

Ist Empathie eine Lüge?

Zur Frage der sozialen Erwünschtheit in der Empathiemessung

PD Dr. Christoph Paulus
Universität des Saarlandes
Fakultät für Empirische Humanwissenschaften und Wirtschaftswissenschaft
www.cpaulus.de
© 2019

Zusammenfassung

In der Studie sollte untersucht werden, ob und ggf. wie stark die Messung von Empathie mittels Fragebogen anfällig gegen Verfälschungstendenzen sei. Dabei wurden sowohl (eher absichtliche) Tendenzen des impression managements (Fremdtäuschung) als auch (eher unabsichtliche) Tendenzen des self-deceptive enhancement (Selbsttäuschung) betrachtet. Insgesamt nahmen 953 Personen im Alter zwischen 12 und 74 Jahren an der Studie teil. Es zeigte sich, dass die deutsche Version des IRI, der Saarbrücker Persönlichkeitsfragebogen zur Messung von Empathie im Bereich des IM recht robust ist. Die Tendenz der Selbsttäuschung zeigte lediglich beim Faktor personal distress einen Effekt. Die allgemeine Tendenz, Antworten bewusst oder unbewusst zu verfälschen, nahm mit zunehmendem Alter der Vpn zu. Es gab keine signifikanten Unterschiede in der Häufigkeit, mit der männliche und weibliche Pbn eine der beiden Strategien benutzen.

Summary

The purpose of the study was to examine whether the measurement of empathy by means of questionnaire is susceptible against falsification. Besides, trends of the impression management as well as tendencies of self-deceptive enhancement were looked at. 953 participants at the age between 12 to 74 years took part at the study. It appeared, that the German version of the IRI, the Saarbruecken Personality Questionnaire of empathy is rather robust at the aspect of impression management. The trend of self-deception showed merely an effect at the factor personal distress. In general, the behaviour of faking increased with increasing age of the subjects. There were no significant differences in the frequency with which male and female subjects used one of both strategies.

Die Messung von Persönlichkeitseigenschaften oder Einstellungen geschieht in der Regel über Fragebögen. Diese werden nach testtheoretischen Verfahren in ihren Gütemaßstäben optimiert und somit wird angenommen, dass, insbesondere bei guten Reliabilitätswerten, die Antworten weitestgehend messfehlerfrei sind. Kommen jetzt noch „optimale“ Testbedingungen dazu, also in anonymer Form, die Antwortskala ist verständlich und nachvollziehbar und es gibt keine direkten Folgen für die Pbn, so geht man davon aus, dass auch wahrheitsgemäß geantwortet wird.

Dabei ist die Beantwortung von Items in einem Persönlichkeitsfragebogen trotz allem noch messfehlerbehaftet. Dies betrifft in erster Linie zunächst einmal die inhaltliche Genauigkeit, also die Frage, wie reflektiert die Versuchsperson bei der Beantwortung der Inhalte ist bzw. wie exakt auch die Differenzierung in der Antwortskala nachvollzogen werden kann. Kaum problematisch dürften 5-stufige Likert-Skalierungen sein, wobei aber auch hier die subjektiven Kategorien wie „oft“ oder „manchmal“ durchaus unterschiedlich sein können. In einer informellen Befragung des Autors mit Studierenden wurde z.B. die Häufigkeit des Begriffes „oft“ auf einer Skala von 0-100% zwischen 30% und 85% benannt.

Wichtiger erscheint allerdings die Frage nach der Ehrlichkeit der Antworten, d.h. gibt es bewusste oder auch unbewusste Tendenzen, die gegebenen Antworten zu schönen. Je mehr persönliche Informationen der Pb von sich preisgeben muss, wie Name, Adresse etc., desto höher ist die Wahrscheinlichkeit eines sozial erwünschten Antwortverhaltens. Gleiches gilt, wenn die Ergebnisse eines Testes persönliche Konsequenzen wie z.B. Selektionsentscheidungen (Ausbildung, Beruf oder Karriere) nach sich ziehen. Im normalen Alltag der empirischen Testungen ist das allerdings nicht der Fall - dort sind die Erhebungen in der Regel anonym bzw. bei Retestungen codiert und persönliche Angaben, die über Alter, Geschlecht und Schulbildung hinausgehen eher selten. Damit wird der Einfluss bewusster Täuschung zwar reduziert, aber nicht ganz ausgeschlossen.

Eine weitere Frage ist die, ob bei vielen psychologischen Konstrukten der Testperson überhaupt bewusst ist, was sozial erwünscht ist bzw. wie sie sich durch ihre Antworten so darstellen kann. Auch hier gibt es natürlich Bereiche, in denen generell eher vorsichtig oder zurückhaltend geantwortet wird, wie z.B. bei Fragen zur Gesundheit, zu Sexualität oder Aggression. Als weniger anfällig hat sich dagegen z.B. das Thema Lernen erwiesen (Paulus, 1999, 2001; Paulus & Lauermaun, 2004).

Schon sehr lange stehen Persönlichkeitstests auch in der Kritik im Hinblick auf bewusste Verfälschungsmöglichkeiten seitens der Pbn. Dieses als „soziale Erwünschtheit“ oder auch „impression management“ bezeichnete Phänomen wird öfters diskutiert mit der Folge, dass sogar teilweise vom Einsatz von Persönlichkeitstests z.B. im Bereich der Personalauswahl ganz abgeraten wurde (Heilmann, 1999 oder Amelang & Bartussek, 2001). Der Effekt dieser bewussten oder auch unbewussten Verfälschung von Testantworten kann im Bereich von einer halben bis weit über einer ganzen Standardabweichung liegen (Allinger & Dwight, 2000; Marcus, 2003). Beide Verfälschungstendenzen, bewusste bzw. unbewusste, werden in der Forschung mit unterschiedlichen Termini betrachtet. Damarin & Messick (1965) benutzten die Begriffe „autistic bias in self-regard“ vs. „propagandistic bias“, Sackheim & Gur (1978) sprachen von „self-deception“ vs. „other-deception“ und Kusyszyn & Jackson (1968) verglichen „desirability“ und „defensiveness“. Paulhus (1984, 1986) entwickelte dazu das „Balanced Inventory of Desirable Responding (BIDR)“, das diese

zwei Komponenten sozial erwünschten Antwortverhaltens erfasst: (bewusste) Eindruckssteuerung (impression management = IM) bzw. (unbewusste) Selbsttäuschung (self-deceptive enhancement = SDE). Impression management wird dabei als die bewusste Verfälschung von Testantworten angesehen mit dem Ziel, sich selbst in einem möglichst günstigen Bild darzustellen; Paulhus (1986) bezeichnet dies als eine Eigenschaft einer „social conventional, dependable persona“ (S. 21). Etwas komplizierter ist die Komponente der Selbsttäuschung, die eher dem Schutz des Selbstbildes und des Selbstwertgefühls dient. In vertretbarem Maße scheint eine Selbsttäuschungstendenz Zeichen eines psychisch gesunden und gut angepassten Individuums zu sein (Paulhus, 1986). Erst wenn der Wert der SDE sehr hoch wird, deutet dies auf eine in starkem Maße verzerrte Selbstwahrnehmung hin. Paulhus (ebd.) konnte zeigen, dass zwischen beiden Tendenzen nur eine sehr geringe Korrelation (zwischen .20 und .32) existiert.

Das BIDR besitzt im englischen Original je 20 Items für beide Skalen mit einer 7-poligen Antwortskala mit den Eckpunkten „völlige Ablehnung“ (= 1) bis zu „völlige Zustimmung“ (=7). Musch et al. (2002) haben eine deutsche Übersetzung dazu getestet, in der wie im Original die Skalenmittelwerte durch Aufsummierung (und ggf. Umpolung der negativen Items) der Itemwerte ermittelt wurden. Nach Auswertung der Testgütekriterien blieben jeweils 10 Items pro Faktor übrig. Auch hier war die Korrelation beider Faktoren mit $r = .19$ sehr gering. Konvergente und diskriminante Validitäten der beiden Skalen konnten durch Korrelationen mit anderen Maßen sozialer Erwünschtheit eindeutig in mehreren Studien belegt werden (ebd). Ebenfalls replizierbar war die hohe negative Korrelation der Neurotizismus-Skala des NEO-FFI mit der Selbsttäuschungsskala des BIDR, was darauf hindeutet, dass Personen mit hohen Selbsttäuschungswerten demnach emotional stabil und ausgeglichen sind. Etwas kritisch zu sehen ist allerdings die mit .64 bzw .66 eher mittlere interne Konsistenz der Skalen (ebd).

Bei der vorliegenden Untersuchung ging es um die Frage, inwieweit das Konstrukt Empathie anfällig für die beiden genannten Verfälschungsarten sei. Empathie wird hier als multidimensionale Eigenschaft angesehen, die sowohl kognitive wie auch affektive Komponenten aufweist. Wir stützen uns dabei auf den Interpersonal Reactivity Index (IRI) von Davis (1983, 1994) in der deutschen Bearbeitung (SPF, s. Anhang) von Paulus (1997, 2001, 2006). Der Saarbrücker Persönlichkeitsfragebogen zur Messung von Empathie (SPF) weist sehr gute Kennwerte in den Bereichen der Reliabilität, der Validität und der Item-Trennschärfe auf (Paulus, 2009).

Davis ging bei der Konstruktion des IRI von vier zueinander in Beziehung stehenden Empathiekonstrukten aus: *perspective taking* (PT), *fantasy* (FS), *empathic concern* (EC) und *personal distress* (PD). PT misst dabei die Fähigkeit, eine Sache aus der Perspektive eines Gegenübers sehen zu können. Die *fantasy scale* (FS) erfasst die Tendenz, sich in die Gefühlswelt von Figuren in Romanen oder Filmen zu versetzen und ist insofern eher unter emotionalen Aspekten zu betrachten (Davis & Franzoi, 1991). Die EC-Skala dient zur Messung fremd-orientierter Gefühle wie Mitleid oder Sorge um Personen in Not, die PD-Skala dagegen soll eigenfokussierte Gefühle wie Unruhe oder Unwohlsein in interpersonalen Situationen messen. Damit enthält der IRI drei emotionale (EC, FS und PD) und einen kognitiven Empathie-Faktor (PT). Obwohl von Davis im Original nicht vorgesehen, konnte Paulus (2012) durch Anwendung des Verfahrens der ROC-Kurven die Bestimmung eines Gesamtscores für Empathie möglich machen. Dieser beinhaltet allerdings nur die drei Faktoren EC, PT und FS, die additiv miteinander verknüpft werden. Die *personal distress-Skala* wird in mehreren anderen Untersuchungen (u.a. Stotland et al., 1978; Davis, 1996; Beaven, 2004 oder Lauterbach et al., 2007) als kritisch angesehen und ihre Rolle im Kontext der Empathie ist bisher noch nicht eindeutig geklärt. Eine kritische Auseinandersetzung mit dieser Skala findet sich u.a. bei Paulus (2014). Die

PD-Skala wird von manchen Autoren (z.B. Baron-Cohen & Wheelwright, 2004) auch als ein Maß zur Emotionsregulation interpretiert.

Die vier Skalen sind faktorenanalytisch sehr klar nachweisbar (vgl. Paulus, 2009).

Methode

Stichprobe

Insgesamt nahmen 953 Personen an der Testung teil, davon waren 353 männlich und 576 weiblich (der Rest ohne Angaben). Das Alter der Vpn reichte von 12 bis zu 74 Jahren mit einem Mittelwert von 30.1 Jahren ($s = 11.55$).

Materialien

Der Saarbrücker Persönlichkeitsfragebogen (SPF)

Die Daten resultieren aus der Onlineversion¹ des SPF, dessen 16 Items zusätzlich noch durch die o.g. sechs Items der BIDR-Kurzform ergänzt wurden. Die Teilnahme war freiwillig, die Vpn erhielten nach der Bearbeitung im Internet eine Rückmeldung in Form von Norm- und PR-Werten sowie einer einfachen Erklärung dazu, eine Auswertung der Items zur sozialen Erwünschtheit wurde nicht gegeben. Die Empathiescores wurden zum besseren Verständnis Z-transformiert (Mittelwert = 100, $s = 10$).

Soziale Erwünschtheit

Zur Messung der sozialen Erwünschtheit verwendeten wir das Balanced Inventory of Desirable Responding (BIDR) nach Paulhus (1984) in der Kurzform von Winkler et al. (2006), das die zwei Dimensionen Selbst- und Fremdtäuschung (SDE und IM) abbildet. Die beiden Skalen bestehen aus jeweils drei Items, die von 1-5 skaliert sind; je höher der Wert, desto höher die Tendenz, sozial erwünscht zu antworten. Als kritischer Punkt wurde bei beiden Variablen der Wert 12 definiert (ebd.), was bedeutet, dass auf allen Variablen ein hoher bis sehr hoher Zustimmungswert angegeben wurde. Darin werden jeweils die folgenden drei Items zu IM bzw. SDE als aussagekräftig und valide angesehen (ebd., S. 19):

- SE1: Mein erster Eindruck von Menschen stellt sich gewöhnlich als richtig heraus (SDE +)
- SE2: Ich bin mir oft unsicher in meinem Urteil (SDE –)
- SE3: Ich weiß immer genau, wieso ich etwas mag (SDE +)
- SE4: Ich habe schon mal zu viel Wechselgeld zurückbekommen und nichts gesagt (IM –)
- SE5: Ich bin immer ehrlich zu anderen (IM +)
- SE6: Ich habe gelegentlich mal jemanden ausgenutzt (IM –)

Bei beiden Testverfahren wurde eine 5-stufige Likertskalierung verwendet, die von „trifft nie zu (1)“ bis „trifft immer zu (5)“ beschriftet war.

¹ <http://bildungswissenschaften.uni-saarland.de/personal/paulus/homepage/empathie.html>

Die Auswertung der Items zur sozialen Erwünschtheit geschah wie folgt: Zunächst wurden die Antworten der jeweils drei zusammengehörigen und z.T. umkodierten Items addiert, danach wurde gem. der Beschreibung bei Winkler et al. (2006) ein kritischer Wert bestimmt, der in unserem Fall mit mindestens 12 Punkten erreicht war: „Nach Paulhus kann in einer Interpretationsvariante nur extremes Antwortverhalten als soziale Erwünschtheit identifiziert werden, also die Werte 4 und 5 auf der fünfstufigen Likert-Skala (Paulhus 1984, 1986). Zu diesem Zweck werden zwei neue, dichotome, Variablen kodiert („SDE_Lügner“ und „IM_Lügner“). Die Ausprägung 1 steht in beiden Fällen für soziale Erwünschtheit. Sie wird erreicht, wenn ein Fall bei der Addition der drei Items, aus dem sich die jeweilige Dimension zusammensetzen, Ausprägungen im Wertebereich von 12 bis 15 aufweist. Der Wert 12 wurde als unterste Grenze genommen, da bei ihm sichergestellt werden kann, dass der Befragte in allen Items eine extreme Antwort gegeben haben muss. Diese Vorgehensweise unterscheidet sich etwas vom eigentlichen „dichotomous scoring“, da nicht jedes einzelne Items dichotomisiert wurde, sondern die additive Skala (im oben zitierten Text wurden die Grenzwerte an die hier verwendete 5er-Skalierung angepasst).

Ergebnisse

Anfälligkeit der Testitems gegen IM-Verfälschung

Tabelle 1: Darstellung der Mittelwerte innerhalb der IM-Gruppe mit Standardabweichung (in Klammern), Signifikanzniveau und Effektstärke

Item Nr	01-e	02-f	03-d	04-p	05-e	06-d	07-f	08-d
IM normal	3.55 (.90)	3.85 (.96)	3.06 (1.02)	3.82 (.98)	3.83 (.88)	2.91 (1.06)	2.86 (1.19)	2.78 (1.04)
IM hoch	3.92 (.78)	4.17 (.86)	2.99 (1.10)	4.21 (.76)	4.13 (.79)	2.70 (1.03)	3.00 (1.18)	2.58 (1.07)
p	***	***		***	***	*		***
Cohen's d	-.41	-.33		-.40	-.34	.20		.19
Item Nr	09-e	10-p	11-e	12-f	13-d	14-p	15-f	16-p
IM normal	3.52 (.94)	3.89 (.90)	3.43 (1.06)	3.76 (.99)	2.44 (1.01)	3.25 (1.02)	3.50 (1.05)	3.35 (.99)
IM hoch	3.79 (.95)	4.20 (.77)	3.72 (.96)	3.96 (.94)	2.28 (.96)	3.48 (.92)	3.75 (.98)	3.77 (.91)
p	***	***	***	***	*	***	***	***
Cohen's d	-.29	-.34	-.27	-.20	.16	-.23	-.24	-.42
***: p<.000 *: p<.05								

Nur zwei Items weisen keine Anfälligkeit gegen IM-Verfälschungen auf, die restlichen Effektstärken liegen im unteren bis mittleren Bereich. Die Mittelwerte der Skalen EC, PT und FS sind im Falle von IM-Verhalten leicht erhöht, bei PD entsprechend etwas niedriger (vgl. Tabelle 2). Bei einer durchschnittlichen Effektstärke von .35 ist die generelle Verfälschungstendenz im Sinne von IM-Verfälschung allerdings nicht sehr

gravierend, so dass der Fragebogen aus Sicht einer IM-Verfälschung als relativ robust angesehen werden kann.

Tabelle 2: Z-normierte Skalenmittelwerte der IM-Gruppe mit Standardabweichung und Effektstärke (alle $p < .000$)

		M	SD	t (df)	Cohen's d
fantasy scale	IM_normal	99,26	9,42		
	IM_hoch	101,82	9,09	-3,88 (919)	-0.28
empathic concern	IM_normal	99,16	9,31		
	IM_hoch	103,54	8,42	-6,84 (919)	-0.45
perspective taking	IM_normal	99,51	9,70		
	IM_hoch	104,01	8,02	-7,39 (676,11)	-0.45
personal distress	IM_normal	99,50	9,39		
	IM_hoch	97,33	8,92	3,31 (919)	0.21

Anfälligkeit der Testitems gegen SDE-Verfälschung

Tabelle 3: Darstellung der Mittelwerte innerhalb der SDE-Gruppe mit Standardabweichung (in Klammern), Signifikanzniveau und Effektstärke

Item Nr	01-e	02-f	03-d	04-p	05-e	06-d	07-f	08-d
SDE normal	3.63 (.88)	3.88 (.94)	3.18 (1.00)	3.90 (.94)	3.92 (.86)	3.04 (1.02)	2.90 (1.16)	2.86 (1.03)
SDE hoch	3.76 (.88)	4.13 (.92)	2.72 (1.06)	4.06 (.89)	3.95 (.87)	2.37 (.98)	2.91 (1.26)	2.37 (1.01)
p	*	***	***	**		***		***
Cohen's d	-.14	-.26	.46	-.18		.65		.48
Item Nr	09-e	10-p	11-e	12-f	13-d	14-p	15-f	16-p
SDE normal	3.57 (.91)	3.95 (.88)	3.56 (.99)	3.78 (.96)	2.53 (.99)	3.27 (.98)	3.55 (1.02)	3.42 (.98)
SDE hoch	3.69 (1.03)	4.09 (.82)	3.43 (1.12)	3.93 (1.04)	2.05 (.94)	3.46 (1.02)	3.67 (1.07)	3.64 (.97)
p		*		*		***		***
Cohen's d		-.16		-.15		-.20		-.22

***: $p < .000$ **: $p < .01$ *: $p < .05$

Die Tendenz zur Selbsttäuschung wirkt sich bei deutlich weniger Testitems aus als im Falle von *impression management*. Mit Ausnahme der PD-Skala sind hier die Mittelwertsunterschiede weniger bedeutsam. Es scheint aber so zu sein, dass man das Empfinden von *personal distress* nicht nur weniger nach außen hin

zugibt, sondern noch stärker sich selbst gegenüber verleugnet (vgl. Tabelle 4). Bei der Auswertung der Skalen sollte also beachtet werden, dass der PD-Wert möglicherweise eher unterschätzt wird.

Tabelle 4: Skalenmittelwerte der SDE-Gruppe mit Standardabweichung, Signifikanzniveau und Effektstärke

SDE-Gruppe		M	SD	p	Cohen's d
fantasy scale	SDE_normal	14.11	3.06		
	SDE_hoch	14.65	3.35	**	-0.17
empathic concern	SDE_normal	14.69	2.65		
	SDE_hoch	14.83	2.89	n.s.	
perspective taking	SDE_normal	14.54	2.91		
	SDE_hoch	15.25	2.88	***	-0.24
personal distress	SDE_normal	11.60	2.91		
	SDE_hoch	9.51	2.87	***	0.71

***: $p < .000$ **: $p < .01$

Täuschungsversuche auf beiden Ebenen gleichzeitig

Die Korrelation zwischen *impression management* und *self-deceptive enhancement* ist, ähnlich wie bei Paulhus (1986) oder Musch et al. (2002) nicht sehr hoch ($r = .16, p < .00, n = 953$), das bedeutet, dass es wohl eher selten der Fall ist, dass beide Täuschungstendenzen gleichzeitig eingesetzt werden. In unserer Stichprobe taten dies lediglich 13% der Pbn; 35,6% verwendeten eine der beiden Täuschungsstrategien, bei 51,3% konnte gar keine Täuschungsabsicht gefunden werden.

Altersunterschiede in der Verwendung der Täuschungsstrategien

Mit zunehmendem Alter der Pbn nimmt das Verwenden von Täuschungsstrategien zu. Während 63,6% der unter 15-jährigen gar keine Strategie verwenden, so tun dies nur noch knapp 40% der Pbn über 40 Jahren.

Tabelle 5: Anteil der Täuschungsstrategien je Altersgruppe in %

Altersgruppe	Anteil der Täuschungsstrategien je Altersgruppe in %		
	Keine	Eine Strategie	beide
unter 15 J.	63.6%	27.3%	9.1%
16-21 J.	56.9%	34.9%	8.1%
22-26 J.	56.2%	31.8%	12.0%
27-33 J.	51.6%	39.1%	9.4%
34-40 J.	52.7%	25.8%	21.5%
41-50 J.	39.6%	43.2%	17.1%
51-60 J.	38.2%	45.5%	16.4%
über 60 J.	40.0%	26.7%	33.3%

Verfälschungstendenzen bei jugendlichen Pbn

Innerhalb der Altersgruppe zwischen 15 und 18 Jahren zeigten sich keine signifikanten Unterschiede im Aspekt der Selbsttäuschung, wohingegen die Tendenz zur Fremdtäuschung bei den 16jährigen Pbn signifikant höher lag als bei den anderen Altersgruppen ($F = 5.82$, $df = 3$, $p < .01$, $part. \eta^2 = .16$).

Tabelle 6: Mittelwerte der Selbst- bzw. Fremdtäuschung innerhalb der 15-18jährigen Altersgruppe

	Alter	Mittelwert	Std.-Abweichung	N
Selbsttäuschung	15	9.33	2.69	9
	16	9.54	2.47	44
	17	9.19	2.44	26
	18	9.45	2.43	20
	Gesamt	9.41	2.44	99
Fremdtäuschung	15	8.44	4.71	9
	16	10.32	2.85	44
	17	7.23	3.53	26
	18	7.70	3.19	20
	Gesamt	8.80	3.53	99

Geschlechtsunterschiede - Lügen Männer anders als Frauen?

Es gibt keine signifikanten Unterschiede in der Häufigkeit, mit der männliche und weibliche Pbn eine der beiden Strategien benutzen (SDE: $\chi^2 = 2.32$, $df = 1$, $p = .13$; IM: $\chi^2 = .75$, $df = 1$, $p = .39$); tendenziell zeigt sich, dass weibliche Pbn etwas häufiger als Männer SDE-Strategien einsetzen. Nur 10% der Männer und 12,3% der Frauen täuschen in beiden Kategorien gleichzeitig. Das bedeutet, dass bei fast allen Pbn eine der beiden Tendenzen vorherrscht.

Jedoch gibt es signifikante Unterschiede innerhalb der SDE-Gruppe zwischen den Geschlechtern. Mit Ausnahme der Variablen *perspective taking* gibt es für die restlichen 3 Faktoren des Fragebogens einen signifikanten Haupteffekt des Geschlechts, allerdings mit sehr niedrigen Effektstärken (vgl. Tabelle 7). Interaktionseffekte konnten keine gefunden werden.

Tabelle 7: MANOVA Haupteffekt Geschlecht innerhalb der SDE-Gruppe

Quelle	Abhängige Variable	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	p	Partielles Eta-Quadrat
sex	FS_norm	2713.09	1	2713.09	31.79	.00	.03
	EC_norm	3296.79	1	3296.79	40.23	.00	.04
	PT_norm	45.00	1	45.00	.51	.47	
	PD_norm	3220.00	1	3220.00	43.30	.00	.04

Ein vergleichbares Ergebnis zeigte sich bei der Nutzung von *impression management* Strategien. Auch hier zeigen weibliche Pbn, die diese Strategie einsetzen, signifikant unterschiedliche Ergebnisse auf den Faktoren *fantasy*, *empathic concern* und *personal distress*, wobei auch hier der letztere Faktor wieder als geringer beschrieben wird als in den Gruppen, die keine IM-Strategie verwenden. Interaktionseffekte sind auch hier nicht vorhanden.

Tabelle 8: MANOVA Haupteffekt Geschlecht innerhalb der IM-Gruppe

QUELLE	ABHÄNGIGE VARIABLE	QUADRAT-SUMME VOM TYP III	DF	MITTEL DER QUADRAT-RATE	F	SIG.	PARTIELLES ETA-QUADRAT
SEX	FS_norm	2303.86	1	2303.86	27.19	.00	.03
	EC_norm	2680.96	1	2680.96	34.34	.00	.03
	PT_norm	6.68	1	6.68	.07	.78	
	PD_norm	3156.36	1	3156.36	38.82	.00	.04

Diskussion

In der Studie sollte untersucht werden, ob und ggf. wie stark die Messung von Empathie mittels Fragebogen anfällig gegen Verfälschungstendenzen sei. Dabei wurden sowohl (eher absichtliche) Tendenzen des impression managements (Fremdtäuschung) als auch (eher unabsichtliche) Tendenzen des self-deceptive enhancement (Selbsttäuschung) betrachtet. Es zeigte sich, dass die deutsche Version des IRI (Davis, 1980, 1983), der Saarbrücker Persönlichkeitsfragebogen zur Messung von Empathie (SPF; Paulus, 2001) im Bereich des IM recht robust ist, also unter den getesteten Bedingungen (anonym) nur wenig anfällig für bewusste Täuschung ist. Die Tendenz der Selbsttäuschung zeigte lediglich beim Faktor personal distress einen Effekt dahingehend, dass Vpn mit tendenz zur SDE niedrigere Angaben machten als Personen, die diese tendenz nicht besaßen (*Cohen's d* = 0,719). Es sollte also beachtet werden, dass der Faktor PD evtl. eher unterschätzt wird. Die Korrelation beider Verfälschungstendenzen lag mit $r = .16$ im gleichen Bereich wie in anderen Studien (Paulhus, 1986 oder Musch et al., 2002). Die allgemeine Tendenz, Antworten bewusst oder unbewusst zu verfälschen, nahm mit zunehmendem Alter der Vpn zu. So benutzen 63,6% der

Pbn im Alter von bis zu 15 Jahren keine Verfälschungsstrategien, wohingegen dies nur noch ca. 40% der über 40-jährigen taten. Es gab keine signifikanten Unterschiede in der Häufigkeit, mit der männliche und weibliche Pbn eine der beiden Strategien benutzen; tendenziell zeigte sich, dass weibliche Pbn etwas häufiger als Männer SDE-Strategien einsetzen.

Bei der Analyse der Ergebnisse muss natürlich berücksichtigt werden, dass die Reduktion der BIDR-Skala von 20 Items im Original auf je 3 in der hier verwendeten Kurzfassung möglicherweise Validitätsprobleme aufweisen könnte. Zudem wurde aus testökonomischen Gründen die Antwortskala von ursprünglich 7 Stufen auf 5 Stufen verkürzt, was nach Erfahrungen des Autors kein problematisches Vorgehen darstellt, solange die Stufenanzahl weiterhin ungerade bleibt.

Literatur

- Alliger, G. M. & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, 60, 59–72.
- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5. aktual. und erw. Aufl.). Stuttgart: Kohlhammer.
- Baron-Cohen, S. & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high-functioning autism and normal sex differences. *Journal of Autism and Developmental Disorder*, 34, 163-175.
- Beaven, J.P., O'Brien-Malone, A., & Hall, G. (2004). Using the Interpersonal Reactivity Index to assess empathy in violent offenders. *International Journal of Forensic Psychology*, 1, 33-41.
- Damarin, E. & Messick, S. (1965). *Response styles as personality variables: A theoretical integration of multivariate research* (Research Bulletin No. 65-10). Princeton, NJ: Educational Testing Service.
- Davis, M. (1983): Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.
- Davis, M. (1994). *Empathy. A Social Psychological Approach*. Madison, WI: Brown & Benchmark's.
- Heilmann, K. (1999). Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP). In E. Fay (Hrsg.): *Tests unter der Lupe II* (S. 19–38). Lengerich: Pabst.
- Kusyszyn, I., & Jackson, D. N. (1968). A multimethod factor analytic appraisal of endorsement and judgment methods in personality assessment. *Educational and Psychological Measurement*, 28, 1047-1061.
- Lauterbach, O. & Hosser, D. (2007). Assessing empathy in prisoners – A shortened version of the Interpersonal Reactivity Index. *Swiss Journal of Psychology*, 66, 91-101.
- Marcus, B. (2003). Persönlichkeitstests in der Personalauswahl: Sind „sozial erwünschte“ Antworten wirklich nicht wünschenswert? *Zeitschrift für Psychologie*, 211, 138-148.
- Musch, J., Brockhaus, R. & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, 48, 121-129.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.

- Paulhus, D. L. (1986). Measurement and control of response bias. In J. P. Robinson, P. Shaver & L. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (Vol. 1, pp. 17–60). New York: Springer.
- Paulus, C. & Laueremann, P.(2004). *Schulische Übergangentscheidung mittels Lernfähigkeitsdiagnose*. Arbeitsbericht Nr. 90 aus der FR Erziehungswissenschaft der Universität Saarbrücken.
- Paulus, C. (1997). Empathie, Kompetenz und Altruismus. <http://www.uni-saarland.de/fak5/ezw/personal/paulus/empathy/empathie.htm>
- Paulus, C. (1999). *Das multidimensionale Lernprofil. Zur Diagnostik von Lernfähigkeit*. Frankfurt: Peter Lang.
- Paulus, C. (2001). Die Messung von Lernfähigkeit. Experimentelle Validierung des multidimensionalen Lernprofils. Arbeitsbericht Nr. 89 der FR Erziehungswissenschaft der Universität Saarbrücken.
- Paulus, C. (2006). Der Saarbrücker Persönlichkeitsfragebogen (SPF-IRI) zur Messung von Empathie. <http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/>
- Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. <http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/>
- Paulus, C. (2012). Ist die Bildung eines Empathiescores in der deutschen Fassung des IRI sinnvoll? <http://scidok.sulb.uni-saarland.de/volltexte/2012/4889/>
- Paulus, C. (2014). <http://psydok.sulb.uni-saarland.de/volltexte/2014/5007/> Personal Distress - das Sorgenkind der Empathiemessung. <http://psdok.sulb.uni-saarland.de/volltexte/2014/5007/>
- Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2, pp. 139-197). New York: Plenum Press.
- Stotland, S., Mathews, K., Sherman, S., Hansson, R. & Richardson, B. (1978). *Empathy, fantasy, and helping*. Beverly Hills, Ca.: Sage.

Anschrift des Autors:

PD Dr. Christoph Paulus

Universität des Saarlandes

Fak. Empirische Humanwissenschaft und Wirtschaftswissenschaften

Geb. A4.2 R. 410

66123 Saarbrücken