

## Reliabilität und Validität von Fragebogenskalen bei Standardreihenfolge versus inhaltshomogener Blockbildung ihrer Items

Günter Krampen\*, Harald Hense\* und Johann F. Schneider\*\*

Universität Trier, FB I — Psychologie\*

und

Universität des Saarlandes, FB 6 — Psychologie\*\*

Empirische Befunde aus zwei Studien zu der seit Jahren verbreiteten, zumeist mit ökonomischen Argumenten begründeten Praxis, Subskalen aus umfassenderen Persönlichkeitsfragebogen zu extrahieren und die Items dieser Subskalen dann als mehr oder weniger inhaltshomogenen Itemblock isoliert zur Beantwortung vorzugeben, werden dargestellt. Exemplarisch für die Halbform A des „Freiburger Persönlichkeitsinventars“ (FPI-A) wurden zwei Datenerhebungen (Gesamt-N = 213) in unterschiedlichen Applikationskontexten (kontrollierte versus nicht kontrollierte Gruppenerhebung) realisiert, in denen neben (1) einem Vergleich der Standard-Fragebogeninstruktion mit einer „soziale Erwünschtheit“ provozierenden Instruktion (als Kontrollbedingung) (2) die Standardreihenfolge der Items im Gesamtfragebogen mit der Darbietung der Items der einzelnen Subskalen im Block (inhaltshomogene Blockbildung), die der Extraktion von Subskalen aus umfassenderen Instrumenten entspricht, verglichen wurde. Die an randomisierten Vergleichsstichproben gewonnenen Resultate belegen signifikante Effekte der Itemblockbildung auf die interne Konsistenz der Skalen, die Skalenmittelwerte und die Skaleninterkorrelationen. Die Befunde ziehen die Zulässigkeit der Subskalenextraktion in Zweifel, da sich die Reliabilitäts- und Validitätsparameter der Skalen z. T. erheblich verändern. Auf alternative Möglichkeiten, dem Bedarf nach „ökonomischen“ Erhebungsinstrumenten in der Persönlichkeitsdiagnostik und -forschung anhand eines rationaleren Vorgehens nachzukommen, wird eingegangen.

Seit Jahren, wenn nicht seit Jahrzehnten, ist es in verschiedensten Forschungs- und Anwendungsbereichen der Psychologie nahezu üblich geworden, Subskalen aus umfassenderen Persönlichkeitsfragebogen zu extrahieren und die Items dieser Subskalen dann als mehr oder weniger inhaltshomogene Itemblocks isoliert zur Beantwortung vorzugeben. Zumeist wird dieses

Vorgehen weder begründet noch reflektiert oder kritisiert, wodurch es unter der Hand akzeptiert wird und mit der Zeit geradezu zu einem Bestandteil des anerkannten Methodeninventars der Psychologie werden kann (evtl. schon geworden ist). Hin und wieder werden ökonomische Gründe angeführt, die sich darauf beziehen, daß das in Frage stehende Persönlichkeitskonstrukt aus Plausibilitätserwägungen (häufiger) oder aus theoretischen Gründen (seltener) im Rahmen der jeweiligen Untersuchung interessiere, jedoch nicht im Rahmen eines umfangreichen Fragebogens erhoben werden könne, da dies die „eigentliche“, ohnehin zumeist für die Probanden aufwendige Untersuchung zu sehr aufblähen würde. Methodische Aspekte des Vorgehens werden somit auch beim Rückgriff auf ökonomische Begründungen kaum reflektiert.

Nun ist die Durchdringung der gesamten Psychologie durch differential- und persönlichkeitspsychologische Konzepte und Ansätze sicherlich zu begrüßen (vgl. hierzu etwa Graumann, 1973, S. 29, der in seinem Bericht „Zur Lage der Psychologie“ auf dem 27. Kongreß der DGPs bereits 1970 als Positivum konstatierte, daß „die Psychologie (...) nicht nur ‚kognitiver‘ und ‚sozialer‘, (sondern) auch differentieller geworden (ist).“). Andererseits wird etwa nicht nur von Herrmann und Lantermann (1985) sowie Pervin (1985) die zunehmende Fragmentierung differential- und persönlichkeitspsychologischer Ansätze in der Forschung beklagt, die sich darin zeigt, daß eine Konzentration auf einige wenige, aktuelle, „moderne“ Konstrukte stattfindet, und die zugleich kaum durch wesentliche theoretische Neuentwürfe oder Weiterentwicklungen abgefangen wird (vgl. hierzu auch Krampen, 1987a). Auch die häufig zu beobachtende Extraktion von Subskalen aus Persönlichkeitsinventaren ist ein Kennzeichen dieser Fragmentierung, bei der aus bestehenden, teststatistisch abgesicherten Persönlichkeitsdiagnostika schlicht und einfach das entnommen wird, was im Bereich der Persönlichkeitsvariablen im Zusammenhang mit der jeweiligen Fragestellung interessieren könnte.

Während bei der Erstellung oder Modifikation eines Interviewbogens die Vernachlässigung der Fragenabfolge, möglicher Ausstrahlungseffekte von Fragen und der Itemkontexte auf dem Hintergrund zahlreicher empirischer Befunde zu ihrer enormen Bedeutung für die Qualität der Daten (vgl. im Überblick etwa Angerer, 1969; Gutjahr, 1985; Scheuch, 1967) als methodischer Kunstfehler verstanden wird, finden sich entsprechende Untersuchungen im Bereich der Fragebogendiagnostik kaum. Eine Ausnahme ist die Arbeit von Schriesheim, Kopelman und Solomon (1989), die im Vergleich inhaltshomogener Itemblocks mit der (verteilten) Item-Standardabfolge für den „Job Diagnostic Survey“ und einen Lebenszufriedenheit-Fragebogen einige Unterschiede in den Reliabilitäts- und Validitätskennwerten feststellten und den Nutzen der Itemblockbildung anhand ihrer Befunde in Frage

stellen. Im deutschen Sprachraum legte lediglich Deusinger (1986) im Rahmen der „Frankfurter Selbstkonzeptskalen“ (FSKN) neben einer Fragebogengesamtversion zugleich „FSKN-Einzelversionen“ vor, die aus jeweils allen Items einer FSKN-Subskala bestehen und im Unterschied zur FSKN-Gesamtversion, die eine nach Subskalen gemischte Itemabfolge aufweist, als inhaltshomogener Block zur Beantwortung vorgegeben werden. Deusinger (1986) erleichtert damit nicht nur die Praxis der Subskalen-Extraktion, sondern propagiert sie geradezu.

Zweifel an dieser Praxis entstehen, da bei Deusinger (1986) dies zunächst mit einem Verweis auf die Untersuchung von Sax und Cromack (1966) begründet wird, wobei allerdings verschwiegen wird, daß sich diese Arbeit *nicht* auf Persönlichkeitsdiagnostik über Fragebogen, sondern auf Leistungsdiagnostik über ein objektives Testverfahren und die Frage der Itemanordnung (nach Itemschwierigkeit) dabei bezieht. Zu betonen ist auch, daß selbst bei Intelligenztests die Frage, ob die Abfolge der Testaufgaben einen Einfluß auf die Daten hat, nicht eindeutig geklärt ist (vgl. hierzu nicht nur Sax & Cromack, 1966, sondern etwa auch French & Greer, 1964; Sax & Carr, 1962). Zusätzlich führt Deusinger (1986) ihre Strategie stützende Untersuchungsbefunde zu Skalenmittelwertsvergleichen bei Vorgabe der Gesamt- versus Einzelversionen der FSKN an, deren Ergebnisse allerdings auch Zweifel offenlassen (vgl. hierzu Krampen, 1988), zumal „es dennoch für möglich (gehalten wird), daß nachfolgende Forschung u. U. belegen kann, daß mit der Anwendung der Einzelversion eine Tendenz zur ‚positiven‘ Selbstdarstellung eher wirksam wird als bei der Anwendung der Gesamtversion“ (Deusinger, 1986, S. 46). Es stellt sich zudem die Frage, ob bei der Äquivalenzprüfung von Gesamtversion und Einzelversionen (mit Itemblocks) Skalenmittelwertsvergleiche ausreichen und nicht auch andere teststatistische Parameter, die sich etwa auf die Reliabilität und Validität beziehen, heranzuziehen sind.

Wenngleich die Interpretationen von Deusinger (1986) dafür sprechen, daß sich in Mittelwertsvergleichen der Skalenwerte aus Fragebogengesamtversionen und Einzelversionen (bei homogenen Itemblocks) „eher“ Tendenzen in Richtung einer positiven Selbstdarstellung in den Einzelversionen zeigen werden, soll im folgenden die Hypothese der Mittelwertsäquivalenz zweiseitig geprüft werden. Die inhaltliche Homogenisierung der Items bei Blockbildung kann nicht nur Tendenzen zur sozialen Erwünschtheit verstärken, sondern auch durch die häufige Wiederholung ähnlicher Frageinhalte und die thematische Perserveration Reaktanz und damit Widerspruch erzeugen oder aber — genau umgekehrt — zur Homogenisierung der Antworten und damit zur Erhöhung der internen Konsistenz beitragen. Aus den gleichen Gründen sollten auch die inferenzstatistischen Prüfungen der Äquivalenz von Fragebogen-Standardversion und inhaltshomogenen

Itemsblocks unter den Aspekten der internen Konsistenz (Reliabilität) und der Interkorrelationen der Subskalen (Konstruktvalidität) zweiseitig vorgenommen werden.

Diesen Fragen nach der Äquivalenz einer Fragebogengesamtversion mit „Einzelversionen“, die aus inhaltshomogenen Blocks der Items einzelner Subskalen bestehen, wurde in den im folgenden dargestellten Studien *exemplarisch* für das im deutschen Sprachraum weit verbreitete und teststatistisch gut abgesicherte „Freiburger Persönlichkeitsinventar“ (FPI, Halbform A; Fahrenberg, Selg & Hampel, 1973) nachgegangen. Als experimentelle Bedingung wurde neben (1) der Standardform des FPI-A mit der gemischten Itemabfolge versus der Darbietung der Items der einzelnen Primärskalen im Block (inhaltshomogene Blockbildung) als *Kontrollbedingung* zusätzlich (2) die Standardinstruktion des FPI mit einer „soziale Erwünschtheit“ provozierenden Instruktion realisiert. Ebenso wie bei Hampel und Klinkhammer (1978) wurde dabei mit der Vorstellung einer Bewerbungssituation im beruflichen Bereich, für die massive Sensibilitäten des FPI im Sinne eines „Syndroms (der) sozial erwünschte(n) Persönlichkeit“ von Hampel und Klinkhammer nachgewiesen wurden, gearbeitet. Zwischen den Studien wurde zusätzlich der Applikationskontext variiert. Während in Studie I die Fragebogen nach dem Verlesen der jeweiligen Instruktion in (instruktionshomogenen) Kleingruppen bei Anwesenheit eines Testleiters bearbeitet wurden, wurden sie in Studie II in (instruktionsheterogenen) Großgruppen ohne Vorlesen der Instruktion bei (passiver) Anwesenheit eines Testleiters ausgefüllt. Zusatzfragen bezogen sich auf das Selbsterleben bei der Fragebogenbearbeitung und die Einstellungen gegenüber dem Fragebogen.

## Studie I

### *Methode*

#### Erhebungsinstrumente

Neben der Standardform des FPI-A wurden drei modifizierte (experimentelle) Versionen des FPI-A verwendet, von denen zwei über eine „Bewerbungsinstruktion“ (vgl. Hampel & Klinkhammer, 1978) und zwei aus der Darbietung der Items jeder FPI-A-Primärskala in (inhaltshomogenen) Blocks bestanden. Die Items der inhaltshomogenen Subskalen wurden in der Originalreihenfolge der FPI-Primärskalen vorgegeben. Damit wurde ein experimentelles 2 (Standard- versus Bewerbungsinstruktion)  $\times$  2 (Standardabfolge versus Blockbildung der Items)-„Design“ realisiert, dem die *Standardversion und drei experimentelle Versionen des FPI-A* zugeordnet sind.

Im Anschluß an die Bearbeitung der jeweiligen FPI-Version wurden Daten mit *Zusatzfragen* erhoben: gefragt wurde nach (1) der (geschätzten) Bearbeitungszeit, (2) der beim Ausfüllen des Fragebogens erlebten Langeweile, (3) den Schwierigkeiten bei der Entscheidung zwischen den beiden Antwortmöglichkeiten bei den Items, (4) den Verständnisschwierigkeiten bei den Items, (5) der Einschätzung, ob man den Fragebogen in einer anderen Situation anders beantwortet hätte, (6) dem Eindruck, ob man wesentliche eigene Charaktereigenschaften in den Items wiedergefunden hat, (7) der „ablehnenden“ versus „aufgeschlossenen“ Einstellungen gegenüber solchen Fragebogen, (8) eigenen Überlegungen zur Absicht des Fragebogens während der Bearbeitung und (9) dem Eindruck, daß man den Fragebogen so beantwortet hat, daß man in einem „günstigen Licht“ erscheint. Für die Fragen (2) bis (9) wurden jeweils sechsstufige Antwortskalen vorgegeben.

Die Datenerhebungen wurden anonym durchgeführt. Soziodemographische Variablen (Geschlecht, Alter, Schulabschluß, Familienstand, Wohnsituation und Beruf) wurden nach dem Standard des FPI erfaßt.

### Stichprobe

An Studie I waren 129 Studierende des zweiten Fachsemesters Psychologie (77 Studentinnen und 52 Studenten;  $M = 22.8$  Jahre,  $SD = 2.8$ ) in acht Kleingruppen zu 13 bis 19 Personen beteiligt (Totalerhebung). Nach dem Zufall wurden die Standardversion des FPI-A und die drei FPI-A-Experimentalversionen jeweils zwei Kleingruppen zugewiesen. Die Datenerhebungen fanden in den Kleingruppen unter Anwesenheit eines Testleiters statt, nachdem die jeweilige Fragebogeninstruktion (Stand- versus Bewerbungsinstruktion) laut verlesen worden war. Die Teilnahme an der Untersuchung war freiwillig (Verweigerungen traten nicht auf) und wurde als eine halbe „Versuchspersonen-Stunde“ im Rahmen der im Vordiplomstudium von Hauptfachstudenten zu erbringenden Untersuchungsteilnahmen angerechnet.

### Ergebnisse

In Tabelle 1 sind die Befunde der varianzanalytischen Mittelwertvergleiche zusammengefaßt. Zunächst ist festzustellen, daß die Mittelwerte der FPI-Primärskalen und -Sekundärskalen bei der Standardversion des FPI-A nahezu vollständig mit den von Fahrenberg et al. (1973) angegebenen Normen für 15- bis 30jährige Frauen und Männer übereinstimmen (jeweils Stabine-Wert = 5). Eine Ausnahme ist die Skala FPI-9 „Offenheit“, auf der die Psychologiestudenten einen im Vergleich zur Normstichprobe deutlich

Tabelle 1

Ergebnisse der (multivariaten) Varianzanalysen mit den Faktoren „Instruktion“, „Itemreihenfolge“ und „Skalen“ mit Meßwiederholung auf dem letzten Faktor für die FPI-Primärskalen, FPI-Sekundärskalen und die Zusatzfragen (Studie I; N = 129)

Abhängige Variable(n)	Mittelwerte				Varianzanalytische Effekte (F-Werte)							
	Standardinstruktion		Bewerbungsinstrukt.		Instruktion (A)	Item-abfolge (B)	Skalen (C)	Interaktionseffekte				
	Stand. Folge	Block	Stand. Folge	Block				AxB	AXC	BxC	AxBxC	
FPI-1: Nervosität	5.8	4.9	3.3	2.4								
FPI-2: Aggressivität	5.1	4.4	3.7	<b>3.0</b>								
FPI-3: Depressivität	7.1	5.5	4.6	<b>3.0</b>								
FPI-4: Erregbarkeit	4.2	3.0	3.0	1.9								
FPI-5: Geselligkeit	8.4	8.9	9.9	10.4	37.4**	11.7**	109.8**	0.2	7.5**	1.5	0.7	
FPI-6: Gelassenheit	3.9	4.5	5.2	<b>6.0</b>								
FPI-7: Dominanzstr.	2.4	2.3	2.1	<b>2.0</b>								
FPI-8: Gehemtheit	4.7	4.1	3.5	<b>3.0</b>								
FPI-9: Offenheit	4.7	4.2	3.5	<b>3.0</b>								
FPI-E: Extraversion	7.3	7.9	7.7	<b>7.9</b>								
FPI-N: emot. Labilität	5.7	4.4	3.6	2.4	4.3*	3.9*	44.9**	0.1	18.8**	2.6	0.8	
FPI-M: Maskulinität	6.2	4.4	3.6	2.4								
Zeitschätzung	9.5	10.2	11.5	<b>12.2</b>	4.4*	0.6	—	0.3	—	—	—	
Langeweile	3.6	3.9	3.4	<b>3.8</b>								
Entscheidungsprobleme	3.0	3.8	2.8	<b>3.7</b>								
Verständnisprobleme	1.7	1.6	1.8	1.7								
Situationsabhängigkeit	2.9	3.6	3.4	4.1	0.8	1.5	56.4**	0.0	4.2**	2.2**	1.4	
Selbsterkennen	3.7	4.2	3.3	3.4								
Einstellung zu Frageb.	3.5	3.2	3.0	<b>2.6</b>								
Fragebogenabsicht	3.7	4.8	4.1	<b>4.5</b>								
Soziale Erwünschtheit	3.4	4.1	4.3	4.4								

\*\*p &lt; .01, \*p &lt; .05

niedrigeren Wert (Stanine-Wert = 2) aufweisen. Die daraus resultierenden Bedenken an der Interpretierbarkeit der Daten müssen im folgenden beachtet werden.

Tabelle 1 informiert in ihrem oberen Teil zunächst über die Ergebnisse einer multivariaten Varianzanalyse (MANOVA) mit den Faktoren „Instruktion“ (Standard- versus Bewerbungsinstruktion), „Itemreihenfolge“ (gemischte Standardreihenfolge versus inhaltshomogene Itemsblocks) und „FPI-A-Primärskalen“ bei Meßwiederholung über den letzten Faktor. Nachgewiesen werden signifikante Haupteffekte aller Faktoren und eine statistisch bedeutsame Interaktion der Faktoren „Instruktion“ und „FPI-A-Skalen“. Nachgeschaltete univariate Mittelwertsvergleiche zeigen, daß die „Bewerbungsinstruktion“ zu signifikant niedrigeren Werten auf den FPI-A-Skalen „Nervosität“, „Aggressivität“, „Depressivität“, „Erregbarkeit“, „Gehemmtheit“ und „Offenheit“ sowie signifikant erhöhten Werten auf den Skalen „Geselligkeit“ und „Gelassenheit“ führt. Dies steht voll und ganz in Einklang mit den Ergebnissen von Hampel und Klinkhammer (1978), was die Zweifel an der Verwertbarkeit der Daten aufgrund der reduzierten Werte auf der FPI-Skala „Offenheit“ abschwächt.

Die inhaltshomogene Itemblockbildung führt im Vergleich zur Standardabfolge der Items zu sehr ähnlichen Mittelwertsunterschieden wie die Bewerbungsinstruktion im Vergleich zur Standardinstruktion (siehe Tabelle 1). Auch wenn der Haupteffekt des Faktors „Itemabfolge“ mit  $\omega^2 = .09$  eine geringere Effektstärke aufweist als der des Faktors „Instruktion“ ( $\omega^2 = .19$ ), sprechen die Mittelwertsunterschiede dafür, daß durch die Itemblockbildung die Effekte der Bewerbungsinstruktion verstärkt werden. Der Haupteffekt des Faktors „Skalen“ ist erwartet und bestätigt die diskriminative Validität der FPI-A-Skalen, die signifikante Interaktion von „Instruktion“ und „Skalen“ weist auf die oben bereits beschriebenen, z. T. gegenläufigen Effekte der Bewerbungsinstruktion auf den verschiedenen Primärskalen des FPI.

In einer analogen MANOVA ließen sich für die Sekundärskalen des FPI-A exakt die gleichen signifikanten Effekte nachweisen (siehe Tabelle 1). Der Mittelwert der Skala „Extraversion“ steigt unter der Bewerbungsinstruktion und bei Itemblockbildung an, die Mittelwerte der Skalen „emotionale Labilität“ und „Maskulinität“ liegen dagegen niedriger als unter der Standardinstruktion und Standard-Itemabfolge.

Für die geschätzte Bearbeitungszeit zeigt sich in der univariaten Varianzanalyse lediglich ein signifikanter Haupteffekt der „Instruktion“, der im Unterschied zu den Ergebnissen von Schneider und Hübner (1980) darauf verweist, daß unter der „Bewerbungsinstruktion“ (zumindest subjektiv) mehr Zeit für die Beantwortung des Fragebogens aufgewendet wird als unter der Standardinstruktion. Für die Itemblockbildung zeigt sich im Ver-

gleich zur Standardabfolge der Items in der Tendenz ein ähnlicher Effekt, der jedoch nicht inferenzstatistisch abgesichert ist.

Die MANOVA der acht Zusatzfragen führte zu einem signifikanten Haupteffekt des Faktors „Skalen“, was die Unterschiedlichkeit der Zusatzfragen bestätigt, und zu signifikanten Interaktionen des Faktors „Skalen“ mit den Faktoren „Instruktion“ und „Itemabfolge“ (siehe Tabelle 1). Nachgeschaltete univariate Mittelwertvergleiche zeigen, daß die „soziale Erwünschtheit“ des eigenen Antwortverhaltens unter der Bewerbungsinstruktion als höher eingeschätzt wird als unter der Standardinstruktion. Dies bestätigt empirisch die Realisierung dieser experimentellen Bedingung (im Sinne eines „tests of the fidelity of treatment implementation“) und reduziert damit weiter die Zweifel an der Datenqualität (s.o.). Für die Itemblockbildung zeigen sich leicht erhöhte Einschätzungen (a) der Langeweile bei der Fragebogenbearbeitung, (b) der Entscheidungsprobleme beim Antworten, (c) der Situationsabhängigkeit des Antwortverhaltens und (d) des Nachdenkens über die Meßabsicht. Die allgemeine Einstellung gegenüber Fragebogen ist bei der Itemblockbildung im Vergleich zur Standardabfolge negativer.

Für alle FPI-A-Skalen wurden die Koeffizienten der internen Konsistenz (Cronbachs Alpha) getrennt für die Standform des FPI-A und seine drei Experimentalformen berechnet (siehe Tabelle 2). Unter der Standardinstruktion und -itemabfolge ergaben sich Reliabilitätsparameter, die nahezu exakt den Angaben in der Handanweisung zum FPI (Fahrenberg et al., 1973, S. 36) entsprechen und dabei auch die Reliabilitätsmängel einiger Skalen dieser FPI-Halbform bestätigen. Im Vergleich zur Bewerbungsinstruktion zeigen sich leichte Schwankungen in der internen Konsistenz der Skalen, die aber nach dem W-Test (nach Feldt, 1969; siehe auch Feldt, Woodruff & Salih, 1987) in keinem Fall zu signifikanten Unterschieden gegenüber der Standardinstruktion führen. Anders ist dies beim Vergleich der Itemblockbildung mit der Standardabfolge der Items. Für fünf Skalen zeigen sich statistisch bedeutsame Unterschiede in der internen Konsistenz, was in dieser Anzahl auf dem Hintergrund der zwölf berechneten Vergleiche selbst nicht als Zufallsbefund bewertet werden kann ( $p < .0001$ ; siehe Feild & Armenakis, 1974). Durch die Blockbildung steigt die interne Konsistenz von drei Skalen („Erregbarkeit“, „Gelassenheit“ und „Gehemmtheit“), sie fällt bei zwei Skalen („Nervosität“ und „Geselligkeit“) ab. Die Befunde sprechen damit weder für eine durchgängige Homogenisierung des Antwortverhaltens durch die Itemblockbildung noch für eine durchgängig zu beobachtende Reaktanz- und Widerspruchsneigung durch sie. Die Extraktion der Items von Subskalen aus dem umfassenderen Persönlichkeitsinventar führt somit zu unsystematischen, jedoch signifikanten Veränderungen in den Reliabilitätsparametern der Skalen.

Tabelle 2

Interne Konsistenzen der FPI-Skalen unter (a) Standard- versus Bewerbungsinstruktion und (b) Standardabfolge versus Blockbildung der Items sowie Vergleiche der internen Konsistenzkoeffizienten (Studie I; N = 129)

FPI-A Skala	interne Konsistenz		W-Test	interne Konsistenz		W-Test
	Standardinstr.	Bewerbungsinstr.	F(63/63)	Standarditemfolge	Item-Blockbildung	F(63/64)
FPI-1: Nervosität	.87	.89	1.18	.87	.61	3.00**
FPI-2: Aggressivität	.68	.68	<b>0.02</b>	<b>.69</b>	.62	<b>1.23</b>
FPI-3: Depressivität	.85	.86	<b>1.07</b>	<b>.86</b>	.82	<b>1.28</b>
FPI-4: Erregbarkeit	.77	.76	<b>1.04</b>	<b>.77</b>	.86	1.64*
FPI-5: Geselligkeit	.82	.82	<b>0.03</b>	<b>.83</b>	.74	1.53*
FPI-6: Gelassenheit	.62	.65	<b>1.09</b>	<b>.61</b>	.76	1.63*
FPI-7: Dominanzstreben	.55	.58	<b>1.07</b>	<b>.56</b>	.41	1.34
FPI-8: Gehemmtheit	.74	.76	<b>1.08</b>	<b>.73</b>	.83	1.59*
FPI-9: Offenheit	.74	.75	<b>1.04</b>	<b>.75</b>	.70	1.20
FPI-E: Extraversion	.67	.66	<b>1.04</b>	.67	.60	<b>1.30</b>
FPI-N: emot. Labilität	.79	.81	<b>1.11</b>	.78	.84	<b>1.38</b>
FPI-M: Maskulinität	.59	.59	0.05	.60	.54	1.15

\*\*p < .01, \*p < .05

Als Aspekt der Konstruktvalidität wurde die Interkorrelationsmatrix der FPI-A-Primärskalen analysiert. Wir beschränken uns hier auf die Befunde zu den mittleren (nach Fishers  $z$ -transformierten absoluten) Interkorrelationen, da mit ihnen die wesentlichen Befunde ökonomisch zusammengefaßt werden können. Unter der Standardinstruktion und -itemabfolge zeigt sich für Studie I (siehe Tabelle 3) kein bedeutsamer Unterschied in der mittleren Interkorrelation der FPI-A-Skalen zu den Angaben von Fahrenberg et al. (1973). Unter der Bewerbungsinstruktion steigt der Wert zwar etwas an, der Unterschied ist aber statistisch nicht bedeutsam. Anders verhält es sich beim Vergleich mit der inhaltshomogenen Itemblockbildung: Die mittlere Skaleninterkorrelation liegt hier signifikant höher als unter der Standarditemabfolge (siehe Tabelle 3). Dies verweist darauf, daß durch die Itemblockbildung — bei veränderten Mittelwerten und veränderten Reliabilitätskennwerten — die Konstruktvalidität der FPI-A-Primärskalen im Sinne einer Konvergenz der Skalen, die Konstrukt differenzierungen empirisch verwischt, beeinträchtigt wird.

## Studie II

### *Methoden*

#### Erhebungsinstrument

In Studie II wurde — im Sinne einer Replikationsstudie mit veränderten Datenerhebungsbedingungen (s. u.) — mit exakt den gleichen Erhebungsinstrumenten wie in Studie I gearbeitet.

#### Stichprobe

An Studie II waren 84 Studierende der Psychologie und der Erziehungswissenschaften des 3. und 4. Fachsemesters (57 Studentinnen und 27 Studenten;  $M = 23.6$  Jahre,  $SD = 4.12$ ) in zwei Großgruppen beteiligt. Den Studierenden wurden die vier unterschiedlichen Erhebungsinstrumente in den Großgruppen nach dem Zufall zugeordnet, so daß in einem Raum gleichzeitig alle vier verschiedenen Fragebogenvarianten bearbeitet wurden. Die Datenerhebungen erfolgten anonym, ohne daß die Fragebogeninstruktionen verlesen wurden. Teilnahmeverweigerungen traten nicht auf; die Untersuchungsteilnahme wurde weder finanziell noch durch „Versuchspersonen-Stunden“ o. ä. vergütet.

Tabelle 3

Mittlere Interkorrelationen der FPI-Primärskalen unter (a) Standard- versus Bewerbungsinstruktion und (b) Standardabfolge versus Blockbildung der Items in verschiedenen Stichproben sowie Ergebnisse der Vergleiche der z-transformierten Korrelationskoeffizienten

Stichprobe	mittlere Interkorrelation bei			mittlere Interkorrelation bei		
	Standardinstr.	Bewerbungsinstr.	d(z)	Standarditemfolge	Item-Blockbildung	d(z)
Normierungsstichprobe (Fahrenberg et al., 1973)	.32	—	—	.32	—	—
N = 129 (Studie I)	.33	.42	.11	.35	.65	.41**
N = 84 (Studie II)	.31	.44	.16	.35	.59	.31*

\*\*p < .01, \*p < .05

## Ergebnisse

Die Befunddarstellung folgt der von Studie I und wird daher knapper gehalten, wobei vor allem auf abweichende Ergebnisaspekte eingegangen wird. In Tabelle 4 sind zunächst die Befunde der ebenso wie in Studie I berechneten MANOVAs und der ANOVA zusammengefaßt. Für die Standardversion des FPI-A zeigen sich hier keinerlei bedeutsame Abweichungen der Skalenmittelwerte von den FPI-A-Normen für 15- bis 30jährige Frauen und Männer (Stanine-Wert = 5 bei allen Skalen, also — im Unterschied zu Studie I — auch für FPI-9 „Offenheit“). Statistisch bedeutsame Haupteffekte der Faktoren „Instruktion“, „Itemabfolge“ und „Skalen“ bestätigen inhaltlich die Befunde zu den FPI-Primärskalen aus Studie I. Die Effektstärke des Faktors „Instruktion“ beträgt  $\omega^2 = .07$ , die des Faktors „Itemabfolge“  $\omega^2 = .10$ . Während die Interaktion zwischen „Instruktion“ und „Skalen“ nicht nachgewiesen werden kann, erweist sich hier die Interaktion der „Itemabfolge“ mit den „Skalen“ als signifikant. Dies belegt, daß die Itemblockbildung auf den FPI-Primärskalen zu unterschiedlichen (gegenläufigen) Effekten führt, die inhaltlich mit den oben dargestellten Ergebnissen aus Studie I konvergieren.

Mit signifikanten Haupteffekten der Itemabfolge und der Skalen sowie einer signifikanten Interaktion von „Instruktion“ und „Skalen“ werden für die Sekundärskalen des FPI-A ähnliche Effekte nachgewiesen wie in Studie I (siehe Tabelle 4). Das gleiche gilt für die geschätzte Bearbeitungszeit, die statistisch bedeutsam unter der Bewerbungsinstruktion, nur im Trend bei der Itemblockbildung ansteigt (Tabelle 4). Für die Zusatzskalen kann neben dem Haupteffekt der „Skalen“ nur die Interaktion der Faktoren „Skalen“ und „Itemabfolge“ repliziert werden. Wichtiger ist hier, daß unter der Bewerbungsinstruktion (zumindest leicht;  $p = .07$ ) erhöhte Einschätzungen der subjektiv erlebten sozialen Erwünschtheit im eigenen Antwortverhalten abgegeben werden.

Tabelle 5 informiert über die internen Konsistenzen der FPI-A-Skalen unter allen experimentellen Bedingungen. Die Werte der Standardversion entsprechen erneut den Angaben in der Handanweisung (Fahrenberg et al., 1973). Unter der Bewerbungsinstruktion sind nur geringfügige Abweichungen in der internen Konsistenz der Skalen zu beobachten, die sich — ebenso wie in Studie I — in keinem Fall als statistisch bedeutsam erweisen. Für die Vergleiche der Standarditemfolge mit der Itemblockbildung zeigen sich dagegen sechs signifikante Unterschiede in den Reliabilitätsparametern, die nur z. T. mit den in Studie I nachgewiesenen konvergieren.

Die Befunde zu den mittleren Interkorrelationen der FPI-A-Primärskalen entsprechen denen aus Studie I (siehe Tabelle 3). Nur für die Itemblockbildung wird ein signifikanter Anstieg der mittleren Skaleninterkorrelation

Tabelle 4

Ergebnisse der (multivariaten) Varianzanalysen mit den Faktoren „Instruktion“, „Itemreihenfolge“ und „Skalen“ mit Meßwiederholung auf dem letzten Faktor für die FPI-Primärskalen, FPI-Sekundärskalen und die Zusatzfragen (Studie II; N = 84)

Abhängige Variable(n)	Mittelwerte				Varianzanalytische Effekte (F-Werte)													
	Standardinstruktion		Bewerbungsinstrukt.		Instruktion	Item-abfolge	Skalen	Interaktionseffekte										
	Stand. Folge	Block	Stand. Folge	Block				(A)	(B)	(C)	AxB	AxC	BxC	AxBxC				
FPI-1: Nervosität	5.2	4.7	4.7	4.3														
FPI-2: Aggressivität	4.2	3.6	3.5	3.3														
FPI-3: Depressivität	7.2	6.2	6.9	6.1														
FPI-4: Erregbarkeit	3.7	3.3	3.6	3.1														
FPI-5: Geselligkeit	8.0	8.7	9.0	8.3	<b>4.3*</b>	5.7*	66.5**	0.8	<b>1.0</b>	<b>2.2*</b>	<b>1.7</b>							
FPI-6: Gelassenheit	3.8	4.1	4.2	4.5														
FPI-7: Dominanzstr.	2.3	1.9	2.7	2.3														
FPI-8: Gehemtheit	<b>4.6</b>	4.2	4.0	4.0														
FPI-9: Offenheit	9.5	9.1	8.2	7.8														
FPI-E: Extraversion	<b>6.6</b>	6.9	<b>6.8</b>	6.0														
FPI-N: emot. Labilität	5.7	5.1	5.2	5.0	<b>2.4</b>	<b>6.7*</b>	31.3**	3.4	4.0*	0.4	2.9							
FPI-M: Maskulnität	5.8	5.2	5.7	6.8														
Zeitschätzung	9.9	10.5	11.5	11.2	<b>4.0*</b>	<b>1.6</b>	-	1.2	-	-	-							
Langeweile	2.9	3.3	2.8	3.2														
Entscheidungsprobleme	2.7	2.9	2.4	2.8														
Verständnisprobleme	1.5	1.6	1.4	1.6														
Situationsabhängigkeit	2.5	2.5	2.5	2.5	0.1	1.9	57.7**	0.9	0.4	3.1*	<b>0.6</b>							
Selbsterkennen	3.5	3.5	3.4	3.5														
Einstellung zu Frageb.	3.5	3.4	3.5	3.4														
Fragebogenabsicht	4.1	4.8	4.0	4.7														
Soziale Erwünschtheit	4.2	4.3	4.5	4.4														

\*\*p < .01, \*p < .05

Tabelle 5  
 Interne Konsistenzen der FPI-Skalen unter (a) Standard- versus Bewerbungsinstruktion und (b) Standardabfolge versus Blockbildung der Items sowie Vergleiche der internen Konsistenzkoeffizienten (Studie II; N = 84)

FPI-A Skala	interne Konsistenz		W-Test	interne Konsistenz		W-Test
	Standardinstr.	Bewerbungsinstr.	F(41/41)	Standarditemfolge	Item-Blockbildung	F(40/41)
FPI-1: Nervosität	.76	.78	<b>1.09</b>	<b>.77</b>	.61	1.70*
FPI-2: Aggressivität	.55	.57	<b>1.05</b>	<b>.56</b>	.54	1.05
FPI-3: Depressivität	.82	.82	<b>0.07</b>	<b>.83</b>	.77	<b>1.35</b>
FPI-4: Erregbarkeit	.75	.75	<b>0.04</b>	<b>.74</b>	.82	<b>1.44</b>
FPI-5: Geselligkeit	.80	.82	<b>1.11</b>	<b>.81</b>	.67	3.66**
FPI-6: Gelassenheit	.67	.66	<b>1.03</b>	<b>.68</b>	.61	1.21
FPI-7: Dominanzstreben	.52	.52	<b>0.04</b>	<b>.52</b>	.17	1.73*
FPI-8: Gehemmtheit	.77	.76	<b>1.04</b>	<b>.77</b>	.61	1.70*
FPI-9: Offenheit	.79	.77	<b>1.10</b>	<b>.79</b>	.65	1.67*
FPI-E: Extraversion	.63	.63	0.01	<b>.63</b>	.63	<b>0.05</b>
FPI-N: emot. Labilität	.75	.74	<b>1.04</b>	<b>.74</b>	.74	0.08
FPI-M: Maskulinität	.67	.65	<b>1.06</b>	<b>.67</b>	.36	1.94*

\*\*p < .01, \*p < .05

nachgewiesen. Der unter der Bewerbungsinstruktion zwar auch zu beobachtende Anstieg erreicht erneut nicht die Signifikanzgrenze.

### *Diskussion*

In den zwei dargestellten Studien konnten exemplarisch für die Halbform A des FPI deutliche Effekte der Itemblockbildung, die der Extraktion der Items von Subskalen aus umfassenderen Persönlichkeitsfragebogen und deren (inhaltshomogene) Darbietung im Block entspricht, auf die Mittelwerte, interne Konsistenzen und Interkorrelationen der Subskalen nachgewiesen werden. Unter Bezug auf die allgemeine Beeinflussung dieser Reliabilitäts- und Validitätsparameter durch die Itemblockbildung sind die Befunde beider Studien (in denen mit der kontrollierten versus nicht kontrollierten Gruppentestung unterschiedliche Formen der Datenerhebung realisiert wurden) konsistent. Die Datenqualität beider Studien wird dadurch belegt, daß für die als Kontrollbedingung mitgeführte (soziale Erwünschtheit provozierende) „Bewerbungsinstruktion“ die von Hampel und Klinkhammer (1978) für den FPI nachgewiesenen Effekte hier vollständig repliziert wurden. Während sich jedoch in den Mittelwertvergleichen für den Faktor „Fragebogeninstruktion“ (mit 7% und 19%) deutlich unterschiedliche Effektstärken ergeben, liegen die Effektstärken des Faktors „Itemabfolge“ übereinstimmend bei etwa 10%.

Ausdrücklich betont sei, daß die ältere FPI-A-Form in der vorliegenden Arbeit nur exemplarisch verwendet wurde, um an einem häufig eingesetzten Verfahren die Fragen nach der Zulässigkeit und den Effekten der Extraktion von Subskalen aus umfassenderen Persönlichkeitsinventaren beispielhaft zu untersuchen. Die Befunde zur Standardform des FPI-A stimmen weitgehend mit den in der Handanweisung (Fahrenberg et al., 1973) gemachten Angaben überein und belegen — selbst knapp 20 Jahre nach der Publikation der 2. Auflage und der Vorlage einer neuen, revidierten Version des FPI (Fahrenberg et al., 1984) — seine Robustheit und psychometrische Qualität, die mit Ausnahme von Reliabilitätsmängeln einiger Skalen der Halbform gut ist.

Nachgewiesen wurden signifikante Effekte der Itemblockbildung auf die Skalenmittelwerte, die internen Konsistenzen und die Interkorrelationen der Primärskalen des FPI-A. Die Standardversion des Persönlichkeitsinventars und die über Subskalenextraktion vorgenommene Itemblockbildung sind unter diesen Aspekten empirisch nicht äquivalent. Interessant ist dabei, daß sich neben dieser Konvergenz der Befunde zwischen den Studien z. T. Unterschiede in den von diesen Effekten betroffenen Subskalen und in der Richtung der Effekte zeigen. Auch wenn so etwa die Hypothese von Deu-

singer (1986), nach der sich die Itemblockbildung in den Skalenmittelwerten durch eine Verstärkung der Tendenz, sich in sozial erwünschter Weise darzustellen, niederschlägt, durch die vorgelegten Befunde zum FPI-A weitgehend bestätigt wird, ist darauf zu verweisen, daß dieses Befundmuster nicht durchgängig ist. So liegt etwa der FPI-E-Skalenwert (Extraversion) in Studie I bei der Itemblockbildung im Vergleich zur Standardabfolge höher, in Studie II dagegen niedriger. Ein ähnliches Bild zeigt sich bei den Skalen FPI-M (Maskulinität) und FPI-5 (Geselligkeit). Dies spricht dafür, daß zwar konsistent Effekte der Itemblockbildung auftreten, daß diese aber kaum einer allgemeinen Richtung folgen. Bestätigt wird dies auch durch die Befunde zu den Effekten der Itemblockbildung auf die interne Konsistenz der Skalen. Nur für die Skala FPI-1 (Nervosität) konnte in beiden Studien übereinstimmend eine Reduktion der internen Konsistenz bei der Itemblockbildung registriert werden. Für die anderen Skalen zeigen sich inkonsistente, z. T. gegenläufige Ergebnisse. Damit kann anhand der vorliegenden Befunde weder die allgemeine Hypothese einer (durchgängigen) Homogenisierung des Antwortverhaltens durch die Subskalenextraktion noch die der (durchgängigen) Reaktanzerhöhung durch die mit der Itemblockbildung verbundenen thematischen Perservationen bestätigt werden. Signifikante Effekte der Subskalenextraktion sind somit nachgewiesen, ohne daß dabei aber unter den Aspekten der Mittelwerts- und Reliabilitätsäquivalenz von Fragebogen-Gesamtform und homogenen Fragebogen-Teilformen ein systematischer Effekt einseitig nachgewiesen werden konnte. Die als Aspekt der Konstruktvalidität untersuchten Skaleninterkorrelationen steigen dagegen in beiden Studien markant an, was darauf verweist, daß die Itemblockbildung dazu beiträgt, Konstrukt differenzierung in der Psychometrie zu verwischen. Es ist zu vermuten, daß dadurch auch die Parameter der konvergenten und diskriminanten (externen) Validität von Persönlichkeitsfragebogenskalen betroffen sein werden, was freilich in Anschlußstudien zu untersuchen bleibt.

In signifikanten Interaktionseffekten der MANOVAs und in den univariaten Vergleichen deuten sich auch bei den Zusatzfragen Effekte der Itemblockbildung an. Die geschätzte Bearbeitungszeit für den FPI-A nimmt bei der Itemblockbildung in der Tendenz zu, die Langeweile bei der Fragebogenbearbeitung wird als höher beurteilt, und es treten eher Entscheidungsprobleme im Antwortverhalten sowie Überlegungen zur Meßabsicht bei der Fragebogenbearbeitung auf. Diese a posteriori erfaßten Aspekte des Fragebogenbeantwortungsprozesses bedürfen freilich der exakteren Analyse (vgl. hierzu etwa Krämer & Schneider, 1987; Krampen, 1981, 1987b; Schneider-Düker & Schneider, 1977), geben jedoch einen ersten Eindruck von den die Fragebogenbeantwortung bei homogenen Itemblocks begleitenden kognitiven und motivationalen Prozessen.

Insgesamt ergibt sich somit ein Befundmuster, das die Extraktion von Subskalen aus umfassenderen Persönlichkeitsinventaren obsolet und zu einem methodischen Kunstfehler macht, da sich dadurch die Reliabilitäts- und Validitätsparameter des standardisierten, teststatistisch abgesicherten Instruments erheblich verändern. Welche alternativen Möglichkeiten sind denkbar, dem Bedarf nach „ökonomischen“ Erhebungsinstrumenten in der Persönlichkeitsforschung und -diagnostik anhand eines rationaleren Vorgehens nachzukommen, wenn die Subskalenextraktion aus umfassenderen Inventaren ausscheidet? Zunächst sei darauf verwiesen, daß man sich an erster Stelle um eine theoretische Anbindung des differential- und/oder persönlichkeitspsychologischen Ansatzes/Konzepts an die jeweilige („eigentliche“) Fragestellung bemühen sollte, die oftmals die Augen für weitere, theoretisch relevante Variablen, die u. U. mit der Standardversion des Inventars erfaßt werden können, öffnet. Gleichzeitig ist festzuhalten, daß der im Vergleich zu Interviewtechniken, Testverfahren etc. ohnehin ökonomischen Methodik der Fragebogenerhebung alleine aus Gründen ihrer Ökonomiemaximierung mit der Subskalenextraktion nicht Gewalt angetan werden sollte, da dann auch ihre Vorteile im Bereich der Objektivität sowie der (erleichterten) Reliabilitäts- und Validitätsprüfungen zugunsten eines isolierten Ökonomiestrebens konterkariert werden und ggf. die (so zwar „ökonomisch“ erhobenen) Daten in ihrer Qualität erheblich beeinträchtigt. Neben der persönlichkeitstheoretischen Anbindung der Fragestellung ist als eine zur Subskalenextraktion aus umfassenden Inventaren rationalere Möglichkeit zunächst die gezielte Suche nach und Auswahl von spezifischen Persönlichkeitsfragebogen, die sich auf einen (nämlich den angezielten) Persönlichkeitsbereich konzentrieren, zu empfehlen. Entsprechende Instrumente liegen inzwischen für viele Variablenbereiche vor (etwa zur Ängstlichkeit, vgl. Becker, 1982, zu Kontrollüberzeugungen, Krampen, 1991, etc.). Daneben erscheint auch die an den teststatistischen Befunden zu einem standardisierten Instrument orientierte Itemselektion (etwa nach den Trennschärfekoeffizienten und Schwierigkeitsindizes der Items oder auch ihren Faktorladungen) im Vergleich zur inhaltshomogenen Itemblockbildung durch Subskalenextraktion als ein rationaleres Vorgehen. Realisiert werden dabei Skalenverkürzungen, deren negative Effekte auf die Reliabilitätskennwerte zwar bekannt sind, allerdings den inhaltsheterogenen Itemkontext erhalten. Veränderungen in spezifischen Ausstrahlungseffekten einzelner Items können dabei freilich nicht ausgeschlossen werden, dürften aber wohl weniger gewichtig sein als die Veränderungen, die durch die unmittelbare thematische Perseveration bei inhaltshomogener Itemblockbildung zustandekommen.

Bleiben die dargestellten Ergebnisse auch auf Effekte der Präsentation mehr oder weniger inhaltshomogener Itemblocks durch die Extraktion von

Subskalen aus umfassenderen Persönlichkeitsinventaren beschränkt, so bedürfen nicht nur diese, hier empirisch nachgewiesenen Effekte der weiteren Analyse (etwa unter Bezug auf andere Reliabilitäts- und Validitätsaspekte sowie andere Persönlichkeitsfragebogen), sondern es sollte auch wieder verstärkt möglichen Reihenfolgeeffekten bei der Beantwortung mehrerer verschiedener Fragebogen empirisch nachgegangen werden (vgl. hierzu etwa Dalbert et al., 1991). Neben der sorgfältigen Fragebogenkonstruktion werden nur solche Analysen des Antwortprozesses und der überlegte Umgang mit diesen Instrumenten auf Dauer dazu beitragen können, die Vorteile und den Nutzen der Fragebogenmethode in der Psychologie zu erhalten.

### Summary

Empirical results of two studies are presented on the widely spread practice (usually founded on economic arguments) to extract subscales from extensive personality inventories and to present their items as more or less homogeneous item-blocks in questionnaire applications. Based on the „Freiburger Persönlichkeitsinventar“ (FPI-A; a German multidimensional personality inventory) two studies (total N = 213 university students) were conducted in different contexts of application (controlled versus not controlled group testing), in which (1) a comparison of the standard questionnaire instruction with a social desirability provoking instruction (control condition) and (2) the standard item-arrangement was contrasted with the presentation of the items of each subscale as a block (homogeneous item-blocking), which matches the extraction of subscales from more extensive inventories. The results confirm significant effects of item-blocking on the reliability, the means, and the intercorrelations of the subscales. The admissibility of subscale-extraction is doubtful because the reliability and validity parameters of the scales change markedly. Alternative possibilities for the development of „economic“ instruments in personality research and psychodiagnosis founded on more rational arguments are discussed.

### Literatur

- Angerer, H. (1969). Befragung und Erhebung. In C. F. Graumann (Hrsg.), *Sozialpsychologie* (Handbuch der Psychologie, Bd. 7/1, S. 567—618). Göttingen:Hogrefe.
- Becker, P. (1982). *Interaktions-Angst-Fragebogen (IAF)*. Weinheim: Beltz.
- Dalbert, C., Strack, K., Boesken, W. H. & Schwenkmezger, P. (1991). Beeinflusst die Auseinandersetzung mit Bewältigungsverhalten das Angsterleben? *Zeitschrift für Klinische Psychologie*, 20, 65—74.

- Deusinger, I. M. (1986). *Die Frankfurter Selbstkonzeptskalen (FSKN)*. Göttingen: Hogrefe.
- Fahrenberg, J., Hampel, R. & Selg, H. (1984). *Das Freiburger Persönlichkeitsinventar* (revidierte Aufl.). Göttingen: Hogrefe.
- Fahrenberg, J., Selg, H. & Hampel, R. (1973). *Das Freiburger Persönlichkeitsinventar* (2. Aufl.). Göttingen: Hogrefe.
- Feild, H. S. & Armenakis, A. A. (1974). On use of multiple tests of significance in psychological research. *Psychological Reports*, 35, 427—431.
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's Alpha or Kuder-Richardson Coefficient Twenty is the same for two tests. *Psychometrika*, 34, 363—373.
- Feldt, L. S., Woodruff, D. J. & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93—103.
- French, J. L. & Greer, D. (1964). Effect of test-item arrangement on physiological and psychological behavior in primary school children. *Journal of Educational Measurement* 1, 151—153.
- Graumann, C. F. (1973). Zur Lage der Psychologie. In G. Reinert (Hrsg.), *Bericht über den 27. Kongreß der DGfPs in Kiel 1970* (S. 19—37). Göttingen: Hogrefe.
- Gutjahr, G. (1985). *Psychologie des Interviews in Theorie und Praxis*. Heidelberg: Sauer.
- Hampel, R. & Klinkhammer, F. (1978). Verfälschungstendenzen beim Freiburger Persönlichkeitsinventar in einer Bewerbungssituation. *Psychologie und Praxis*, 22, 58—69.
- Herrmann, T. & Lantermann, E.-D. (1985). Einleitung: Psychologie der Persönlichkeit — Status und Perspektiven. In T. Herrmann & E.-D. Lantermann (Hrsg.), *Persönlichkeitspsychologie: Ein Handbuch in Schlüsselbegriffen* (S. VII—XVII). München: Urban & Schwarzenberg.
- Krämer, H.-J. & Schneider, J. F. (1987). Validität von Fragebogendaten in Abhängigkeit von Antwortzeit-Instruktion und der intraindividuellen Variabilität der Probanden. *Psychologische Beiträge*, 29, 458—468.
- Krampen, G. (1981). Soziale Orientierungslagen als Prozeßbedingungen bei der Beantwortung von Persönlichkeitsfragebogen-Items. *Diagnostica*, 27, 127—139.
- Krampen, G. (1987a). Ein handlungstheoretisches Partialmodell der Persönlichkeit. In M. Amelang (Hrsg.), *Bericht über den 35. Kongreß der DGfPs in Heidelberg 1986* (Bd. 2, S. 345—355). Göttingen: Hogrefe.
- Krampen, G. (1987b). Analysen der Antwortprozesse bei der Bearbeitung von Fragebogenitems. *Diagnostica* 33, 289—300.
- Krampen, G. (1988). Testbesprechung zu „Die Frankfurter Selbstkonzeptskalen“ (FSKN). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 9, 69—72.
- Krampen, G. (1991). *Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK)*. Göttingen: Hogrefe.
- Pervin, L. A. (1985). Personality: Current controversies, issues, and directions. *Annual Review of Psychology*, 36, 83—114.
- Sax, G. & Carr, A. (1962). An investigation of response sets on altered parallel forms. *Educational and Psychological Measurement*, 22, 371—376.
- Sax, G. & Cromack, T. R. (1966). The effects of various forms of item arrangements on test performance. *Journal of Educational Measurement*, 3, 309—311.
- Scheuch, E. K. (1967). Das Interview in der empirischen Sozialforschung. In R. König (Hrsg.), *Handbuch der empirischen Sozialforschung* (Bd. 1, S. 348—384). Stuttgart: Enke.
- Schneider, J. F. & Hübner, R. (1980). Einfluß von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebogen. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 27, 565—579.
- Schneider-Düker, M. & Schneider, J. F. (1977). Untersuchungen zum Beantwortungsprozeß bei psychodiagnostischen Fragebogen. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 24, 282—302.

Schriesheim, C. A., Kopelman, R. E. & Solomon, E. (1989). The effect of grouped versus randomized questionnaire format on scale reliability and validity. *Educational and Psychological Measurement*, 49, 487—508.

Anschrift der Verfasser: Prof. Dr. Günter Krampen, Universität Trier, FB I-Psychologie, Postfach 3825, 5500 Trier.