

Effekte der Computeradministration bei der Symptom-Checkliste (SCL-90-R) unter besonderer Berücksichtigung der Itemreihenfolge

Gabriele Helga Franke

Zusammenfassung. Anhand der revidierten Version der Symptom-Checkliste (SCL-90-R) wurde bei 400 Studierenden (1) eine Papier-Bleistift-Version mit einer Computerversion sowie (2) die Originalitemreihenfolge mit der Präsentation der Items jeder Subskala als inhaltshomogener Block verglichen. Die Computeradministration beeinflusste die Mittelwerte der Subskalen nicht; es zeigte sich keine Interaktion mit dem Faktor „Geschlecht (G)“. Die Itemblockbildung erhöhte die Mittelwerte. Weder die PC-Darbietung noch die Itemreihenfolge veränderten die Varianzen. Die Veränderungen der internen Konsistenzen waren unsystematisch und schwer interpretierbar. Papier-Bleistift- und PC-dargebotene Versionen der SCL-90-R sind vergleichbar.

Schlüsselwörter: SCL-90-R, computerunterstützte Darbietung, Itemreihenfolge, Reliabilität, Validität

Effects of computer administration on the Symptom Checklist (SCL-90-R) with a special focus on the item sequence

Abstract. Based on the revised version of the Symptom Checklist (SCL-90-R), a study was conducted involving 400 university students. (1) A paper-pencil version was compared with a computer-administered version, and (2) the original item arrangement was contrasted with the presentation of the blocked items of each subscale. Computer administration did not influence the means of the subscales. No interactions were shown with the factor "gender." Item-blocking led to higher means. Neither PC administration nor item sequence changed the variances. Changes in internal consistencies were unsystematic and hardly interpretable. Paper-pencil and PC versions of the SCL-90-R are comparable.

Key words: SCL-90-R, computer administration, item sequence, reliability, validity

Psychometrische Äquivalenz von Papier-Bleistift- und PC-unterstützter Psychodiagnostik

Die Ergebnisse psychometrischer Äquivalenzuntersuchungen bei persönlichkeitsdiagnostischen Verfahren sind uneinheitlich. Die psychometrische Äquivalenz der computerunterstützten Form des *Minnesota Multiphasic Personality Inventory (MMPI)* mit herkömmlicher Papier-Bleistift-Anwendung kann seit Mitte der 80er Jahre als gesichert gelten (Lambert et al., 1987; Russell, Peace & Mellsop, 1986; White, Clements & Fowler, 1985). Sowohl

Brielmeier (1995) als auch Franke (1997b) konnten für die revidierte Version des *Freiburger Persönlichkeitsinventares (FPI-R)* sowie Franke (1998) für das *FPI-A1* deutliche Hinweise auf die Äquivalenz finden.

Hinweise auf die psychometrische Äquivalenz beim *Eysenck Personality Questionnaire (EPQ)* (French & Beaumont, 1989), beim *16 Persönlichkeits-Faktoren-Test (16-PF)* (Harrell & Lombardo, 1984), dem *Stuttgarter Bogen für Abwehrkategorien* (Gitzinger, 1990) sowie dem *Gießen-Test* (Bader, Hofmann & Kubinger, 1993) gilt es, durch Replikationsstudien zu prüfen.

Erst in den 80er Jahren wurde mit der Erforschung der psychometrischen Äquivalenz psychodiagnostischer Verfahren, die zeitlich variable Konstrukte erfassen, begonnen. Heute kann nur die psychometrische Äquivalenz bei der *Hamilton Depression Rating Scale* (Akthar et al., 1984; Kobak et al., 1990, 1994) als gesichert gelten. Bei anderen Verfahren fehlen Replikationsstudien (Lukin et al., 1985; Malcolm et al., 1989; Schwenkmezger & Hank, 1993; Skinner & Allen, 1983).

Ich danke Dipl.-Ing. Ralf Baethke-Franke, Dipl.-Psych. Hellmut Hasenburger und Mara Stegemann für ihre Mitarbeit sowie allen Studierenden für ihre Studienteilnahme. Ebenso danke ich Prof. K.-H. Stäcker für die Unterstützung dieser Studie durch das Institut für Medizinische Psychologie am Universitätsklinikum in Essen. Ein abschließender Dank gilt einem anonymen Gutachter für seine wertvollen Hinweise.

Die PC-unterstützte Anwendung von psychodiagnostischen Instrumenten wird heute durch die immer selbstverständlicher werdende Nutzung von Personalcomputern für den Praktiker immer attraktiver. Sowohl das Testkuratorium der Föderation Deutscher Psychologenvereinigungen (1986) als auch die Amerikanische Psychologenvereinigung (1986) fordern, daß computerunterstützte Psychodiagnostik neu normiert werden muß. Ein Nachweis der Äquivalenz würde die Nutzer der computerunterstützten Versionen des Verfahren versichern, daß der Rückgriff auf vorhandene Normierungsdaten statthaft ist.

Die deutlichsten Vorteile in der Anwendung computerunterstützter Psychodiagnostik liegen in der Entlastung des Diagnostikers von Routinearbeit (Weise, 1994). Inwieweit die Erhebung zusätzlicher Daten wie die Bearbeitungszeit bei persönlichkeitspsychologischen Verfahren einen praktischen diagnostischen Fortschritt erbringen, ist bis heute nicht vollständig geklärt. Nützlich ist diese Datenquelle aber für die differentielle Erforschung der Auswirkungen von Zeitdruckinstruktionen, wie Neubauer und Malle (1997) zeigten, sowie für das adaptive Testen (Beckman, Guthke & Vahle, 1997; Hornke, 1997).

Effekte inhaltshomogener Item-Blockung

Die Auswirkungen einer inhaltshomogen geblockten Skalenpräsentation aus umfassenden Persönlichkeitsinventaren auf die Validität und die psychometrischen Kennwerte der Skalen wird seit einiger Zeit auch im deutschsprachigen Raum diskutiert.

Krampen, Hense und Schneider (1992) bezweifeln die Zuverlässigkeit der Subskalenextraktion beim *FPI-A*, da sich die Mittelwerte, die Reliabilitäts- und Validitätsparameter einiger Skalen veränderten. Allerdings waren die gefundenen Veränderungen unsystematische. Ein solches Ergebnis könnte deshalb auch als *Zufallsbefund* bewertet werden, wie es Rost und Hoberg (1997) nahelegen.

Krampen (1993) fand auch beim *Fragebogen zu Kompetenz- und Kontrollüberzeugungen* in zwei Studien mit einer Gesamtstichprobe von $N = 412$ Belege für Itemblockbildungseffekte. Er interpretierte die statistisch signifikant niedrigeren Skalenwerte für das Selbstkonzept, die Internalität und die fatalistische Externalität als Abschwächung zustimmender Antworten, die durch thematische Perseveration zustande gekommen seien. Die statistisch signifikant höheren Skalenwerte für soziale Externalität interpretierte er als Verstärkung zustimmender Antworten (im Sinne einer höheren Bereitschaft, solche sozialen Einflüsse auf das eigene Leben zu konzessionieren) bei Aussagen, in denen Einflüsse anderer Menschen auf personenrelevante Ereignisse beschrieben werden. Die Itemblockbildung führte in Krampens Studie (bis auf eine Ausnahme) zu einer Reduktion der Skalenreliabilität. Die Erhöhung der gemittelten Interkorrelationen der Skalen unter der Itemblockbildung interpretierte Krampen

(1993, S.105) als empirische Verwischung der Konstrukt-differenzierung. Diese Interpretation scheint theoretisch noch nicht genügend fundiert.

Schriesheim, Kopleman und Solomon (1989) stellten für den *Job Diagnostic Survey* fest, daß keine der untersuchten Itemanordnungen (randomisiert vs. Zufallsreihenfolge) der anderen überlegen ist. In früheren Studien fanden hingegen Schriesheim und DeNisi (1980) eine durch Itemblockbildung verursachte verringerte konvergente und diskriminante Validität bei Fragebogen zum Führungsverhalten. Auch Solomon und Kopleman (1984) berichten Veränderungen von Skalenhomogenitäten durch die Itemblockbildung bei Items zur Lebens- und Berufszufriedenheit.

In einer eigenen Studie wurde beobachtet, daß eine andere zufällige Itemreihenfolge als die originale bei der *SCL-90-R* nicht zu güteverändernden Effekten führte (Franke, 1996). Inhaltshomogene Itemdarbietungen führten zu unsystematischen Veränderungen der Mittelwerte sowie zu *unsystematischen – vermutlich zufallsbedingten* – Veränderungen bei den internen Konsistenz (Franke & Stäcker, 1995; Franke, 1997a; Hoberg & Rost, 1997).

Rost und Hoberg (1997) diskutieren ausführlicher die Auswirkungen der inhaltshomogenen Itemblockbildung als „tolerierbare Praxis“. In einer eigenen Studie präsentierten sie zwei Stichproben von Schülern (Gesamt $N = 600$) 18 Zielitems (je sechs Items zum schulischen Selbstkonzept, die für die Fächer Mathematik, Deutsch und Erdkunde gleich formuliert waren), entweder zufällig oder in sechs Blöcken zu je drei formulierungsmäßig äquivalenten Items unter 107 Distraktoritems gemischt. Der in der MANOVA statistisch bedeutsame Haupteffekt „Fragebogenversion“ zeitigte eine Varianzaufklärung von 4.8%. Zwar stiegen die Skalenmittelwerte bei zwei Skalen statistisch signifikant an, Rost und Hoberg (1997, S. 108) interpretierten die univariaten Varianzaufklärungen von 1.2% bzw. 2.4% unter Berücksichtigung der Interkorrelationen der Skalen als „von vergleichsweise geringer praktischer Bedeutung“, da die Effekte auf praktisch-diagnostische Entscheidungen zu vernachlässigen sind. Die statistische Signifikanz der bei diesen beiden Skalen ebenfalls vorliegenden Homogenitätsdifferenzen ist aber bei großen Stichproben und hohen Homogenitäten „praktisch nicht relevant“ (ebd., vgl. auch Hoberg & Rost, 1997, S. 222-223).

Weiterhin stellten sie überzeugend dar, daß die von ihnen realisierte Itemblockbildung die Konstrukt-differenzierung nicht verwischt, da sich faktorenanalytisch unter beiden Itemdarbietungen die jeweiligen Skalen optimal reproduzieren ließen.

Rost und Hoberg (1997) realisierten nur eine teilweise inhaltshomogene Itemblockbildung, die zu, ihrer Meinung nach, vernachlässigbaren Effekten führte. Es wäre interessant zu untersuchen, welche Effekte die vollständige Itemblockbildung der von ihnen untersuchten Skalen, ohne die Nutzung von Distraktoritems, hätte. Bezüglich des *FPI-A* sowie der *SCL-90-R* ist festzustellen, daß unsystematische Auswirkungen der Itemblockbildung

nachzuweisen sind, die vermutlich Zufallseffekte darstellen. Es scheint somit notwendig, die Effekte der inhaltshomogenen Itemblockbildung noch weiter zu erforschen.

Methode

In der folgenden Studie soll die psychometrische Äquivalenz zwischen der Papier-Bleistift- und der Computerversion bei der *SCL-90-R* zusammen mit Auswirkungen der inhaltshomogenen Itemblockbildung analysiert werden. Bisher wenig untersucht wurde, ob Blockbildungseffekte auch bei PC-unterstützten Verfahren zu finden sind. In einer eigenen Studie konnten Blockbildungseffekte beim *FPI-R* unter PC-unterstützter Darbietung (und auch bei der herkömmlichen Papier-Bleistift-Darbietung) kaum nachgewiesen werden (Franko, 1997b). Die beiden Forschungsfragen nach den Auswirkungen der Computerdarbietung und den Auswirkungen der Itemblockbildung wurden ergänzt durch die Frage, ob sich statistisch signifikante Interaktionen zwischen den beiden experimentell manipulierten Faktoren und dem Faktor „Geschlecht“ finden lassen, da die Befundlage zu geschlechtsspezifischen Unterschieden heterogen ist (Lamberti, 1991; Sanitioso & Reynolds, 1992; Scissons, 1976).

Stichprobenbeschreibung

An dem Versuch nahmen 400 Studierende teil (50% männlich, 50% weiblich). Ihr Durchschnittsalter betrug 24.3 Jahre ($SD = 5$; Median = 23 J.).

Untersuchungsmaterial

Den Probanden wurde die *Symptom-Checkliste (SCL-90-R)* in der originalen Itemreihenfolge (Normal) oder in einer itemgeblockten Darbietung (Block) dargeboten. Diese inhaltshomogen geblockte Version berücksichtigt alle 90 Items, die in acht inhaltshomogenen Gruppen dargeboten werden. Die erste Itemgruppe umfaßte 15 Items zur „Selbstwertproblematik“, die zweite 14 Items zu „Körperlich manifester Angst“, die dritte 10 Items aus dem Bereich „Traurigkeit bis Depressivität“, gefolgt von 12 Items zur „Körperlichen Belastung“, 12 Items zu „Entfremdung/Depersonalisation“, 11 Items zu „Furcht“ und 6 Items zu „Aggressivität“.

Die Originalskalen lauten Somatisierung (SOMA), Zwanghaftigkeit (ZWAN), Unsicherheit im Sozialkontakt (UNSI), Depressivität (DEPR), Ängstlichkeit (ANGS), Aggressivität/Feindseligkeit (AGGR), Phobische Angst (PHOB), Paranoides Denken (PARA) sowie Psychotizismus (PSYC). Zu ihrer Berechnung werden nur 83 der 90 Items herangezogen. Drei globale Kennwerte geben Auskunft über alle 90 Items. GSI mißt die globale psychische Belastung, PSDI mißt die Intensität der Antworten und PST gibt Auskunft über die Anzahl der belastenden Symptome.

Untersuchungsablauf

Jeweils die Hälfte der Probanden ($N = 200$, 50% männlich, 50% weiblich) beantworteten die Papier-Bleistift-Version (Papier), die andere Hälfte bearbeitete die computergestützte Darbietung (Computer). Die Gruppenzuweisung erfolgte nach dem Randomisierungsprinzip. Innerhalb dieser beiden Gruppen beantworteten (ebenfalls zufällig zugeordnet) je 100 Probanden die *SCL-90-R* bei originaler Itemreihenfolge (Normal), je 100 bei itemgeblockter Darbietung nach den Ergebnissen einer früheren Faktorenanalyse (Block).

Die Instruktion der PC-Version entsprach weitgehend dem Original. Die Antworten der Probanden konnten sowohl durch ein Anklicken der jeweiligen Schaltfläche mit der linken Maustaste als auch durch ein Drücken der entsprechenden Buchstabentaste auf der Tastatur gegeben werden. Jede Frage wurde einzeln präsentiert. Es war möglich, schon beantwortete Fragen durch Zurückblättern erneut anzuschauen und zu korrigieren bzw. Fragen auszulassen. Von der Auslassungsmöglichkeit machten nur wenige Probanden Gebrauch; diese Probanden wurden nicht in die Auswertung aufgenommen. Nach jeder Antworteingabe wurde das folgende Item präsentiert. Das Experiment umfaßte noch verschiedene andere Fragebereiche, auf die hier aus Platzgründen nicht eingegangen wird.

Auswertung

Die Auswertung wird anhand der T-transformierten Skalenwerte der Originalskalen der *SCL-90-R* vorgenommen.

(1) Mit dem Box-Test erfolgte die Prüfung der Homogenität der Varianz-Kovarianz-Matrizen pro Skala [Gruppierungsvariablen: Testform (A), Itemreihenfolge (I), Geschlecht (G)].

(2) Mittelwertsdifferenzen wurden in einer vierfaktoriellen univariaten Varianzanalyse mit Meßwiederholung [Gruppierungsvariablen: Testform (A), Itemreihenfolge (I), Geschlecht (G), Meßwiederholung bezüglich der Skalen (S)] und gegebenenfalls post-hoc überprüft. Zur Prüfung der Voraussetzungen der Unabhängigkeit und Varianzhomogenität der transformierten Variablen wurde der Mauchly-Sphericity-Test eingesetzt. Streuungsdifferenzen wurden mit Hilfe des F-Tests von Levene überprüft.

(3) Zur Konsistenzschätzung wurde Cronbachs α herangezogen. Der statistische Vergleich der α -Koeffizienten erfolgte nach Feldt, Woodruff und Salih (1987); nachgeschaltete Paarvergleiche wurden mit Hilfe des F-Tests von Feldt (1969) durchgeführt.

Das Signifikanzniveau wurde bei $p = 0.01$ festgelegt. Die bekannten Probleme des multiplen Testens fanden unter Rückgriff auf Feild und Armenakis (1974) bzw. auf die α -Fehleradjustierung Berücksichtigung.

Ergebnisse

Skalenzusammenhänge

Laut Box-Test unterschieden sich die Varianz-Kovarianz-Matrizen der verschiedenen Versionen univariat pro Skala statistisch nicht signifikant voneinander.

Mittelwertsunterschiede

Auf eine Modifikation der Freiheitsgrade wurde verzichtet, da der Korrekturfaktor ϵ nach Greenhouse-Geisser ($\epsilon = 0.89$) sowie der korrigierte ϵ -Wert nach Huyn-Feldt ($\epsilon = 0.92$) jeweils $\epsilon > 0.75$ lagen (bei $W = 0.60$, $p < 0.01$ im Mauchly-Sphericity-Test). In der varianzanalytischen Prüfung hinsichtlich der neun T-transformierten Skalen (SOMA bis PSYC) fanden sich ein statistisch signifikanter Haupteffekt im Faktor „Itemreihenfolge (I)“ mit einer Varianzaufklärung von 3.1% sowie der erwartete Haupteffekt im Meßwiederholungsfaktor (6.2% Varianzaufklärung). Weiterhin fanden sich keinerlei Interaktionen mit dem Faktor „Geschlecht (G)“. Zwei Interaktionen mit dem Meßwiederholungsfaktor erreichten die statistische Signifikanzgrenze ($I \times S$, $\eta^2 = 0.013$, $G \times S$, $\eta^2 = 0.012$). Die Interaktion zwischen dem Faktor „Geschlecht“ und dem Meßwiederholungsfaktor interessiert im hier diskutierten Zusammenhang nicht. Die Interaktion zwischen dem Faktor „Itemreihenfolge“ und dem Meßwiederholungsfaktor weist darauf hin, daß die Itemblockbildung die einzelnen Skalenmittelwerte unterschiedlich beeinflusste. Die inhaltshomogene Itemblockbildung führte nur bei SOMA

($\eta^2 = 0.033$), UNSI ($\eta^2 = 0.057$) sowie PARA ($\eta^2 = 0.046$) zur statistisch signifikanten Erhöhung der Mittelwerte.

Die PC-unterstützte Darbietung führte also nicht zu einer Veränderung auf der Ebene der Mittelwerte, wohl aber die geblockte Darbietung der Items. Die Itemblockbildungseffekte zeigten sich sowohl bei der Papier-Bleistift-Darbietung als auch unter der PC-Bedingung.

Prüfung der internen Konsistenzen

Tab. 2 gibt eine Übersicht über die Koeffizienten der internen Konsistenz (Cronbachs α). Im χ^2 -Test wurde nachgewiesen, daß sich die Homogenitätsparameter bei den Skalen SOMA, DEPR, AGGR, PHOB, PARA, PSYC und GSI unter den verschiedenen experimentellen Bedingungen unterschieden (7 von 10 Tests, statistisch signifikantes Ergebnis nach Feild & Armenakis, 1974). Angeregt durch Rost und Hoberg (1997) soll die statistische Signifikanz von Homogenitätsdifferenzen bei SOMA, DEPR und GSI als praktisch nicht relevant bewertet werden, da sie hauptsächlich auf das große N und die guten Homogenitäten ($\alpha > 0.74$) zurückzuführen ist. Bei AGGR und PARA zeigten nachgeschaltete Paarvergleiche auf, daß sich allein die Gruppe 1 von den drei anderen unterschied. Bei PHOB unterschied sich die Gruppe 4 von den anderen drei, auch unterschieden sich die Gruppen 2 und 3. Bei PSYC hingegen unterschied sich sowohl die Gruppe 1 als auch die Gruppe 4 von den Gruppen 2 und 3. Zusammengefaßt läßt sich keine Systematik erkennen; der Verdacht auf Zufallseffekte liegt nahe.

Tabelle 1. Ergebnisse der vierfaktoriellen univariaten varianzanalytischen Prüfung in den Faktoren „Testform (A)“, „Itemreihenfolge (I)“, „Geschlecht (G)“ und „Skalen (S)“ mit Meßwiederholung im letzten Faktor hinsichtlich der neun T-transformierten Skalen der SCL-90-R.

Faktoren	df	F	p	η^2
Testform (A)	1,392	0.95	.33	.002
Itemreihenfolge (I)	1,392	12.67	.01	.031
Geschlecht (G)	1,392	1.01	.32	.003
A x I	1,392	3.25	.07	.008
A x G	1,392	1.85	.18	.005
I x G	1,392	2.42	.12	.006
A x I x G	1,392	0.90	.34	.002
Skalen (S)	8,3136	25.80	.01	.062
A x S	8,3136	2.13	.03	.005
I x S	8,3136	5.10	.01	.013
G x S	8,3136	4.67	.01	.012
A x I x S	8,3136	1.22	.28	.003
A x G x S	8,3136	1.00	.43	.003
I x G x S	8,3136	0.93	.49	.002
A x I x G x S	8,3136	1.85	.06	.005

Neun Skalen der SCL-90-R: Somatisierung, Zwanghaftigkeit, Unsicherheit im Sozialkontakt, Depressivität, Ängstlichkeit, Aggressivität/Feindseligkeit, Phobische Angst, Paranoides Denken und Psychotizismus.

Tabelle 2. Vergleich der internen Konsistenzen der SCL-90-R-Skalen.

Gruppe	1	2	3	4	
Testform	Papier	Papier	Computer	Computer	χ^2 -Test
Itemreihenfolge	Normal N = 100	Block N = 100	Normal N = 100	Block N = 100	
Somatisierung	.85	.76	.74	.85	11.42**
Zwanghaftigkeit	.82	.72	.82	.79	4.71
Unsicherheit im Sozialkontakt	.82	.73	.78	.84	6.41
Depressivität	.89	.82	.85	.90	9.02*
Ängstlichkeit	.83	.72	.78	.81	5.70
Aggressivität/Feindseligkeit	.79	.59	.60	.50	15.12**
Phobische Angst	.58	.45	.66	.82	26.08**
Paranoides Denken	.77	.46	.63	.57	14.65**
Psychotizismus	.81	.60	.67	.80	16.16**
Globaler Kennwert GSI	.97	.94	.96	.97	15.85**

* = $p < 0.05$, ** = $p < 0.01$

Diskussion

Die vorliegende Studie überprüfte die psychometrische Äquivalenz bei herkömmlicher Papier-Bleistift- und einer PC-unterstützt dargebotenen Version bei der *SCL-90-R*. Neben diesem Hauptfaktor „Testform“ wurde weiterhin die Reihenfolge der Items variiert („Itemreihenfolge“).

- Zusammenfassend kann festgestellt werden, daß die PC-Version der *SCL-90-R* in dieser Studie hinsichtlich des Geschlechtes keine Auswirkungen zeigt. Dieses Ergebnis korrespondiert mit früheren Befunden (Sanitioso & Reynolds, 1992) sowie mit eigenen Befunden hinsichtlich des *FPI* (Franke, 1997b, 1998).
- Weiterhin zeigte sich kein Einfluß der Testform auf die Skalenmittelwerte.
- Die inhaltshomogen geblockte Itemreihenfolge führte bei einer Varianzaufklärung von $\eta^2 = 3.1\%$ zu einem kleinen Effekt bezüglich der Mittelwerte.
- Beide experimentell variierte Faktoren beeinflussten weiterhin nicht die Varianzen der Skalen.
- Die Befunde zu den internen Konsistenzen sind *recht unsystematisch und inhaltlich nicht interpretierbar*. Bei den Skalen SOMA, DEPR und GSI sollten die gefundenen Unterschiede, der Argumentation von Rost und Hoberg (1997) folgend, aufgrund des Stichprobenumfangs sowie der guten Homogenität ($\alpha > 0.74$) als praktisch irrelevant eingeschätzt werden. Bei den verbleibenden Skalen AGGR, PHOB, PARA und PSYC folgten die Veränderungen in den internen Konsistenzen keinem inhaltlich interpretierbaren Muster; sie sind ebenfalls in Anlehnung an die Argumentation von Rost und Hoberg (1997) bzw. Hoberg & Rost (1997) als *Zufallseffekte* einzustufen.

Es bleibt die Frage, ob die gefundene Varianzaufklärung von 3.1% bei der Betrachtung der Auswirkung der Itemreihenfolge auf alle neun Skalen nun relevant ist oder nicht. Rost und Hoberg (1997) interpretieren die von ihnen gefundene Varianzaufklärung des Reihenfolgefaktors mit 4.8% als nicht relevant, da die Auswirkungen auf diagnostische Entscheidungen in der Praxis, der Argumentation von Rost und Hoberg (1997) folgend, in den meisten Fällen vernachlässigt werden können. Die statistisch bedeutsamen Veränderungen der Homogenitätsparameter sind in drei Fällen praktisch nicht bedeutsam und folgen in vier Fällen keiner inhaltlichen Logik. Diese unklaren Ergebnisse bedürfen noch weiterer Prüfung.

Weiterhin bleibt zusammenfassend festzustellen, daß die PC-unterstützte Version der Papier-Bleistift-Version bei der *SCL-90-R* annähernd äquivalent ist. Dieses Ergebnis korrespondiert mit den schon in der Einleitung zitierten Befunden zum *MMPI*, zum *FPI* und zum *Gießen-Test*.

Bei der Auswertung PC-unterstützt erhobener Daten der *SCL-90-R* kann unbedenklich auf die Normierungsdaten zurückgegriffen werden. Die Forderungen sowohl der Amerikanischen Psychologengemeinschaft (1986) als auch des Testkuratoriums der Föderation Deutscher Psychologengemeinschaften (1986) nach der zwingenden Neunormierung computerunterstützter Psychodiagnostik sind nicht für jedes Verfahren berechtigt.

Literatur

- Akhthar, M. J., Davey, A., Cox, H. E. & Ancill, R. J. (1984). A double-blind study comparing mianserin and dothiepin: An application for computers in clinical psychology. *The British Journal of Clinical Practice*, 9, 316-319.

- American Psychological Association (APA) (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: APA.
- Bader, P., Hofmann, K. & Kubinger, K. D. (1993). Zur Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel des Gießen-Tests. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *14*, 129-135.
- Beckmann, J. F., Guthke, J. & Vahle, H. (1997). Analysen zum Zeitverhalten bei computergestützten adaptiven Intelligenz-Lerntests. *Diagnostica*, *43*, 40-62.
- Brielmeier, I. (1995). *Bewirkt der Computer einen Unterschied? Eine experimentelle Studie zur Äquivalenz von Papier-Bleistift- und Computerversion des FPI-R*. Marburg: Diplomarbeit im Fach Psychologie, Fachbereich Psychologie der Phillips-Universität Marburg.
- 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.
- Franke, G. H. (1995). *SCL-90-R. Die Symptom-Checkliste von Derogatis – Deutsche Version*. Göttingen: Beltz Test.
- Franke, G. H. (1996). Effekte von Typographie und Itempositionierung in der Fragebogendiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *17*, 187-200.
- Franke, G. H. (1997a). The whole is more than the sum of its parts: The effect of grouping and randomizing items on the reliability and validity of questionnaires. *European Journal of Psychological Assessment*, *13*, 67-74.
- Franke, G. H. (1997b). Über die Möglichkeit der computerunterstützten Darbietung beim revidierten Freiburger Persönlichkeitsinventar. Zwei experimentelle Studien. *Zeitschrift für Experimentelle Psychologie*, *44*, 332-356.
- Franke, G. H. (1998). Auswirkungen der Computerapplikation beim Freiburger Persönlichkeitsinventar unter besonderer Berücksichtigung der Itemreihenfolge. Eine erste experimentelle Studie. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *19*, 187-199.
- Franke, G. H. & Stäcker, K.-H. (1995). Reliabilität & Validität der Symptom-Checkliste [SCL-90-R, Derogatis, 1986] bei Standardreihenfolge versus inhaltshomogener Itemblockbildung. *Diagnostica*, *41*, 349-73.
- French, C. C. & Beaumont, J. G. (1989). A computerized form of the Eysenck Personality Questionnaire: A clinical study. *Personality and Individual Differences*, *10*, 1027-1032.
- Gitzinger, I. (1990). Akzeptanz der Darbietung eines Tests auf dem Personalcomputer von stationären Patient/innen. *Psychotherapie, Psychosomatik, medizinische Psychologie*, *40*, 143-5.
- Harrell, T. H. & Lombardo, T. A. (1984). Validation of an automated 16PF administration procedure. *Journal of Personality Assessment*, *48*, 638-42.
- Hoberg, K. & Rost, D. H. (1997). Zufallsreihenfolge vs. Blockbildung: Beeinflusst die Itemanordnung die Kennwerte von Fragebogenskalen? In Thonhauser, J. & Riffert, F. (Hrsg.), *Evaluation heute – zwölf Antworten auf aktuelle Fragen* (S. 215-226). Braunschweig: Braunschweiger Studien zur Erziehungs- und Sozialarbeitswissenschaft.
- Hornke, L. F. (1997). Untersuchung von Itembearbeitungszeiten beim computergestützten adaptiven Testen. *Diagnostica*, *43*, 27-39.
- Kobak, K. A., Reynolds, W. M. & Greist, J. H. (1994). Computerized and clinician assessment of depression and anxiety: Respondent evaluation and satisfaction. *Journal of Personality Assessment*, *63*, 173-180.
- Kobak, K. A., Reynolds, W. M., Rosenfeld, R. & Greist, J. H. (1990). Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. *Psychological Assessment*, *2*, 56-63.
- Krampen, G. (1993). Effekte von Bewerbungsinstruktionen und Subskalensextraktion in der Fragebogendiagnostik. *Diagnostica*, *39*, 97-108.
- Krampen, G., Hense, H. & Schneider, J. (1992). Reliabilität und Validität von Fragebogenskalen bei Standardreihenfolgen versus inhaltshomogener Blockbildung ihrer Items. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *34*, 229-248.
- Lambert, M. E., Andrews, R. H., Rylee, K. & Skinner, J. R. (1987). Equivalence of computerized and traditional MMPI administration with substance abusers. *Computers in Human Behavior*, *3*, 139-143.
- Lamberti, G. (1991). Persönlichkeitsdiagnostik: Computergestützte oder Paper-Pencil-Vorgabe. *Zeitschrift für Neuropsychologie*, *2*, 125-129.
- Lukin, M. E., Dwod, E. T., Plake, B. S. & Kraft, R. G. (1985). Comparing computerized versus traditional psychological assessment. *Computers in Human Behavior*, *1*, 49-58.
- Malcolm, R., Sturgis, E. T., Anton, R. F. & Williams, L. (1989). Computer-assisted diagnosis of alcoholism. *Computers in Human Services*, *5*, 163-170.
- Neubauer, A. C. & Malle, B. F. (1997). Questionnaire response latencies: Implications for personality assessment and self-schema theory. *European Journal of Psychological Assessment*, *13*, 109-117.
- Rost, D. & Hoberg, K. (1997). Itempositionsveränderungen in Persönlichkeitsfragebogen: Methodischer Kunstfehler oder tolerierbare Praxis? *Diagnostica*, *43*, 97-112.
- Russell, G. K. G., Peace, K. A. & Mellsop, G. W. (1986). The reliability of a micro-computer administration of the MMPI. *Journal of Clinical Psychology*, *42*, 120-122.
- Sanitioso, R. & Reynolds, J. H. (1992). Comparability of standard and computerized administration of two personality questionnaires. *Personality and Individual Differences*, *13*, 899-907.
- Schriesheim, C. A. & DeNisi, A. S. (1980). Item presentation as an influence on questionnaire validity: A field experiment. *Educational and Psychological Measurement*, *40*, 175-182.
- Schriesheim, C. A., Kopelman, R. E. & Solomon, E. (1989). The effect of grouped versus randomized questionnaire format on scale reliability and validity: A three-study investigation. *Educational and Psychological Measurement*, *49*, 487-508.
- Schwenkmezger, P. & Hank, P. (1993). Papier-Bleistift- versus computerunterstützte Darbietung von State-Trait-Fragebogen: eine Äquivalenzüberprüfung. *Diagnostica*, *39*, 189-210.
- Scissons, E. H. (1976). Computer administration of the California Psychological Inventory. *Measurement and Evaluation in Guidance*, *9*, 22-5.
- Skinner, H. A. & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use. *Journal of Consulting and Clinical Psychology*, *51*, 267-275.
- Solomon, E. & Kopelman, R. E. (1984). Questionnaire format and scale reliability: An examination of three modes of item presentation. *Psychological Reports*, *54*, 447-452.

- Testkuratorium der Föderation Deutscher Psychologervereinigungen (1986). Richtlinien für den Einsatz elektronischer Datenverarbeitung in der psychologischen Diagnostik. *Psychologische Rundschau*, 37, 163-5.
- Weise, G. (1994). Gütekriterien – speziell die Objektivität – von Tests in Abhängigkeit von der Art der Auswertung: Manuelle vs. Computerauswertung. In D. Bartussek & M. Amelang (Hrsg.), *Fortschritte der Differentiellen Psychologie und Psychologischen Diagnostik – Festschrift zum 60. Geburtstag von Prof. Dr. Kurt Pawlik* (S. 307-317). Göttingen: Hogrefe.
- White, D. M., Clements, C. B. & Fowler, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. *Computers in Human Behavior*, 1, 153-162.

Priv.-Doz. Dr. Gabriele Helga Franke

Institut für Medizinische Psychologie
Universitätsklinikum Essen
Hufelandstraße 55
D-45122 Essen
Tel. 02 01/7 23-42 82
Fax. 02 01/7 23-59 48
E-Mail: gabriele.franke@uni-essen.de
<http://www.uni-essen.de/~omp.020>