzt zu sein. Dies kann dazu führen, dass das Weglassen einzelner Primärfaktoren, bereits zu deutlichen Verschiebungen bei den Faktorenanalysen zweiter Ordnung führt. Da insgesamt nur wenige substantielle Zusammenhänge zwischen Primärfaktoren auftreten, kann die geringfügige aber wiederholte Einbeziehung oder Eliminierung einzelner Zusammenhänge wiederholt zu Unterschieden bei den Analysen der Faktoren zweiter Ordnung führen. Wenn dieser Effekt mit den in Kapitel 3 aufgezeigten methodologischen Unverbindlichkeiten interagiert, erscheint die systematische Koexistenz verschiedener Modelle wahrscheinlicher als die empirische Integration in ein Gesamtmodell. Vermutlich hat eine frühzeitige Fokussierung auf Faktoren höherer Ordnung die Erarbeitung einer tragfähigen Integration von Traitmodellen erschwert.

Nicht ohne Ironie kann man darüber hinaus feststellen, dass der Betrag der durch die Faktoren zweiter Ordnung erklärten Varianz ziemlich genau in dem Bereich liegt, der von Mischel (1968) als Obergrenze für Persönlichkeitsfaktoren angegeben wurde. Mischel (1968) behauptete, dass Korrelationen, die die transsituationale Konsistenz von Traits angeben, kaum über .30 liegen. Da sich der Persönlichkeitskoeffizient auf die transsituationale Konsistenz bezieht, wäre es jedoch angemessener, nicht die erklärte Varianz ganzer Faktorenlösungen mit dem Persönlichkeitskoeffizienten zu vergleichen,

sondern die durchschnittliche erklärte Varianz der einzelnen Faktoren der Lösungen. Die Faktoren erklären ja unabhängige Varianzanteile, weshalb die Summe der erklärten Varianzen über die Faktoren nicht die durchschnittliche transsituationale Konsistenz repräsentiert. Der Mittelwert der erklärten Varianz der Faktoren indiziert jedoch deren durchschnittliche transsituationale Konsistenz in Bezug auf die Items als unterschiedliche Messgelegenheiten. Da die Lösungen zweiter Ordnung hier meist fünf Faktoren umfassen und etwa 8% Varianz erklären, ergibt sich für die durchschnittliche Varianz, die ein Faktor zweiter Ordnung an den Items erklärt, etwa ein Betrag von 1.6%. Dies liegt deutlich unter der von Mischel angegebenen Obergrenze. Um die ironische Betrachtung weiter fortzusetzen, könnte dieser Wert dahingehend interpretiert werden, dass Mischels Einschätzung der transsituationalen Konsistenz von Traits zu optimistisch war.

Es ergibt sich jedoch auch eine andere Variante der Bewertung dieses Aspektes: Möglicherweise ist die Interaktionismus-Debatte zu sehr aus der Perspektive der als hochgradig generell konzipierten Persönlichkeitsdimensionen geführt worden. Jedenfalls betonten Mischel (1968) und Mischel und Peake (1982) die Bewertung der Gesamtkonsistenz des Verhaltens. Demgegenüber trat die Bedeutung der Kombination einer größeren Anzahl moderat transsituational konsistenter Persönlichkeitsmerkmale vermutlich in den Hintergrund. Vielleicht wären der Forschung lange, teilweise ideologisch anmutende Diskussionen erspart geblieben, wenn man von vornherein stärker auf Traits moderater Generalität, wie beispielsweise Primärfaktoren, fokussiert hätte. Dabei liegt auch für die Lösungen erster Ordnung die durchschnittliche Varianzaufklärung der Faktoren unter 2% und damit deutlich unter der von Mischel (1968) angegebenen Obergrenze für die transsituationale Konsistenz von Persönlichkeitsmaßen. Wenn die transsituationalen Konsistenzen einzelner Persönlichkeitsdimensionen im Durchschnitt auf derartig geringe Werte begrenzt sind, liegt es jedenfalls nahe, eine größere Zahl moderat konsistenter Merkmale zur Verhaltensvorhersage zu verwenden. Umgekehrt könnte man auch argumentieren, dass der Nachweis größerer transsituationaler Konsistenzen als in Mischel und Peake (1982) mit den Ergebnissen der psychometrischen Traitforschung im Widerspruch gestanden hätte. Dies kann als ein weiterer Hinweis darauf gelten, dass die Interaktionismus-Debatten nicht immer ausreichend fundiert geführt wurden.

Im Einklang mit der Betonung der Primärfaktoren berichteten Mershon und Gorsuch (1988) sowie Schneider, Hough und Dunnette (1996) höhere Kriteriumsvaliditäten für eine größere Zahl spezifischerer Traitdimensionen. Allerdings sollte die Perspektive, die sich aus der Berücksichtigung der Brunswik-Symmetrie der Aggregations-Niveaus von Prädiktoren und Kriterien ergibt (Wittmann, 1988; Wittmann & Süß, 1999), dazu beitragen, hier von einseitigen Wertungen der Nützlichkeit der Generalität Abstand zu nehmen. Aus dieser

Perspektive folgt, dass Generalität an sich nicht notwendig ein Qualitätsmerkmal darstellt. Dies gilt auch, wenn, wie bereits ausgeführt, die Beziehung zwischen Aggregationsniveau und Generalität von vielen Faktoren beeinflusst wird.

- 10) Auf der Ebene der Faktoren zweiter Ordnung konnten in beiden Analysen die modellübergreifenden Faktoren „Extraversion“ und „Neurotizismus“ repliziert werden. Die Replikation dieser Faktoren konnte natürlich mit den hier eingesetzten Inventaren erwartet werden. Der Wert derartiger Replikationen von Faktoren in einer Vielzahl von Untersuchungen (z.B. Amelang & Borkenau, 1982; Norman, 1967; Goldberg, 1990; Costa & McCrae, 1992a) sollte jedoch nicht am geringen Neuheitswert der Information festgemacht werden. Vielmehr stellen über Jahrzehnte wiederholt replizierte Befunde der Differentiellen Psychologie die wesentliche Grundlage für diagnostische Anwendungen dar.

Auch für die Faktoren „Verträglichkeit“ und „Psychotizismus“ gab es deutliche Hinweise. Faktoren, die Varianzquellen im Bereich „Gewissenhaftigkeit“ und „Offenheit für Erfahrungen“ abbilden, konnten mitunter nachgewiesen werden, allerdings waren die Ergebnisse hier inkonsistent. In Bezug auf die Repräsentativität der Traitmodelle kann somit festgehalten werden, dass die drei Dimensionen von Eysencks Persönlichkeitstheorie in den meisten Analysen deutlich wurden, wobei der „Psychotizismus“-Faktor teilweise stark mit Aspekten von „Verträglichkeit (minus)“ zusammenhing. Zugleich ergab sich aber auch, dass Extraversion, Neurotizismus und Psychotizismus nicht ausreichen, um die systematische Varianz in den Datensätzen zu erklären. Es konnten durchgängig Faktoren in den Bereichen Gewissenhaftigkeit und Offenheit für Erfahrungen aufgezeigt werden. Dabei ergaben sich für diese Faktoren jedoch unterschiedliche Strukturierungen. Die Ergebnisse sprechen insgesamt dafür, dass mit dem Fünf-Faktoren-Modell der Persönlichkeit relevante, über das Modell von Eysenck hinausgehende, Varianzquellen erschlossen werden, obgleich die Frage nach der Strukturierung dieser Varianzquellen auf der Basis der bisherigen Befunde nicht als entschieden angesehen werden kann. Auch verdrängen diese zusätzlichen Varianzquellen nicht den Psychotizismus-Faktor, was der Vorstellung von Eysenck (1992), nach der sich Psychotizismus aus der Zusammenfassung von Verträglichkeit und Gewissenhaftigkeit ergibt, widerspricht. Varianzanteile von „Verträglichkeit (minus)“ und „Gewissenhaftigkeit (minus)“ konstituierten zwar durchaus einen „Psychotizismus“-Faktor, aber es verblieben noch Varianzanteile von Gewissenhaftigkeit und Verträglichkeit, die offenbar nicht unter Psychotizismus subsumiert werden können.

Neben den Faktoren „Extraversion“ und „Neurotizismus“ konnten keine weiteren mit den Faktoren des 16 PF-R direkt kompatiblen Faktoren zweiter Ordnung aufgezeigt werden. Auch das Alternative Fünf-Faktorenmodell setzte sich in den Analysen höherer Ordnung nicht durch, obgleich der Faktor „Offenheit für Erfahrungen“ in deutlichem Umfang Aspekte von Sensation Seeking integrierte. Insofern setzten sich das Persönlichkeitsmodell von Eysenck und das Fünf-Faktoren-Modell der Persönlichkeit gegenüber dem Modell von Cattell (in der mit dem 16 PF-R erfassten Form) und dem Alternativen Fünf-Faktoren-Modell auf der Ebene der Faktoren zweiter Ordnung durch.

Die Überlegenheit der Strukturierungen i.S. von Eysencks Persönlichkeitsmodell und dem Fünf-Faktorenmodell gegenüber dem Cattellschen Modell und dem Alternativen Fünf-Faktorenmodell bezieht sich allein auf die Faktorenanalysen zweiter Ordnung. Jedoch wird die meiste persönlichkeitsrelevante Varianz von den Faktoren erster Ordnung repräsentiert, bei denen keine eindeutige Überlegenheit eines der untersuchten Persönlichkeitsmodelle festgestellt werden konnte. Vielmehr wiesen auf dieser Ebene alle Fragebogen bis auf den EPQ-R modellspezifische Varianzen auf, was dafür spricht, dass sie nicht ohne weiteres durch die Fragebogen der jeweils anderen Modelle substituiert werden können.

Darüber hinaus sprechen die geringen Interkorrelationen zwischen den Sekundärfaktoren insgesamt gegen eine Etablierung von Faktoren dritter Ordnung, wie sie beispielsweise Digman (1997) vorschlägt. In dem einen Fall, in dem Faktoren dritter Ordnung sinnvoll berechnet werden konnten (siehe Anhang, Tabelle A 6.31), entsprachen diese nicht den von Digman (1997) ermittelten Faktoren.

- 11) Aus methodischer Perspektive ist festzuhalten, dass sich in den Lösungen erster Ordnung mit 21 bis 32 Faktoren die Hauptladungen in den Oblimin-Lösungen deutlich günstiger auf die Faktoren verteilten als in den Varimax-Lösungen bzw. den auf einer Varimax-Vorrotation aufbauenden Trasid-Lösungen. Dies kann als ein Hinweis darauf gelten, dass die Sättigung der Faktoren nur in Abhängigkeit von der Rotationsmethode interpretiert werden sollte. Dieser Befund relativiert die Bemühungen, über die Anzahl der Hauptladungen pro Faktor die Anzahl der zu extrahierenden Faktoren zu bestimmen. Allerdings sollte auch die spezifische Bedeutung dieser Befunde für die Faktorenrotation nicht vernachlässigt werden: Die Varimax-Vorrotation, wie sie bei Promax und Trasid durchgeführt wird, sollte bei großen Faktorenzahlen durch eine andere Vorrotation ersetzt werden. Nach der in Kapitel 3.2.2.3 diskutierten Literatur könnte sich eine Kombination der Trasid-Rotation mit einer Parsimax-Vorrotation in Zukunft als besonders geeignete Methode zur Erreichung optimaler Einfachstruktur erweisen. Im weiteren sollte bei hierarchischen Lösungen im Bereich der Traitforschung auch der Nutzen der Schmid-Leiman (1957) Transformation

exploriert werden, obgleich die geringen Varianzaufklärungen der Faktoren höherer Ordnung dabei nur geringe Verschiebungen der Lösungen erwarten lassen.

Ein weiterer Hinweis auf einen deutlichen Einfluss der Rotationsmethode auf die Ergebnisse ergibt sich daraus, dass viele Faktoren in den Oblimin- und Trasid-rotierten Lösungen unterschiedlich interpretiert werden mussten (z.B. konnten im Variablensatz aus EPQ-R, 16 PF-R und NEO-PI-R in der ungewichteten Stichprobe nur acht der 21 Faktoren erster Ordnung und vier der sechs Faktoren zweiter Ordnung äußerst ähnlich interpretiert werden). Dieser beachtliche Einfluss der Rotationsmethode kann mit der großen Zahl von Faktoren und Variablen, also mit der insgesamt großen Ladungsmatrix (z.B. für den Variablensatz aus EPQ-R, 16 PF-R und NEO-PI-R: 21 Faktoren x 494 Variablen = 10374 Faktorladungen bzw. 32 Faktoren x 494 Variablen = 15808 Faktorladungen) zusammenhängen, da die Komplexität der Ladungsmatrix dazu führen kann, dass kleine Unterschiede in den optimierten Rotationskriterien zu großen Unterschieden im Ergebnis führen. Der Befund von Goldberg (1990), dass die Rotationsmethode einen geringen Einfluss auf die Ergebnisse hat, wurde anhand einer um ein Vielfaches kleineren Ladungsmatrix (5 Faktoren x 75 Variablen = 375 Ladungen) gewonnen. Insofern deutet dieser Befund darauf hin, dass – vor allem bei Analysen mit umfangreichen Variablen und Faktorenzahlen – stärker als bisher die genaue theoretische Bedeutung der jeweils in den verschiedenen Rotationsmethoden operationalisierten Konzepte von Einfachstruktur reflektiert werden sollte (z.B. Beauducel, 1996, 1997; Schweizer, 2001).

- 12) Häufig beschränkten sich psychometrische Traitstudien auf studentische Stichproben (z.B. Benet-Martinez & Waller, 1997; Carver & White, 1994; Cattell & Delhees, 1973; Church & Burke, 1994; Goldberg, 1990). In der vorliegenden Arbeit konnte jedoch ein deutlicher Einfluss der Stichprobe auf die Anzahl der zu extrahierenden Faktoren und somit auf die gesamte hierarchische Analyse nachgewiesen werden. Durch die Gewichtung der Stichprobe wurden im Variablensatz aus EPQ-R, 16 PF-R und NEO-PI-R gemäß der Parallelanalyse 11 Faktoren mehr extrahiert als in der ungewichteten Stichprobe. Im Variablensatz aus EPQ-R, NEO-PI-R, SSS-V und ZKPQ wurden anhand der Parallelanalyse in der gewichteten Stichprobe 9 Faktoren mehr extrahiert als in der ungewichteten Stichprobe. Dies ist ein deutlicher Hinweis darauf, dass auf die Stichprobe als Einflussfaktor gerade auch bei psychometrischen Analysen im Persönlichkeitsbereich deutlicher als bisher geachtet werden sollte.

7.3 Ergebnisse des empirischen Generalitätsvergleichs

Die Ergebnisse zu den Inklusions-Indizes haben eher exemplarischen Charakter, da sie im Rahmen dieser Arbeit auf die Faktoren Neurotizismus und Extraversion beschränkt wurden. Die Berechnungen wurden aus Aufwandsgründen nur für die gewichteten Stichproben der Gruppen A und B durchgeführt.

- 13) Zunächst wurde die Generalität der Sekundärfaktoren „Neurotizismus“ und „Extraversion“ mit der Generalität der jeweils vier am höchsten auf den Sekundärfaktoren ladenden Primärfaktoren verglichen (siehe Kapitel 6.4.1). Dies diente vor allem dem Zweck, erste Hinweise auf die Validität der Inklusions-Indizes zu gewinnen. Für den „Neurotizismus“-Faktor zeigten die Inklusions-Indizes I_{f2} , I_{f3} und I_{f4} durchgängig eine höhere Generalität des Faktors zweiter Ordnung an. Lediglich der Index I_{f5} zeigte in der Gruppe B eine höhere Generalität für zwei der vier Primärfaktoren an. Beim Sekundärfaktor „Extraversion“ deuteten alle berechneten Inklusions-Indizes (I_{f2} bis I_{f5}) auf eine höhere Generalität des Sekundärfaktors im Vergleich zu den Primärfaktoren hin. Die Ergebnisse werden als Hinweis auf die Validität der Inklusions-Indizes I_{f2} , I_{f3} und I_{f4} gewertet. Der Inklusions-Index I_{f5} sollte auf der Basis der vorliegenden Ergebnisse mit größerer Vorsicht interpretiert werden. Insgesamt sind auch diese Ergebnisse mit der Einschränkung zu versehen, dass die Ähnlichkeit der Generalität der Primärfaktoren als Voraussetzung für eine höhere Generalität der Sekundärfaktoren im Vergleich zu den Primärfaktoren nicht geprüft werden konnte. Allerdings ergeben sich aus den Ladungsmustern Hinweise auf die Richtigkeit dieser Vermutung (siehe Kapitel 6.4.1). Im übrigen erlauben die vorliegenden Ergebnisse keine weitergehende Beurteilung der Validität der Indizes. Diese könnte im Rahmen weiterführender Simulationsstudien erfolgen.
- 14) Als nächstes wurden die modellspezifischen Neurotizismus-Faktoren des EPQ-R, NEO-PI-R und ZKPQ verglichen (siehe Kapitel 6.4.2). Gemäß aller hier ermittelten Inklusions-Indizes (I_{f2} bis I_{f5}) wies der Neurotizismus-Faktor des NEO-PI-R die höchste Generalität im Vergleich zu den Neurotizismus-Faktoren des EPQ-R und ZKPQ auf, wobei die Neurotizismus-Faktoren des EPQ-R und ZKPQ eine ähnliche Generalität aufwiesen. Einschränkend muss allerdings festgehalten werden, dass der Neurotizismus-Faktor des NEO-PI-R in nennenswertem Umfang Nebenladungen der Markiertvariablen anderer Faktoren aufweist, die möglicherweise zur höheren Generalität dieses Faktors beitragen.
- 15) Ebenso wurden die modellspezifischen Extraversionen-Faktoren des EPQ-R und NEO-PI-R sowie die Faktoren Aktivität und Soziabilität des ZKPQ verglichen (siehe Kapitel 6.4.3). Dabei ergaben sich für die Inklusions-Indizes I_{f2} und I_{f3} andere Generalitätsunterschiede als für die Inklusions-Indizes I_{f4} und I_{f5} . Die Inklusions-Indizes I_{f2} und I_{f3} deuten darauf hin, dass der Extraversionen-Faktor des EPQ-R die höchste Generalität aufweist, gefolgt von dem Extraver-

sions-Faktor des NEO-PI-R. Die Faktoren Aktivität und Soziabilität weisen gemäß I_{f2} und I_{f3} die geringste Generalität auf. Demgegenüber weist der Faktor Extraversion des NEO-PI-R gemäß I_{f4} und I_{f5} die höchste Generalität auf, gefolgt von dem Faktor Aktivität des ZKPQ. Die Extraversion des EPQ-R weist gemäß I_{f4} und I_{f5} geringere Generalität auf und der Faktor Soziabilität des ZKPQ weist die geringste Generalität auf.

Die Unterschiede zwischen den Generalitätsschätzungen I_{f2} und I_{f3} im Vergleich zu I_{f4} und I_{f5} hängen vermutlich damit zusammen, dass I_{f2} und I_{f3} auf die wechselseitig erklärten Varianzen fokussieren, während I_{f4} und I_{f5} daneben auch die Verteilung der erklärten Varianzen einbeziehen. Es sei auch nochmals darauf verwiesen, dass der Index I_{f5} mit besonderer Vorsicht zu interpretieren ist (siehe Punkt 13). Möglicherweise hängt die höhere Generalität der Extraversion des NEO-PI-R in den Indizes I_{f4} und I_{f5} damit zusammen, dass für den Extraversions-Faktor im NEO-PI-R die sechs Facetten-Skalen zugrunde gelegt wurden. Allerdings kann auf der Basis der vorliegenden Ergebnisse nicht geklärt werden, ob die höhere Generalität des Extraversions-Faktors des NEO-PI-R auf den Effekt eines höheren Aggregationsniveaus durch die Verwendung der Facetten-Skalen zurückgeführt werden kann. Dieser Aspekt wäre in weiterführenden Analysen auf der Basis der NEO-PI-R Items zu klären. Im übrigen deuten die Unterschiede zwischen den Indizes darauf hin, dass verschiedene Aspekte von Generalität unterschieden werden können. Auch eine mögliche Differenzierung verschiedener Generalitätskonzepte müsste in weiterführenden theoretischen und methodischen Untersuchungen eingehender geklärt werden. Die bisherige Verwendung des Generalitätsbegriffs in der Literatur lässt demgegenüber eher auf eine eindimensionale Generalitäts-Konzeption schließen.

Aus der inhaltlichen Perspektive kann festgehalten werden, dass der Faktor Soziabilität im ZKPQ gemäß aller hier ermittelten Generalitäts-Indizes zu den Faktoren zählt, die die geringste Generalität aufweisen. Da die Indizes verschiedene Aspekte umgreifen, ist die Generalität des Soziabilitäts-Faktors des ZKPQ aus verschiedenen Perspektiven als gering einzuschätzen. Dies deutet darauf hin, dass die Soziabilität nicht als Faktor angesehen werden kann, der gleichberechtigt neben den Faktoren des FFM und den Faktoren in Eysencks Persönlichkeitsmodell steht.

- 16) Beim Generalitätsvergleich der hier ermittelten „Neurotizismus“-Faktoren zweiter Ordnung mit den modellspezifischen Neurotizismus-Faktoren des EPQ-R und NEO-PI-R deuteten I_{f2} und I_{f3} auf eine höhere Generalität der modellspezifischen Faktoren, während I_{f4} und I_{f5} auf eine höhere Generalität der Sekundärfaktoren hinwiesen. Dieser Befund ist auf die geringe Varianzaufklärung der Faktoren zweiter Ordnung im Vergleich zu den modellspezifischen Primärfaktoren zurückzuführen (siehe Punkt 9). Da die Inklusions-Indizes I_{f2} und I_{f3} die

wechselseitig erklärte Varianz repräsentieren, erscheint es naheliegend, dass die Faktoren zweiter Ordnung, die in deutlich geringerem Umfang Varianz erklären als die modellspezifischen Primärfaktoren, hier eine geringere Generalitätsschätzung aufweisen. Andererseits ist die Verteilung der Korrelationen der „Neurotizismus“-Faktoren zweiter Ordnung mit den Items offenbar breitgipfliger als die der modellspezifischen Primärfaktoren, was zu der höheren Generalitätsschätzung der Sekundärfaktoren in I_{f4} und I_{f5} führt.

- 17) Beim Generalitätsvergleich der „Extraversions“-Faktoren zweiter Ordnung mit den modellspezifischen Extraversionsfaktoren des EPQ-R und NEO-PI-R deuteten die Inklusions-Indizes I_{f2} und I_{f3} durchgängig auf eine höhere Generalität der modellspezifischen Faktoren des EPQ-R und NEO-PI-R hin. Auch der Index I_{f5} wies überwiegend auf eine höhere Generalität der modellspezifischen Extraversionsfaktoren hin, wobei dieser Index mit Zurückhaltung interpretiert werden sollte. Allein der I_{f4} -Index deutete auf eine höhere Generalität des „Extraversions“-Faktors zweiter Ordnung im Vergleich zum Extraversions-Faktor des EPQ-R. Zugleich deutete I_{f4} auf eine höhere Generalität des Extraversions-Faktors des NEO-PI-R im Vergleich zum Sekundärfaktor hin. Die Tatsache, dass der Index I_{f4} auf eine höhere Generalität des modellspezifischen Extraversions-Faktors des NEO-PI-R hindeutet, kann nicht allein auf das geringe Ausmaß erklärter Varianz der Sekundärfaktoren zurückgeführt werden, da I_{f4} nicht allein die wechselseitig erklärte Varianz repräsentiert. Eine Erklärung für dieses Ergebnis könnte sich allerdings aus der hohen faktoriellen Komplexität der Facetten-Skalen bzw. der Items des Extraversions-Faktors im NEO-PI-R ergeben. Eine Inspektion der Ladungsmatrizen der Faktorenanalysen zum NEO-PI-R zeigt, dass sowohl die Facetten-Skalen (siehe Anhang, Tabelle A 6.39) als auch die Items im Bereich Extraversion viele substantielle Nebenladungen aufweisen (siehe Anhang, Tabellen A 6.40 und A 6.41). Da die Extraktionskriterien in den modellspezifischen Analysen des NEO-PI-R nicht auf die Extraktion von fünf Faktoren hinwiesen (siehe Kapitel 6.2), war eine suboptimale Einfachstruktur der Fünf-Faktorenlösungen zu erwarten. Demgegenüber sind die Nebenladungen der Extraversions-Items im EPQ-R deutlich geringer (siehe Anhang, Tabelle A 6.38). Es kann somit nicht ausgeschlossen werden, dass die hohe Generalitätseinschätzung des „Extraversions“-Faktors des NEO-PI-R durch die hohe faktorielle Komplexität der Facetten-Skalen bzw. Items dieses Faktors zustande kommt.

Im vorliegenden Fall konnten die Nebenladungen der Extraversions-Facetten bzw. -Items des NEO-PI-R nicht als modellkonform gewertet werden, weshalb die Ergebnisse der Generalitätsvergleiche hier nur als erste vorläufige Hinweise aufgefasst werden können. Dieses Ergebnis verweist auf ein grundsätzliches Problem der Generalitätsschätzung. In Kapitel 4.3.2 wurde bereits auf das Erfordernis hingewiesen, dass die Faktoren, für die ein Generalitätsvergleich durchgeführt wird, modellkonform sein sollten. Modellkonformität ist jedoch ein graduelles

Phänomen und auch bei nennenswerten Nebenladungen werden Faktorenstrukturen mitunter noch als modellkonform angesehen. Dabei ist nicht auszuschließen, dass gerade die faktoriell unreinen Variablen zur Generalität eines Faktors beitragen. Man könnte derartige Verzerrungen beim Generalitätsvergleich jedoch durch die Entwicklung von Indizes auffangen, die die Abweichung von der Einfachstruktur einer Lösung quantifizieren. Dabei stellt sich jedoch die Frage nach der intendierten Varianz. In manchen Ansätzen wird die faktorielle Komplexität von Variablen bzw. eine geringe Einfachstruktur durchaus zugelassen (z.B. Hofstee et al., 1992).

7.4 Grenzen dieser Arbeit, verbleibendes Arbeitsprogramm und Perspektiven

Als wesentliche Begrenzung dieser Arbeit ist zunächst festzuhalten, dass die vorstehenden exploratorischen Faktorenanalysen im Einklang mit gängigen Kriterien (z.B. Brocke, 2000; Eysenck, 1991; Holz-Ebeling, 1995) zunächst der Replikation bedürfen. Allerdings sind einige der hier berichteten Ergebnisse bereits im Rahmen dieser Arbeit repliziert worden. Dies betrifft diejenigen Ergebnisse, die über die beiden Variablensätze (EPQ-R, 16 PF-R und NEO-PI-R sowie EPQ-R, ZKPQ, SSS-V und NEO-PI-R) replizierbar und damit partiell sicher auch generalisierbar sind. Für diese Ergebnisse gilt, dass sie nicht nur in unabhängigen Probandenstichproben, sondern auch in leicht unterschiedlichen Variablengruppen nachweisbar waren. Zu den wichtigsten in diesem Sinne replizierbaren Ergebnissen zählen folgende: Die äußerst geringen Interkorrelationen zwischen den Primärfaktoren und die damit verbundene geringe Varianzaufklärung der Sekundärfaktoren (Punkt 9). Die Replikation von Extraversion und Neurotizismus, der Nachweis von Psychotizismus-ähnlichen Faktoren sowie der Nachweis von über diese Faktoren hinausgehender Varianz, die jedoch in den vorliegenden Analysen nicht eindeutig strukturiert werden konnte (Punkt 10). Der nennenswerte Einfluss der Rotationsmethode, verbunden mit den Problemen der Varimax-Rotation bei großen Faktorenzahlen konnte in beiden Stichproben nachgewiesen werden (Punkt 11). Auch der erhebliche Einfluss der Gleich-Gewichtung der Stichprobe in Bezug auf Geschlecht und Alter auf die Anzahl extrahierter Faktoren war in beiden Stichproben nachweisbar (Punkt 12). Überdies wurden die Generalitätsvergleiche hier auf die Extraversions- und Neurotizismus-Faktoren beschränkt, die in beiden Stichproben nachgewiesen werden konnten. Da die Generalitätsvergleiche in beiden Stichproben durchgeführt wurden und da alle Generalitätsunterschiede auf der Basis der Inklusion zwischen Faktoren in beiden Stichproben für die Indizes I_{12} , I_{13} und I_{14} gleich ausfielen, können auch diese Befunde als replizierbar eingestuft werden. Den eindeutigsten, in diesem Sinne replizierbaren Befund stellt die besonders hohe Generalität des Neurotizismus-Faktors im NEO-PI-R dar.

Es ergeben sich allerdings Einschränkungen der Interpretierbarkeit der Befunde zur Generalität: Die wesentliche Einschränkung ist, dass, sowohl bei den Generalitätsvergleichen zwischen den modellspezifischen Extraversions- und Neurotizismus-Faktoren als auch bei den Vergleichen der Sekundärfaktoren mit den modellspezifischen Faktoren, die modellspezifischen Faktoren nach den in Kapitel 3 erarbeiteten Kriterien für die Durchführung der Faktorenanalyse nicht als bestätigt gelten konnten. Insofern muss bei der Interpretation der Generalitätsunterschiede der Einfluss der geringen Einfachstruktur der NEO-PI-R Neurotizismus- und Extraversions-Faktoren berücksichtigt werden. Es erscheint problematisch, dass hohe Generalität möglicherweise auch durch Unschärfen bei der Konzeption von Persönlichkeitstraits erreicht werden kann. Dies ist nicht nur ein Problem der Generalitäts-Indizes, sondern hängt damit zusammen, dass hohe Generalität eines Traits dazu führen kann, dass der Trait nur zu wenigen anderen Traits diskriminante Validitäten aufweist. Insofern besteht die Gefahr, dass der Versuch, möglichst generelle Traits zu entwickeln, zu einer Vernachlässigung der diskriminanten Validitäten führt bzw. dass hohe Generalität auf Kosten geringer diskriminanter Validität angestrebt wird. Der Einfluss der Einfachstruktur der Lösungen auf die Generalitätsschätzungen sollte somit in weiterführenden Studien untersucht werden, um zu prüfen, ob regelmäßig eine negative Kovariation zwischen Einfachstruktur und Generalität besteht.

Eine weitere Einschränkung bei der Interpretation der Generalitäts-Indizes ergibt sich aus den Fragen der Parametrisierung dieser Indizes bzw. der Verteilung. Die Verteilungen der komplexen Indizes I_{f2} , I_{f3} , I_{f4} und I_{f5} sind von einer Vielzahl von Aspekten abhängig: Von der Variablenanzahl in beiden zu vergleichenden Faktoren, vom Verhältnis der Variablenanzahl in beiden zu vergleichenden Faktoren, von der Extraktions- und Rotationsmethode, der Höhe der Ladungen, dem Grad der Einfachstruktur sowie der Anzahl der Probanden. Diese große Zahl von Abhängigkeiten ist nicht als inhaltlicher Nachteil dieser Indizes anzusehen, da davon ausgegangen werden kann, dass die Generalität von Faktoren ebenfalls von den genannten Merkmalen abhängt. Nur für die Parametrisierung entsteht damit eine besondere Problematik. Aufgrund der Komplexität der Einflussquellen erscheint es angemessen, anhand der Simulation von Zufalls- und systematischen Faktoren die Verteilungen der Inklusions-Indizes zu untersuchen. Aufgrund der zahlreichen Aspekte, die bei den Simulationen systematisch variiert werden müssten, konnten derartige Simulationsstudien im Rahmen dieser Arbeit nicht durchgeführt werden. Grundsätzlich könnten anhand von Simulationsstudien mit hierarchischen Faktorenanalysen auch Validitätsvergleiche für die hier vorgeschlagenen Inklusions-Indizes durchgeführt werden, indem versucht wird, anhand der Indizes die Hierarchie-Ebene von Faktoren zu präzisieren. Simulationsstudien zur Verteilung der Generalitätsindizes gehören somit zum erforderlichen Arbeitsprogramm, wenn es darum geht, die Generalität von Faktoren verbindlicher zu bestimmen.

Das weitere Arbeitsprogramm zur verbindlicheren Gestaltung der psychometrischen Traitforschung ergibt sich auch aus weiterführenden Überlegungen. Wie ausführlich dargestellt wurde, ist die Varianzaufklärung der Sekundärfaktoren äußerst gering, da die Primärfaktoren in zwei verschiedenen obliquen Rotationsmethoden sehr gering interkorrelierten. Dieses Ergebnis legt für die zukünftige Forschung eine deutlichere Betonung der Primärfaktoren nahe. Einer Betonung der Primärfaktoren steht bisher jedoch die Uneindeutigkeit der Strukturen auf dieser Ebene entgegen. Diese Uneindeutigkeit kann mit der geringen Aufmerksamkeit, die Primärfaktoren in der psychometrischen Traitforschung bisher erfahren haben, zusammenhängen: Den Cattellschen Primärfaktoren liegen leider eine Reihe kaum nachvollziehbarer, subjektiver Entscheidungen zugrunde (siehe Eysenck, 1971), weshalb es nicht verwundert, dass auch noch im 16 PF-R auf eine etwas fragwürdige Voraggregation der Skalen rekurriert wurde (Schneewind & Graf, 1998), um die intendierte Struktur nachzuweisen. Die 30 Facetten-Skalen des NEO-PI-R konnten in exploratorischen Faktorenanalysen selbst bei Vorgabe von 30 Faktoren und orthogonaler Procrustes-Rotation (Schönemann, 1966) nicht zufriedenstellend nachgewiesen werden (siehe Costa et al., 1991), da 34 der 240 Items nicht ihre höchste Ladung auf dem intendierten Faktor aufwiesen. Der ZKPQ und dem EPQ-R beinhalten keine Ebene, die der von Faktoren erster Ordnung im System von Cattell oder den Facetten-Skalen des NEO-PI-R entspricht. Die SSS-V ist zwar zur Erfassung von vier Faktoren auf der Ebene von Primärfaktoren angelegt, allerdings ist die psychometrische Qualität dieser Faktoren eher gering. Insofern war das hier verwendete Material für eine optimale Identifikation von Primärfaktoren nur bedingt geeignet. Da somit die Variablenauswahl nicht im Hinblick auf die optimale Identifikation von Primärfaktoren getroffen wurde, ist es nicht verwunderlich, dass sich die Primärfaktoren der verschiedenen Lösungen deutlich unterscheiden.

In Bezug auf eine optimale Variablenauswahl zur Identifikation von Primärfaktoren kann folgendes festgehalten werden: Es wären weiterführende modellübergreifende Analysen mit Inventaren, die im Gegensatz zum ZKPQ und dem EPQ-R eine den Primärfaktoren von Cattell bzw. den Facetten-Skalen des NEO-PI-R vergleichbare Aggregationsebene aufweisen, interessant. Hier würde sich die Kombination des 16 PF-R und NEO-PI-R mit dem Eysenck Personality Profiler (Eysenck, Wilson & Jackson, 1998), dem Tridimensional Personality Questionnaire (Cloninger, 1987; Weyers, Krebs & Janke, 1995) und dem Multidimensional Personality Questionnaire (Tellegen, 1996) anbieten. Im Prinzip erscheint auch der Trierer Persönlichkeitsfragebogen von Becker (1989) für Untersuchungen von Primärfaktoren interessant. Da dieser Fragebogen jedoch nur sechs itembasierte Faktoren aufweist, ist zu erwarten, dass diese Faktoren ein etwas höheres Generalitätsniveau aufweisen als beispielsweise die 30 Facetten-Skalen des NEO-PI-R. Auch das Hamburger Persönlichkeitsinventar von Andresen, Peter, Hand, Maibach und Jankuhn (2001) könnte für derartige Untersuchungen herangezogen werden, wobei die 36 Skalen dieses Instruments offenbar bisher nicht als Faktoren nach-

gewiesen wurden, weshalb hier ähnliche Probleme zu erwarten sind wie beim NEO-PI-R. Es muss allerdings eingeräumt werden, dass Costa et al. (1991) für den NEO-PI-R Faktorenanalysen auf Item-Ebene berichten, auch wenn diese als wenig erfolgreich anzusehen sind.

Eine Untersuchung anhand des 16 PF-R, NEO-PI-R, Eysenck Personality Profiler, Tridimensional Personality Questionnaire und Multidimensional Personality Questionnaire würde bereits 998 Items umfassen. Die Fragebogen müssten an mehreren Tagen ausgefüllt werden und es müssten deutlich mehr als 1000 Probanden untersucht werden. Um robuste Ergebnisse zu erzielen, bzw. die Replizierbarkeit der Primärfaktoren zu prüfen, müssten deutlich über 2000 Probanden untersucht werden. Für die Analyse von etwa 1000 Items liegen in der psychometrischen Traitforschung bisher keine Erfahrungen vor. Auch die Software für derartige Analysen müsste erst bereitgestellt werden. Eine nennenswerte Veränderung der untersuchten Probanden und Variablenzahlen kann auch Einfluss auf die Qualität der Methoden haben. Die Probleme der Identifikation von Faktoren anhand der Varimax-Rotation bei Lösungen mit ca. 30 Faktoren sind ein Beispiel für derartige Einflüsse.

Im übrigen ergaben sich gerade die deutlichen Unterschiede zwischen den verschiedenen Lösungen erster Ordnung zum einen aus dem ungünstigen Einfluss der Varimax-Vorrotation bei der Trasid-Rotation und zum anderen aus dem massiven Einfluss der Gewichtung der Stichprobe auf die Faktorenanzahl. Weiterführende Analysen sollten insofern auf der Basis möglichst repräsentativer Stichproben (ausschließlich studentische Stichproben sind hier unzureichend) und auf der Basis von Rotationsmethoden, die auch bei großer Faktorenzahl robuste Ergebnisse liefern (z.B. Oblimin-Rotation), durchgeführt werden. Erst wenn diese Bedingungen, zusammen mit den in Kapitel 3 ermittelten Spezifikationen, eingehalten werden, kann auch erwartet werden, auf der Ebene von Primärfaktoren replizierbare Strukturen zu erhalten.

Aus dieser Perspektive betrachtet, befindet sich die psychometrische Traitforschung erst in der Anfangsphase. Gemessen am oben skizzierten Umfang von Untersuchungen zur umfassenden Identifikation von Primärfaktoren haben die bisher durchgeführten Untersuchungen in diesem Bereich eher fragmentarischen Charakter. Diese Einschätzung schließt die vorliegende Untersuchung ein, obgleich in der bisherigen Traitforschung bisher kaum der hier verwendete Variablenumfang mit 400 bis 500 Variablen erreicht wurde. Gemessen an der hier festgestellten geringen Varianzaufklärung der Sekundärfaktoren und dem erheblichen noch zu leistenden Aufwand, der mit einer umfassenden Identifikation von Primärfaktoren verbunden ist, erscheinen manche Schlussfolgerungen von Traitforschern kaum noch vertretbar. So stellten Costa und McCrae (1997, S. 271) in Bezug auf das FFM fest: „Because it is comprehensive, conclusions about the five domains of Neuroticism, Extraversion, Openness,

Agreeableness, and Conscientiousness can confidently be treated as conclusions about the full range of personality traits.“

Derartige Aussagen müssen anhand der Befunde dieser Arbeit und den skizzierten weiterhin erforderlichen Untersuchungen als verfrühte und unangemessene Übergeneralisierungen gewertet werden. Dabei ist zu betonen, dass es sich hier nicht um eine Kritik *an* der psychometrischen Perspektive wie beispielsweise in Mischel und Peake (1982), sondern um eine Kritik *aus* der psychometrischen Perspektive handelt. Die zweifellos ironisch gemeinte Überschrift in Brody (1998, S. 58) „The big five: Are we finished?“ kann auf der Basis der Befunde und Konsequenzen aus dieser Arbeit eindeutig beantwortet werden:

We are just beginning.

Zusammenfassung

In der vorliegenden Arbeit werden zwei wesentliche Problembereiche, die zu einer Unverbindlichkeit der Befunde der psychometrischen Persönlichkeits- bzw. Traitforschung führen, herausgearbeitet. Der erste Problembereich betrifft die *empirisch-methodische Unverbindlichkeit*, die vor allem auf einen suboptimalen Einsatz der Faktorenanalyse zurückgeführt wird. Der suboptimale Einsatz der faktorenanalytischen Verfahren wird vor allem auf eine mangelnde Orientierung an der verfügbaren Literatur zur Qualität verschiedener Methoden zurückgeführt. Dementsprechend wird in der vorliegenden Arbeit die verfügbare Literatur gesichtet und es werden Vorschläge für eine optimale Durchführung konfirmatorischer und exploratorischer Faktorenanalysen im Bereich der psychometrischen Traitforschung erarbeitet.

Der zweite hier identifizierte Problembereich der psychometrischen Traitforschung, betrifft die *theoretisch-interpretative Unverbindlichkeit* der Aussagen zur Generalität von Traitdimensionen. Nach einer Sichtung typischer Generalitätszuschreibungen, werden Indizes zur Quantifizierung der Generalität von Skalen und Faktoren vorgeschlagen.

Die Vorschläge zur optimalen Durchführung konfirmatorischer und exploratorischer Faktorenanalysen sowie die Vorschläge zur Quantifizierung der Generalität von Faktoren werden auf zwei Datensätze exemplarisch angewandt. Beim ersten Datensatz (Probandengruppe A, N = 554) wurden Fragebogen zur Erfassung der Modelle von Cattell (16 PF-R), Eysenck (EPQ-R) sowie Costa und McCrae (NEO-PI-R) eingesetzt. Der zweite Datensatz (Probandengruppe B, N = 620) umfasst Fragebogen zur Erfassung der Modelle von Eysenck (EPQ-R), Costa und McCrae (NEO-PI-R) sowie Zuckerman (SSS-V, ZKPQ).

Zunächst wurde in modellspezifischen Faktorenanalysen die Struktur der jeweiligen Fragebogen geprüft. Anhand der hier vorgenommenen Prüfstrategie konnten die Primärfaktoren aller eingesetzten Fragebogen anhand der konfirmatorischen Faktorenanalyse weitgehend bestätigt werden. Bei den hierarchischen Analysen ergaben sich deutliche Probleme für die Sekundärfaktoren des 16 PF-R Fragebogens und Einschränkungen bezüglich der Orthogonalität beim NEO-PI-R. Anhand der exploratorischen Faktorenanalyse konnte nur die Struktur des EPQ-R und des ZKPQ bestätigt werden.

In den modellübergreifenden, hierarchischen Faktorenanalysen ergaben sich Hinweise auf 21 bis 32 Faktoren erster Ordnung. Die Primärfaktoren unterschieden sich deutlich in Abhängigkeit von der Stichprobe und Methode der Faktorenrotation. Die Interkorrelationen zwischen diesen Faktoren erster Ordnung waren äußerst gering. Demzufolge erklärten die fünf bis sechs Faktoren zweiter Ordnung extrem wenig Varianz.

Neurotizismus und Extraversion konnten als Faktoren zweiter Ordnung in beiden Stichproben repliziert werden. Es zeigten sich auch Hinweise auf einen Faktor, der vor allem Aspekte von Verträglichkeit und Psychotizismus beinhaltet. Neben diesem Faktor traten Faktoren der Bereiche Offenheit für Erfahrungen und Gewissenhaftigkeit in unterschiedlichen Varianten auf. Die Faktoren entsprachen jedoch nicht durchgängig den im Rahmen des Fünf-Faktorenmodells der Persönlichkeit postulierten Faktoren.

Auf der Basis der Generalitäts-Indizes wurde für den Neurotizismus-Faktor des NEO-PI-R eine größere Generalität festgestellt als für die Neurotizismus-Faktoren des EPQ-R und ZKPQ, die sich wiederum kaum in der Generalität unterschieden. Für den Extraversion-Faktor des EPQ-R wurde anhand von zwei Generalitäts-Indizes die höchste Generalität ermittelt. Demgegenüber deuteten zwei andere Generalitäts-Indizes darauf hin, dass der Extraversions-Faktor des NEO-PI-R die höchste Generalität aufweist. Auch bei einem Vergleich der in hierarchischen Faktorenanalysen aufgezeigten Neurotizismus-Faktoren zweiter Ordnung mit den modellspezifischen Neurotizismus-Faktoren des NEO-PI-R und EPQ-R zeigten zwei Generalitäts-Indizes eine höhere Generalität der modellspezifischen Faktoren und zwei Generalitäts-Indizes eine höhere Generalität der Faktoren zweiter Ordnung an. Da die Generalitäts-Indizes unterschiedliche Aspekte von Generalität quantifizieren, deuten diese Ergebnisse darauf hin, dass die verschiedenen Aspekte der Generalität unterschiedlich variieren können. Bei einem analogen Vergleich für die Extraversions-Faktoren zeigte sich eine höhere Generalität der modellspezifischen Faktoren im Vergleich zu den Faktoren zweiter Ordnung.

Die geringen Korrelationen zwischen den Primärfaktoren, die zu einer geringen Varianzaufklärung von Faktoren zweiter Ordnung führten, sprechen für die Relevanz der Primärfaktoren im Vergleich zu den Sekundärfaktoren. Dieses Ergebnis relativiert somit die Bedeutung der Modelle, die auf Sekundärfaktoren fokussieren (z.B. Fünf-Faktorenmodell; Eysencks deskriptive Persönlichkeitstheorie). Andererseits wurden deutliche Abhängigkeiten der Primärfaktoren von den Stichproben und Rotationsmethoden festgestellt. Daher wurden Vorschläge für eine systematische und umfangreiche Untersuchung von Primärfaktoren unterbreitet, die in dieser Form in der bisherigen Forschung noch nicht durchgeführt wurde. Bei einer konsequenten Umsetzung der hier verfolgten Methodologie, kann davon ausgegangen werden, dass entgegen den Erwartungen mancher Fachvertreter, noch gravierende Neustrukturierungen im Bereich der psychometrischen Traitforschung möglich sind.

Literatur

- Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G.W. & Odbert, H.S. (1936). Trait names: A psychlexical study. *Psychological Monographs*, 47,1/Whole No.211.
- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5. Aufl.). Stuttgart: Kohlhammer.
- Amelang, M. & Borkenau, P. (1982). Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 119-145.
- Amelang, M., Schmidt-Rahtjens, C. & Yousfi, S. (2001). Not a decisive test, but some informative results with different modalities of questionnaire administration: The effects of group versus individual testing. In R. Riemann, Spinath, F. & Ostendorf, F. (Eds.), *Personality and Temperament: Genetics, Evolution, and Structure*. Lengerich: Pabst Science Publishers.
- Amelang, M. & Ullwer, U. (1990). Untersuchungen zur experimentellen Bewährung von Eysencks Extraversionstheorie. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 11, 127-148.
- Amelang, M. & Ullwer, U. (1991). Ansatz und Ergebnisse einer (fast) umfassenden Überprüfung von Eysencks Extraversionstheorie. *Psychologische Beiträge*, 33, 23-46.
- Andresen, B. (1995). Risikobereitschaft (R) - Der sechste Basisfaktor der Persönlichkeit: Konvergenz multivariater Studien und Konstruktextplikation. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 16, 210-236.
- Andresen, B. (2000). Six Basic Dimension of Personality and a Seventh Factor of Generalized Dysfunctional Personality: A Diathesis System Covering all Personality Disorders. *Neuropsychobiology*, 41, 5-23.
- Andresen, B., Peter, H., Hand, I., Maibach, H. & Jankuhn, A. (2001). Sechs klinisch akzentuierte Basisdimensionen der Persönlichkeit unter besonderer Berücksichtigung schizotyper Variablen. In Andresen, B. & Mass, R. (Eds.), *Schizotypie. Psychometrische Entwicklungen und biopsychologische Forschungsansätze* (S. 45-97). Göttingen: Hogrefe.
- Asendorpf, J. (1991). *Die differentielle Sichtweise in der Psychologie*. Göttingen: Hogrefe.
- Asendorpf, J.B. (1995). Persönlichkeitspsychologie: Das empirische Studium der individuellen Besonderheit aus spezieller und differentieller Perspektive. *Psychologische Rundschau*, 46, 235-247.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (1996). *Multivariate Analysemethoden*. Berlin: Springer.
- Barbaranelli, C. & Caprara, G.V. (1996). How many dimensions to describe personality? A comparison of Cattell, Comrey, and the Big Five taxonomies of personality traits. *Revue Européenne de Psychologie Appliquée*, 46, 15-24.
- Bargmann, R. (1955). Signifikanzuntersuchungen der einfachen Struktur in der Faktoren-Analyse. *Mittbl. math. Statist.*, 7, 1-24.
- Barrett, P. & Kline, P. (1982). An item and radial parcel factor analysis of the 16PF questionnaire. *Personality and Individual Differences*, 3, 259-270.
- Barrett, P.T. & Kline, P. (1981). Radical parcel factor analysis. *Personality and Individual Differences*, 2, 311-318.
- Barrett, P.T. & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study and Group Behavior*, 1, 23.
- Bartlett, M.S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28, 97-104.
- Bartlett, M.S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, 3, 77-85.
- Bartlett, M.S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology*, 4, 1-2.
- Bartussek, D. (1973). Zur Interpretation der Kernmatrix in der dreimodalen Faktorenanalyse von L.R. Tucker. *Psychologische Beiträge*, 15, 169-184.
- Bartussek, D. (1988). Beurteilung der deutschen Form des 16 PF-Testes. *Diagnostica*, 34, 367-379.
- Bartussek, D., Becker, G., Diedrich, O., Naumann, E. & Maier, S. (1996). Extraversion, neurotizismus, and event-related brain potentials in response to emotional stimuli. *Personality and Individual Differences*, 20, 301-312.
- Beauducel, A. (1996). *Zur Optimierung der schiefwinkligen Rotation zur Einfachstruktur durch Berücksichtigung der faktoriellen Komplexität von Variablen*. In Unveröff. Diss., Freie Universität Berlin, FB Erziehungswissenschaft, Psychologie und Sportwissenschaft (Ed.).
- Beauducel, A. (1997). Transformation-matrix-Search and -Identification (Trasid): A new method for oblique rotation to simple structure. *Methods of Psychological Research Online*, 2, 113-138.

- Beauducel, A. (2000). Extraktionskriterien: Probleme und Optimierungen der Parallelanalyse. *42. Kongreß der Deutschen Gesellschaft für Psychologie an der Friedrich-Schiller-Universität Jena*, 24.-28.9.2000. Abstract-CD-ROM: Pabst Science Publishers.
- Beauducel, A. (2001a). On the generalizability of factors: The influence of changing contexts of variables on different methods of factor extraction. *Methods of Psychological Research*, 6, 1-24.
- Beauducel, A. (2001b). Problems with parallel analysis in data sets with oblique simple structure. *Methods of Psychological Research Online*, 6, 141-157.
- Beauducel, A., Debener, S., Brocke, B. & Kayser, J. (2000). On the reliability of augmenting/reducing: Peak amplitudes and principal component analysis of auditory evoked potentials. *Journal of Psychophysiology*, 14, 226-240.
- Beauducel, A., Brocke, B. & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences*, 30, 977-994.
- Beauducel, A., Brocke, B., Strobel, A. & Strobel, A. (1999). Construct validity of sensation seeking: A psychometric investigation. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20, 155-171.
- Beauducel, A. & Kersting, M. (in Druck). Fluid and crystallized intelligence and the Berlin Model of Intelligence Structure (BIS). *European Journal of Psychological Assessment*.
- Becker, P. (1989). *Der Trierer Persönlichkeitsfragebogen TPF*. Göttingen: Hogrefe.
- Becker, P. (1999). Beyond the Big Five. *Personality and Individual Differences*, 26, 511-530.
- Benet, V. & Waller, N.G. (1995). The Big Seven Factor Model of Personality Description: Evidence for Its Cross-Cultural Generality in a Spanish Sample. *Journal of Personality and Social Psychology*, 69, 701-718.
- Benet-Martinez, V. & Waller, N.G. (1997). Further evidence for the cross-cultural generality of the big seven factor model: Indigenous and imported Spanish personality constructs. *Journal of Personality*, 65, 567-598.
- Bentler, P.M. (1977). Factor simplicity index and transformations. *Psychometrika*, 42, No. 2, 277-295.
- Bentler, P.M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
- Bentler, P.M. & Yuan, K.H. (1998). Tests for linear trend in the smallest eigenvalues of the correlation matrix. *Psychometrika*, 63, 131-144.
- Bernstein, I.H. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, 105, 467-477.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117(2), 187-215.
- Boesser, T. (1979). Phi-Interkorrelationen seltener Symptome. *Psychologische Beiträge*, 21, 343-348.
- Borg, I. (1981). *Anwendungsorientierte Multidimensionale Skalierung*. Berlin: Springer-Verlag.
- Borkenau, P. & Ostendorf, F. (1989). Untersuchungen zum Fünf-Faktoren-Modell der Persönlichkeit und seiner diagnostischen Erfassung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 10, 239-251.
- Borkenau, P. & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, 11, 515-524.
- Borkenau, P. & Ostendorf, F. (1991). Ein Fragebogen zur Erfassung fünf robuster Persönlichkeitsfaktoren. *Diagnostica*, 37, 29-41.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler (5. Aufl.)*. Berlin: Springer.
- Brand, C.R. (1994). Open to experience - closed to intelligence: why the 'Big Five' are really the 'Comprehensive Six'. *European Journal of Personality*, 8, 299-310.
- Brocke, B. (2000). Das bemerkenswerte Comeback der Differentiellen Psychologie: Glückwünsche und Warnungen vor einem neuen Desaster [The remarkable comeback of personality psychology: Congratulations and warnings against a new desaster]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 21(1), 5-30.
- Brocke, B. & Battmann, W. (1985). Die Aktivierungstheorie der Persönlichkeit. Eine systematische Darstellung und partielle Rekonstruktion. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 6, 189-213.
- Brocke, B. & Battmann, W. (1992). The arousal-activation theory of extraversion and neuroticism: A systematic analysis and principal conclusions. *Advances in Behaviour Research and Therapy*, 14, 211-246.
- Brocke, B., Beauducel, A., John, R., Debener, S. & Heilemann (2000). Sensation seeking and affective disorders: Characteristics in the intensity dependence of acoustic evoked potentials. *Neuropsychobiology*, 41, 24-30.
- Brocke, B., Beauducel, A. & Tasche, K.G. (1999). Biopsychological bases and behavioral correlates of sensation seeking: Contributions to a multilevel validation. *Personality and Individual Differences*, 26, 1103-1123.
- Brocke, B., Tasche, K.G. & Beauducel, A. (1996). Biopsychological foundations of extraversion: Differential effort reactivity and the differential P300 effect. *Personality and Individual Differences*, 21, 727-738.

- Brocke, B., Tasche, K.G. & Beauducel, A. (1997). Biopsychological foundations of extraversion: Differential effort reactivity and state control. *Personality and Individual Differences*, 22, 447-458.
- Brody, N. (1998). *Personality psychology: the science of individuality*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Browne, M.W. (1968). A note on lower bounds for the number of common factors. *Psychometrika*, 33(2), 233-236.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (S. 136-162). Beverly Hills, CA: Sage.
- Buley, J.L. (1995). Evaluating Exploratory Factor Analysis: Which Initial-Extraction Techniques Provide the Best Factor Fidelity. *Human Communication Research*, 21, 478-493.
- Burns, K.J. (1998). Beyond classical reliability: using generalizability theory to assess dependability. *Research in Nursing & Health*, 21, 83-90.
- Buse, L. & Pawlik, K. (1991). Zur State-Trait-Charakteristik verschiedener Meßvariablen der psychophysiologischen Aktivierung, der kognitiven Leistung und der Stimmung in Alltagssituationen. *Zeitschrift für experimentelle und angewandte Psychologie*, 38, 521-538.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56, 81-105.
- Caprara, G.V., Barbaranelli, C., Borgogni, L. & Perugini, M. (1993). The Big Five Questionnaire: A new questionnaire for the measurement of the five factor model. *Personality and Individual Differences*, 15, 281-288.
- Carlson, R. & Mulaik, S. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Hrsg.), *Testing structural equation models* (S. 136-162). Newbury Park, CA: Sage.
- Carroll, J.B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23-28.
- Carroll, J.B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carstensen, C.H. (2000). *Mehrdimensionale Testmodelle mit Anwendungen aus der pädagogisch-psychologischen Diagnostik*. In Unveröff. Diss., Christian-Albrecht-Universität, Kiel (Ed.).
- Carver, C.S. & White, T.L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, 67, 319-333.
- Cattell, R.B. (1957). *Personality and motivation: Structure and measurement*. New York: Yonkers, World Book Comp.
- Cattell, R.B. (1962). The basis of recognition and interpretation of factors. *Educational and Psychological Measurement*, 22, 667-697.
- Cattell, R.B. (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally & Co.
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cattell, R.B. (1972). The 16PF and basic personality structures: A reply to Eysenck. *Journal of Behavioral Science*, 1, 169-187.
- Cattell, R.B. (1978). *The scientific use of factor analysis*. New York: Plenum.
- Cattell, R.B. (1987). *Intelligence: its structure, growth, and action*. Amsterdam: Elsevier Science Publishers B.V.
- Cattell, R.B. (1988). The meaning and strategic use of factor analysis. In J.R. Nesselroade & R.B. Cattell (Eds.), *Handbook of multivariate experimental psychology (2nd Edition)* (pp. 131-201). New York: Plenum Press.
- Cattell, R.B. (1994a). Constancy of global, second-order personality factors over a twenty-year-plus period. *Psychological Reports*, 75, 3-9.
- Cattell, R.B. (1994b). A cross-validation of primary personality structure in the 16 P.F. by two parcelled factor analyses. *Multivariate Experimental Clinical Research*, 10, 181-190.
- Cattell, R.B. & Cattell, A.K.S. (1955). Factor rotation for proportional profiles: Analytic solution and an example. *British Journal of Statistical Psychology*, 8, 83-86.
- Cattell, R.B. & Delhees, K.H. (1973). Seven missing normal personality factors in the questionnaire primaries. *Multivariate Behavioral Research*, 8, 173-194.
- Cattell, R.B. & Dickman, K.W. (1962). A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin*, 59, 389-400.
- Cattell, R.B. & Dreger, R.M. (1977). *Handbook of modern personality theory*. New York: Wiley & Sons.
- Cattell, R.B. & Jaspers, J. (1967). A general plasmode (No. 30-10-5-2) for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, No. 76-3.
- Cattell, R.B. & Muerle, J.L. (1960). *The "Maxplane" program for factor rotation to oblique simple structure* (Vol. 20, pp. 569-590).

- Cattell, R.B. & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales. *Educational and Psychological Measurement*, 24, No. 1, 3-30.
- Chan, R. & Joseph, S. (2000). Dimensions of personality, domains of aspiration, and subjective well-being. *Personality and Individual Differences*, 28, 347-354.
- Church, A.T. & Burke, P.J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology*, 66, 93-114.
- Church, A.T., Katigbak, M.S. & Reyes, J.A. (1998). Further exploration of Filipino personality structure using the lexical approach: do the big-five or big-seven dimensions emerge? *European Journal of Personality*, 12, 249-269.
- Clarkson, D.B. & Jennrich, R.I. (1988). Quartic rotation criteria and algorithms. *Psychometrika*, 53, 251-259.
- Cliff, N. (1970). The relation between sample and population characteristic vectors. *Psychometrika*, 35(2), 163-178.
- Cliff, N. (1970). The relation between sample and population characteristic vectors. *Psychometrika*, 35, 163-178.
- Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, 103, 276-279.
- Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry*, 44, 573-588.
- Cloninger, C.R., Przybeck, T.R. & Svrakic, D.M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. *Psychological Reports*, 69, 1047-1057.
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences (2. Aufl.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooley, W.S. & Lohnes, P.R. (1976). *Evaluation research in education*. New York: Irvington Publishers.
- Corballis, M. (1968). Some difficulties with difficulty: note on Horst's 'Matrix factoring and test theory'. *Psychological Reports*, 22, 15-22.
- Corulla, W.J. (1988). A further psychometric investigation of the Sensation Seeking Scale Form-V and its relationship to the EPQ-R and the I.7 Impulsiveness-Questionnaire. *Personality and Individual Differences*, 9, 77-87.
- Costa, P.T., Jr., Busch, C.M., Zonderman, A.B. & McCrae, R.R. (1986). Correlations of MMPI factor scales with measures of the five factor model of personality. *Journal of Personality Assessment*, 50, 640-650.
- Costa, P.T., Jr. & McCrae, R.R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, 55, 258-265.
- Costa, P.T., Jr. & McCrae, R. (1995). Solid ground in the wetlands of personality: A reply to Block. *Psychological Bulletin*, 2, 216-220.
- Costa, P.T., Jr. & McCrae, R.R. (1992a). *The Revised Neo Personality Inventory (NEO-PI-R) and NEO- Five Factor Inventory (NEO-FFI) Professional Manual*. Odessa: Psychological Assessment Resources.
- Costa, P.T., Jr. & McCrae, R.R. (1992b). Four ways five factors are basic. *Personality and Individual Differences*, 13, 653-665.
- Costa, P.T., Jr., McCrae, R.R. & Dye, D. . (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory (*Personality and Individual Differences* (12.; 12, 887-898).
- Costa, P.T., Jr. & McCrae, R.R. (1995). Primary traits of Eysenck's P-E-N system: Three- and five-factor solutions. *Journal of Personality and Social Psychology*, 69, 308-317.
- Costa, P.T. & McCrae, R.R. (1997). Longitudinal stability of adult personality. In Hogan, R., Johnson, J. & Briggs, S. (Eds.), *Handbook of personality psychology (S.269-290)*. San Diego: Academic Press.
- Cota, A.A., Longman, R.S., Holden, R.R., Fekken, G.C. & Xinaris, S. (1993). Interpolating 95th percentile eigenvalues from random data: An empirical example. *Educational and Psychological Measurement*, 53, 585-596.
- Crawford, C.B. (1975). Determining the number of interpretable factors. *Psychological Bulletin*, 82, 226-237.
- Crawford, C.B. & Ferguson, G.A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, 35, 321-332.
- Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Cronbach, L.J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Cureton, E.E. & Mulaik, S.A. (1975). The weighted Varimax rotation and the Promax rotation. *Psychometrika*, 40, No. 2, 183-195.
- Debener, S. (2001). Individuelle Unterschiede in der frontalen EEG-Alphaasymmetrie: Emotionalität und intra-individuelle Veränderungen. Berlin: Dissertation.de.

- Di Blas, L. & Forzi, M. (1998). An alternative taxonomic study of personality-descriptive adjectives in the Italian language. *European Journal of Personality*, 12, 75-101.
- Digman, J.M. (1997). Higher-Order Factors of the Big Five. *Journal of Personality and Social Psychology*, 73, 1246-1256.
- Draycott, S.G. & Kline, P. (1995). The Big Three or the Big Five - the EPQ-R vs the NEO-PI: A research note, replication and elaboration. *Personality and Individual Differences*, 18, 801-804.
- Embretson, S.E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56, 495-515.
- Everett, J.E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research*, 18, 197-218.
- Everett, J.E. & Entreekin, L.V. (1980). Factor comparability and the advantages of multiple group factor analysis. *Multivariate Behavioral Research*, 15, 165-180.
- Eysenck, H.J. (1950). Criterion analysis - an application of the hypothetico-deductive method in factor analysis. *Psychological Review*, 57, 35-62.
- Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: C. C.Thomas.
- Eysenck, H.J. (1971). On the choice of personality tests for research and prediction. *Journal of Behavioural Science*, 1, 85-89.
- Eysenck, H.J. (1981). *A model for personality*. Berlin: Springer.
- Eysenck, H.J. (1991). Dimension of personality: 16, 5 or 3?- Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 8, 773-790.
- Eysenck, H.J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, 13(6), 667-673.
- Eysenck, H.J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, 73, 1224-1237.
- Eysenck, H.J. & Eysenck, S.B.G. (1975). *Manual of the Eysenck Personality Questionnaire (junior & adult)*. London: Hodder & Stoughton.
- Eysenck, H.J. & Eysenck, M.W. (1985). *Personality and individual differences*. New York: Plenum Press.
- Eysenck, H.J. & Eysenck, S.B.G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult)*. London: Hodder and Stoughton.
- Eysenck, H.J., Wilson, G.D. & Jackson, C.J. (1998). *Eysenck Personality Profiler - Deutsche Fassung*. Frankfurt am Main: Swets.
- Eysenck, S.B.G., Daum, I., Schugens, M.M. & Diehl, J.M. (1990). A cross-cultural study of impulsiveness, venturesomeness and empathy: Germany and England. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 11, 209-213.
- Eysenck, S.B.G., Eysenck, H.J. & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, 6, 21-29.
- Fahrenberg, J. (1987). Zur psychophysiologischen Methodik: Konvergenz, Fraktionierung oder Synergismen? *Diagnostica*, 33, Heft 3, 272-287.
- Fava, J.L. & Velicer, W.F. (1992). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research*, 27, 301-322.
- Feger, H. (1988). Spatial representation of feature patterns. In H.H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 431-437). Amsterdam: North Holland.
- Feger, H. (2000). Konfigurationsfrequenzanalyse und Feature Pattern Analysis: Einige vergleichende Betrachtungen. *Psychologischen Beiträge*, 42, 448-468.
- Ferguson, G.A. (1941). The factorial interpretation of test difficulty. *Psychometrika*, 4, 323-329.
- Fischer, G.H. (1996). IRT-Modelle als Forschungsinstrumente der Differentiellen Psychologie. In Pawlik, K. (Ed.), *Enzyklopädie der Psychologie. Differentielle Psychologie und Persönlichkeitsforschung, Band 1. Grundlagen und Methoden der Differentiellen Psychologie (S. 673-729)*. Göttingen: Hogrefe.
- Fruchter, B. (1954). *Introduction to Factor Analysis*. Princeton: D. van Nostrand Company.
- Fürntratt, E. (1969). Zur Bestimmung der Anzahl interpretierbarer gemeinsamer Faktoren in Faktorenanalysen psychologischer Daten. *Diagnostica*, 15, 62-75.
- Gebert, A. (1977). Jäger's Phi(G) als Item-Interkorrelationsmaß für Faktorenanalysen. *Psychologische Beiträge*, 19, 336-339.
- Geweke, J.F. & Singleton, K.J. (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Stat. Association*, 75, 133-137.

- Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55, 377-393.
- Goldberg, L.R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.
- Goldberg, L.R., Sweeney, D., Merenda, P.F. & Hughes, J.E. (1998). Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. *Personality and Individual Differences*, 24, 393-403.
- Gorsuch, R.L. (1973). Using Bartlett's significance test to determine the number of factors to extract. *Educational and Psychological Measurement*, 33, 361-364.
- Gorsuch, R.L. (1983). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Gorsuch, R.L. & Nelson, J. (1981). CNG scree test: An objective procedure for determining the number of factors. *Presented at the annual meeting of the Society of Multivariate Experimental Psychology*.
- Gourlay, N. (1951). Difficulty factors arising from the use of tetrachoric correlations between items or between tests. *British Journal of Psychology, Statistical Section*, 4, 65-76.
- Grau, E. & Ortet, G. (1999). Personality traits and alcohol consumption in a sample of non-alcoholic women. *Personality and Individual Differences*, 27, 1057-1066.
- Gray, J.A. (1973). Causal theories of personality and how to test them. In J.R. Royce (Ed.), *Multivariate analysis and psychological theory* (S. 409-462). New York: Academic Press.
- Gray, J.A. (1983). Where should we search for biologically based dimensions of personality. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 4, 165-176.
- Gray, J.A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament* (pp. 105-128). New York: Plenum Press.
- Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guilford, J.P. (1975). Factors and factors of personality. *Psychological Bulletin*, 82(5), 802-814.
- Guilford, J.S., Zimmermann, W.S. & Guilford, J.P. (1976). *The Guilford-Zimmermann temperament survey handbook*. San Diego, Cal.: Edits Publ.
- Gustafsson, J.E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179-203.
- Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18(4), 277-296.
- Guttman, L. (1954). A new approach to factor analysis: The radex. In P.F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. Glencoe, IL: Free Press.
- Hakstian, A.R. (1971). A comparative evaluation of several prominent methods of oblique factor transformation. *Psychometrika*, 36, No. 2, 175-193.
- Hakstian, A.R. & Abell, R.A. (1974). A further comparison of oblique factor transformation methods. *Psychometrika*, 39, 429-444.
- Hakstian, A.R. & Boyd, W.M. (1972). An empirical investigation of some special cases of the general "orthomax" criterion for orthogonal factor transformation. *Educational and Psychological Measurement*, 32, 3-22.
- Hakstian, A.R. & Muller, V.J. (1973). Some notes on the number of factors problem. *Multivariate Behavioral Research*, 8, 461-475.
- Hakstian, A.R., Rogers, W.T. & Cattell, R.B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, 17, 193-219.
- Hammond, K.R. (Hrsg.). (1966). *The psychology of Egon Brunswik*. New York: Holt, Rinehart & Winston.
- Harman, H.H. (1960). *Modern factor analysis*. Chicago: The University of Chicago Press.
- Harman, H.H. (1967). *Modern factor analysis*. Chicago: The University of Chicago Press.
- Harris, C.W. & Kaiser, H.F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 29, 347-362.
- Harris, M.L. & Harris, C.W. (1971). A factor analytic interpretation strategy. *Educational and Psychological Measurement*, 31, 589-606.
- Haynes, C.A., Miles, J.N.V. & Clements K. (2000). A confirmatory factor analysis of two models of sensation seeking. *Personality and Individual Differences*, 29, 823-839.
- Heath, A.C., Cloninger, C.R. & Martin, N.G. (1994). Testing a model for the genetic structure of personality: A comparison of the personality systems of Cloninger and Eysenck. *Journal of Personality and Social Psychology*, 66, 762-775.
- Hendrickson, A.E. & White, P.O. (1964). Promax: a quick method for rotation to oblique simple structure. *The british journal of statistical psychology*, 17, 65-70.

- Hess, R. (1998). Type nouns and the five factor model of personality description. *European Journal of Personality*, 12, 57-71.
- Herrmann, T. (1969). *Lehrbuch der empirischen Persönlichkeitsforschung*. Göttingen: Hogrefe.
- Heubeck, B.G., Wilkinson, R.S. & Cologon, J. (1998). A second look at Carver and White's (1994) BIS/BAS scales. *Personality and Individual Differences*, 25, 785-800.
- Hodapp, V. (1984). *Analyse linearer Kausalmodelle*. In Bern, Stuttgart (Ed.).
- Hofstee, W.K.B., De Raad, B. & Goldberg, L.R. (1992). Integration of the Big Five and the circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63, 146-163.
- Hojat, M. (1982). Loneliness as a function of selected personality variables. *Journal of Clinical Psychology*, 38, 137-141.
- Holling, H. (1993). Zur Anwendung von Sturkturgleichungsmodellen in der psychologischen Forschung. In W. Bungard & T. Herrmann (Hrsg.), *Arbeits- und Organisationspsychologie im Spannungsfeld zwischen Grundlagenorientierung und Anwendung*. Bern: Huber.
- Holz-Ebeling, F. (1995). Faktorenanalysen und was dann? Zur Frage der Validität von Dimensionsinterpretationen. *Psychologische Rundschau*, 46, 18-35.
- Horn, J. (1963). Second-order factors in questionnaire data. *Educational and Psychological Measurement*, 23, No. 1, 117-134.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Horn, J.L. & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, 14, 283-300.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 321-377.
- Hough, L.M. (1992). The "Big Five" personality variables - construct confusion: description versus prediction. *Human Performance*, 5, 139-155.
- Howard, K.I. & Gordon, R.A. (1963). Empirical note on the "number of factors" problem in factor analysis. *Psychological Report*, 12(1), 247-250.
- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hubbard, R. & Allen, S.J. (1987). An empirical comparison of alternative methods for principal component extraction. *Journal of Business Research*, 15, 173-190.
- Humphreys, L.G. (1981). *The primary mental ability*. In M.P. Friedman, J.P. Das & M. O'Connor, (Eds.), *Intelligence and learning*, New York: Plenum Press.
- Hurley, J.R. & Cattell, R.B. (1962). The Procrustes program: producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7, 258-262.
- Jäger, A.O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica, Band XXVIII, Heft 3*, 195-225.
- Jäger, A.O., H.M. Süß & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. BIS-Test, Form 4*. Göttingen: Hogrefe.
- Jäger, R.S. (1976). Phi(G) - ein statistischer Test zur Bestimmung der Differenzierungsfähigkeit psychologischer Skalen. *Psychologische Beiträge*, 18, 214-223.
- Jennrich, R.I. & Sampson, P.F. (1966). Rotation for simple loadings. *Psychometrika*, 31, 313-323.
- Jöreskog, K.G. & Sörbom, D. (1981). *LISREL V. Analysis of linear structural relationships by maximum likelihood and least squares methods*. In Research Report 81-8. University of Uppsala, Department of Statistics, P.O. Box 513, S-75120 Uppsala, Sweden (Ed.).
- Jöreskog, K.G. & Sörbom, D. (1993). *LISREL 8-User's reference guide*. Chicago, IL: Scientific Software International.
- Jung, C.G. (1960). *Psychologische Typen. Gesammelte Werke, VI*. Zürich: Rascher.
- Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kaiser, H.F. (1970). A second generation little jiffy. *Psychometrika*, 35, 401-415.
- Kaiser, H.F. & Caffrey, J. (1965). Alpha factor analysis. *Psychometrika*, 30(1), 1-14.
- Kallina, H. & Hartmann, A. (1976). Ein Vergleich von Hauptkomponentenanalyse und klassischer Faktorenanalyse. *Psychologische Beiträge*, 18, 84-98.
- Katz, J.O. & Rohlf, F.J. (1974). Functionplane - a new approach to simple structure rotation. *Psychometrika*, 39, No. 1, 37-51.

- Katz, J.O. & Rohlf, F.J. (1975). *Primary product functionplane: An oblique rotation to simple structure* (Vol. 10, pp. 219-232). In *Multivariate Behavioral Research* (Ed.).
- Kiers, H.A.L. (1994). Simplimax: Oblique rotation to an optimal target with simple structure. *Psychometrika*, 59, 567-579.
- Kirkcaldy, B.D., Cooper, C.L., Eysenck, M. & Brown, J. (1994). Anxiety and coping. *Personality and Individual Differences*, 17, 681-684.
- Klages, L. (1926). *Die Grundlagen der Charakterkunde*. Leipzig: Bart.
- Knuth, D.E. (1981). *The art of computer programming. Vol. II (2. Aufl.)*. Reading, Mass.: Addison-Wesley.
- Krolak-Schwerdt, S. (1991). Modelle der dreimodalen Faktorenanalyse: formale Eigenschaften, theoretische Zusammenhänge und ihre Implikationen für das Konzept individueller Differenzen. *Psychologische Beiträge*, 33, 314-346.
- Lanning, K. (1996). Robustness is not dimensionality: On the sensitivity of component comparability coefficients to sample size. *Multivariate Behavioral Research*, 31, 33-46.
- Lautenschlager, G.J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, 24, 365-395.
- Lawley, D.N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 60, 64-82.
- Lawley, D.N. (1941). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 61, 176-185.
- Lawley, D.N. & Maxwell, A.E. (1963). *Factor analysis as a statistical method*. London: Butterworth.
- Lee, H.B. & Comrey, A.L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research*, 14, 301-321.
- Lienert, G.A. & Sporer, S.L. (1982). Interkorrelationen seltener Symptome mittels Nullfeld-korrigierter Yule-Koeffizienten. *Psychologische Beiträge*, 24, 411-418.
- Linn, R.L. (1968). A Monte Carlo approach to the number of factor problem. *Psychometrika*, 33(1), 37-72.
- Lynn, R., Hampson, S. & Lee, M. (1988). The intelligence of chinese children in Hong Kong. *School Psychology International*, 9, 29-32.
- Macrosson, W.G.K. (1999). Scree plots, data structure, and random variance. *Psychological Reports*, 84, 533-540.
- Matthews, G. & Oddy, K. (1993). Recovery of major personality dimensions from trait adjective data. *Personality and Individual Differences*, 15, 419-431.
- McCrae, R.R. & Costa, P.T., Jr. (1986). Clinical assessment can benefit from recent advances in personality psychology. *American Psychologist*, 41, 1001-1003.
- McCrae, R.R. & Costa, P.T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- McCrae, R.R. & Costa, P.T., Jr. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, 49, 710-721.
- McCrae, R.R. & Costa, P.T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- McCrae, R.R. & Costa, P.T., Jr. (1995). Positive and negative valence within the five-factor model. *Journal of Research in Personality*, 29, 443-460.
- McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H. & Paunonen, S.V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, 70, 552-566.
- McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale: Lawrence Erlbaum Associates.
- McDonald, R.P. & Ahlward, K.S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99.
- Mershon, B. & Gorsuch, R.L. (1988). Number of factors in the personality sphere: does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology*, 55, 675-680.
- Michel, L. & Conrad, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K.J. Groffmann & L. Michel (Eds.), *Grundlagen pädagogischer Diagnostik (Ezyklopädie der Psychologie)* (Vol. Bd. B. II.1, pp. 1-129). Göttingen: Hogrefe.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W. & Peake, P.K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review*, 89, 730-755.

- Mischel, W. & Peake, P.K. (1983). Some facets of consistency: Replies to Epstein, Funder & Bem. *Psychological Review*, 90, 394-402.
- Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11, 3-31.
- Montanelli, R.G., Jr. & Humphreys, L.G. (1976). Latant roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, 41, 1976.
- Moosbrugger, H. & Frank, D. (1992). *Clusteranalytische Methoden in der Persönlichkeitsforschung*. Bern: Huber.
- Mosier, C.I. (1943). On the reliability of a weighted composite. *Psychometrika*, 8, 161-168.
- Mulaik, S.A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Netter, P., Hennig, J. & Rohrmann, S. (1999). Psychobiological differences between the aggression and psychoticism dimension. *Pharmacopsychiatry*, 32, 5-12.
- Nunnally, J.C. (1978). *Psychometric theory* (2. Aufl.). New York: McGraw-Hill.
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O. & Wittmann, W.W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences*, 29, 1017-1045.
- Ones, D. S. & Viswesvaran, C. . (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, 17, 609-626.
- Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur. Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit*. Regensburg: S. Roderer Verlag.
- Ostendorf, F. & Angleitner, A. (1994). A comparison of different instruments proposed to measure the big five. *European Review of Applied Psychology*, 44, 45-53.
- Ostendorf, F. & Angleitner, A. (in Druck). *NEO-Persönlichkeitsinventar nach Costa und McCrae revidierte Form (NEO-PI-R)*. Göttingen: Hogrefe.
- Overall, J.E. & Klett, C.J. (1972). *Applied multivariate analysis*. New York: McGraw-Hill.
- Parkes, K.R. (1985). Dimensionality of Rotter's Locus of Control Scale: an application of the 'Very Simple Structure' technique. *Personality and Individual Differences*, 6, 115-119.
- Paunonen, S.V. (1997). On chance and factor congruence following orthogonal procustes rotation. *Educational and Psychological Measurement*, 57, 33-59.
- Paunonen, S.V., Jackson, D.N., Trzebinski, J. & Forsterling, F. (1992). Personality structure across culture: A multimethod evaluation. *Journal of Personality and Social Psychology*, 62, 447-456.
- Pawlik, K. (1968). *Dimensionen des Verhaltens. Eine Einführung in die Methodik und Ergebnisse faktorenanalytischer psychologischer Forschung*. Bern: Hans Huber.
- Raykov, T. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences*, 24, 291-293.
- Revelle, W. & Rocklin, T. (1979). Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14, 403-414.
- Revenstorff, D. (1976). *Lehrbuch der Faktorenanalyse*. Stuttgart: Kohlhammer.
- Revenstorff, D. (1980). *Faktorenanalyse*. Stuttgart: Kohlhammer.
- Rodenhausen, T. (1999). Der 16-Persönlichkeits-Faktoren-Test Revidierte Fassung (16 PF-R). In Fay, E. (Ed.), *Tests unter der Lupe II* (S. 116-129). Lengerich: Pabst Science Publishers.
- Rost, J., Carstensen, C.H. & von Davier, M. (1999). Sind die Big Five Rasch-skalierbar? Eine Reanalyse der NEO-FFI-Normierungsdaten. *Diagnostica*, 45, 119-127.
- Rozeboom, W.W. (1991). Hyball: A method for subspace-constrained factor rotation. *Multivariate Behavioral Research*, 26, 163-177.
- Rozeboom, W.W. (1991). Theory & practice of analytic hyperplane optimization. *Multivariate Behavioral Research*, 26, 179-197.
- Ruch, W. (1999). Die revidierte Fassung des Eysenck Personality Questionnaire und die Konstruktion des deutschen EPQ-R bzw. EPQ-RK. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20, 1-24.
- Saucier, G. . (1997). Effects of Variable Selection on the Factor Structure of Person Descriptors. *Journal of Personality and Social Psychology*, 73, 1296-1312.
- Saucier, G. & Goldberg, L.R. (1998). What is beyond the big five? *Journal of Personality*, 66, 495-524.
- Saucier, G. & Ostendorf, F. . (1999). Hierarchical Subcomponents of the Big Five Personality Factors: A Cross-Language Replication. *Journal of Personality and Social Psychology*, 76, 613-627.
- Saunders, D.R. (1962). Transvarimax: Some properties of the ratiomax and equamax criteria for blind orthogonal rotation. Paper delivered at the American Psychological Association meeting.
- Schmid, J. & Leiman, J.M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61.

- Schmitz, B. (1993). Stabilität als Metatrait? Zur Erfassung von intraindividuelle Stabilität als Teilkonstrukt der Koonsistenz. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 1-26.
- Schmitz, B. (2000). Auf der Suche nach dem verlorenen Individuum: Vier Theoreme zur Aggregation von Prozessen. *Psychologische Rundschau*, 51, 83-92.
- Schneewind, K.A. & Graf, J. (1998). *Der 16-Persönlichkeits-Faktoren-Test. Revidierte Fassung*. Göttingen: Verlag Hans Huber.
- Schönemann, P.H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31, 1-10.
- Schönemann, P.H. (1981). Factorial definitions of intelligence: Dubious legacy of dogma in data analysis . In I. Borg (Ed.), *Multidimensional data representations: When and why*. Ann Arbor, MI: Mathesis Press.
- Schönemann, P.H. & Wang, M.M. . (1972). Some new results on factor indeterminacy. *Psychometrika*, 37, 61-91.
- Schneider, R.J., Hough, L.M. & Dunnette, M.D. (1996). Broad-sided by broad traits: how to sink science in five dimension or less. *Journal of Organizational Behavior*, 17, 639-655.
- Schweizer, K. (2001). The MR- and other criteria for orthogonal rotation of factors. *Methods of Psychological Research Online*, 6, 1-21.
- Schweizer, K., Boller, E. & Braun, G. (1996). Der Einfluß von Klassifikationsverfahren, Stichprobengröße und strukturellen Datenmerkmalen auf die Klassifizierbarkeit von Variablen. *MPR-online: Methods for Psychological Research*, 1(4), 93-105.
- Smedslund, J. (1987). The epistemic status of inter-item correlations in Eysenck's Personality Questionnaire: the a priori versus the empirical in psychological data. *Scandinavian Journal of Psychology*, 28, 42-55.
- Snook, S.C. & Gorsuch, R.L. (1989). Component Analysis Versus Common Factor Analysis: A Monte Carlo Study. *Psychological Bulletin*, 106, 148-154.
- Spearman, C. (1927). *The abilities of Man: Their nature and measurement*. New York: Macmillan.
- SPSS, Inc. (1999). SPSS for Windows, Release 9.0. Chicago: Author.
- Statistisches Bundesamt. (1999). *Statement von Präsident Johann Hahlen zum Pressegespräch "Hochschulstandort Deutschland"*.
- Statistisches Bundesamt. (2001). *Informationelle Grundversorgung*. Wiesbaden: Statistisches Bundesamt.
- Stelmack, R.M. (1997). Toward a Paradigm in Personality: Comment on Eysenck's (1977) View. *Journal of Personality and Social Psychology*, 73, 1238-1241.
- Stelmack, R.M. & Pivik, R.T. (1996). Extraversion and the effects of exercise on spinal motoneuronal excitability. *Personality and Individual Differences*, 21, 69-76.
- Sternberg, R.J. & Powell, J.S. (1982). Theories of intelligence. In Sternberg, R. J. (Ed.), *Handbook of human intelligence*. Cambridge: University Press.
- Stewart, D. & Love, W. (1968). A general canonical correlation index. *Psychological Bulletin*, 70, 160-163.
- Streiner, D.L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry*, 39, 135-140.
- Strobel, A., Beauducel, A., Debener, S. & Brocke, B. (in Druck). Psychometrische und strukturelle Merkmale einer deutschsprachigen Version des BIS/BAS-Fragebogens von Carver und White. *Zeitschrift für Differentielle und Diagnostische Psychologie*.
- Strobel, A., Wehr, A., Michel, A. & Brocke, B. (1999). Association between the dopamine D4 receptor (DRD4) exon III polymorphism and measures of Novelty Seeking in a German population. *Molecular Psychiatry*, 4, 378-384.
- Tabachnick, B.G. & Fidell, L.S. (1989). *Using multivariate statistics (2nd Ed.)*. New York: Harper & Row.
- Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorder. *Psychological Inquiry*, 4, 122-130.
- Tellegen, A. (1996). *Multidimensional Personality Questionnaire*. Minneapolis: University of Minnesota Press.
- Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: Chicago University Press.
- Thurstone, L.L. (1940). Experimental study of simple structure. *Psychometrika*, 5, 153-168.
- Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Thurstone, L.L. & Thurstone, T.G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.
- Todt, E. & Fürntratt, E. (1963). Einige technische Hilfsmittel für die Eichung von Leistungsverfahren. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 10, 540-553.
- Trendafilov, N.T. (1994). A simple method for procrustian rotation in factor analysis using majorization theory. *Multivariate Behavioral Research*, 29, No. 4, 385-408.
- Trendafilov, N.T. (1996). Iterative majorizing rotation to orthogonal simple structure solution. *Multivariate Behavioral Research*, 31, 271-287.

- Tucker, L.R. (1960). *Determination of generalized learning curves by factor analysis*. In ONR Technical Report (Ed.), Princeton, N.J.: Educational Testing Service.
- Tucker, L.R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederikson & H. Gulliksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart and Winston.
- Turner, N.E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, 58, 541-568.
- Tucker, L.R., Koopman, R.F. & Linn, R.L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34, 421-460.
- Überla, K. (1968). *Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler*. New York: Springer.
- Underwood, B.J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, 30, 128-134.
- van Heck, G.L., Perugini, M., Caprara, G.V. & Fröger, J. (1994). The Big Five as tendencies in situations. *Personality and Individual Differences*, 16, 715-731.
- Vassend, O. & Skrondal, A. (1997). Validation of the NEO Personality Inventory and the five-factor model. Can findings from exploratory and confirmatory factor analysis be reconciled? *European Journal of Personality*, 11, 147-166.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Velicer, W.F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *Multivariate Behavioral Research*, 12, 3-22.
- Velicer, W.F., Eaton, C.A. & Fava, J.L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R.D. Goffin & Helmes, E. (Eds.), *Problems and solutions in human assessment. Honoring Douglas N. Jackson at Seventy (p. 41-71)*. Boston: Kluwer Academic Publishers.
- Velicer, W.F. & Fava, J.L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3, 231-251.
- Velicer, W.F. & Jackson, D.N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25, 1-28.
- Walkey, F.H. & McCormick I.A. (1985). Multiple replication of factor structure: A logical solution to the number of factors problem. *Multivariate Behavioral Research*, 20, 57-67.
- Waller, N.G., Tellegen, A., McDonald, R.P. & Lykken, D.T. (1996). Exploring Nonlinear Models in the Personality Assessment: Development and Preliminary Validation of the Negative Emotionality Scale. *Journal of Personality*, 64, 545-576.
- Watson, D., Clark, L.A. & Harkness, A.R. (1994). Structures of personality and their relevance to psychopathology. *Journal of Abnormal Psychology*, 103, 18-31.
- Westmeyer, H. (1992). *The structuralist program in psychology: Foundations and applications*. Toronto: Hogrefe & Huber Publishers.
- Westmeyer, H. (1995). Persönlichkeitspsychologie zwischen Realismus und Konstruktivismus. In Pawlik, K. (Ed.), *Bericht über den 39. Kongreß der Deutschen Gesellschaft für Psychologie in Hamburg 1994* (pp. 748-753). Göttingen: Hogrefe.
- Weyers, P., Krebs, H. & Janke W. (1995). Reliability and construct validity of the German version of Cloninger's tridimensional personality questionnaire. *Personality and Individual Differences*, 19, 853-861.
- Wherry, R.I. & Gaylord, R.H. (1944). Factor pattern of test items and tests as a function of the correlation coefficient: content, difficulty and constant error factors. *Psychometrika*, 9, 237-244.
- Widaman, K.F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28, 263-311.
- Wiggins, J.S. & Trapnell, P.D. (1997). Personality structure. The return of the big five. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of personality psychology (S. 737-765)*. New York: Academic Press.
- Wilson, G.D., Gray, J.A. & Barrett, P.T. (1990). A factor analysis of the Gray Wilson Personality Questionnaire. *Personality and Individual Differences*, 11, 1037-1045.
- Wittmann, W. & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik symmetry. In P.L. Ackerman, P.C. Kyllonen & R.D.E. Roberts (Eds.), *Learning and individual differences. process, trait, and content determinants* (pp. 77- 104).. Washington, D.C.: APA Books.

- Wittmann, W.W. (1978). Drei Klassen verschiedener faktorenanalytischer Modelle und deren Zusammenhang mit dem Konzept der Alpha-Generalisierbarkeit der klassischen Testtheorie. *Psychologische Beiträge*, 20, 456-470.
- Wittmann, W.W. (1987). Grundlagen erfolgreicher Forschung in der Psychologie: Multimodale Diagnostik, Multiplismus, multivariate Reliabilitäts- und Validitätstheorie. *Diagnostica*, 33, 209-226.
- Wittmann, W.W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J.R. Nesselroade & R.B. Cattell (Eds.), *Handbook of multivariate experimental psychology*. 2nd ed (pp. 505-560). New York: Plenum.
- Wittmann, W.W. & Hampel, R. (1976). Faktorenanalyse, Rotationsproblem und Implikationen für die Testkonstruktion. *Archiv für Psychologie*, 128, 219-239.
- Wittmann, W.W. & Süß, H.M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik symmetry. In P.L. Ackerman, P.C. Kyllonen & R.D.E. Roberts (Eds.), *Learning and individual differences. process, trait, and content determinants* (pp. 77- 104).. Washington, D.C.: APA Books.
- Wood, C.C. & McCarthy, G. (1984). Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. *Electroencephalography and clinical Neurophysiology*, 59, 249-260.
- Wrigley, C. (1960). A procedure for objective factor analysis. *Paper presented at the first annual meeting of the Society of Multivariate Experimental Psychology*.
- Wrigley, C. & Neuhaus, J.O. (1955). The matching of two sets of factors. Contract Report, No. A-32, Task A. Urbana Illinois: University of Illinois. In Zitiert nach H.H. Harman (1967), *Modern factor analysis* ((2. Aufl.) Chicago: University of Chicago Press.
- Yeomans, K.A. & Golder, P.A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. *Statistician*, 31, 221-229.
- Zelenski, J.M. & Larsen, R.J. (1999). Susceptibility to affect: A comparison of three personality taxonomies. *Journal of Personality*, 67, 761-791.
- Zoski, K.W. & Jurs, S. (1993). Using multiple regression to determine the number of factors to retain in factor analysis. *Journal of Multiple Linear Regression Viewpoints*, 20, 5-9.
- Zoski, K.W. & Jurs, S. (1996). An objektive counterpart to the visual scree test for factor analysis: The standard error scree. *Educational ans Psychological Measurement*, 56, 443-451.
- Zuckerman, M. (1989). Personality in the third dimension: A psychobiological approach. *Personality and Individual Differences*, Vol. 10, No. 4, 391-418.
- Zuckerman, M. (1990). The psychophysiology of sensation seeking. *Journal of Personality*, 58,1, 313-345.
- Zuckerman, M. (1992). What is a basic factor and which factors are basic? Turtles all the way down. *Personality and Individual Differences*, 13(6), 675-681.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge: Cambridge University Press.
- Zuckerman, M. (1994). Impulsive Unsocialized Sensation Seeking: The biological foundations of a basic dimension of personality. In J.E. Bates & Th.D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 219-255). Washington, D.C.: APA.
- Zuckerman, M., Eysenck, H.J. & Eysenck, S.B.G. (1978). Sensation Seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology*, 46, 139-149.
- Zuckerman, M., Kuhlman, D.M. & Camac, C. (1988). What lies beyond E and N? Factor analyses of scales believed to measure basic dimensions of personality. *Journal of Personality and Social Psychology*, 54, 96-107.
- Zuckerman, M., Kuhlman, D.M., Joireman, J., Teta, P. & Kraft. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of Personality and Social Psychology*, 65, 757-768.
- Zuckerman, M., Kuhlman, D.M., Thornquist, M. & Kiers, H. (1991). Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences*, 12, 929-941.
- Zuckerman, M., Kuhlman, M. & Thornquist, M. (1991). Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences*, 12(9), 929-941.
- Zwick, W.R. & Velicer, W.F. (1980). Factors influencing four rules for determining the number of components to retain. *Unpublished, University of Rhode Island*.
- Zwick, W.R. & Velicer, W.F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, 17, 253-269.

- Zwick, W.R. & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.
- Zysno, P.V. (1997). Die Modifikation des Phi-Koeffizienten zur Aufhebung seiner Randverteilungsabhängigkeit. *Methods for Psychological Research-online*, 1, 41-53.

Erklärung

Ich versichere, dass ich diese Habilitationsschrift und die anderen vorgelegten wissenschaftlichen Veröffentlichungen selbst und ohne andere als die darin angegebenen Hilfsmittel angefertigt sowie die wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Dresden, 8. November 2001