

Effekte von Bewerbungsinstruktionen und Subskalenextraktion in der Fragebogendiagnostik

Günter Krampen

Exemplarisch für den „Fragebogen zu Kompetenz- und Kontrollüberzeugungen“ (FKK) wurden die Effekte einer Bewerbungsinstruktion und die der Darbietung skalenhomogener Itemblocks, die der Extraktion von Subskalen aus umfassenderen Persönlichkeitsinventaren entspricht, in zwei Stunden experimentell untersucht. Realisiert wurden Gruppentestungen (Studie I, N=165 Studenten) bzw. Individualtestungen (Studie II, N=247 Erwachsene), in denen neben (1) einem Vergleich der Standardinstruktion des FKK mit einer sozialen Erwünschtheit provozierenden Bewerbungsinstruktion (2) die (gemischte) Standardreihenfolge der Items im Gesamtfragebogen mit der Darbietung der Items der FKK-Primärskalen im Block (inhaltshomogene Blockbildung) verglichen wurde. Alle Informanden bearbeiteten zusätzlich die „Marlow-Crowne Social Desirability Scale“ (MCSD). Die Ergebnisse belegen signifikante Effekte der Itemblockbildung und der Bewerbungsinstruktion auf die Mittelwerte, internen Konsistenzen, Interkorrelationen und Außenkorrelationen der FKK-Primärskalen. Implikationen dieser Ergebnisse für die Anwendung von Persönlichkeitsfragebogen in Selektionstestungen sowie die verbreitete Praxis der Extraktion von Subskalen aus umfassenderen Persönlichkeitsinventaren werden diskutiert.

The effects of a social desirability provoking instruction and of the presentation of homogenous item-blocks, which is in accord with the extraction of subscales from more extensive personality inventories, on the reliability and validity of a personality inventory were analysed in two studies. Subjects responded to four different versions of the „Fragebogen zu Kompetenz- und Kontrollüberzeugungen“ (FKK; a German inventory for the assessment of self-concept, internality, powerful others externality, and chance control in locus of control beliefs) and to the „Marlow-Crowne Social Desirability Scale“ (MCSD). Besides a comparison of the standard instruction of the FKK with a social desirability provoking instruction (job applicant screening) the (scale-mixed) standard item arrangement of the FKK was compared with the presentation of the items in homogeneous blocks referring to the items of the four subscales. The results show that there are significant effects of instruction and item-blocking on the means, internal consistencies, intercorrelations, and external correlations of the FKK Scales. Implications of these results are discussed with reference to the application of personality inventories in selection tests and with reference to the widely spread practice of subscale-extraction from more extensive inventories.

In der psychologischen Forschung und psychodiagnostischen Praxis sind Persönlichkeitsfragebogen wegen ihrer – etwa im Vergleich zu Interviews und objektiven Testverfahren – Ökonomie und Datenerhebungsobjektivität weit verbreitet. Andererseits werden immer wieder Zweifel an ihrer Validität geäußert, die nicht nur aus der bei ihrem Einsatz häufig bestehenden Distanz zwischen dem Diagnostiker und dem Informanden resultieren, sondern auch aus Bedenken gegenüber Antwortverzerrungen, die bei ihrer Verwendung möglich sind. Obwohl es bislang dafür etwa aus Vergleichen mit der Interviewmethodik kaum empirische Belege gibt, wird dabei

insbesondere befürchtet, daß es Informanden in Fragebogen leicht(er) fallen könnte, sich selbst in sozial erwünschter Weise darzustellen. Nur am Rande sei vermerkt, daß Fragebogen hier im Vergleich zum Interview zumindest über den Vorteil verfügen, daß solche Antwortverzerrungen mit der gleichen Methodik (durch Skalen zur „sozialen Erwünschtheit“, „Offenheit“, „Lügentendenz“ etc.) überprüft werden können. Persönlichkeitsfragebogen weisen aber auch weitere Sensitivitäten auf, die mit dem Applikationskontext und mit dem Bearbeitungsprozeß zusammenhängen.

Wenig berücksichtigt wird so etwa bei der Anwendung von Fragebogen in der psychodiagnostischen Praxis, daß die meisten Persönlichkeitsinventare im Rahmen anonym durchgeführter Untersuchungen entwickelt und normiert werden. In der psychodiagnostischen Praxis ist die anonyme Fragebogenbearbeitung dagegen der Ausnahmefall. Die Entwicklung und der Einsatz von Fragebogen in der Praxis (das gleiche gilt freilich etwa auch für Testverfahren) unterscheiden sich aber nicht nur in der Anonymität versus Nicht-Anonymität bei der Bearbeitung, sondern auch darin, daß den Datenerhebungen in der Forschung (bei „Versuchspersonen“ oder Informanden) zumeist kaum persönliche Relevanz und Folgen zukommen, die in psychodiagnostischen „Ernstsituationen“ (sei es im Rahmen von Selektionsuntersuchungen oder im Rahmen der interventionsorientierten Diagnostik und Evaluation) dagegen in der Regel in hohem Maße bestehen bzw. auftreten. Empirische Untersuchungsbefunde weisen so etwa für breit angelegte, mehrdimensionale Persönlichkeitsinventare durchgängig darauf, daß alleine durch Instruktionen, die den „Ernstfall“ simulieren (etwa eine „Bewerbungsinstruktion“), selbst bei bestehender Anonymität die Tendenz, in sozial erwünschter Weise zu antworten, massiv und ähnlich zu wirklichen diagnostischen Situationen in der Praxis beeinflusst werden kann (vgl. etwa Hampel & Klinkhammer, 1978; Hoeth & Gregor, 1964; Hoeth, Kucklick & Simmat, 1965; Krampen, Hense & Schneider, 1992; Schneider & Hübner, 1980). Ähnliche Ergebnisse liegen etwa auch für Instrumente zur Erfassung generalisierter Kontrollüberzeugungen vor (vgl. etwa Davidson & Bailey, 1978; Kestenbaum & Hammersla, 1976), wobei zudem deutlich wurde, daß solche Verfälschungen unter „Bewerbungsinstruktionen“ o. ä. nicht durch (konstrukt-irrelevante) Füllitems verringert werden können.

Außer dem Anwendungskontext, der neben der Anonymität versus Nicht-Anonymität und der persönlichen Bedeutung der Psychodiagnostik etwa mit vorhandenen versus nicht vorhandenen zeitlichen Beschränkungen (vgl. etwa Krämer & Schneider, 1987; Schneider & Schneider-Düker, 1981), möglichen Einflußnahmen von Außen, Störungen etc. weitere diagnostisch relevante Merkmale aufweisen kann, wird der Bearbeitungsprozeß nicht nur durch interindividuell unterschiedliche (konstrukt-irrelevante) Motivationen und Kognitionen (vgl. hierzu etwa Krampen, 1987), sondern auch durch die Gestaltung und den Aufbau des Fragebogens selbst beeinflusst. Neben optisch-gestalterischen Aspekte, die das „Layout“ und die Lesbarkeit betreffen (vgl. etwa Champion & Sear, 1968/69), den sprachlichen Formulierungen (vgl. etwa Löhr & Angleitner, 1980), der Polung der Items und den vorgegebenen Antwortalternativen oder -skalen (vgl. etwa Wegener, Faulbaum & Maag, 1982) ist hier vor allem auch die bislang in der Forschung weitgehend vernachlässigte Frage nach der Anordnung der Items zu nennen.

Während bei objektiven Leistungstests eine nach Aufgabenart und -schwierigkeit gestaffelte Itemordnung allgemein üblich geworden ist (obwohl auch dazu keine einheitlichen Befunde vorliegen; siehe etwa French & Greer, 1964; Sax & Carr, 1962; Sax & Cromack, 1966), weisen umfassendere Persönlichkeitsinventare übereinstimmend eine nach ihren Subskalen gemischte (also inhaltlich heterogene) Itemabfolge auf. Dies ist zumeist die Folge einer früh bei der Fragebogenentwicklung getroffenen Entscheidung, die dann bis hin zur Normierung aufrechterhalten bleibt und schließlich auch das teststatistisch abgesicherte Instrument charakterisiert.

Zumeist mit ökonomischen Argumenten, methodisch jedoch kaum reflektiert, ist es nun in Teilen der psychologischen Forschung und Praxis nahezu üblich geworden, aus solchen umfassenderen Persönlichkeitsinventaren die Items der Subskala (oder Subskalen), die aus irgendwelchen Gründen heraus für die eigene Fragestellung für relevant gehalten werden, zu extrahieren und als mehr oder weniger inhaltlich homogene Itemblocks zur Bearbeitung vorzulegen. Während die Vernachlässigung des Fragekontextes und die möglicher Ausstrahlungseffekte von Fragen in der Interviewmethodik allgemein als Kunstfehler bewertet wird (vgl. etwa Gutjahr, 1985; Scheuch, 1967), werden die möglichen Effekte dieser Itemblockbildung auf die Reliabilität und Validität der Daten kaum reflektiert. Für die Halbform A des „Freiburger Persönlichkeitsinventars“ (FPI; Fahrenberg, Selg & Hampel, 1973) konnten nun aber signifikante Effekte der Itemblockbildung auf die Mittelwerte, internen Konsistenzen und Interkorrelationen der FPI-Skalen bei Krampen, Hense und Schneider (1992) empirisch nachgewiesen werden. Interessant ist dabei, daß sich in zwei Studien zwar übereinstimmend deutliche Effekte nachweisen ließen, daß sich die Richtung der Effekte aber sowohl innerhalb als auch zwischen den Studien unterschied. Die Äquivalenz von FPI-Standardversion und der inhaltshomogenen Blockdarbietung seiner Items ist damit empirisch widerlegt, ohne daß zugleich eine allgemeine Hypothese der (durchgängigen) Homogenisierung des Antwortverhaltens durch die Itemblockbildung oder eine allgemeine Hypothese der (durchgängigen) Reaktanzbildung im Antwortverhalten (bedingt durch die thematischen Perseverationen bei der Itemblockbildung) bestätigt werden konnte.

Im folgenden werden die Befunde zweier Studien mit unterschiedlichen Applikationsbedingungen (kontrollierte Gruppen- versus unkontrollierte Individualtestung) zu den Effekten einer sozialen Erwünschtheit provozierenden „Bewerbungsinstruktion“ und zu den Effekten der (inhaltshomogenen) Itemblockbildung auf die Reliabilität und Validität des „Fragebogens zu Kompetenz- und Kontrollüberzeugungen“ (FKK; Krampen, 1991) dargestellt. Der FKK erfaßt als theoretisch und empirisch ausgerichtete Neuentwicklung, die in der Tradition des „IPC-Fragebogens zu Kontrollüberzeugungen“ (Krampen, 1981) steht, auf vier Primärskalen generalisierte selbstbezogene Kognitionen des Selbstkonzepts eigener Fähigkeiten (FKK-SK), der Internalität (FKK-I), der sozial bedingten Externalität (FKK-P; „powerful others control“) und der fatalistischen Externalität (FKK-C; „chance control“), ist bei Jugendlichen (ab 14 Jahren) sowie Erwachsenen (bis in das höhere Lebensalter) einsetzbar und wurde in einer repräsentativen Stichprobe deutscher Erwachsener normiert. Im

Anschluß an die von Krampen et al. (1992) für den FPI-A gewonnenen Befunde werden signifikante Effekte der Bewerbungsinstruktion und der Itemblockbildung auf die Mittelwerte, internen Konsistenzen und Interkorrelationen der FKK-Skalen erwartet. Darüber hinaus soll untersucht werden, ob sich die für die Standardversion des FKK nicht signifikanten Beziehungen zur Tendenz, in sozial erwünschter Weise zu antworten (vgl. Krampen, 1991), unter den experimentellen Bedingungen der Bewerbungsinstruktion und der Itemblockbildung verändern.

Methode

Erhebungsinstrumente

In zwei Studien wurde mit dem gleichen Erhebungsinstrumentarium, das neben vier Varianten des „Fragebogens zu Kompetenz- und Kontrollüberzeugungen“ (FKK; Krampen, 1991) die „Marlow-Crowne Social Desirability Scale“ (MCSD) und einige Zusatzfragen enthielt, gearbeitet. Außer der Standardform des FKK wurden experimentelle Versionen verwendet, die eine soziale Erwünschtheit provozierende Bewerbungsinstruktion („Bitte stellen Sie sich vor, die Beantwortung dieses Fragebogens hätte – etwa im Zusammenhang mit einer Bewerbung für eine von Ihnen angestrebten Arbeitsstelle – große Bedeutung für Sie. Antworten Sie so, als befänden Sie sich in dieser Bewerbungssituation“; vgl. Hampel & Klinkhammer, 1978) und die Darbietung der Items jeder FKK-Primärskala als Block umfaßten. Es resultiert ein 2 (Standard- versus Bewerbungsinstruktion) x 2 (gemischte Standardabfolge der Items versus Itemblockbildung) – „Design“, dem die vier verwendeten FKK-Versionen eindeutig zugeordnet sind.

Unmittelbar nach der Bearbeitung der jeweiligen FKK-Version wurden die Untersuchungsteilnehmer gebeten, *Zusatzfragen* nach (1) der (geschätzten) Bearbeitungszeit, (2) der beim Ausfüllen des Fragebogens erlebten Langeweile, (3) den Schwierigkeiten bei den Antwortentscheidungen, (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, zu beantworten (vgl. auch Krampen et al., 1992). Für die Fragen (2) bis (9) wurden jeweils sechsstufige Antwortskalen vorgegeben. Im Anschluß daran wurde die „Marlow-Crowne Social Desirability Scale“ (MCSD) nach der Standardinstruktion bearbeitet.

Die Datenerhebungen wurden anonym durchgeführt. Soziodemographische Variablen (Geschlecht, Alter, höchster Schulabschluß und Beruf) wurden nach dem Standardformat des FKK erfaßt.

Stichproben

An *Studie I* waren 165 Studienanfänger des Faches Psychologie (97 Studentinnen und 68 Studenten; $M=21.1$ Jahre, $SD=2.6$) in acht Kleingruppen (zu 16 bis 23 Studierenden) beteiligt. Nach dem Zufall wurde jeweils zwei dieser Kleingruppen die Standardform bzw. eine der drei Experimentalversionen des FKK zugewiesen. Die Instruktion der jeweiligen FKK-Form wurde von einem Testleiter laut vorgelesen, bevor die Fragebogen in der Kleingruppe (bei Anwesenheit des Testleiters) bearbeitet wurden. Für die Mitarbeit wurde den Studierenden eine halbe „Versuchspersonenstunde“ im Rahmen der von ihnen im Vordiplomstudium zu erbringenden Untersuchungsteilnahmen bescheinigt. Teilnahmeverweigerungen traten in dieser Totalerhebung nicht auf.

An *Studie II* waren 247 Erwachsene (145 Frauen und 102 Männer; $M=42.7$ Jahre, $SD=5.4$) beteiligt. Die Stichprobe wurde über Studierende der Psychologie nach einem „Schneeballsystem“ mit der Auflage, keine Studierenden um Mitarbeit zu bitten, rekrutiert. Die Standardform und die drei Experimentalversionen des FKK wurden den Personen nach dem Zufall zugewiesen. Von 320 ausgegebenen Fragebogen wurden 258 (80.6%) postalisch in einem Freiumschlag zurückgesandt, von denen 11 unvollständig oder nicht ausgefüllt waren (bereinigter Rücklauf: 77.2%). Soziodemographisch kann diese heterogene Zufallsstichprobe wie folgt charakterisiert werden: Es dominieren Verheiratete (69%) gegenüber Ledigen (17%), Geschiedenen (11%) und Verwitweten (3%); die Mehrzahl der Informanten ist berufstätig (76%; „Nur“- Hausfrau/Hausmann: 19%; im Ruhestand: 5%), wobei Tätigkeiten als Angestellte (40%) und Beamte (23%) am häufigsten sind; als höchste Schulabschlüsse sind Realschul-/Handelsschulabschluß (39%), Hauptschul-/Volksschulabschluß (26%) und Abitur (32%) am häufigsten vertreten. Die Teilnahme an der Untersuchung wurde nicht vergütet.

Ergebnisse

Für die Bedingung „Bewerbungsinstruktion“ wurde zunächst anhand der direkten Zusatzfrage (9) nach der „sozialen Erwünschtheit“ im eigenen Antwortverhalten bei der Bearbeitung des FKK und anhand der MCSD-Skalenwerte die Wirkung dieser experimentellen Manipulation (im Sinne eines „tests of the fidelity of treatment implementation“) in univariaten Mittelwertsvergleichen überprüft. Die Ergebnisse bestätigen entsprechende Effekte nicht nur für die Zusatzfrage (Studie I: $t(163)=7.63$, $p<0.1$; Studie II: $t(245)=6.47$, $p<.01$), sondern auch – in leicht abgeschwächter Form – für die mit der Standardinstruktion (also ohne Instruktionsvariation) vorgegebene MCSD (Studie I: $t(163)=3.01$, $p<.01$; Studie II: $t(245)=2.12$, $p<.05$). Dies kann als Nachwirkung des mit der Bewerbungsinstruktion beim FKK verbundenen kognitiven Sets interpretiert werden und bestätigt die Wirkung dieser experimentellen Variation.

Für die vier FKK-Skalen wurden multivariaten Varianzanalysen (MANOVA) mit den Faktoren „Instruktion“ (Standard- versus Bewerbungsinstruktion), „Itemabfolge“

(Standardabfolge versus Itemblockbildung) und „Skalen“ bei Meßwiederholung über den letzten Faktor berechnet. Die in Tabelle 1 zusammengefaßten Befunde belegen signifikante Haupteffekte aller drei Faktoren und einige signifikante Interaktionsterme in beiden Studien.

Für den Faktor „Skalen“ wird mit hohen Effektstärken von $\omega^2 = .31$ (Studie I) bzw. $\omega^2 = .37$ (Studie II) die im FKK vorgenommene Konstrukt differenzierung empirisch bestätigt. Die Effektstärken des Faktors „Itemabfolge“ liegen bei $\omega^2 = .14$ (Studie I) bzw. $\omega^2 = .15$ (Studie II), die des Faktors „Instruktion“ mit $\omega^2 = .07$ bzw. $\omega^2 = .10$ etwas niedriger. Die ebenfalls in Tabelle 1 aufgeführten Skalenmittelwerte weisen für die Standardform des FKK in beiden Studien mit Prozenträngen im Bereich $45 < PR < 55$ eine sehr hohe Übereinstimmung mit den Werten der Normierungsstichprobe (vgl. Krampen, 1991) auf. Unter der Bewerbungsinstruktion liegen die Skalenwerte für das Selbstkonzept eigener Fähigkeiten, die Internalität und die soziale Externalität signifikant höher, der Skalenwert für die fatalistische Externalität dagegen signifikant niedriger, was mit einer Tendenz zur sozialen Erwünschtheit im Antwortverhalten in Einklang steht. Unter der Itemblockbildung sind die Skalenwerte für das Selbstkonzept, die Internalität und die fatalistische Externalität signifikant niedriger, während der für die soziale Externalität höher liegt. Augenscheinlich führt die thematische Perseveration bei der Selbstzuschreibung von Fähigkeiten und Handlungskonsequenzen sowie bei fatalistischen Aussagen zu einer Abschwächung zustimmender Antworten und bei Aussagen, in denen Einflüsse anderer Menschen auf personrelevante Ereignisse beschrieben werden, zu einer Verstärkung zustimmender Antworten (im Sinne einer höheren Bereitschaft, solche sozialen Einflüsse auf das eigene Leben zu konzipieren). Die signifikante Interaktion von „Instruktion“ und „Itemabfolge“ bestätigt die z. T. gegenläufigen Effekte beider Faktoren auf den FKK-Skalen. Während diese Befunde in beiden Studien übereinstimmen, erreichen die signifikanten Wechselwirkungen des Faktors „Skalen“ mit den beiden anderen (experimentellen) Faktoren nur in Studie I die übliche Signifikanzgrenze.

Für den (geschätzten) zeitlichen Aufwand bei der Bearbeitung der FKK-Version weisen univariate Varianzanalysen (ANOVAs) mit den Faktoren „Instruktion“ und „Itemabfolge“ weder auf statistische bedeutsame Haupteffekte noch eine statistisch bedeutsame Interaktion. In MANOVAs der acht weiteren Zusatzfragen ergaben sich dagegen in beiden Studien signifikante Haupteffekte für den Faktor „Itemabfolge“. Univariate Einzelmittelwertvergleiche deuten darauf, daß die Itemblockbildung zu einer signifikanten ($p < .05$) Erhöhung der Entscheidungsprobleme bei der Fragebogenbeantwortung, zu einer Zunahme der Verständnisprobleme, zu einer Verringerung im Wiedererkennen eigener Eigenschaften und zur mehr Überlegungen über die Frageabsicht führen als die gemischte Standarditemabfolge.

Tabelle 2 gibt die internen Konsistenzen (Cronbachs Alpha) der FKK-Skalen unter den vier Untersuchungsbedingungen wieder. Für die Standardversion des FKK ergeben sich Koeffizientenwerte, die weitgehend den Angaben in der Handanweisung (Krampen, 1991) entsprechen. Unter der Bewerbungsinstruktion zeigen sich einige geringfügige Schwankungen, die nach dem W-Test (nach Feldt, 1969) aber

Tabelle 1: Ergebnisse der multivariaten Varianzanalysen mit den Faktoren „Instruktion“, „Itemabfolge“ und „FKK-Primärskalen“ mit Meßwiederholung auf dem letzten Faktor in Studie I und II

Abhängige Variable	Mittelwerte				Varianzanalytische Effekte (F-Werte)						
	Standardinstruktion		Bewerbungsinstrukt.		Instruktion (A)	Itemabfolge (B)	Skalen (C)	Interaktionseffekte			
	Stand. Folge	Block	Stand. Folge	Block				AxB	AxC	BxC	AxBxC
Studie I (N=165)											
FKK-SK: Selbstkonzept	31.4	29.4	34.5	32.6							
FKK-I: Internalität	32.9	31.0	36.2	34.8							
FKK-P: soz. Externalität	25.3	28.1	27.5	29.3							
FKK-C: fatal. Externalit.	27.7	23.5	21.8	20.0	7.5**	9.8**	20.1**	5.0**	3.3*	2.7*	0.8
Studie II (N=247)											
FKK-SK: Selbstkonzept	30.9	28.8	33.1	31.3							
FKK-I: Internalität	31.6	30.4	35.8	32.9							
FKK-P: soz. Externalität	26.0	28.3	28.2	29.4							
FKK-C: fatal. Externalit.	28.4	25.3	24.2	22.2	4.9*	7.1**	33.8**	5.2*	1.9	2.7	1.3

** $p < .01$, * $p < .05$

keine statistisch bedeutsamen Unterschiede zur Standardinstruktion aufweisen. Für die Itemblockbildung weichen dagegen die Koeffizienten in Studie I dreimal und in Studie II viermal signifikant von denen der (gemischten) Standarditemfolge ab. Im Unterschied zu den Befunden zum FPI-A, für dessen Skalen sowohl Zunahmen als auch Abnahmen der internen Konsistenz unter Itemblockbildung beobachtet wurden (Krampen et al., 1992), weist das vorliegende Befundmuster konsistent auf erhebliche Reduktionen in der Reliabilität der Skalen zu generalisierten Kompetenz- und Kontrollüberzeugungen bei Blockbildung ihrer Items, was ebenso wie die oben dargestellten Mittelwertsunterschiede auf entsprechende Veränderungen in den Itemkennwerten zurückgeht.

Tabelle 2: Interne Konsistenzen der FKK-Primärskalen unter (a) Standard- versus Bewerbungsinstruktion und (b) Standardabfolge versus Blockbildung der Items sowie Vergleiche der internen Konsistenzkoeffizienten für Studie I und II

FKK-Skala	interne Konsistenz		W-Test	interne Konsistenz		W-Test
	Standard-instruktion	Bewerbungs-instruktion	F(81/83)	Standard-itemfolge	Item-Blockbildung	F(82/83)
Studie I (N=165)						
FKK-SK:						
Selbstkonzept	.79	.77	1.10	.78	.41	2.68**
FKK-I: Internalität	.71	.75	1.16	.70	.47	1.77**
FKK-P:						
soz. Externalität	.81	.80	1.05	.81	.71	1.53*
FKK-C:						
fatal. Externalität	.82	.80	1.11	.82	.85	1.20
Studie II (N=247)			F(122/124)	F(122/124)		
FKK-SK:						
Selbstkonzept	.76	.78	1.09	.75	.47	2.12**
FKK-I: Internalität	.74	.74	0.03	.76	.51	2.04**
FKK-P:						
soz. Externalität	.79	.77	1.10	.78	.66	1.55**
FKK-C:						
fatal. Externalität	.80	.75	1.25	.79	.71	1.38*

** $p < .01$, * $p < .05$

Als Aspekt der Konstruktvalidität der FKK-Skalen wurden ihre Interkorrelationen unter den verschiedenen experimentellen Bedingungen analysiert. Da sich für die Standardversion in beiden Studien Korrelationsmuster ergaben, die statistisch nicht bedeutsam von früheren Erfahrungen (vgl. Krampen, 1991, S. 53) abweichen, wird hier auf die ausführliche (tabellarische) Darstellung verzichtet. Auch unter der Bewerbungsinstruktion sind – ähnlich wie beim FPI-A (Krampen et al., 1992) – kaum signifikante Veränderungen zu beobachten (lediglich die Korrelation zwischen den I- und C-Skalen wird negativer; $p < .05$). Für die Itemblockbildung wurden dagegen

bei drei (Studie I) bzw. zwei (Studie II) der sechs (möglichen) FKK-Primärskaleninterkorrelationen im Vergleich ihrer z-transformierten Werte mit denen der Standardreihenfolge statistisch bedeutsame Zunahmen ($d(z) > .32, p < .05$) registriert. Diese Zunahme der gemeinsamen Varianz der Skalen spricht dafür, daß durch die Itemblockbildung Konstruktendifferenzierungen empirisch verwischt werden.

In Tabelle 3 sind die Korrelationen der mit der MCSD erfaßten Tendenz, in sozial erwünschter Weise zu antworten, mit den vier Primärskalen des FKK unter den vier Untersuchungsbedingungen zusammengefaßt. Erneut zeigen sich für die Standardform des FKK (bei anonymer Datenerhebung) keine bedeutsamen Zusammenhänge mit der „sozialen Erwünschtheit“ (siehe auch Krampen, 1991). Unter der Bewerbungsinstruktion steigen die Korrelationskoeffizienten dagegen deutlich an, was für die Skalen zum Selbstkonzept eigener Fähigkeiten und zur Internalität in beiden Studien inferenzstatistisch abgesichert werden kann. Auch die Itemblockbildung führt zu stärkeren Beziehungen der FKK-Skalen zur Tendenz, in sozial erwünschter Weise zu antworten, was in beiden Studien für die Skala zur fatalistischen Externalität (negative Beziehung zur MCSD) und in Studie II auch für die Skala zur sozialen Externalität (positive Beziehung zur MCSD) inferenzstatistisch im Vergleich der Koeffizienten zwischen den Bedingungen belegt werden kann.

Tabelle 3: Korrelationen der FKK-Primärskalen mit der sozialen Erwünschtheit (MCSD) unter (a) Standard- versus Bewerbungsinstruktionen und (b) Standardabfolge versus Blockbildung der Items sowie Ergebnisse der Vergleiche der z-transformierten Korrelationskoeffizienten in Studie I und II

Stichprobe	Korrelation mit MCSD unter			Korrelation mit MCSD unter		
	Standard- instruktion	Bewerbungs- instruktion	d(z)	Standard- itemfolge	Itemblock- bildung	d(z)
Studie I (N=165)						
FKK-SK:						
Selbstkonzept	.08	.39**	0.33*	.06	.09	0.03
FKK-I: Internalität	.07	.42**	0.28*	.08	-.12	0.09
FKK-P:						
soz. Externalität	.14	.23**	0.09	.11	.28**	0.18
FKK-C:						
fatal. Externalität	-.03	-.19*	0.21	-.02	-.49**	0.52**
Studie II (N=247)						
FKK-SK:						
Selbstkonzept	-.02	.27**	0.30*	.01	.14*	0.13
FKK-I: Internalität	.03	.45**	0.46**	.06	-.08	0.14
FKK-P:						
soz. Externalität	.09	.21**	0.12	.10	.37**	0.28**
FKK-C:						
fatal. Externalität	.04	-.10	0.14	-.01	-.41**	0.43**

** $p < .01, p < .05$

Diskussion

Die Befunde beider Studien belegen signifikante Effekte einer sozialen Erwünschtheit provozierenden Bewerbungsinstruktion und der Subskalenextraktion aus einem umfassenden Persönlichkeitsfragebogen auf die Reliabilität und Validität der Skalen. Die externe Validität der Ergebnisse kann dadurch, daß weitgehend übereinstimmende Befunde in Stichproben aus unterschiedlichen Populationen und unter verschiedenen Datenerhebungsbedingungen (kontrollierte Gruppentestung versus unkontrollierte Individualtestung) gewonnen wurden, als gut bezeichnet werden, wenngleich zu bedenken ist, daß auch hier die Datenerhebungen im Unterschied zur psychodiagnostischen Praxis anonym erfolgten. Bedenkt man jedoch, daß durch die soziale Erwünschtheit provozierende Bewerbungsinstruktion, die nach Hampel und Klinkhammer (1978) zumindest in der Tendenz zu einem Antwortverhalten führt, das in die Richtung von diagnostischen Ernstsituationen geht, deutlich Effekte auf die Skalenmittelwerte und die Korrelationen der FKK-Skalen mit der MCSD empirisch nachgewiesen werden konnten, so erscheint auch unter diesem Aspekt eine Annäherung an die Realität der psychodiagnostischen Praxis erreicht worden zu sein.

Ebenso wie andere, zum großen Teil umfassendere Persönlichkeitsinventare ist somit der FKK unter bestimmten Bedingungen für die Tendenz, in sozial erwünschter Form zu antworten, sensibel, was Zweifel an seinen (durch entsprechende „Kontrollskalen“ nicht geprüften) Einsatzmöglichkeiten in Selektionstestungen bestätigt. Gleichzeitig belegen die Befunde, daß die Effekte solcher Antwortverzerrungen auf die Erfassung generalisierter selbstbezogener Kognitionen (wie Kompetenz- und Kontrollüberzeugungen) in hohem Maße vom Applikationskontext (hier den Instruktionen) abhängen. Bei freiwilligen, ggfs. auch durch Leidensdruck motivierten Datenerhebungen im Rahmen der interventionsorientierten Diagnostik und Intervention oder der Forschung treten Verfälschungstendenzen dagegen seltener auf (vgl. auch Krampen, 1982, S. 121 ff.).

Auch im Vergleich zu anderen Befunden (Krampen et al., 1992) relativ stabil ist mit ca. 14% die Effektstärke der Itemabfolge, was zeigt, daß die im Originalfragebogen festgelegte Abfolge der Fragen eines seiner zentralen Charakteristika ist, das nicht ohne Effekte auf seine Reliabilitäts- und Validitätsparameter verändert werden kann. Damit ist ebenso wie für den FPI-A empirisch nachgewiesen, daß die Gesamtversion eines Persönlichkeitsinventars nicht mit der über Subskalenextraktion zustandekommenen Darbietung (inhaltshomogener) Itemblocks äquivalent ist. Dokumentiert wurden deutliche Effekte auf die Skalenmittelwerte, die nicht durchgängig mit den Effekten der sozialen Erwünschtheit provozierenden Instruktion identisch sind, sowie solche auf die internen Konsistenzen, Interkorrelationen und Außenkorrelationen der Skalen zur MCSD. Die hier im Unterschied zu den FPI-A-Skalen (bei Krampen et al., 1992) durchgängig festgestellte Reduktion der internen Konsistenz der Skalen bei Itemblockbildung sprechen dabei eher für die Hypothese der Reaktanzbildung im Antwortverhalten durch die mit der inhaltlichen Homogenisierung der Items verbundenen thematischen Perseverationen. Vielleicht führt die Häufung inhaltlich

ähnlicher Fragestellungen gerade im Bereich der selbstbezogenen Kognitionen eher zu Widerspruch und damit inkonsistentem Antwortverhalten. Darauf deuten zumindest die Ergebnisse zu den hier anhand der Zusatzfragen nur grob analysierten Antwortprozesse, da unter der Itemblockbildung über mehr Entscheidungsprobleme, Verständnisprobleme und Überlegungen zur Fragebogenabsicht sowie ein geringes Ausmaß des Wiedererkennens eigener Merkmale berichtet wird als unter der (gemischten) Standardreihenfolge der FKK-Items.

Anstelle der nach den dargestellten Befunden obsolet gewordenen Subskalenextraktion aus umfassenderen Persönlichkeitsinventaren, die ebenso wie die Vernachlässigung von Fragenkontext und Ausstrahlungseffekten in der Interviewmethodik auch in der Fragebogenmethodik als Kunstfehler anzusehen ist, sollten dann, wenn unbedingt mit „ökonomischen“ Erhebungsmethoden gearbeitet werden muß, besser spezifisch für das interessierende Konstrukt entwickelte und erprobte Fragebogen in ihrer Originalform eingesetzt werden. Allenfalls ist noch die an teststatistischen Parametern des Originalfragebogens orientierte Skalenverkürzung (durch Itemselektion) als rationales Vorgehen anzusehen, da dabei (1.) zumindest grob bekannt ist, was man sich in etwa damit einhandelt (nämlich Reliabilitätsabfälle), und (2.) spezifische Ausstrahlungseffekte einzelner Items zwar nicht ausgeschlossen werden können, dafür aber zumindest eine inhaltlich heterogene Reihenfolge der Items erhalten bleibt.

Literatur

- Champion, D. J. & Sear, A. M. (1968/69). Questionnaire response rate: A methodological analysis. *Social Forces*, 47, 335–339.
- Davidson, K. M. & Bailey, K. G. (1978). Effects of "status sets" on Rotter's locus of control scale. *Journal of Consulting and Clinical Psychology*, 46, 186.
- Fahrenberg, J., Selg, H. & Hampel, R. (1973). *Das Freiburger Persönlichkeitsinventar* (2. Aufl.). Göttingen: Hogrefe.
- 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.
- 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.
- 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.
- Hoeth, F. & Gregor, H. (1964). Guter Eindruck und Persönlichkeitsfragebogen. *Psychologische Forschung*, 28, 64–88.
- Hoeth, F., Kucklick, G. & Simmat, W. E. (1965). Experimentelle Untersuchung zum Problem des „guten Eindrucks“. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 12, 59–84.
- Kestenbaum, J. M. & Hammersla, J. (1976). Filler items and social desirability in Rotter's locus of control scale. *Journal of Personality Assessment*, 40, 162–168.

- 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). *IPC-Fragebogen zu Kontrollüberzeugungen*. Göttingen: Hogrefe.
- Krampen, G. (1982). *Differentialpsychologie der Kontrollüberzeugungen*. Göttingen: Hogrefe.
- Krampen, G. (1987). Analysen der Antwortprozesse bei der Bearbeitung von Fragebogenitems. *Diagnostica*, 33, 289–300.
- Krampen, G. (1991). *Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK)*. Göttingen: Hogrefe.
- Krampen, G., Hense, H. & Schneider, J. F. (1992). Reliabilität und Validität von Fragebogenskalen bei Standardreihenfolge versus inhaltshomogener Blockbildung ihrer Items. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 39, 229–248.
- Löhr, F.-J. & Angleitner, A. (1980). Eine Untersuchung zu sprachlichen Formulierungen der Items in deutschen Persönlichkeitsfragebogen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 217–235.
- 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, J. F. & Scheider-Düker, M. (1981). Warum sollen eigentlich Persönlichkeitsfragebogen „schnell und zügig“ beantwortet werden? In L. Tent (Hrsg.), *Erkennen, Wollen, Handeln: Festschrift für Heinrich Düker zum 80. Geburtstag* (S. 375–385). Göttingen: Hogrefe.
- Wegener, B., Faulbaum, F. & Maag, G. (1982). Die Wirkung von Antwortvorgaben bei Kategorialskalen. *ZUMA-Nachrichten*, 10, 3–20.

Anschrift des Verfassers:

Prof. Dr. Günter Krampen
 Universität Trier
 FB I – Psychologie
 Postfach 3825
 5500 Trier
 (Fax: 06 51 / 30 99 15)