

The Double Cross:
Individual differences between respondents
with different response sets and styles on questionnaires



Beatrice Gerber-Braun

München 2010

The Double Cross:
Individual differences between respondents
with different response sets and styles on questionnaires

Dissertation
an der Fakultät für Psychologie und Pädagogik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Beatrice Gerber-Braun

München, den 18.03.2010

Gutachter

Erstgutachter: Prof. Dr. Markus Bühner

Zweitgutachter: Prof. Dr. Matthias Ziegler

Tag der mündlichen Prüfung: 30.04.2010

© Beatrice Gerber-Braun, 2010

Alle Rechte liegen bei der Autorin, außer für Elemente aus anderen Veröffentlichungen

“When a person takes an objective test,
he may bring to the test a number of test-taking habits which affect his score.”

(Cronbach, 1950, p.3)

Acknowledgements

I would like to use this page to express my gratitude to a number of people, without whom my dissertation would not have been finished.

First of all I would like to thank Prof. Dr. Markus Bühner and Prof. Dr. Matthias Ziegler for always supporting me and believing in me – even in my darkest moments.

Furthermore, I would like to thank Wolfgang Moderegger and Dr. Tobias Wagner who have been both obliging and patient with my working times and energy.

Also my students deserve my gratitude for keeping me “on the ball” when I was in danger of getting lost.

I would also like to thank Heide and Bob Hughes, who were kind enough to proof-read my work.

Of course I want to thank my family, especially my husband, for always supporting me, even when this meant high charges and restrictions. Last but not least, I would like to thank Valentina for motivating and reminding me, that everything has to come to an end – naturally.

Beatrice Gerber-Braun

Zusammenfassung

Fragebogenverfahren sind ein häufig eingesetztes Instrumentarium, sowohl in der (organisationspsychologischen) Forschung als auch in der praktischen Anwendung im Unternehmen – beispielsweise im Zuge des Personalauswahlprozesses.

Welche Konstrukte werden im Rahmen der Personalauswahl erfasst? Neben Intelligenz werden in Deutschland mit Hilfe von nicht-kognitiven Fragebogenverfahren vor allem Persönlichkeitseigenschaften in Personalauswahlsituationen erhoben. Warum Persönlichkeitseigenschaften? Mit Hilfe von umfangreichen Meta-Analysen hat sich gezeigt, dass wenn man Persönlichkeit mit Hilfe der Big 5 erfasst, sich signifikante Zusammenhänge beispielsweise zwischen Berufserfolg und der Persönlichkeitseigenschaft Gewissenhaftigkeit nachweisen lassen und dies unabhängig von der Branche, dem Rang der zu besetzenden Position oder dem Land, in dem die Vakanz besteht. Ein kurzer Abriss sowohl der Entstehungsgeschichte der Big 5 als auch über gefundene Zusammenhänge mit berufsrelevanten Kriterien ist in der Einleitung dieser Arbeit gegeben.

Was müssen Personen, die beispielsweise einen Persönlichkeitsfragebogen mit geschlossenem Antwortformat / Likertskalen ausfüllen, zur Beantwortung tun? Sie müssen „einfach“ nur durch das Ankreuzen von Antwortalternativen angeben, in wie weit sie einzelnen Aussagen auf einer Skala, die meistens von *stimme nicht zu* bis *stimme völlig zu* gelabelt ist, zustimmen. Doch so leicht wie es scheint, ist es leider nicht, da der dahinterstehende kognitive Prozess sehr komplex ist: zuerst muss die Frage interpretiert werden, dann das entsprechende Verhalten etc. aus dem Gedächtnis abgerufen werden, ein Urteil

muss gebildet werden und schließlich muss das Urteil mit dem vorliegenden Antwortformat abgeglichen werden und eventuell aufgrund von sozialer Erwünschtheit oder anderen Faktoren angepasst werden. Leider können in jeder dieser Stufen Antwortverzerrungen auftreten – ein kurzer Abriss sowohl des kognitiven Prozesses bei der Beantwortung von Fragebogenverfahren als auch möglicher Antwortverzerrungen, ist ebenfalls in der Einleitung gegeben. Zusammenfassend lässt sich feststellen, dass es ein „Kreuz mit dem (Antwort-) Kreuz ist“: Antwortstile, wie beispielsweise die vom Inhalt unabhängige übermäßige Bevorzugung extremer Antwortkategorien auf Likert-Skalen (extreme response style, ERS) sowie die Anpassung der Antworten aufgrund sozialer Erwünschtheit (socially desirable responding / faking) können Antworten und somit Ergebnisse von nicht-kognitiven Fragebogenverfahren beeinflussen. Trotz jahrzehntelanger Forschung auf dem Gebiet der Antwortverzerrungen bei Fragebogenverfahren konnten bisher nicht alle Fragen (konfliktfrei) beantwortet werden. Wo bestehen derzeit noch offene Fragen und konfliktäre Befunde?

Zwischenzeitlich konnten Fragen wie, was ist ein Antwortstil, was sind die Effekte / Folgen und zugrunde liegende Faktoren von Antwortstilen (fast) beantwortet werden. Eine Zusammenfassung ist in Studie 1 gegeben. Jedoch gibt es in Bezug auf Antwortstile auch noch uneindeutige Befunde und offene Fragen, wie beispielsweise die Frage, was eine Person charakterisiert, die extreme Antwortkategorien unabhängig vom Frageninhalt bevorzugt. In welchen individuellen Variablen (Persönlichkeitseigenschaften, kognitive Fähigkeiten, Alter, Geschlecht) unterscheiden sich Personen mit unterschiedlichem Antwortstil? Dieser Frage wird in Studie 1 nachgegangen –

mit dem Schwerpunkt auf ERS (extreme response style), also der vom Inhalt unabhängigen und übermäßigen Bevorzugung extremer Antwortkategorien. Ähnliche ungeklärte Fragen existieren auch in Bezug auf sozial erwünschtes Antwortverhalten (faking): obwohl Fragen wie was ist eine sozial erwünschte Antworttendenz, was sind die Effekte / Folgen, wann treten sozial erwünschte Antworten auf und antworten alle Personen mit selben Ausmaß an sozialer Erwünschtheit (fast) beantwortet sind (Zusammenfassung siehe Studie 2), stellt sich auch hier die Frage, in welchen Persönlichkeitseigenschaften (und anderen individuellen Variablen wie Alter, Geschlecht und kognitive Fähigkeiten) sich Personen mit unterschiedlichem Ausprägungsgrad in sozial erwünschtem Antwortverhalten unterscheiden. Dieser Frage wird in Studie 2 nachgegangen.

Obwohl Antwortstile und sozial erwünschtes Antwortverhalten Ergebnisse von Fragebogenverfahren beeinflussen können, werden nicht-kognitive Fragebogenverfahren beispielsweise im Rahmen der Personalauswahl eingesetzt. Um zu überprüfen, ob die psychometrische Qualität von nicht-kognitiven Fragebogenverfahren trotz des Auftretens von unterschiedlichen Antwortstilen und sozial erwünschtem Antwortverhaltens erhalten bleibt, wird in Studie 3 die psychometrische Qualität des ersten deutschsprachigen Integrity Tests untersucht – und zwar in einer neutralen Situation, in der das Auftreten von Antwortstilen sehr wahrscheinlich ist, als auch in einer simulierten Einstellungssituation, in der mit hoher Wahrscheinlichkeit sozial erwünschte Antworten abgegeben werden. Ein kurzer Abriss, was Integrity Tests sind und warum sie im Rahmen von Personalauswahlverfahren insbesondere in den USA eingesetzt werden, ist in Studie 3 ebenfalls gegeben. Was sind somit die Hauptziele dieser Arbeit?

Insgesamt soll in dieser Arbeit vor allem durch die Untersuchung von Unterschieden in Persönlichkeitseigenschaften – sowohl auf Faktoren- als auch auf Facetten-Ebene – der Frage nachgegangen werden, warum Personen in nicht-kognitiven Fragebogenverfahren dort kreuzen, wo sie kreuzen (Studie 1, 2) und ob die psychometrische Qualität von Fragebogenverfahren trotz zu erwartender Antwortstile und sozial erwünschter Antworten gegeben ist (Studie 3). Welche Befunde konnten sich in den einzelnen Studien nachweisen lassen?

Studie 1. In dieser Studie wurden individuelle Unterschiede zwischen Personen mit verschiedenen Antwort-Stilen untersucht. Hierzu wurden 312 zumeist weibliche Studenten gebeten einen Breitband-Persönlichkeitstest sowie einen Intelligenztest auszufüllen. Zusätzlich wurde das Alter sowie das Geschlecht erfasst.

Mit Hilfe von Rasch / Mixed-Rasch-Modellen war es möglich zwei latente Gruppen zu identifizieren, die sich in ihrem Antwortstil unterscheiden. In wie weit sich diese Gruppen signifikant in Persönlichkeitsfaktoren, Persönlichkeitsfacetten, ihren kognitiven Fähigkeiten oder dem Alter unterscheiden, wurde mit Hilfe von t-Test für unabhängige Stichproben und Effektstärkemaßen errechnet.

Um zu vermeiden, dass ein und derselbe von Antwortstilen „kontaminierte“ Persönlichkeitsscore herangezogen wird, um sowohl die latenten Klassen der Mittel- und Extremkreuzer als auch die Unterschiede in den Persönlichkeitseigenschaften zwischen den Klassen zu bestimmen, werden Personenparameter aus einer 2 Klassenlösung des Mixed-Rasch Modells herangezogen. Auf diese Art und Weise konnte der Einfluss des Antwortstiles

kontrolliert werden. Somit ist der große Vorteil dieser Studie, sowohl die Vorteile der Identifikation von Mittel- und Extremkreuzern mit Hilfe von Rasch / Mixed-Rasch Modellen zu nutzen und als auch eine Verzerrung der Persönlichkeitsunterschiede zwischen den Klassen der Mittel- und Extremkreuzer durch die Antwortstile an sich zu vermeiden. Welche Ergebnisse konnten in dieser ersten Studie gefunden werden?

Die Ergebnisse dieser Studie replizieren das Auftreten von verschiedenen Antwortstilen (Mittel- und Extremkreuzer) sowie den Befund, dass die Gruppe der Extremkreuzer – in diesem Fall mit 32% – stets die kleinere Gruppe darstellt (Austin, Deary, & Egan, 2006; Eid & Rauber, 2000). Jedoch muss angemerkt werden, dass Mittel- und Extremkreuzer nicht in allen Facetten gefunden werden konnten. In den Facetten, in denen sie jedoch nachzuweisen waren, traten sie konstant über die Facetten hinweg auf. Darüber hinaus konnte gezeigt werden, dass sich Extremkreuzer von Mittelkreuzern vor allem in ihrem Ausprägungsgrad von Extraversion unterscheiden: auf *Faktorebene* konnte hier der stärkste Effekt gefunden werden (moderate Effektstärke). Extremkreuzer tendieren somit dazu, herzlicher (E1) zu sein sowie einen größeren Frohsinn (E6) zu verbreiten. Darüber hinaus sind sie eher Aktiv (E4) und besitzen einen größeren Erlebnishunger (E5) sowie eine größere Durchsetzungsstärke (E3). Somit bestätigen diese Ergebnisse die auch von Austin (2006) sowie Meiser und Machunsky (2008) gefundenen Zusammenhänge von ERS und Extraversion. Im Gegensatz zu diesen Untersuchungen konnte jedoch in dieser Studie ebenfalls ein Zusammenhang zwischen ERS und Offenheit sowie Verträglichkeit sowohl auf Faktoren als auch auf Facettenebene gefunden werden: Extremkreuzer sind auch hier

wiederum aktiver, nämlich in Bezug auf ein aktiveres Phantasieerleben (Offenheit für Phantasie, O1), auf die Deutlichkeit, mit der sie Gefühle erleben (Offenheit für Gefühle, O3), in Bezug auf das Bedürfnis neue Handlungsweisen zu erproben (Offenheit für Handlungen, O4), sich mit theoretischen Fragen auseinander zu setzen (Offenheit für Ideen, O5) und aktiver in Bezug auf die in Fragestellung von sozialen, politischen, ethischen etc. Normen (Offenheit des Normen- und Wertesystems, O6).

Außerdem engagieren sie sich aktiver für das Wohlergehen anderer (Altruismus, A3) und bringen ihnen ein größeres Maß an Vertrauen (A1), Sympathie und Anteilnahme (Gutherzigkeit, A6) entgegen.

Die stärksten Effekte auf *Facettenebene* zeigen sich zwischen Mittel- und Extremkreuzern aber im Bereich der Gewissenhaftigkeit: Extremkreuzer haben eine höhere Überzeugung bezüglich ihrer Kompetenz (C1) und einen höheren Anspruch an ihr Leistungsstreben (C4). Während hier moderate Effektstärken gefunden wurden, wurden schwache bis moderate Effektstärken für die signifikanten Unterschiede in Pflichtbewusstsein (C3) und Selbstdisziplin (C5) nachgewiesen.

Extremkreuzer schätzen sich darüber hinaus als weniger ängstlich ein (N1) und neigen weniger dazu, sich die Schuld für etwas zu geben, sich entmutigt, traurig und einsam zu fühlen (Depression, N3).

Zusammenfassend kann man feststellen, dass sich Mittelkreuzer im Vergleich zu Extremkreuzern dadurch charakterisieren lassen, dass sie selbstbezogener, weniger aktiv, weniger zielstrebig und gewissenhaft sind, sondern eher ängstlich sowie depressiv. Extremkreuzer sind dagegen in vielerlei Hinsicht aktiver, zielstrebiger und durchsetzungsfähiger.

Geschlechtsunterschiede oder Unterschiede in den generellen kognitiven Fähigkeiten konnten nicht gefunden werden, jedoch wiesen Extremkreuzer niedrigere Werte in verbaler Intelligenz auf und waren eher jünger. Die nicht-gefundenen Zusammenhänge zwischen ERS und Geschlecht sowie genereller Intelligenz lassen jedoch nicht eindeutig den Schluss zu, dass diese Zusammenhänge nicht existieren. Die sehr homogene Stichprobe, die zumeist aus weiblichen Studenten bestand, die aufgrund des universitären Auswahlverfahrens nach Intelligenz bereits vorselektiert waren, könnte auch der Grund für die nicht gefundenen Zusammenhänge sein. Darüber hinaus führte die geringe Itemanzahl auf Facettenebene (8 Items) u.a. dazu, dass Schätzprobleme bei den WINMIRA-Analysen auftraten. Zusammen mit dem Problem, dass einige Items inhaltlich unterschiedlich von den Studienteilnehmern interpretiert wurden, führte dies dazu, dass lediglich 12 der 30 Facetten in die Latente Klassen Analyse einfließen konnten, um festzustellen, in wie weit der Antwort-Stil konstant über die Facetten ist.

Abschließend lässt sich bezüglich der Ergebnisse von Studie 1 festhalten, dass die Ergebnisse dieser Studie einen Beitrag dazu leisten, warum Personen auf nicht-kognitiven Fragebogenverfahren mit Likert-Skalen dort kreuzen, wo sie kreuzen, wenn kein situationaler Druck gegeben ist: insbesondere aktive Personen mit hoher Ausprägung in Extraversion, Leistungsstreben und hoher Kompetenzüberzeugung neigen dazu, unabhängig vom Iteminhalt überproportional häufig extreme Antwortkategorien anzukreuzen. Ob diese Unterschiede in den Persönlichkeitseigenschaften die Ursache oder ein Symptom des jeweiligen Antwortstiles sind, müssen weitere Studien zeigen. Bei der gängigen Praxis über das Aufaddieren von Itemwerten zu

Summenscores und somit zu Aussagen zu gelangen, kann es bei Fragebögen mit durchwegs positiv kodierten Items dazu kommen, dass Personen mit höheren Werten, beispielsweise in Extraversion, höhere Ausprägungen in dem untersuchten Konstrukt (Leistungsmotivation, Depression etc.) zugeschrieben werden. Sofern also bei Befragungen keine positiv und negativ ausbalancierten Itemantworten verwendet werden, wäre eine kombinierte Auswertungsmethode, mit vorheriger Klassifizierung des Antwortstils und anschließenden Gruppenvergleichen eine geeignete Methode, wenn Schlussfolgerungen aus Gruppenunterschieden gezogen werden sollen. Antwortstile, wie die hier untersuchte, vom Inhalt unabhängige, überproportionale Bevorzugung extremer Antwortkategorien, sind eher unbewusste Antwortverzerrungen. In welchen Persönlichkeitseigenschaften (und anderen individuellen Variablen) unterscheiden sich aber Personen, die bewusst ihre Antwort verzerren, beispielsweise aufgrund von sozialer Erwünschtheit? Dieser Frage wurde in Studie 2 nachgegangen.

Studie 2. In dieser Studie wurden individuelle Unterschiede zwischen Personen mit unterschiedlichem Ausprägungsgrad in sozial erwünschtem Antwortverhalten (Faking-Stile) untersucht. Um die Unterschiede zwischen Personen mit verschiedenen Faking-Stilen unabhängig von dem jeweiligen Antwort-Stil zu untersuchen, wurde der Einfluss des Antwortstiles bei dieser Untersuchung kontrolliert. Hierzu wurden insgesamt 312 Personen zufällig auf zwei Gruppen verteilt. Während die Kontrollgruppe das NEO-PI-R zweimal mit einer Instruktion gemäß Testhandbuch beantwortete, erhielt die Experimentalgruppe beim zweiten Durchgang eine Faking-Instruktion. Zusätzlich wurde die fluide Intelligenz mit Hilfe des IST-2000-R erfasst sowie

das Geschlecht und das Alter. Durch dieses Design war es nicht nur möglich, die „wahren“ Persönlichkeitseigenschaften der teilnehmenden Personen zu erfassen, sondern auch festzustellen, in welchen Persönlichkeitseigenschaften sich Personen mit verschiedenen Faking-Stilen unterscheiden – unter Kontrolle des Antwortstiles.

Die Ergebnisse dieser Untersuchung zeigen, dass nicht die Antworten aller Skalen des NEO-PI-R verfälscht wurden – sozial erwünschtes Antwortverhalten (Faking) ist demnach nicht skalenunabhängig. Darüber hinaus konnte repliziert werden, dass unterschiedliche Faking-Stile existieren und dass nicht jeder, der aufgrund der Situation sozial erwünscht antworten sollte, dies auch tut bzw. nicht jeder ehrlich antwortet, von dem dies erwartet wird. Insgesamt konnten verschiedene Faking-Stile identifiziert werden, wenn auch nicht in allen Persönlichkeitsfacetten. 83% der Personen, die sozial erwünscht antworten, behalten jedoch ihren Faking-Stil konstant über unterschiedliche Persönlichkeitsfacetten hinweg bei. Nachgewiesene Faking-Stile sind hierbei: slight faking (geringe Anpassung der Antwort an soziale Erwünschtheit), extreme faking (starke Anpassung an soziale Erwünschtheit) sowie das „switchen“ zwischen ehrlichen und sozial erwünschten Antworten.

Was kennzeichnet nun Angehörige verschiedener Faking-Stile? Bei Betrachtung der um die Antwort-Stile kontrollierten „wahren“ Persönlichkeitseigenschaften der Studienteilnehmer zeigte sich, dass sich die Personen, die zwischen ehrlichen und sozial erwünschten Antworten „switchen“ (Switcher) vor allem dadurch auszeichnen, dass sie in allen Facetten von Gewissenhaftigkeit signifikant niedrigere Werte aufweisen als slight und extreme Faker – bei durchschnittlich moderaten Effektstärken.

Switcher und extreme Faker unterscheiden sich darüber hinaus signifikant in ihrem Ausprägungsgrad der Facetten Ängstlichkeit (N1), Depression (N3), Verletzlichkeit (N6), Aktivität (E4), Offenheit für Handlungen (O4) und Altruismus (A3). *Switcher* sind daher weniger aktiv, was ihr Aktivitätsniveau (E4) an sich betrifft, als auch ihre Abenteuerlust (O4) oder ihre Bereitschaft aktiv anderen zu helfen (A3) – sie sind eher ängstlich (N1), depressiv und verletzlich (N3).

Von den *slight Fakern* unterscheiden sich die *Switcher* vor allem durch ihre signifikant höheren Werte in Verträglichkeit (A) und niedrigeren Werte in Gewissenhaftigkeit (C).

Wie unterscheiden sich *slight und extreme Faker*? Sie unterscheiden sich hauptsächlich in ihrem Ausprägungsgrad von Freimütigkeit (A2), Altruismus (A3), Entgegenkommen (A4) sowie Offenheit für Handlungen (O4): *extreme Faker* sind aktiver in Bezug auf ihre Hilfsbereitschaft für andere (A3), ihre Bereitschaft bei Konflikten nachzugeben (A4) und bezüglich ihrer Abenteuerlust (O4). Die nicht signifikanten, aber fast moderaten Effektstärken bei den Persönlichkeitsfacetten Kompetenz (C1) und Leistungsstreben (C4) weisen darauf hin, dass *extreme Faker* gewissenhafter sind.

Zusammenfassend lässt sich feststellen, dass Angehörige der *Switcher*-Klasse eher weniger gewissenhaft, weniger aktiv und ängstlicher sind. Je aktiver jemand ist – und dies kann sowohl die Aktivität an sich, die Offenheit für Neues als auch die aktive Hilfeleistung für andere sein – desto höher ist die Wahrscheinlichkeit, dass diese Person zu stärkeren sozial erwünschten Antworten tendiert (*extreme faking*).

Individuelle Unterschiede bezüglich Intelligenz (Reasoning), Alter oder Geschlecht konnten nicht nachgewiesen werden, was jedoch an der sehr homogenen Stichprobe (zumeist weibliche Studenten eines nach Intelligenz selektierten Studienganges) liegen könnte. Erwartungsgemäß haben jedoch extreme Faker leicht höhere Werte in Reasoning und waren eher jünger. Slight Faker waren eher weiblich.

Einschränkend muss jedoch, wie bereits angedeutet, festgehalten werden, dass es sich um eine studentische Stichprobe handelte, die mehrere Nachteile mit sich brachte: große Homogenität bezüglich Intelligenz, Alter und Geschlecht. Darüber hinaus führte die geringe Itemanzahl auf Facettenebene (8 Items) u.a. dazu, dass Schätzprobleme bei den WINMIRA-Analysen auftraten. Zusammen mit dem Problem, dass einige Items inhaltlich unterschiedlich von den Studienteilnehmern interpretiert wurden, führte dies dazu, dass lediglich 13 der 30 Facetten in die Latente Klassen Analyse einfließen konnten, um festzustellen, in wie weit der Faking-Stil konstant über die Facetten ist.

Zusammenfassend lässt sich bezüglich Studie 2 festhalten, dass die Ergebnisse dieser Studie einen Beitrag dazu leisten, warum Personen auf nicht-kognitiven Fragebogenverfahren mit Likert-Skalen dort kreuzen, wo sie kreuzen, wenn situationaler Druck gegeben ist: insbesondere gewissenhafte, aktivere Personen neigen dazu, konstant extreme Antwortkategorien anzukreuzen. Ob diese Unterschiede in den Persönlichkeitseigenschaften die Ursache oder ein Symptom des jeweiligen Faking-Stiles sind, müssen weitere Studien zeigen.

Studie 3. In dieser letzten Studie dieser Arbeit wurde untersucht, in wie weit ein Test zur Personalauswahl über psychometrische Qualität verfügt –

trotz der Antwortstile und trotz des Fakings, die zu erwarten sind. Aus diesem Grund wurden 134 Auszubildende der chemischen Industrie gebeten, ein Persönlichkeitstest sowie den ersten deutschsprachigen Integrity Test unter zwei verschiedenen Versuchsbedingungen auszufüllen: zuerst in einer neutralen Situation und später in einer simulierten Einstellungssituation, bei der sich die Auszubildenden in ihren eigenen Einstellungstest zurückversetzen sollten. Sowohl für die neutrale Situation als auch für die simulierte Einstellungssituation werden die Reliabilitäten, die Faktorielle-, die Konstrukt- sowie die Kriteriumsvalidität untersucht. Darüber hinaus wurde überprüft, in wie weit das IBES inkrementell über einen Intelligenz- und einen Persönlichkeitstests hinaus einen Beitrag zur Varianzaufklärung von Leistungskriterien liefern kann.

Die Ergebnisse dieser Studie zeigen, dass das IBES faktorielle Validität besitzt, auch wenn die Ergebnisse der Konfirmatorischen Faktorenanalyse belegen, dass einige Subskalen schlecht abschneiden – wie auch schon im Handbuch von Marcus (2006). Die höhere Korrelation der IBES-Skalen untereinander in der simulierten Bewerbungssituation, die auch in dieser Studie repliziert werden konnte, wird von dem Autor des Tests als Beleg für die innere Struktur des Tests angesehen.

Das IBES hat sich darüber hinaus in beiden Situationen als weitgehend reliables Instrument erwiesen. Lediglich 2 Subskalen weisen – auch wiederum in Übereinstimmung mit dem Test-Handbuch – Werte auf, die als zu niedrig erachtet werden müssen.

Zum Nachweis der Konstruktvalidität wurden sowohl konvergente als auch divergente Validitäten in beiden experimentellen Bedingungen bestimmt.

Geringe Korrelationen des IBES mit konstruktfernen Variablen, wie beispielsweise Intelligenz und deutlich höhere Korrelationen mit konstruktnahen Variablen, wie beispielsweise Gewissenhaftigkeit, sprechen für eine Unabhängigkeit des Konstruktes mit Nähe zu den Big 5 und belegen die Konstruktvalidität in beiden Versuchsbedingungen.

Korrelationen des IBES mit Leistungskriterien, wie Vorgesetztenbeurteilungen, belegen darüber hinaus in beiden experimentellen Settings die Kriteriumsvalidität des Verfahrens. Korrelationen des IBES mit Berufsschulnoten fielen erwartungs- und konstruktgemäß niedriger aus, da kontraproduktive Verhaltensweise wie Diebstahl zwar sehr wohl in die Vorgesetztenbeurteilungen mit einfließen, nicht jedoch in Schulnoten.

Abschließend wurde überprüft, in wie das IBES über einen Intelligenz- und einen Persönlichkeitstest hinaus einen Beitrag zur Varianzaufklärung des Kriteriums liefern kann. In keinen der beiden Versuchsbedingungen ist dem IBES dies gelungen. Somit bestätigt sich der Vorschlag von Marcus (2006), dass IBES als erstes Instrument in einem mehrstufigem Auswahlprozess einzusetzen.

Was ist mit dem Einfluss von sozial erwünschten Antworten / Faking? In der simulierten Bewerbungssituation waren die erzielten IBES-Werte zumeist höher und die Varianzen niedriger als in der neutralen Situation. Darüber hinaus hat sich gezeigt, dass die Korrelationen des IBES mit konstruktnahen Persönlichkeitsfaktoren und mit den IBES-Subskalen an sich steigen. Jedoch konnte auch für die simulierte Bewerbungssituation – trotz sozial erwünschtem Antwortverhaltens – sowohl die Reliabilität als auch die Validität des Verfahrens nachgewiesen werden.

In wie weit konnten durch diese Ergebnisse die Ziele dieser Arbeit erreicht werden? Die Ziele dieser Arbeit bestanden darin, individuelle Unterschiede zwischen Personen mit verschiedenen Antwortstilen sowie verschiedenem Ausprägungsgrad in sozial erwünschtem Antwortverhalten zu explorieren und zu überprüfen, ob trotz dieser Antwortverzerrungen die psychometrische Qualität eines Test, der zum Einsatz in der organisationspsychologischen Praxis entwickelt wurde, bestehen bleibt. Studie 1 war die erste, die nachweisen konnte, dass Personen, die unabhängig vom Frageninhalt überproportional häufig extreme Antwortkategorien präferieren, sich in vielerlei Persönlichkeitseigenschaften von den Personen unterscheiden, die mittlere Antwortkategorien bevorzugen. Durch die Verwendung von Personenparametern, die aus einem 2 Klassen Mixed Rasch Modell gewonnen wurden, konnte bei der Exploration der Persönlichkeitsunterschiede auch der Einfluss des Antwortstiles an sich kontrolliert werden. Studie 2 konnte zeigen, dass auch Personen mit unterschiedlichem Ausprägungsgrad in sozial erwünschtem Antwortverhalten sich in verschiedenen Persönlichkeitseigenschaften signifikant unterscheiden – auch wenn man den Antwortstil kontrolliert. Trotz wahrscheinlich stattgefundener Antwortverzerrungen konnte jedoch die psychometrische Qualität eines organisationspsychologischen Testes bestätigt werden. Ob jedoch die gefundenen Unterschiede in den Persönlichkeitseigenschaften *der* Grund dafür sind, warum Personen kreuzen, wo sie kreuzen oder ob es sich bei den individuellen Unterschieden lediglich um ein weiteres Symptom der jeweiligen Art der Antwortverzerrung handelt, müssen weitere Untersuchungen zeigen. Da jedoch Fragebogenverfahren häufig eingesetzte Instrumente sowohl in der

Zusammenfassung

Forschung als auch in der organisationspsychologischen und klinischen Praxis sind, lohnt es sich, weiterhin der Frage nach zu gehen, warum Personen dort kreuzen, wo sie kreuzen. Diese Arbeit war hierzu nicht der erste, aber hoffentlich ein bemerkenswerter Schritt.

1	<i>Introduction</i>	4
1.1	The usage of questionnaires in the personnel selection process	4
1.2	Measuring Personality	6
1.2.1	What is personality?	6
1.2.2	The history of the BIG 5	7
1.2.3	Why is personality assessed?	8
1.3	Respondent's process of answering a questionnaire question	11
1.3.1	Phases of answering a closed-ended question	11
1.3.2	Response distortions	13
1.4	Goals of the present project	17
1.4.1	Extreme response style and individual differences	17
1.4.2	Socially desirable responding and individual differences	18
1.4.3	Psychometric quality of a questionnaire used in personnel selection	18
1.4.4	Summary and outlook	19
2	<i>The extreme response style (ERS) and individual differences</i>	22
2.1	Present Study	22
2.1.1	What is a response style?	22
2.1.2	What are the effects of the extreme response style?	23
2.1.3	Response styles and underlying factors	23
2.1.4	ERS and individual differences	25
2.1.5	Ways detecting ERS	27
2.1.6	Goals of the present study	28
2.2	Method	29
2.2.1	Procedure	29
2.2.2	Participants	30
2.2.3	Measures	30
2.3	Results	32
2.3.1	Searching for Subgroups	34
2.3.2	Response styles and individual differences	40
2.4	Discussion	43
2.4.1	The occurrence of the extreme response style (ERS)	43
2.4.2	ERS and individual differences in personality factors / facets	45
2.4.3	ERS and individual differences in fluid intelligence	47
2.4.4	ERS and individual differences in age and gender	48
2.4.5	Limitations	48
2.4.6	Implications and future directions	49
3	<i>Socially desirable responding (SDR) and individual differences</i>	52
3.1	Present Study	52
3.1.1	What is socially desirable responding?	52
3.1.2	What are the effects of SDR / Faking?	53
3.1.3	When do people answer in a socially desirable way?	55
3.1.4	What do people fake?	55
3.1.5	Do all fake to the same extend?	56
3.1.6	Socially desirable responding and individual differences	57
3.1.7	Goals of the present study	58
3.2	Method	59
3.2.1	Procedure	59
3.2.2	Participants	59
3.2.3	Measures	60

3.3	Results	63
3.3.1	The occurrence of socially desirable responding	64
3.3.2	Searching for Subgroups	67
3.3.3	Socially desirable responding and individual differences	74
3.4	Discussion	78
3.4.1	The occurrence of socially desirable responding (SDR)	78
3.4.2	The occurrence of different faking styles	78
3.4.3	SDR and individual differences in personality factors / facets	79
3.4.4	SDR and individual differences in fluid intelligence	80
3.4.5	SDR and individual differences in age and gender	81
3.4.6	Limitations	81
3.4.7	Implications and future directions	82
4	<i>Psychometric quality of a questionnaire used in personnel assessment</i>	85
4.1	Present Study	85
4.1.1	What is an Integrity Test?	85
4.1.2	IBES – field of application	86
4.1.3	History of Integrity Tests	86
4.1.4	Integrity and counterproductive working criteria	88
4.1.5	Integrity and job-performance criteria	88
4.1.6	Goals of the present study	89
4.2	Method	91
4.2.1	Procedure	91
4.2.2	Participants	91
4.2.3	Measures	92
4.3	Results	97
4.3.1	Factor Validity	98
4.3.2	Reliability	100
4.3.3	Construct Validity	101
4.3.4	Criterion Validity	106
4.3.5	Incremental Validity	107
4.4	Discussion	109
4.4.1	Factor Validity	110
4.4.2	Reliability	110
4.4.3	Construct Validity	111
4.4.4	Criterion Validity	111
4.4.5	Incremental Validity	112
4.4.6	Situational Influence	112
4.4.7	Limitations	113
4.4.8	Implications and future directions	113
5	<i>Summaries / Abstracts and Conclusions</i>	116
5.1	Summaries / Abstracts of the studies presented	116
5.1.1	Study 1	116
5.1.2	Study 2	116
5.1.3	Study 3	117
5.2	Conclusions regarding the goals of the present project	117
5.2.1	Extreme response style and individual differences	117
5.2.2	Socially desirable responding and individual differences	119
5.2.3	Psychometric quality of a questionnaire used in personnel selection	121
5.2.4	Summary and outlook	127
6	<i>References</i>	130

1. Introduction

1 Introduction

1.1 The usage of questionnaires in the personnel selection process

In 1991 approximately 21 million employees were working in the service industry in Germany. In 2008 the amount of people working in the service industry rose to 26 million (Federal Statistical Office, 2009), so a lot of new employees had to be selected – and this figure does not even include the number of people only changing their jobs. However, how to find the ideal employee? In the history of personnel selection, which began in the late 19th and early 20th centuries (Scroggins, Thomas, & Morris, 2008), a lot of recruiting strategies were developed to help employers to choose the ideal employee for the vacant position. What are recruiting strategies? According to Rynes (1991), recruiting includes “all organizational practices and decisions that affect either the number, or types, of individuals who are willing to apply for, or to accept, a given vacancy” (p. 429). These methods include e.g. the analyses of CVs, reference checks, work sample tests, employment interviews, assessment centres, graphology, GMA tests, and questionnaires. What method / methods are the ones to use? The methods used in the personnel selection practice depend on many company specific variables like the capacity / capability the company has to carry out the selection process, on the amount of vacant jobs, on the image the company wants to transfer (Rynes, 1993), of the knowledge the HR manager has about effective selection tools (Hirsh, 2009), and – of course – on company non-specific variables like legal restrictions, or the validity of selection procedures. Why is the validity of selection procedures so important? As Van Iddekinge and Ployhart summarize (2008; p. 871/872):

The use of validated employee selection and promotion procedures is crucial to organizational effectiveness. For example, valid selection procedures can lead to higher levels of individual, group, and organizational performance (Barrick, Stewart, Neubert, & Mount, 1998; Huselid, 1995; Schmidt & Hunter, 1998; Wright & Boswell, 2002). Valid procedures are also essential for making legally defensible selection decisions. Indeed, selection procedures that have been properly validated should be more likely to withstand the legal scrutiny associated with employment discrimination suits (Sharf & Jones, 2000) and may even reduce the likelihood of litigation in the first place.

Due to the fact that self-report questionnaires / tests are easy to administer (Peterson, Griffith, & Converse, 2009), they are a cost-effective way to test applicants, even group wise (Casillas, Robbins, McKinniss, Postlethwaite, & Oh, 2009; Peterson, et al., 2009), they are not rated negatively by applicants (Marcus, 2003; Sackett & Wanek, 1996), and due to the high validity some sort of questionnaires / tests have proven (Ones, Viswesvaran, & Schmidt, 1993; Schmidt & Hunter, 1998), such questionnaires are an appropriate and often used method in the selection process (Roberts, Harms, Caspi, & Moffitt, 2007).

However, as summarized in the studies following later on, different effects question the validity of non-cognitive self-report questionnaires and not all problems and questions concern why people cross where they cross, are answered. Shedding further light on this topic is the aim of this work. Therefore, the construct, often measured with the help of self-report questionnaires, is portrait first and then the cognitive processes a respondent has to undergo when answering a question on a non-cognitive self-report questionnaire is summarized including, the different kinds of response distortion which might occur. The question in which individual variables, respondents using different ways of response distortions, are different and whether the psychometric quality of a test used as a personnel selection tool is still given despite response distortion will be examined in subsequent studies.

But first let us have a look on the construct, often measured with the help of self-report questionnaires, in the personnel selection process as well as in research: personality. In applicant settings, personality questionnaires try to identify those applicants, who will have a higher probability in showing required characteristics like conscientiousness and have a lower probability in showing less favourable attitudes like counterproductive working behaviour (Ones, et al., 1993). These kinds of questionnaires are called personality tests. Why are personality traits like conscientiousness the one to measure?

1.2 Measuring Personality

1.2.1 What is personality?

Personality can be seen as “...the unique, dynamic organization of characteristics of a particular person, physical and psychological, which influence behaviour and responses to the social and physical environment. Of these characteristics, some will be entirely unique to the specific person (i.e. memories, habits, mannerisms) and others will be shared with a few, many, or all other people” (Liebert & Liebert, 1998; p. 5-6). A more operational and measurable description of personality is to describe the personality of people as well as interpersonal differences with the help of the Five Factor Model of personality (FFM; BIG 5). These dimensions describing personality are neuroticism, extraversion, and openness to experience, agreeableness, and conscientiousness. According to Costa and McCrae (1992), neuroticism expresses the amount of anxiety, angry hostility, depression, self-consciousness, impulsiveness, and vulnerability a person has. Extraversion shows the extent to which a person is introverted / extroverted. It summarizes the degree of warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotion. Trust, straightforwardness, altruism, compliance, modesty, and tender-mindedness are the

variables, which indicate the extent of agreeableness a person has. With the help of the personality dimension openness to experience it is possible to specify the magnitude of a person's active imagination, aesthetic sensitivity, attentiveness to inner feelings, preference for variety, intellectual curiosity, and independence of judgement. Conscientiousness describes the quantity of competence, order, dutifulness, achievement striving, self-discipline, and deliberation a person comes up with.

1.2.2 The history of the BIG 5

How did research in personality evolve? According to Klages (1926) all prominent characteristics / individual differences of a person will become encoded into language. The greater the difference, the more likely is the difference to become expressed as a single word. To get a taxonomy of personality, Allport and Odberg (1936) analyzed all personality-relevant terms in the English language by extracting a list of words that distinguish between the behaviour of people out of a Dictionary (Webster's New International Dictionary, 1925), which contained about 550,000 terms. This lexical approach resulted in a list of almost 18,000 words relating to personality descriptions. Cattell (1943a, 1943b, 1945) shortened this list to a more manageable size of 35 bipolar variables and with the help of factorial studies he was able to identify 12 personality dimensions. Cattell's work stimulated other researchers like Fiske (1949) as well as Tupes and Christal (1961) to examine the structure of trait ratings. The comprehensive research of Tupes and Christal (1961; p. 14) resulted in "five relatively strong and recurrent factors and nothing more of any consequence". This five factor structure could be replicated (Borgatta, 1964; Norman, 1963; Smith, 1967). Later on, variables like the influence of situation on behaviour stopped the first wave of research on personality dimensions (Digman, 1990). With Goldberg's (1981) work on lexical analysis in a second wave of research, the robustness of the five factors could be proven and the

declaration of the five personality factors as BIG 5 was born: “it should be possible to argue the case that any model for structuring individual differences will have to encompass – at some level – something like these ‘big five’ dimensions” (p. 159). Costa and McCrae (1985) developed in several steps an inventory to assess the personality dimensions by developing self-report questionnaires with whole sentences instead of lists of adjectives: They started with the personality dimensions extraversion and neuroticism (BIG 2), integrated openness (BIG 3; NEO), and finally also agreeableness and conscientiousness. Consequently, the Neuroticism Extraversion Openness Personality Inventory (NEO-PI; today used in the revised form: NEO-PI-R) was developed. As Marcus, Höft and Riediger (2006; p. 121) mention, “the NEO-PI-R is currently the most widely used and the most researched marker of the FFM, and it has been demonstrated to outperform alternative instruments in comparative analyses”. Therefore, this instrument is used in this work to shed light on individual differences in personality factors and facets between respondents with different response sets and styles.

1.2.3 Why is personality assessed?

In the meantime it was not only possible to verify that the Five Factor Model of personality holds true independent of the inventory, language or culture (Digman & Shmelyov, 1996; Kallasmaa, Allik, Realo, & McCrae, 2000), but also that personality dimensions are stable over time (Block, 1971; Costa & McCrae, 1988). Why is it important for personality dimensions to be stable over time? As already Berg and Collier (1953; p. 166) noted, “if response sets are not stable over a period of time, it is idle to consider them as even remotely useful measures of personality characteristics”. Why should personality be an important variable to measure in all languages / cultures? As several meta-analytic studies demonstrated, personality – measured with the

construct of the FFM – is able to predict job related outcome variables like job performance or counterproductive workplace behaviours as summarized in the following paragraphs (Barrick & Mount, 1991; Berry, Ones, & Sackett, 2007; Salgado, 2002; Tett, Jackson, & Rothstein, 1991).

1.2.3.1 Personality and job-related performance criteria

As Ones, Dilchert, Viswesvaran and Judge (2007; p. 1001) noted, “self-report personality scale scores assessing the Big Five are useful for a broad spectrum of criteria and variables in organizational settings”. What does this mean? Meta-analyses in the personality domain demonstrated the validity of personality in the prediction of job performance (Barrick & Mount, 1991; Tett, et al., 1991).

As Schmidt and Hunter (1998) were able to show, personality-related variables account for the most variance in job performance after cognitive ability, which is the best single predictor. Which personality trait accounts for what? Of the Big Five personality dimensions, conscientiousness has the highest validities across organizations and occupational groups (Barrick & Mount, 1991; Salgado, 1997). As Barrick, Mount and Judge (2001) in their summary of 15 meta-analyses state: “The results for conscientiousness underscore its importance as a fundamental individual difference variable that has numerous implications for work outcomes. Conscientiousness appears to be the trait-oriented motivation variable that industrial-organizational psychologists have long searched for, and it should occupy a central role in theories seeking to explain job performance” (p. 21). So conscientiousness is not only able to be a predictor for job success, training success and team work (Barrick, et al., 2001), but also for higher task performance, contextual performance, and motivation (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Judge & Ilies, 2002) – and applicable across countries (Ones & Viswesvaran, 2001). Other personality dimensions are not as relevant as

conscientiousness (Barrick, et al., 2001). For example, extraversion is a valid predictor for teamwork and for training success, openness only for training success and agreeableness only for teamwork or in context of carrying / helping jobs. Neuroticism is a valid predictor for general performance and teamwork. Judge and Ilies (2002) were able to prove that neuroticism and conscientiousness are also valid predictors for performance motivation and are therefore also able to explain motivational aspects within occupational settings. Because these various studies were able to show that personality measures within the construct of the Big 5 are useful predictors of job performance across occupations (Barrick, et al., 2001; Salgado, 2002), personality itself is one criterion often measured in the personnel selection process. However, the Big 5 were not only found to be valid predictors for performance criteria, but also for counterproductive working behaviours.

1.2.3.2 Personality and counterproductive working behaviours

What are counterproductive working behaviours (CWB)? They are “volitional acts by members of an organization that violate the legitimate interests of the organization or its individual members” (Marcus & Wagner, 2007, p. 161), like absenteeism, alcoholism, drug abuses or theft – mainly measured with the help of Integrity Tests. Accordingly, which personality factors of the BIG five were found to be able to predict which kinds of counterproductive working behaviours? Different measures of personality and different systems of describing personality traits were able to show that low agreeableness and low conscientiousness are key correlates of diverse counterproductive working behaviours (Berry, Ones, et al., 2007; Colbert, Mount, Harter, Witt, & Barrick, 2004; Ones & Viswesvaran, 2003; Salgado, 2002). Inverted, high levels of agreeableness and conscientiousness were found to “weaken the within-person relations of daily negative emotions with daily CWB directed at the organization

and individuals” (Yang, 2009, p. 259). Moreover, neuroticism was found to predict counterproductive working behaviours like substance abuse at work (Ones & Viswesvaran, 2001) and according to Salgado (2002) all five personality traits are able to predict fluctuation in organizations. Therefore, the Big 5 personality measures are not only valid predictors for performance criteria, but also for counterproductive working behaviours.

1.3 Respondent’s process of answering a questionnaire question

What do respondents have to do when answering a questionnaire like the NEO-PI-R? Only one task: to mark their answer of an attitude question on a Likert-type scale with a cross. Sounds easy – but it is not. Respondents have to undergo complex cognitive processes before cross setting and unfortunately misreporting takes place – explicitly and implicitly. What are these cognitive processes and what misreporting can take place?

1.3.1 Phases of answering a closed-ended question

Since the early 1980s, research into cognitive aspects of survey methods (CASM) “has made considerable progress in illuminating the cognitive and communicative processes underlying survey responding” (Schwarz, 2007; p. 277). Answering a survey question “involves several cognitive steps as described in the well known four-step model of survey response” (Ongena & Dijkstra, 2007; p. 145).

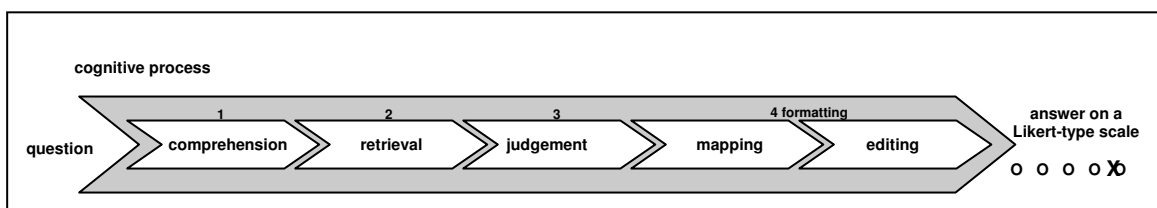


Figure 1
The extended four-step model of survey response on close-ended attitude questions

What are these cognitive steps in answering a non-cognitive self-report survey with close-ended questions? Firstly, the question has to be interpreted to deduce its intent (comprehension phase). Secondly, relevant information has to be searched for (retrieval phase). Thirdly, the information has to be integrated into a judgement (judgement phase). Finally – in the formatting phase – this judgement has to be translated into one of the response options (mapping phase) and eventually adjusted (editing phase) for social desirability or consistency (Krosnick, 1999; Schwarz, 2007; Tourangeau & Rasinski, 1988).

What happens in detail in these cognitive steps while a respondent is answering a survey question? In the comprehension phase, the respondent tries to understand the meaning of the question and tries to determine which information he or she should provide. Respondents try to find out the literal as well as the pragmatic meaning of a question (Sudman, Bradburn, & Schwarz, 1996).

In the retrieval phase, the respondent is searching the memory for relevant information. Here the question type has to be distinguished: In attitude questions, the respondent recalls a previously formed attitude from memory, or – more often – he has to form a new judgement, based on the information he has access to at this moment (Smith & Conrey, 2007). In behavioural frequency questions, the respondent has to identify the relevant behaviour and review the number the relevant behaviour occurred in the specified reference period, like “last week” (recall-and-count). But often only estimations based upon general impressions are given (Tourangeau, Rips, & Rasinski, 2000). A review of the specified period takes only place when the behaviour is rare and important (Menon, 1994).

After having retrieved and judged the relevant information, respondents have to map their answers to the response format available to them in the survey. This step is called “mapping an answer”. However, before crossing, respondents eventually adapt their

answer to criteria such as consistency, social desirability, intrusiveness or politeness (Ongena & Dijkstra, 2007), which is called “editing”. Taken together, mapping and editing are the phases in which respondents format (formatting phase) their answers (Tourangeau, et al., 2000) and in which intended response distortion takes place. What is response distortion and what effects might occur?

1.3.2 Response distortions

1.3.2.1 Response style vs. response set

Editing an answer means that people’s responses are also influenced by content-irrelevant factors. These non-content-based forms of responding are referred to as response styles, response sets, or response bias. Whereas some authors use the terms interchangeably, because “the distinction is not widely accepted and the terms are used in different senses” (Baumgartner & Steenkamp, 2001; p. 143), the following distinction can be made: According to Paulhus (2002; p. 49) response biases are “any systematic tendency to answer questionnaire items on some basis that interferes with accurate self-reports”, distinguishing between “response styles – biases that are consistent across time and questionnaires – from response sets – short-lived response biases attributable to some temporary distraction or motivation.” Examples for response styles are acquiescence (yea-saying), or extreme responding (tendency to disproportionately favour extreme categories of a Likert-type scale). Socially desirable responding is a response set. Due to the fact that response styles are consistent across time and questionnaires response styles can also be seen as “a manifestation of a deep-seated personality syndrome” (Couch & Keniston, 1960; p. 151) or a manifestation of basic personality traits (Berg & Collier, 1953).

1.3.2.2 *Optimizing vs. Satisficing*

To answer a single question honestly by setting a cross to one of the response alternatives given by a survey requires a lot of cognitive work, because all the phases in answering a question have to be executed when giving an optimal answer is the goal. Desires for self-expression, intellectual-challenge or feelings of altruism are motives, which may encourage respondents to spend considerable cognitive effort (Warwick & Lininger, 1975). Giving an optimal answer is called optimizing. Unfortunately, not all people are willing to always give an optimal answer (Krosnick, 1999). Even when starting with high cognitive effort, respondents may change their response strategies for example due to tiredness and conduct all phases, but with less cognitive effort (Krosnick, 1991). This response behaviour is called weak satisficing (Simon, 1957). To reduce the cognitive effort even more, people can also interpret each question superficially and select the first reasonable or a random answer, thus skipping the retrieval and judgement step. This answering strategy without referring to any internal psychological cues relevant to the attitude, belief or event of interest is called strong satisficing (Krosnick, 1999).

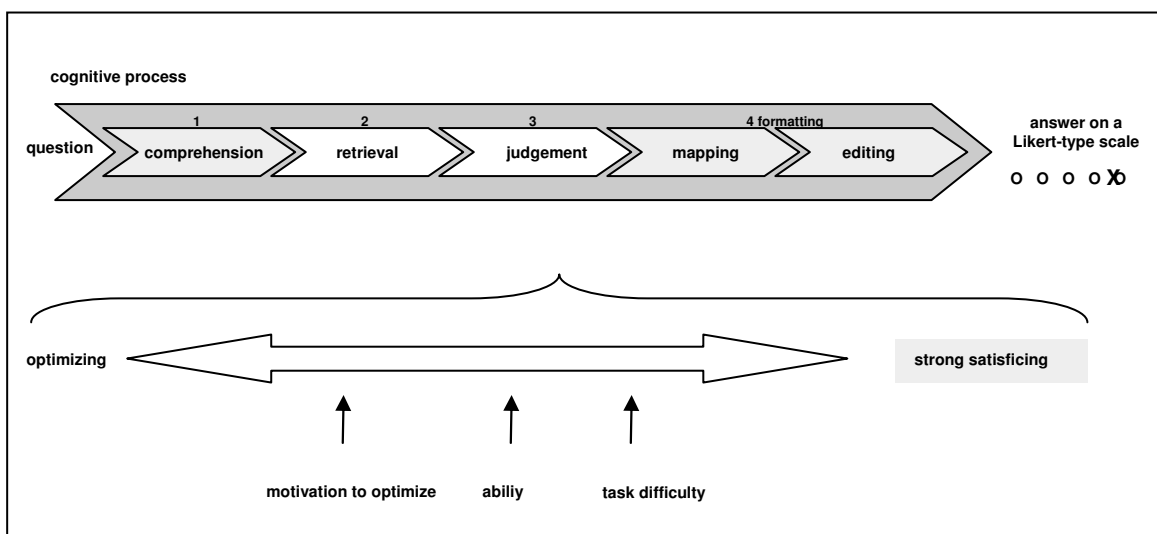


Figure 2

The extended four-step model of survey response influenced by the continuum between optimizing and strong satisficing (only grey collared steps take place)

Which strategy do people use when answering a question? Optimizing or satisficing? According to Krosnick (1999) these answering strategies can be seen as a continuum with an optimizing and a strong satisficing end of scale and intermediate levels of satisficing in between. The higher the task difficulty, the lower the ability of the respondent and the lower the motivation to optimize is, the higher is the risk of satisficing (Krosnick, 1991).

However, the answer a person gives depends not only on the level of optimizing or satisficing. Unfortunately, in all of these steps mentioned before, response biases can happen, too.

1.3.2.3 Response bias in the process of answering a survey question

In each of the four cognitive steps of answering a survey question (comprehension, retrieval, judgement and formatting), response bias can take place.

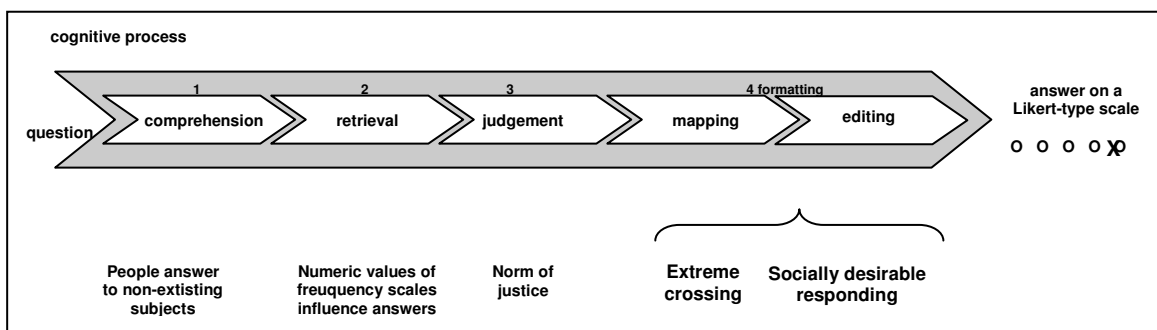


Figure 3

The cognitive process of answering a question and examples for response distortions occurring the individual phases

Some examples are given: In the first step, in the comprehension phase, respondents take the rules of conversation into account when answering a question (Grice, 1975; Ongena & Dijkstra, 2007). They try to answer with a relation to the topic asked (maxim of relation), avoid to mention things twice (maxim of quantity), try to avoid untrue statements or generate answers of topics that do not exist (maxim of quality) and take always the obvious meaning for given (maxim of manner). Respondents try to

comprehend the questions asked, even when the topic does not exist. Therefore, a high rate of answers concerning a non-existing topic can be found (Strack, Schwarz, & Wanke, 1991). In cases where the literal meaning of a question is quite easy to understand (“What have you done today?”), but the context is not, respondents have need for clarification. When this clarification is not available, either because of interviewer restriction (“It means whatever it means to you”) or because of the absence of an interviewer (self-report), respondents take contextual information into account for their answer and the answers of the respondents differ only because of the different context, like the title or the sponsor of a questionnaire (Galesic & Tourangeau, 2007).

What can go wrong in the retrieval phase? Possible response distortions are priming (the memories are guided in a certain direction) and wrong frequency estimations. Respondents do, for example, make systematic use of features of the questionnaire, like the numeric values of frequency scales, to arrive at a plausible estimate (Schwarz, Hippler, Deutsch, & Strack, 1985).

What about the other phases? In the judgement phase, according to Schuman and Presser (1981), people take norms of justice into account and in the formatting phase, people tend to adjust their answer to criteria such as consistency or social desirability. Over-reporting of admirable attitudes and behaviour and under-reporting those that are not socially respected is one well-known phenomenon in research (Krosnick, 1999), also referred to as socially desirable responding (SDR) or faking. However, even when no intentional response distortion takes place, an unintentional might very well: Whereas some people tend to disproportionately favour extreme categories of Likert-type scales (extreme crossers), others prefer middle categories for an answer on a non-cognitive self-report questionnaire (middle crossers). This response style is therefore called extreme response style (ERS). Summing up, different response biases might take place, which might lead to a contamination of questionnaire results. Therefore this topic

is a long researched topic, as the study of Berg and Collier concerning ERS shows, which was already published in 1953. Nevertheless, there are still unanswered questions. This work tries to shed light on unanswered questions and on conflicting former results as far as the extreme response style and socially desirable responding is concerned in an attempt to try to help to find an answer as to why people cross where they cross on a self-report questionnaire.

1.4 Goals of the present project

Research on cognitive aspects of survey methods (CASM) made considerable progress in illuminating the cognitive and communicative processes underlying survey responding (Belli, Conrad, & Wright, 2007). The steps of interpretation, retrieval, judgement and formatting are as well documented as influencing factors like task difficulty, ability and motivation (Krosnick, 1999; Schwarz, 2007). However, as far as response distortion is concerned, not all questions are yet answered.

1.4.1 Extreme response style and individual differences

Questions such as “What are response styles?, What are the effects and causing factors of response styles?” are already investigated and a short summary on response styles is given in this work, with the focus on the extreme response style (ERS). However, there are still conflicting results and open questions: Although already Berg and Collier (1953; p. 164) “hypothesized that tendencies to choose the extremes of an affective continuum when responding to a series of ambiguous test items are stable and that these tendencies reflect certain personality and group differences”, up to now it is not clear in what personality traits and personality facets extreme crossers differ from middle crossers. Results concerning personality factors are scanty and conflicting, results concerning differences on the level of personality facets are not searched for. Therefore, study 1 tries to shed light on these topics, taking into account that also

personality scores itself are contaminated by the response style and are therefore no good measure to determine differences between respondents with different response styles.

1.4.2 Socially desirable responding and individual differences

The usage of self-report questionnaires in personnel selection processes has often been criticized due to the fact that people are not only able to answer in a socially desirable way but that they also do so (Alliger & Dwight, 2000; Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). This effect, also called faking, is a long researched topic. Questions like what is socially desirable responding (SDR / faking), what are the effects of SDR, when do people answer in a socially desirable way, what do they fake and do they all fake to the same extend were investigated and a short summary is given in this work. However, the question whether and how people with different faking styles have individual differences in personality traits, intelligence, age, and gender is still not answered properly. Therefore, study 2 tries to examine these topics, controlling the first time for different response styles (middle / extreme crossing).

1.4.3 Psychometric quality of a questionnaire used in personnel selection

Although not all questions concerning response biases are answered, yet, non-cognitive self-report questionnaires are widely used and new questionnaires are developed. To investigate whether response sets like socially desirable responding and response styles like the extreme response style distort the psychometric quality of a questionnaire, the psychometric quality of the first German Integrity Test is examined in study 3. Therefore, the factorial validity, the reliability, the construct validity and the criterion validity of the job-related attitudes and self-evaluations inventory (IBES; Marcus, 2006) will be examined – in a neutral situation, where response styles are likely

and in a simulated applicant setting, where socially desirable responding probably takes place.

1.4.4 Summary and outlook

Non-cognitive self-report questionnaires are an often used tool, for personnel selection as well as for research purpose. Unfortunately, people do not always cross where they cross on Likert-type scales of non-cognitive self-report questionnaires due to their “real” trait. Intentional (socially desirable responding) and unintentional (extreme response style) response distortion takes place. The question in which individual variables like personality factors and facets, fluid intelligence and its facets, age or gender respondents with different response styles (ERS) and response sets (SDR) differ is not completely answered, yet: Studies concerning individual differences in personality factors, fluid intelligence, age and gender are scanty and partly conflicting (especially as far as ERS is concerned). Individual differences in personality facets are not researched for, yet. Questions like “do people with a higher level of activity (a facet of Extraversion) prefer disproportionately extreme categories of Likert-type scales?” or “do people using different faking styles differ significantly in their level of activity?” cannot be answered until now. Why is this important? It is important to shed further light on the question why people cross where they cross on a non-cognitive self-report questionnaire. If for example, a respondent chooses the extreme category on a non-cognitive self-report questionnaire his answer might be caused by his “real” level of the intended-to-measure trait. However, if for example the association between activity (E4) and ERS holds true, the extreme answer might also be caused by his level of activity. So whatever associations are searched for with the help of questionnaires, which do not have counterbalanced items, ERS might distort responses, leading to higher scores and therefore better results in the intended-to-measure trait for extreme

responders. Thus, the correlation between customer satisfaction and employee motivations might be influenced by the level of activity respondents have – and not only by the customer satisfaction and employee motivation itself. The same holds true for the personnel selection process: Due to the situational pressure socially desirable responding might occur, influencing applicants results. Also here, individual differences between respondents engaging extreme in socially desirable responding and those who only slightly engage in SDR would be interesting. Perhaps applicants with higher levels of activity also engage in more extreme socially desirable responding. Accordingly, results of Integrity Tests – tests used in the personnel selection process to identify applicants with higher probabilities for counterproductive working behaviour – might also be influenced by respondent's level of activity and not only by their "true" trait of integrity.

In the following chapters three different studies are described, attempting to answer the questions just stated. The first study (chapter 2) investigates the individual differences between middle and extreme crossers. To identify individual differences between slight and extreme fakers irrespective of their response style, study 2 (chapter 3) is conducted. In study 3 (chapter 4), it was tested whether response styles and response sets distort the psychometric quality of a test used in the personnel selection process. After the three studies have been described and discussed, chapter 5 will provide a summary of the results in form of abstracts for each study and conclusions regarding the goals of the present project to complete this work.

**2. The extreme response style (ERS)
and individual differences**

2 The extreme response style (ERS) and individual differences

2.1 Present Study

2.1.1 What is a response style?

“When a person takes an objective test, he may bring to the test a number of test-taking habits which affect his score” (Cronbach, 1950; p. 3). Such test-taking habits or response styles can be defined as tendencies to respond systematically to items on another basis than what the items were specifically designed to measure (Cronbach, 1946, 1950; Paulhus, 1991). Examples for such response styles are the tendency to acquiescence / yea-saying (acquiescence response style, ARS) or the tendency to use the middle category of a rating scale, also known as midpoint responding (MPR; see Baumgartner and Steenkamp (2001) for an review of response styles). However, also the opposite was found: Between 25% and 30% of all respondents prefer the end of Likert-type scales and are so called extreme crossers (Austin, et al., 2006; Eid & Rauber, 2000). This “extreme response style (ERS) refers to the tendency to disproportionately favour the endpoints or extreme categories of ordinal response or Likert-type scales, irrespective of particular item content” (Naemi, Beal, & Payne, 2009; p. 261). In contrast to socially desirable responding, which comes along with situational pressure for respondents, ERS remains relatively consistent over time (Berg & Collier, 1953; Greenleaf, 1992; Hamilton, 1968). Therefore, response styles can be seen as type of nuisance dimension that interferes with the measurement of topic (Roussos & Stout, 1996) or as systematic measurement error, which is a source of concern, because it threatens the validity of empirical findings by contaminating respondents answers (Baumgartner & Steenkamp, 2001).

2.1.2 What are the effects of the extreme response style?

Response styles have several effects: If for example the endorsement of items leads in a survey to higher rankings of a trait, the person's tendency to yea-saying or using extreme categories will result in a higher total score of the person and thus in a higher level of the intended-to-be-measured trait (Bolt & Johnson, 2009). So response styles like the acquiescent response style (ARS) and the extreme response style (ERS - if items are coded in only one direction) can lead to a bias in the total score of a respondent, leading to a bias on respondent level.

Due to the fact that response styles tend to vary across different respondent groups another bias also might occur: When individual items function differently in their measurement for different groups, effects can be a result of a group specific response style, not being based on any other differences and so leading to misinterpretations, because mean differences cannot be validly compared (Bolt & Johnson, 2009; Eid & Rauber, 2000).

2.1.3 Response styles and underlying factors

Why do people use scales differently? Why do some people prefer extreme categories, some middle categories and others give greatly differentiated judgements, especially on larger Likert-type scales? How can this difference in response styles be explained?

First of all, people may differ in their judgement complexity: Whereas some might have differentiated attitudes and perceptions using the whole scale for describing an answer, others may think in rather global categories such as good / bad which is also referred to as simplistic thinking having less differentiated cognitive structures and poorly developed schemas (Baumgartner & Steenkamp, 2001; Eid & Rauber, 2000; Naemi, et al., 2009). "People with a more simple attitude structure evaluating objects

with broad categories (good or bad) might be overwhelmed by a 9-point response scale. Therefore, these individuals might prefer the two extreme categories of a scale and avoid the other ones, whereas people with more complex attitude structures might use the whole scale” (Eid & Rauber, 2000; p. 21).

Secondly, people might tend to do satisficing, because a differentiated judgement is too time consuming or the sense of the study cannot be seen (Eid & Rauber, 2000; Krosnick, 1999) and they do not want to “waste” their time differing between categories. So they use only the extreme ones. For example, employees working for longer than 10 years in the same position in a company were found to engage more in an extreme response style when a satisfaction questionnaire is presented (Eid & Rauber, 2000).

Thirdly, the difference in response styles reflects also the difference in rigidity, intolerance of ambiguity and dogmatism (Baumgartner & Steenkamp, 2001). Someone who is very certain about his attitudes in general and has only a little tolerance of ambiguity might avoid the middle category if the category indicates indifference, while people who are consistently uncertain and have a high tolerance of ambiguity might prefer the middle category (Eid & Rauber, 2000).

Fourthly, concerning items, higher ERS levels were found for stimuli / questions that are important or involving respondents (Baumgartner & Steenkamp, 2001). So involvement might be “a mediating explanatory variable” (Warr & Coffman, 1970; p. 108) for the relationship between individual differences and extreme responses, showing that personality and extreme responding are linked when involvement is sufficiently high. Involvement is defined as “a compound formed by the relevance of construct dimensions and the perceived importance of the stimulus to be judged” (Warr & Coffman, 1970; p. 117). So, more extreme ratings were found when construct relevance is given (Schroder, Driver, & Streufert, 1967; Warr & Coffman, 1970) and /

or construct dimensions are used, which are chosen by the subjects' and not supplied by another person (Cromwell & Caldwell, 1962; Landfield, 1968; Mitsos, 1961).

Finally, concerning individuals, response tendencies might be due to stable personality traits: Higher ERS levels were found for people with higher levels of anxiety (Baumgartner & Steenkamp, 2001; Berg & Collier, 1953; Lewis & Taylor, 1955), and higher scores in extraversion and conscientiousness (Austin, et al., 2006). Therefore, research has tried to identify dispositional antecedents of the extreme response style by exploring the relationship between ERS and individual difference variables (Austin, et al., 2006; Hamilton, 1968; Naemi, et al., 2009). Which relationships were found?

2.1.4 ERS and individual differences

As mentioned above, personality constructs that have been investigated are tolerance of ambiguity, simplistic thinking, and decisiveness. According to Naemi et al. (2009) those who quickly complete surveys and are intolerant of ambiguity or simplistic thinkers are most likely to exhibit ERS: "In short, simply rushing through a questionnaire is not sufficient to lead to ERS; one must also be highly intolerant of ambiguity, decisive, or inclined toward simplistic thinking" (Naemi, et al., 2009; p. 279).

What are further variables influencing the response style? Concerning age, children and adolescents tend to give more extreme responses than adults aged 20 - 59 years. Elderly adults aged 60 - 83 years, respond in a manner like that of children aged 9 - 10 (Austin, et al., 2006; Hesterly, 1963; Light, Zax, & Gardiner, 1965). Therefore, the extreme response style may be curvilinear: ERS decreases through childhood and adolescence from its very high level in earlier childhood, is stable throughout the

middle-age and begins to rise with older age groups (Das & Dutta, 1969; Hamilton, 1968).

Unfortunately, results concerning individual differences like gender and cognitive ability are ambiguous: ERS can differ by sex with females engaging more in ERS than men (Austin, et al., 2006; Berg & Collier, 1953; Eid & Rauber, 2000) – or no differences in sex were found (Brenzelmann, 1960b; Greenleaf, 1992; Light, et al., 1965; Marin, Gamba, & Marin, 1992; Naemi, et al., 2009).

The results regarding cognitive ability are ambiguous, too: Some studies show a negative relationship between ERS and cognitive ability, meaning that lower cognitive ability individuals engage in more ERS (Brenzelmann, 1960a; Das & Dutta, 1969; Light, et al., 1965) and others do not find any relationship (Kerrick, 1954; Zuckerman & Norton, 1961). However, it has to be mentioned that different intelligence tests were used, questioning the comparability of the results. A relationship between education and ERS was also found with ERS being more common among lower-educated (less than twelve years of formal education) respondents (Greenleaf, 1992; Marin, et al., 1992). This is supported by the result of Eid and Rauber (2000): The leadership level was found to distinguish between extreme responders and non-extreme responders with secretaries, typists, and workman using more extreme categories than heads of departments or leaders of working groups. Moreover, also the acquiescence response style was more often found by less educated respondents (Ross, Steward, & Sinacore, 1995). All these results support the negative relationship between cognitive ability and ERS.

What about ERS and personality factors and facets? “Under appropriate conditions...it might be possible to utilize response sets as personality measures by assuming that such sets are manifestations of basic traits” (Berg & Collier, 1953; p. 164). What are the personality traits ERS is thought to be a behavioural manifestation

of? Extraversion was found to increase significantly with an extreme response style (Austin, et al., 2006; Meiser & Machunsky, 2008). The relationship of ERS and higher scores in conscientiousness could not always be proven (Austin, et al., 2006) and for agreeableness, openness and neuroticism no results were found at all (Austin, et al., 2006; Meiser & Machunsky, 2008). The results concerning neuroticism are controversial, because relationships were found between ERS and anxiety (Berg & Collier, 1953; Lewis & Taylor, 1955), with high-anxiety respondents being more likely to use endpoints of Likert-type scales. However, not all studies did find differences between high- and low-anxiety groups in ERS (Hamilton, 1968). The relationship between ERS and personality facets is not searched for, yet.

2.1.5 Ways detecting ERS

These conflicting results might be caused by different methods used to identify extreme crossers. One of the easiest ways to detect extreme crossers is to count the amount of end-scale usage (Berg & Collier, 1953; Harzing, 2006; Johnson, Kulesa, Llc, Cho, & Shavitt, 2005): Questions sharing the same Likert-type scale were recoded so that selection of one of the endpoints received a code 1 and the middle values received a code 0. Higher values on this ERS index (either in absolute or in relative numbers) reflected more extreme responding. Another possibility to measure ERS was to use the deviation from the middle of scales, irrespective of direction (Warr & Coffman, 1970). Due to the fact that all these questions are often measuring the same construct, it was hard to divide between people answering extreme to all the questions because of an extreme response style or because of an extreme attitude toward the underlying concept. To find a remedy, scales to measure ERS were created, which consisted of uncorrelated items (Greenleaf, 1992). This prevented confounding ERS with specific item content. To find associations between an extreme response style and individual variables like

gender or education, regression analyses were conducted to test for significant predictive relationships (Greenleaf, 1992). However, only observed groups could be tested for differences in ERS. Another way for testing ERS is to use item response theory (IRT). Using Mixed-Rasch Models, latent classes of extreme responders and classes of non-extreme responders can be detected (Eid & Rauber, 2000; Meiser & Machunsky, 2008; Rost, Carstensen, & von Davier, 1999). Due to a successive latent class analysis, the stability of the class membership can be tested (Rost, et al., 1999), too. Thus, a method was found to identify extreme crossers irrespective of their personality trait score. Using this method, individual differences of middle and extreme crossers in variables like personality factors, personality facets, intelligence, or age were investigated to close research gaps.

2.1.6 Goals of the present study

Research has shown that people have different response styles when answering non-cognitive self-report questionnaires: Whereas some disproportionately favour endpoints or extreme categories of ordinal response or Likert-type scales, irrespective of particular item content, others prefer middle categories. Although earlier studies examined individual differences between individuals of both groups (middle and extreme crossers), results especially concerning individual differences in personality traits are inconclusive and rare. Furthermore, most studies – even those using Mixed-Rasch-Models to identify classes of middle and extreme crossers – have to struggle with contaminated results: Scores on personality tests were used to identify the response style and the personality traits of middle and extreme crossers.

Independent from this methodical problem and conflicting results on factor / domain level, individual differences in personality facets between middle and extreme crossers are completely unknown, although personality facets have proven to bring

further insights for diverse criteria (Costa & McCrae, 1995; Lounsbury, Sundstrom, Loveland, & Gibson, 2002). To close this gaps this study tries to shed light on individual differences between middle and extreme crossers, taking the problems of contamination of personality test scores by response styles into account. Therefore it is hypothesized that respondents use different response styles when answering a non-cognitive self-report questionnaire (H1), and that the response style is stable across different personality facets (H2). Moreover, middle and extreme crossers should not only differ in the personality facets of neuroticism (H3a), extraversion (H3b), openness (H3c), and agreeableness (H3d), but also in conscientiousness (H3e). To avoid contamination between scores of personality tests used to identify classes of middle and extreme crossers and personality test scores used to identify differences between middle and extreme crossers in personality traits, person parameters (thetas) derived from a two-class MRM were used to identify differences between middle and extreme crossers in personality factors and facets. Thus, a contamination can be avoided. Moreover, it is hypothesized that middle and extreme crossers also differ in their intelligence scores (H4). Finally, extreme and middle crossers should differ in gender (H5) and age (H6), too. Due to the fact that most previous results are conflicting, the hypotheses are non-directed.

2.2 Method

2.2.1 Procedure

The data were collected in a German university. Participants worked on a personality measure, on a cognitive ability test and on several other tests, like a lexical knowledge test, which are not reported in this study. Computer versions of all tests were used.

2.2.2 Participants

Participants in this study were $N = 326$ undergraduate psychology students of the Ludwig-Maximilians-University (LMU) in Munich. Due to technical problems with the computer versions of the tests, data sets were lost so that only results of 312 participants were examined (304 as far as cognitive abilities are investigated). 247 participants (79%) were female. Mean age was 25 years ($SD = 5.5$) with a range from 20.5 to 53. All students received study participation credits for their participation in the study.

2.2.3 Measures

2.2.3.1 Instruments

Personality was assessed with the NEO-PI-R (Costa & McCrae, 1992) in a German adaptation (Ostendorf & Angleitner, 2004), which allows a comprehensive assessment of general personality. The NEO-PI-R contains 240 items, measuring the five factor model (FFM): neuroticism, extraversion, openness, agreeableness, and conscientiousness. Respondents with a minimum age of 16 can rate their statements in the questionnaire on a five-point Likert-type scale of endorsement, ranging from *strongly disagree* to *strongly agree*.

Cognitive ability was tested with the help of the basic module of the Intelligence Structure Test 2000 R (Amthauer, Brocke, Liepmann, & Beauducel, 2001). This module consists of 180 items measuring in nine subtests verbal, figural, and numerical reasoning. Combined, the verbal, figural and numerical score build the reasoning score (see Beauducel, Broke and Liepmann, 2001, for details concerning the theoretical basis and factor structure).

2.2.3.2 *Statistical analyses*

To calculate personality scores of participants, negative formulated items will be recoded in accordance with the manual of the NEO-PI-R.

To compute the internal consistency for the personality and the intelligence test SPSS 17.0 is used.

To avoid response bias concerning the middle category of the five-point Likert-type scale of the NEO-PI-R, the scale will be collapsed into a four-point Likert-type scale for all IRT analyses as proposed by Rost, Carstensen and von Davier (1999) and exercised by Austin et al. (2006).

Furthermore, the computer program Windows Mixed Rasch Model Analysis 2001 (WINMIRA 2001; Davier, 2001) will be used to identify distinct subpopulations, like middle or extreme crossers. WINMIRA requires that the frequency of the lowest answering category of all items is unequal zero, meaning that for every item at least one participant has to choose the lowest answering possibility of the scale named *strongly disagree*. If this is not the case, all items of the affected facet (eight items) will be reverse-coded. This dataset will be used to calculate different class analyses for each facet of the NEO-PI-R with the help of WINMIRA. Class solutions will be calculated until all information criteria rise again. Information criteria used are the Akaike's Information Criterion (AIC), Schwartz's Best Information Criterion (BIC) and Bozdogan's Consistent AIC (CAIC). For an overview of these coefficients, see Bozdogan (1987) or Read and Cressie (1988).

In line with other studies using Rasch Models to detect subgroups even in latent classes (Eid & Rauber, 2000; Rost, et al., 1999), the partial credit model (PCM; Masters, 1982) will be used. Profiles of item locations of the two- and three-class solutions will be analyzed to check if the content of items caused different classes (Rost, et al., 1999). Moreover, plots of option thresholds for each facet will be examined,

searching for items with estimation problems. In both cases, items causing problems will be eliminated. If less than five items per facet will remain, the whole facet will be eliminated, because scale length is an important factor in the accurate identification of classes within WINMIRA (Zickar & Burnfield, 2003). For these corrected facets, a second run of class analyses will be conducted. The class where an information criteria fits and which is clearly interpretable (class size, option thresholds, mean, content), will be chosen to determine the number of classes needed to fit the data. All interpretable facets and classes will be coded with a dichotomous indicator variable (0 = middle crossers; 1 = extreme crossers) and a latent class analysis (LCA) will be carried out to check whether this response style is consistent across facets. Finally, to detect individual differences between middle and extreme crossers, t-tests for independent samples will be calculated searching for differences in personality factors, facets, cognitive ability, and age. To avoid contamination between scores of personality tests used to identify classes of middle and extreme crossers and personality test scores used to identify differences in personality traits between the classes, person parameters (thetas) derived from a two-class MRM will be used in the t-tests to identify differences between middle and extreme crossers in personality factors and facets. To determine whether the effects are strong, the effect size Hedge's g will be examined. According to Cohen (1988), an effect size of .20 indicates a small effect, whereas a Hedge's g of .50 signalizes a moderate and .80 a strong effect. Differences in gender will be determined using a χ^2 -test.

2.3 Results

As can be seen in Table 1, the internal consistencies of the NEO-PI-R ranged from $\alpha = .46$ to $\alpha = .88$ for the personality facets and from $\alpha = .86$ to $\alpha = .93$ for the personality factors. So not all internal consistencies were acceptable, but comparable to those of the

test-handbook (Costa & McCrae, 1992). Cronbachs Alphas for the intelligence test IST-2000-R were between $\alpha = .85$ and $\alpha = .95$.

Table 1
Cronbachs Alphas for NEO-PI-R and IST-2000-R scales

Dimension	Facet / Subscales					
N	N1	N2	N3	N4	N5	N6
.93	.85	.75	.86	.76	.60	.81
E	E1	E2	E3	E4	E5	E6
.89	.75	.78	.82	.69	.59	.80
O	O1	O2	O3	O4	O5	O6
.86	.75	.80	.79	.63	.82	.46
A	A1	A2	A3	A4	A5	A6
.88	.88	.68	.71	.65	.75	.61
C	C1	C2	C3	C4	C5	C6
.91	.72	.73	.71	.67	.81	.80
Reasoning	Numeric Reasoning	Verbal Reasoning	Figural Reasoning			
.95	.94	.85	.88			

Notes. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-conscientiousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation.

All facets of the NEO-PI-R, in which no participant used the lowest category *strongly disagree* to answer a single item, were recoded to enable the usage of WINMIRA. Altogether, 13 facets from 4 factors had to be recoded (see Table 2).

Table 2
Recoded facets of the NEO-PI-R

Neuroticism	Extraversion	Openness to Experience	Agreeableness	Conscientiousness
N1 Anxiety	E1 Warmth	O1 Fantasy	A1 Trust	C1 Competence
N2 Angry hostility	E2 Gregariousness	O2 Aesthetics	A2 Straight-forwardness	C2 Order
N3 Depression	E3 Assertiveness	O3 Feelings	A3 Altruism	C3 Dutifulness
N4 Self-conscientiousness	E4 Activity	O4 Actions	A4 Compliance	C4 Achievement striving
N5 Impulsiveness	E5 Excitement seeking	O5 Ideas	A5 Modesty	C5 Self-discipline
N6 Vulnerability	E6 Positive emotions	O6 Values	A6 Tender-mindedness	C6 Deliberation

Notes. black coloured facets had to be recoded.

2.3.1 Searching for Subgroups

2.3.1.1 *Winmira 1st Run*

To prove the occurrence of different response styles (H1), Rasch / Mixed Rasch Analyses had to be conducted. Analyzing the results of these class analyses, seven facets showed one-class solutions, five facets three-class solutions and for the remaining 18 facets, two classes were needed to fit the data (see Table 3 for an overview). Plots of the item locations for the two- and three-class solutions showed, that participants interpreted some items differently. So not all item locations were (almost) parallel and items had to be removed from facets as can be seen in Figure 4.

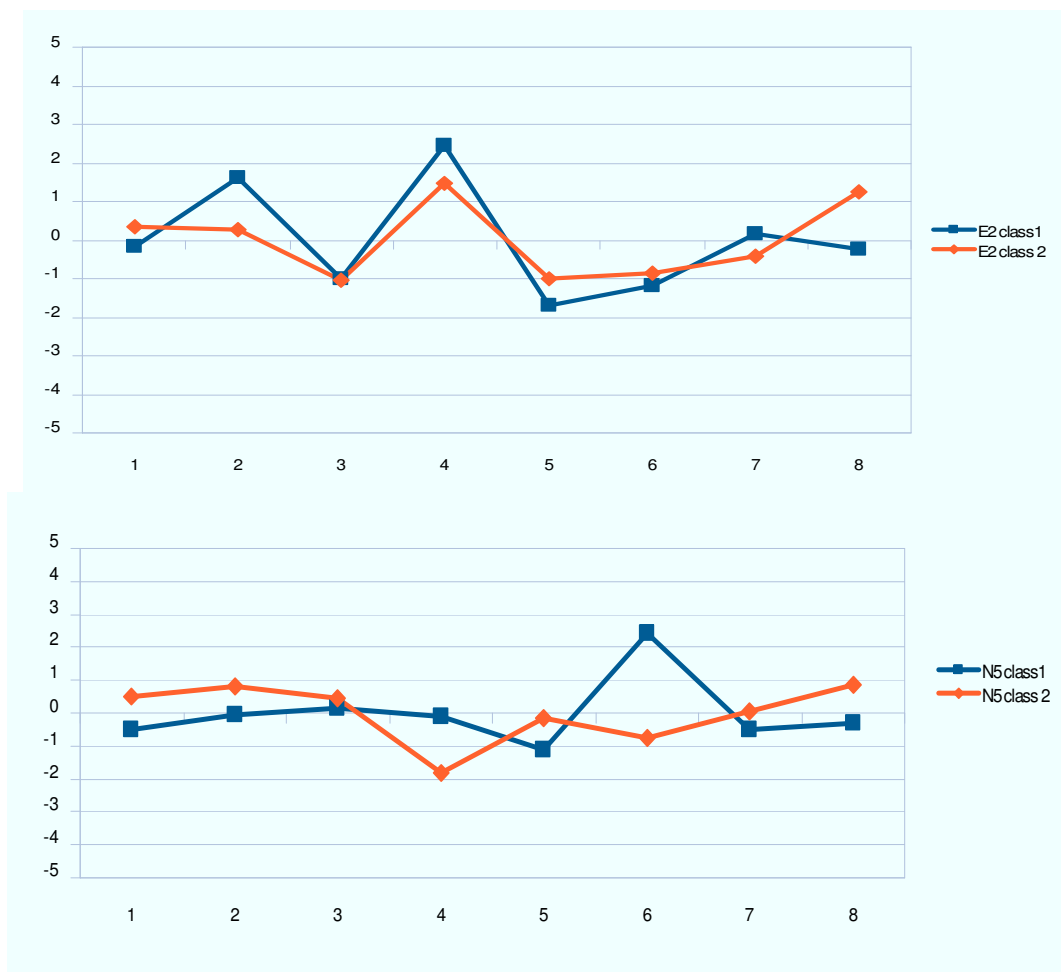


Figure 4
(Almost) parallel item locations for E2; variations in item locations for N5

Examining the plots of option thresholds for all facets, two generalizations can be made: First, in eight facets of the two-class solutions threshold estimates were interpretable, meaning that thresholds had (almost) a correct ordering, smaller thresholds were found in classes with higher means indicating extreme crossers, and class sizes were not extreme little representing a small group of outliers (see Figure 5 for an example). So in all eight facets, extreme crossers have a higher mean and are the smaller class. Exceptions are only facet A4, in which middle crossers have a higher – but not significantly higher mean ($M_{K1} = 16.87$, $SD = 4.18$; $M_{K2} = 15.59$, $SD = 4.63$; $t(86) = 1.97$, $p = .052$, Hedge's $g = .28$) – and facet E5, in which extreme crossers are the larger class consisting of 53% of the respondents.

Second, estimating problems (item locations > 16l) made an interpretation in some facets impossible (see Figure 6 for an example). In such cases, items causing estimating problems were removed and WINMIRA analyses were repeated.

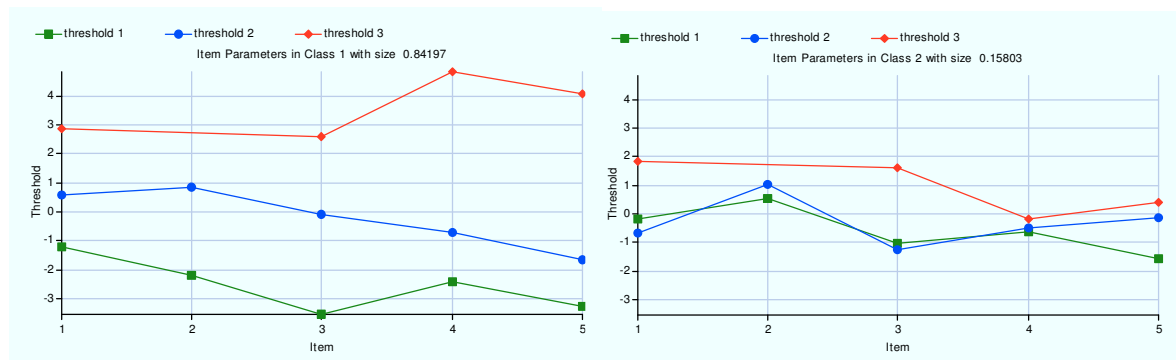


Figure 5
Threshold estimates for facet O6 class 1 and 2 – an interpretable example

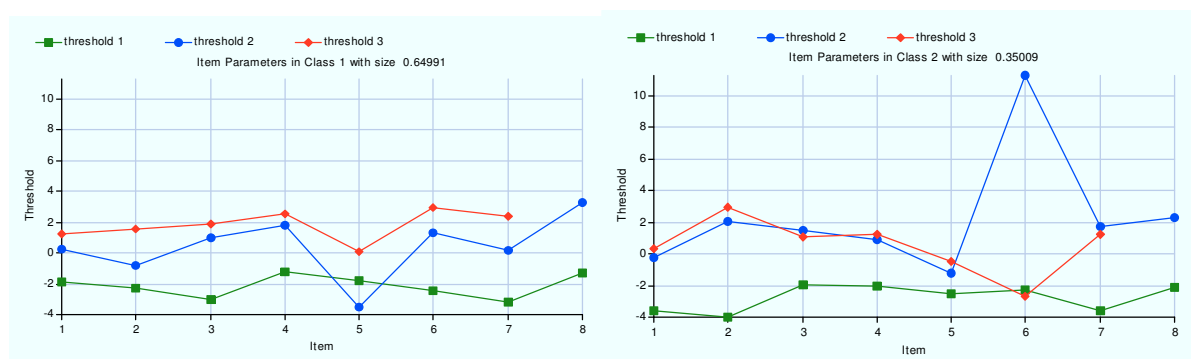


Figure 6
Threshold estimates for facet A2 class 1 and 2 – an example with estimation problems

2.3.1.2 Winmira 2nd Run

After eliminating items due to estimation problems or due to difficulties in item interpretation, one-, two-, and three-class solutions were calculated again using WINMIRA 2001. Four facets (C2, O3, E6, N3) were not included in the second run, because after item elimination less than five items would have remained.

This time, five facets showed a one-class solution, two facets three-class solutions and four facets two-class solutions (see Table 3 and Table 4 for an overview). 4 out of these 11 facets (A2, N4, O4, O6) had classes clearly interpretable as middle and extreme crossing classes: The class of the extreme crossers were smaller, had thresholds with smaller distances and had higher means. The only exceptions are the means of O6, which is in the extreme crossing class lower, but not significantly lower, than in the class of the middle crossers ($M_{K1} = 7.00$, $SD = 1.74$; $M_{K2} = 6.90$, $SD = 2.51$; $t(46) = .24$, $p = .82$, Hedge's $g = .04$), and the class size of A2 (63% of the respondents are in the extreme crosser class instead of less than 50%). Although being a three-class solution according to information criteria, O4 is integrated in the following LCA, because the two-class solution is clearly interpretable and the three-class solution stems from a different interpretation of single items of this facet.

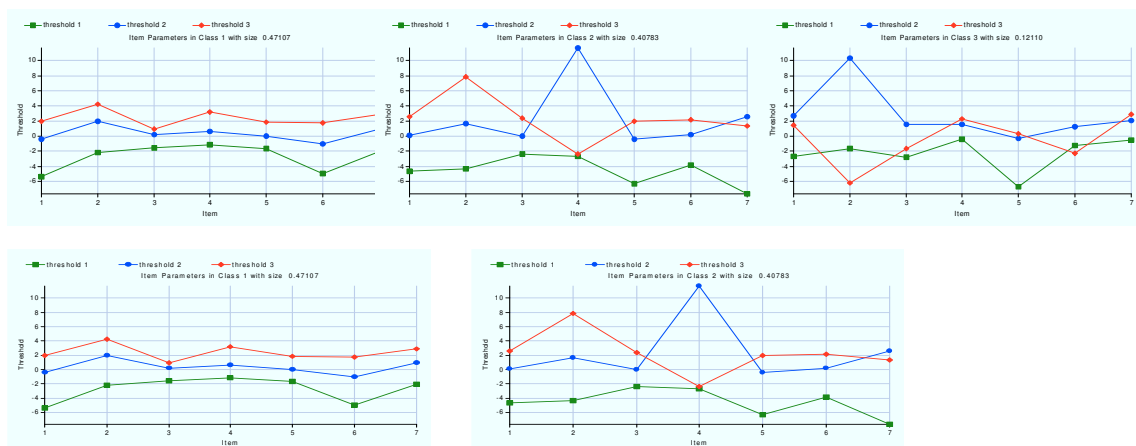


Figure 7
Two- and three-class solutions of facet O4

Thus, eight facets from the first run and four facets of the second WINMIRA run were included in the next step – the latent class analysis.

Table 3
Class solutions for NEO-PI-R scales

WINMIRA 1st and 2nd run						
Facet	Nr class 1st run	Item locations 1st run	Thres-holds 1st run	Nr class 2nd run	Item locations 2nd run	Thres-holds 2nd run
A1	2	parallel	clear-cut	-	-	-
A2	2	parallel	not clear	2	variations	clear-cut
A3	1	-	clear-cut	-	-	-
A4	2	variations	clear-cut	-	-	-
A5	2	variations	not clear	1	-	clear-cut
A6	2	variations	clear-cut	-	-	-
C1	1	-	clear-cut	-	-	-
C2	2	variations	not clear	--	--	--
C3	2	variations	clear-cut	-	-	-
C4	2	variations	clear-cut	-	-	-
C5	1	-	clear-cut	-	-	-
C6	2	parallel	not clear	1	-	clear-cut
E1	1	-	clear-cut	-	-	-
E2	2	parallel	clear-cut	-	-	-
E3	2	parallel	not clear	2	variations	not clear
E4	1	-	clear-cut	-	-	-
E5	2	variations	clear-cut	-	-	-
E6	3	variations	not clear	--	--	--
N1	2	parallel	not clear	1	-	clear-cut
N2	3	variations	not clear	1	-	clear-cut
N3	3	variations	not clear	--	--	--
N4	3	variations	not clear	2	parallel	clear-cut
N5	2	variations	not clear	1	-	clear-cut
N6	2	variations	clear-cut	-	-	-
O1	3	variations	not clear	3	variations	not clear
O2	1	-	clear-cut	-	-	-
O3	2	variations	not clear	--	--	--
O4	2	parallel	not clear	3	variations	clear-cut
O5	1	-	clear-cut	-	-	-
O6	2	variations	not clear	2	variations	clear-cut

Notes. N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-consciousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation. The information criteria AIC was used. All facets in bold letters were used in the following latent class analysis; -- indicate facets not included in the second run, because less than five items would have remained after item elimination.

Table 4
Information criteria for each facet and each class of WINMIRA 1st and 2nd run

Class	A1			A2			A3			A4			A5			A6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	4594.96	4770.88	4817.88	4338.07	4491.53	4532.53	<i>4688.76</i>	<i>4864.68</i>	<i>4911.68</i>	4976.79	5152.71	5199.71	2865.30	2973.85	3002.85	<i>4693.67</i>	4869.59	4916.59
K2	4548.37	4896.47	4989.47	4323.33	4626.52	4707.52	<i>4689.12</i>	<i>5037.22</i>	<i>5130.22</i>	4960.35	5308.45	5401.45	2889.25	<i>3102.60</i>	<i>3159.60</i>	4655.58	5003.68	5096.68
K3	4554.90	5075.18	5214.18	4340.08	4792.99	4913.99	<i>4713.00</i>	<i>5233.27</i>	<i>5372.27</i>	5052.42	5572.69	5711.69	2910.53	<i>3228.68</i>	<i>3313.68</i>	<i>4719.54</i>	5239.82	5378.82
K4	4592.70	5285.15	5470.15	4463.39	5066.02	5227.02	<i>4814.46</i>	<i>5506.92</i>	<i>5691.92</i>	5074.98	5767.43	5952.43	<i>2894.46</i>	<i>3317.42</i>	<i>3430.42</i>	<i>4825.36</i>	<i>5517.82</i>	<i>5702.82</i>
	C1			C2			C3			C4			C5			C6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	4612.69	4788.61	4835.61	5125.48	5301.40	5348.40	<i>4889.39</i>	5065.31	5112.31	5020.37	5196.30	5243.30	4628.04	4803.96	4850.96	2856.49	2965.03	2994.03
K2	<i>4630.31</i>	<i>4978.41</i>	<i>5071.41</i>	5077.87	5425.97	5518.97	4840.29	<i>5188.39</i>	<i>5281.39</i>	4977.61	5325.71	5418.71	<i>4639.11</i>	<i>4987.21</i>	<i>5080.21</i>	2857.20	3070.55	3127.55
K3	<i>4660.99</i>	<i>5181.27</i>	<i>5320.27</i>	5085.09	5605.37	5744.37	<i>4844.82</i>	<i>5365.10</i>	<i>5504.10</i>	4999.54	5519.81	5658.81	<i>4724.73</i>	<i>5245.01</i>	<i>5384.01</i>	2960.60	3278.76	3363.76
K4	<i>4714.73</i>	<i>5407.19</i>	<i>5592.19</i>	5100.41	5792.87	5977.87	<i>4914.26</i>	<i>5606.71</i>	<i>5791.71</i>	5079.35	5771.81	5956.81	<i>4751.40</i>	<i>5443.85</i>	<i>5628.85</i>	2884.02	3306.98	3419.98
	E1			E2			E3			E4			E5			E6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	4720.80	4896.72	4943.72	<i>5057.47</i>	5233.39	5280.39	4030.21	4183.68	4224.68	4825.32	5001.24	5048.24	5878.02	6053.94	6100.94	4849.69	5025.61	5072.61
K2	<i>4811.75</i>	<i>5159.85</i>	<i>5252.85</i>	5006.99	<i>5355.09</i>	<i>5448.09</i>	4020.22	4323.41	4404.41	<i>4914.41</i>	<i>5262.51</i>	<i>5355.51</i>	5844.30	6192.40	6285.40	4734.08	5082.18	5175.18
K3	<i>4737.11</i>	<i>5257.39</i>	<i>5396.39</i>	<i>5009.76</i>	<i>5530.04</i>	<i>5669.04</i>	4072.41	4525.32	4646.32	<i>4874.73</i>	<i>5395.00</i>	<i>5534.00</i>	5874.55	6394.83	6533.83	4719.98	5240.26	5379.26
K4	<i>4834.45</i>	<i>5526.91</i>	<i>5711.91</i>	<i>5086.94</i>	<i>5779.40</i>	<i>5964.40</i>	4118.38	4721.01	4882.01	<i>4957.92</i>	<i>5650.37</i>	<i>5835.37</i>	5884.67	6577.13	6762.13	4765.65	5458.11	5643.11
	N1			N2			N3			N4			N5			N6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	3133.23	3241.77	3270.77	2961.76	3070.31	3099.31	4690.99	4866.91	4913.91	3731.78	3862.78	3897.78	3647.53	3778.54	3813.54	4178.97	4354.89	4401.89
K2	3139.67	3353.02	3410.02	2963.90	3177.25	3234.25	4763.04	5111.14	5204.14	3715.94	3974.21	4043.21	3695.77	3954.04	4023.04	4144.88	4492.98	4585.98
K3	3135.66	3453.82	3538.82	2963.11	3281.27	3366.27	4677.36	5197.64	5336.64	3760.53	4146.06	4249.06	3612.09	3997.62	4100.62	4171.97	4692.25	4831.25
K4	3146.09	3569.05	3682.05	3027.85	3450.81	3563.81	4691.21	5383.66	5568.66	3710.43	4223.22	4223.22	3647.62	4160.41	4297.41	4248.07	4940.53	5125.53
	O1			O2			O3			O4			O5			O6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	3303.75	3412.29	3441.29	5151.39	5327.31	5374.31	<i>4573.96</i>	4749.88	4796.88	4311.87	4465.33	4506.33	4951.19	5127.11	5174.11	<i>3138.98</i>	3247.53	3276.53
K2	<i>3274.79</i>	<i>3488.14</i>	<i>3545.14</i>	5206.03	5554.13	5647.13	4572.79	<i>4920.89</i>	<i>5013.89</i>	4297.76	4600.95	4681.95	<i>5011.44</i>	<i>5359.54</i>	<i>5452.54</i>	3128.08	<i>3341.43</i>	<i>3398.43</i>
K3	3230.30	<i>3548.46</i>	<i>3633.46</i>	5250.02	5770.30	5909.30	<i>4614.15</i>	<i>5134.43</i>	<i>5273.43</i>	4288.60	4741.51	4862.51	<i>5123.15</i>	<i>5643.43</i>	<i>5782.43</i>	<i>3151.40</i>	<i>3469.56</i>	<i>3554.56</i>
K4	<i>3249.60</i>	<i>3672.56</i>	<i>3785.56</i>	5378.90	6071.35	6256.35	<i>4624.46</i>	<i>5316.91</i>	<i>5501.91</i>	4358.73	4961.35	5122.35	<i>5039.51</i>	<i>5731.96</i>	<i>5916.96</i>	<i>3180.14</i>	<i>3603.10</i>	<i>3716.10</i>

Notes. K1 = one-class solution; K2 = two-class solution; K3 = three-class solution; K4 = four-class solution; grey facets indicate results from WINMIRA 1st run; black facets indicate results from WINMIRA 2nd run; italic letters indicate recoded facets; bold numbers indicate the best solution according to the particular information criteria.

2.3.1.3 Latent Class Analysis

To test whether the extreme response style is scale specific or a consistent personality trait throughout all facets (H2), a latent class analysis was executed. Therefore, an indicator variable was created for each of the 12 remaining facets indicating middle crossers (coded with 0) or extreme crossers (coded with 1). As Table 5 shows, two classes were needed to fit the data according to the BIC and CAIC and three according to the AIC.

Table 5
Latent class analysis with 12 facets of the NEO-PI-R

model	class nr.	class size	sum-score	Information Criteria		
				AIC	BIC	CAIC
1	1	1	4.27	4537.13	4582.04	4594.04
2	1	.67	3.21	4426.02	4519.60	4544.60
	2	.33	6.40			
3	1	.40	3.47	4424.70	4566.93	4604.93
	2	.31	6.54			
	3	.29	2.95			
4	1	.35	3.39	4425.18	4616.07	4667.07
	2	.23	5.23			
	3	.23	2.32			
	4	.19	6.99			

Notes. Model = number of classes chosen in the LCA; sum score = sum of the means of the facets included in the LCA; bold class solutions indicate the best solution according to the particular criteria.

The data used in the LCA were dichotomous variables indicating middle and extreme crossers. Accordingly, the two classes of the LCA can be described as classes of middle and extreme crossers. This result is confirmed by analyzing the item locations of the two-class solution of the LCA: The item profiles of the two-class solution are (almost) parallel, indicating that the reaction of the middle and extreme crossers are similar independent of the facet (see Figure 8).

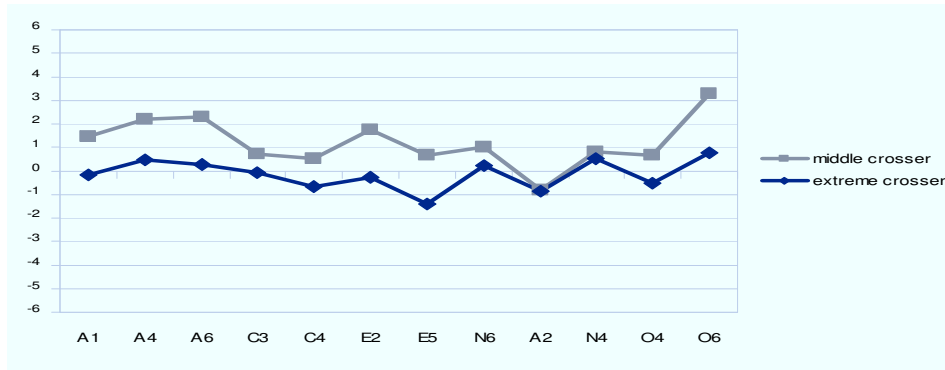


Figure 8

Item locations for the two-class solution of the 12 facets in the LCA

A1 = Trust; E2 = Gregariousness; A2 = Straightforwardness; C3 = Dutifulness; N4 = Self-conscientiousness; O4 = Actions; A4 = Compliance; C4 = Achievement striving; E5 = Excitement seeking; N6 = Vulnerability; O6 = Values; A6 = Tender-mindedness.

Due to the fact that three classes were not interpretable with regard to the content, the class sizes with their means, and items locations, the two-class solution will be used. The estimated mean probabilities of assigning a person to the right class were 90.4% for class 1 and 83.52% for class 2. All probabilities of assigning a person to a wrong class were below 16.5%. So class assignment can be regarded as good.

2.3.2 Response styles and individual differences

To test for individual differences between middle and extreme crossers in variables like personality factors and facets (H3), intelligence (H4), and age (H6) t-tests for independent samples and effect sizes were calculated. Differences in gender (H5) were tested using a χ^2 -test.

As Table 7 shows, middle and extreme crossers differ significantly in every personality factor, with neuroticism as exception and the effects are up to medium-sized. As far as personality facets are regarded, extreme crossers have significantly higher traits in five facets of extraversion and openness, in four facets of conscientiousness, in three facets of agreeableness, and significantly lower traits in two facets of neuroticism. Extreme crossers were also significantly younger and had significantly lower scores in verbal intelligence as can be seen in Table 7. All effects

were up to medium-sized ($g < .60$). Significant differences in gender were not found ($\chi^2 (1, N = 312) = 3.40, p = .07$; see Table 6 for the distribution of gender across classes).

Table 6
Gender distinguished by LCA class

class	gender	frequency	percent
middle crosser	male	38	17.9
	female	174	82.1
	sum	212	100.0
extreme crosser	male	27	27.0
	female	73	73.0
	sum	100	100.0

Table 7
Individual differences between middle and extreme crossers

	M_{Perspara} MC	SD_{Perspara} MC	M_{Perspara} EC	SD_{Perspara} EC	t ($df = 310$)	Hedge's g
N	-0.28	0.96	-0.50	1.26	1.56	.19
E	0.60	0.76	1.10	0.96	-4.55	-.55***
O	0.68	0.65	1.08	0.88	-4.05	-.49***
A	0.36	0.69	0.68	0.85	-3.26	-.40***
C	0.44	0.79	0.82	1.05	-3.18	-.39***
N1	0.31	1.24	-0.02	1.49	1.97	.24*
N2	-0.38	1.32	-0.53	1.81	0.71	.09
N3	-0.60	1.43	-1.02	1.76	2.11	.26*
N4	-0.30	1.14	-0.50	1.74	1.07	.13
N5	0.29	0.95	0.39	1.38	-0.68	-.08
N6	-0.99	1.56	-1.30	1.70	1.58	.19
E1	1.32	1.11	1.89	1.35	-3.69	-.45***
E2	0.88	1.22	1.02	1.40	-0.91	-.11
E3	-0.34	1.39	0.18	1.74	-2.60	-.32**
E4	0.05	0.87	0.58	1.19	-4.04	-.49***
E5	-0.05	0.79	0.35	0.85	-4.12	-.50***
E6	1.73	1.55	2.54	1.93	-3.67	-.45***
O1	0.32	1.00	0.66	1.10	-2.74	-.33**
O2	1.47	1.23	1.65	1.57	-1.03	-.12
O3	1.50	1.20	1.98	1.55	-2.77	-.34**
O4	-0.05	1.01	0.40	1.09	-3.60	-.44***
O5	0.72	1.30	1.39	1.92	-3.14	-.38***
O6	0.13	1.00	0.41	1.04	-2.23	-.27*
A1	0.56	1.32	0.94	1.50	-2.26	-.27*
A2	0.42	0.94	0.66	1.33	-1.68	-.20
A3	0.77	1.14	1.54	1.40	-4.81	-.58***
A4	0.08	1.00	0.07	1.14	0.09	.01
A5	-0.16	1.12	-0.05	1.28	-0.80	-.10
A6	0.53	0.92	0.92	0.96	-3.48	-.42***
C1	0.44	1.06	1.18	1.30	-4.96	-.60***
C2	0.65	0.95	0.87	1.28	-1.57	-.19
C3	0.78	1.04	1.24	1.27	-3.36	-.41***
C4	0.83	0.94	1.37	1.03	-4.57	-.55***
C5	0.06	1.22	0.45	1.76	-1.99	-.24*
C6	-0.09	1.79	-0.19	2.31	0.37	.05
Verbal R.	39.56	6.01	37.92	7.10	1.97	-.24*
Num. R.	41.80	9.48	41.38	1.11	.35	-.04
Figural R.	35.68	7.75	35.48	7.98	.20	-.02
Reasoning	117.04	18.46	114.79	19.69	.96	-.12
Age (month)	307.25	74.51	286.65	39.77	3.18	-.39**

Notes. Perspara = person parameters; MC = middle crosser; EC = extreme crosser; $N = 312$; $n_{\text{MC}} = 212$, $n_{\text{EC}} = 100$; * $p < .05$, ** $p < .01$, *** $p < .001$; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-conscientiousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation; Verbal R. = verbal reasoning; Num. R. = numeric reasoning; Figural R. = figural reasoning; bold letters indicate significant results.

2.4 Discussion

Research concerning individual differences in personality factors and facets between middle and extreme crossers is comparatively rare and conflicting. Using Mixed-Rasch Models, individual differences of middle and extreme crossers in variables like personality factors, personality facets, intelligence, or age were investigated to close research gaps. To avoid contamination between scores of personality tests used to identify classes of middle and extreme crossers and personality test scores used to identify differences in personality traits between the classes, person parameters (thetas) derived from a two-class MRM were used to identify differences between middle and extreme crossers in personality factors and facets. Thus, this study is one of the few avoiding contamination of personality traits by response style.

2.4.1 The occurrence of the extreme response style (ERS)

Using Rasch / Mixed-Rasch Models, classes of middle and extreme crossers could be found (H1). However, classes of middle and extreme crossers were not found in all facets, but in facets of all personality factors. Therefore, hypothesis 1 could only be partly verified. What might be the reason for not finding the response styles in all facets? On the one hand, there might be a methodical explanation: Because of estimation problems (not all rating scales categories were used, each facet consisted of only eight items) and participants interpreting some items differently, items had to be eliminated. An item, which was interpreted quite differently, is for example item 152 of the NEO-PI-R: “It is easy for me to smile and to get along with strangers”¹. Probably some respondents find it easy to smile *and* to get along with strangers, but some others find it only easy to smile. Thus, the questions arises where to cross when

¹ „Ich finde es leicht zu lächeln und mit Fremden gut auszukommen“ (translation of the author)

only half of the attitude can be agreed for? Therefore, this item (and some other items) was interpreted differently: Some respondents only answered to the first half of the attitude-sentence and others to the whole or to the second half. So from the originally eight items per facet sometimes only seven or six remained because of different item interpretation. Due to the fact that scale length is an important factor in the accurate identification of classes within WINMIRA (Zickar & Burnfield, 2003), not finding the classes might be partly a methodical problem. On the other hand, the combination of only a five-point Likert-type scale and a homogeneous sample consisting of psychology students in their first years might be another explanation: Every facet in which only a one-class solution was found consisted of items, which had a strong content for the specific sample. Maybe the five-point Likert-type scale was for those items not differentiating enough. So NEO-PI-R items like “I try to be friendly to everybody I meet”² (item 44, A3), “I’m proud of my good ability to judge”³ (item 125, C1), or “I’m open minded and tolerant for the lifestyle of others”⁴ (item 178, O5) are items to which future psychotherapists / people working with other people probably agree and on a five point Likert-type scale there are only two possibilities to express agreement – perhaps not differentiating enough between middle and extreme crossers when the content is so meaningful to the sample.

However, for the facets where middle and extreme crossers were found, the stability of ERS across personality facets (H2) could be confirmed. Due to the fact that middle and extreme crossers were found in facets of all factors the existence of ERS widely irrespective of item content could be replicated (Austin, et al., 2006; Eid & Rauber, 2000; Rost, et al., 1999).

² „Ich versuche zu jedem, dem ich begegne freundlich zu sein“ (translation of the author)

³ „Ich bin stolz auf mein gesundes Urteilsvermögen“ (translation of the author)

⁴ „Ich bin aufgeschlossen und tolerant für die Lebensgewohnheiten anderer Menschen“ (translation of the author)

Further analysis showed that 67% of the respondents were found to be in the class of the middle crossers, whereas 33% preferred extreme answering categories resulting in an extreme response style. This is also in accordance with Austin et al. (2006) or Eid and Rauber (2000), who also found the extreme crossers to be the smaller class with 29% of the respondents.

2.4.2 ERS and individual differences in personality factors / facets

To test differences between middle and extreme crossers in variables like personality factors, personality facets, intelligence, and age t-tests for independent samples and effect sizes were calculated. To avoid contamination between scores of personality tests used to identify classes of middle and extreme crossers and personality test scores used to identify differences in personality traits between the classes, person parameters (thetas) derived from a two-class MRM were used to identify differences between middle and extreme crossers in personality factors and facets. Results show that extreme crossers have significant higher values (with almost modest effect sizes) in the personality factors extraversion, openness, agreeableness, and conscientiousness, but no significant lower mean in neuroticism was found. What about differences on the level of personality facets? Middle crossers have significant higher means in anxiety (N1) and depression (N3) by small effect sizes, indicating that they tend to blame oneself, worry more, feeling lonely and sad. Due to the fact that middle and extreme crossers did significantly differ in anxiety (N1), the results of Berg and Collier (1953) as well as Lewis and Taylor (1955), who found relationships between ERS and anxiety could be replicated. However, the results also are in line with Austin et al. (2006) as well as Meiser and Machunsky (2008) who found no significant correlations between ERS and neuroticism on the factor level.

The results of Austin et al. (2006) as well as Meiser and Machunsky (2008) could be replicated in another point, too: Middle and extreme crossers differ significantly and with moderate effect sizes in their personality scores of extraversion. In fact, extraversion is the factor they differ with the highest effect size on *factor* level (Hedge's $g = .55$). What about the *facet* level? Extreme crossers tend to be more open-hearted (E1) and are more often light-hearted and frolic (E6). Moreover, extreme crossers tend to have a higher level of activity (E4), taking centre stage (E3), seeking for adventures (E5). Furthermore, extreme crossers prefer diversification / changes (O4), discussions, and have a higher level of curiosity (O5). They are also more active with regard to recognizing their feelings (O3), questioning social, ethical or politic norms (O6) or being open for fantasy (O1). Extreme responders tend to trust others more easily (A1), engage themselves more in helping others (A3), feeling higher sympathy for others (A6). Whereas former studies were not able to find differences between middle and extreme crossers in openness and agreeableness (Austin, et al., 2006; Meiser & Machunsky, 2008), the results of this study show significantly and with small to moderate effect sizes that differences clearly exist – but not in all personality facets.

Results concerning differences in conscientiousness were conflicting up to now. The results of this study show that a correlation clearly exists – but again – not in all facets: Whereas middle and extreme crossers differ significantly and with moderate effect sizes in their level of competence (C1) and achievement striving (C4) – and with small to moderate effect sizes in their level of dutifulness (C3) and self-discipline (C5) – no differences at all were found in their level of deliberation (C6). To conclude, in all factors in which previously no differences between middle and extreme crossers were found (consistently), there was at least one facet, in which no significant difference

exists, but also at least two facets with significant differences between middle and extreme crossers and small to moderate effect sizes ($g = .24$ to $.60$). Therefore, these results confirm the hypotheses that middle and extreme crossers differ in their personality traits (H3). Summing up the differences, *middle crossers* are less active concerning a lot of different personality traits: They are less active concerning their level of activity (E4), their need for changes (E5), their hunger for adventure (O4), their level of curiosity (O5), their tendency to question social or political norms (O6) or the level of engagement for others (A3). They are less open-hearted (E1) and frolic (E6). Instead, they are more self-referential (A3), more realistic (O1), more sceptical (A1) and unsentimental (A6), being less open for their feelings (O3). They are more concerned having a higher level of anxiety (N1, N3). Furthermore, they have a lower faith in their competences (C1) and a lower level of dutifulness (C3), achievement striving (C4) and self-discipline (C5). In contrast, *extreme crossers* are more active in many ways, are more in contact with their surrounding, and have a higher level of assertiveness and conscientiousness. However, if these differences in personality traits are the reason or a symptom of middle or extreme crossing has to be investigated further.

2.4.3 ERS and individual differences in fluid intelligence

Concerning cognitive ability, middle and extreme crossers did not significantly differ in their scores for numeric and figural reasoning. However, the difference in verbal reasoning was significant with a small effect size. Accordingly, the reasoning score, which is the sum of verbal, numeric and figural scores, did not differ (H4 rejected). Due to fact that the sample is preselected by intelligence because of the allocation of university places, it is interesting that especially the verbal scores were those being different between middle and extreme crossers. Why? Because people with

rather global categories, less differentiated cognitive structures, and poorly developed schemas were found to prefer extreme categories, due to the fact that they “might be overwhelmed by a 9-point response scale” (Eid & Rauber, 2000; p. 21). So the judgement complexity might be one underlying factor which causes people to prefer extreme categories as Baumgartner and Steenkamp (2001), Eid and Rauber (2000), as well as Naemi et al. (2009) mentioned, too. However, due to the fact that ERS does not always occur – at least in this study – this cannot be the one and only underlying factor.

2.4.4 ERS and individual differences in age and gender

In this study, extreme crossers were significant younger by small to moderate effect sizes – so differences in age were found (H6). Due to the fact that this was a student sample, the result of rather younger respondents being in the class of extreme crossers is in line with earlier research (Austin, et al., 2006; Eid & Rauber, 2000; Hesterly, 1963), which also found younger respondents to favour extreme categories, but it is also in line with the findings of ERS being curvilinear: ERS decreases through childhood and adolescence from its very high level in earlier childhood, is stable throughout the middle-age and begins to rise with older age groups (Das & Dutta, 1969; Hamilton, 1968). Gender differences were not found (H5 rejected), which might be a consequence of the sample with 79% of the participants being female.

2.4.5 Limitations

Using a student sample limited the results in several ways: Because of the small variance and the high homogeneity the student sample has in intelligence and gender, differences could hardly be found. Moreover, due to the high homogeneity in age, the influence of age on the relationship between ERS and individual differences in

personality traits and facets was not tested. Further research should investigate whether the conflicting results of ERS and individual differences not only stem from different operations of ERS and different study designs, but from the curvilinear correlation of age and ERS, which should be betrayed as a covariate.

However, not only the sample yielded in limitations, but also statistical and content based limitations have to be mentioned: Due to the fact that personality facets are only measured with the help of eight items, estimation problems occurred in WINMIRA analyses, because not all rating categories were used to answer an item. Moreover, respondents interpreted items differently. These two effects yielded – maybe among other things as mentioned before – in the amount of only 12 facets in which middle and extreme crossers were found and which could be used in the LCA to test whether the response style is consistent across personality facets.

2.4.6 Implications and future directions

This study is the first proving that middle and extreme crossers differ in a wide range of personality traits, accounting for the contamination of personality scores by response styles. Results show that middle and extreme crossers differ significantly and with up to moderate effect sizes in personality facets of all five personality factors with extreme crossers having higher scores in extraversion, openness, agreeableness, and conscientiousness and lower in neuroticism. Furthermore, the response style has proven to be consistent across personality facets. Thus, this study sheds further light on the process of why people cross where they cross on non-cognitive self-report questionnaires with Likert-type scales. However, if these differences in personality traits cause the response style or are only one more symptom has to be investigated further. On the one hand, results clearly indicate that extreme responders have a higher level of different kinds of activity and extraversion, indicating that differences in

personality traits may be the reason for extreme crossers to prefer disproportionately extreme categories on Likert-type scales. On the other hand, the fact that ERS could not be found in all facets conflict with this assumption. Therefore, future research is necessary trying to prove whether the reason for not finding middle and extreme crossers in all facets is really due to this unique combination of methodical problems (estimation problems), homogenous sample and comparatively small Likert-type scale. If this hold true, differences in personality traits might be (the) one underlying factor for the response style – in combination with a lower level of (verbal) reasoning.

Moreover, the results advise for caution: The standard method of summing item scores to receive trait scores will end in higher scores for extreme responders with comparable levels of the underlying trait than non-extreme responders. If for example achievement motivation or depression is investigated with the help of questionnaires, which consist only of positive formulated items, respondents higher in extraversion will receive higher scores in the intended-to-be-measured trait, because respondents higher in extraversion have a higher tendency to disproportionately favour extreme categories. Accordingly, results might be contaminated by the personality traits of a respondent. Therefore, when no questionnaires with counterbalanced positive and negative items are used, a combined assessment strategy with first classifying respondents according to their response tendencies and than comparing different groups might be the more appropriate way, when conclusions are drawn from group differences in self-report questionnaires.

**3. Socially desirable responding (SDR)
and individual differences**

3 Socially desirable responding (SDR) and individual differences

3.1 Present Study

The usage of self-report questionnaires in personnel selection processes has often been criticized due to the fact that people are not only able to answer in a socially desirable way but that they also do so (Alliger & Dwight, 2000; Birkeland, et al., 2006). This effect, also called faking, is a long researched topic. Therefore, questions like what is socially desirable responding (SDR / faking), what are the effects of SDR, when do people answer in a socially desirable way, what do they fake and do all fake to the same extend were (almost) answered and a short summary is given. However, the question whether and how people with different faking styles have individual differences in personality traits, intelligence, age, and gender is still not answered properly. Hence, this study tries to shed light on these topics, controlling the first time for different response styles (middle / extreme crossing).

3.1.1 What is socially desirable responding?

Paulhus (1986) distinguished two aspects of socially desirable responding: Impression management, which refers to the conscious response distortion to present oneself in a positive light and self-deception, an unconscious tendency to see oneself in a favourable manner. Later on, Paulhus (2002) proposed a two tier system (content-level: exaggeration of positive attitudes / reduction of negative attitudes; process level: conscious / unconscious adaption of the answer), splitting up self-deception in self-deceptive enhancement (tendency to exaggerate one's social and intellectual status) and self-deceptive denial (tendency to deny socially-deviant impulses). Impression management was divided into communion management (deliberate minimization of

faults) and agency management (deliberate promoting of competence)⁵, the part of impression management, which job applicants show in assessment situations.

Due to the fact that self-deception is not intentional, but part of a person's personality (Paulhus, 1991) and is not affected by situational cues (McFarland & Ryan, 2000), impression management / agency management is the part of socially desirable responding searched for in faking studies: People over-report in non-cognitive self-report questionnaires admirable attitudes and behaviour and underreport less socially respected ones (Krosnick, 1999) – and this systematically. Therefore, socially desirable responding (SDR) can be seen as a systematic measurement error resulting from the interaction between the situational demand and the person (Ziegler & Buehner, 2009). These measurement errors, also referred to as spurious measurement error (Schmidt, Le, & Ilies, 2003), are systematic because it is assumed that they do not always occur, but always under identical circumstances (Ziegler & Buehner, 2009).

3.1.2 What are the effects of SDR / Faking?

Socially desirable responding leads to over-reporting of admirable attitudes like conscientiousness and underreporting of less admirable attitudes like neuroticism in non-cognitive self-report measures (Viswesvaran & Ones, 1999). As a result, the mean under faking conditions rises up to .93 standard deviations in personality inventories and integrity tests as meta-analyses showed (Alliger & Dwight, 2000; Viswesvaran & Ones, 1999). Variance rises, too (Zickar & Robie, 1999), which leads to higher correlations in faking situations – when no sealing effect occurs and when the additional variance is systematic as assumed by Ziegler and Buehner (2009). A sealing effect due to limited answering categories of Likert-type scales can cause correlations

⁵ Self-deceptive enhancement and agency management are betrayed as egoistic bias (the conscious and unconscious exaggeration of positive attitudes), whereas self-deceptive denial and communion management are seen as moralistic bias (the conscious and unconscious reduction of negative attitudes).

to be unaltered. Hence, faking instructions can lead to higher correlations between theoretically unrelated personality measures in experimental faking studies (Ellingson, Sackett, & Hough, 1999; Pauls & Crost, 2005b; Zickar & Robie, 1999), but also in applicant settings (Schmit & Ryan, 1993). Whereas some authors have argued that this response distortion can affect the construct and criterion-related validity (Schmit & Ryan, 1992; Zickar & Drasgow, 1996), much research has indicated that the criterion-related validities of self-report personality and integrity measures are not significantly affected (Barrick & Mount, 1996; Smith & Ellingson, 2002; Viswesvaran & Ones, 1999; Viswesvaran, Ones, & Hough, 2001; Ziegler & Buehner, 2009) and construct validity is still given (Ellingson, Smith, & Sackett, 2001). Nonetheless, not all concerns and questions are answered properly. For example, as Ziegler, Danay, Schoelmerich and Buehner (2010) were able to prove, the criterion validity on domain level was not influenced by faking, but the criterion validity on facet level: Criterion validities on facet level were affected in size and sometimes in direction, too. Therefore, further research in this field is needed.

Furthermore, applicant faking resulted in significant rank order changes, which influenced hiring decisions (Birkeland, et al., 2006; Converse, Peterson, & Griffith, 2009; Peterson, et al., 2009): When personality was the only predictor in a personnel assessment process (single-predictor selection), overall decision consistency (the extent to which different individuals are hired when selection occurs in the presence vs. absence of faking) was about 70%. This demonstrates that most individuals hired in the absence of faking would be hired in the presence of faking, too. For multiple-predictor selection – the method, which is primarily used in selection practice – the overall decision consistency rose up to approximately 90%, indicating that very similar decisions will be made in the presence vs. absence of faking (Converse, et al., 2009).

Summing up, rank order changes are a cause of concern. However, as long as non-cognitive self-report measures are used to reduce the pool of applicants and not to exclusively and definitively choose an applicant, non-cognitive self-report questionnaires are an appropriate tool in the personnel selection process.

3.1.3 When do people answer in a socially desirable way?

People adjust their answers especially in situations with a high situational pressure: In imagined applicant settings there are higher means in integrity tests than in “straight-take” conditions (Jackson, Wroblewski, & Ashton, 2000) and in real-life application settings responses are more distorted than in non-applicant settings (Birkeland, et al., 2006). This situational pressure is determined by the context in which the self-report takes place. Is the survey part of an applicant testing battery and the respondent would like to get this job or is the outcome of the survey important to get a looked-for therapy (Blanchard, McGrath, Pogge, & Khadivi, 2003)? This fake-good or fake-bad effect could be reported for experimental settings as well as for real application settings (Rothstein & Goffin, 2006). However, even in situations without external situational pressure 7.2% to 22.9% of the respondents answer in a socially desirable way (Zickar, Gibby, & Robie, 2004).

3.1.4 What do people fake?

People do not fake everything in an assessment: Testing personality measures of their fakability revealed that respondents adjusted their answers due to fake-good instructions for conscientiousness, agreeableness and neuroticism, but they did not fake openness (Furnham, 1997; McFarland & Ryan, 2000). In contrast, a meta-analysis conducted by Birkeland et al. (2006) revealed that respondents faked extraversion, emotional stability (neuroticism), conscientiousness, and openness but not

agreeableness. Why? Because of the implicit idea of the respondent what is important for the special issue (e.g. job, therapy). This is the reason why people with different knowledge applying for different jobs fake different items when trying to achieve the same goal like getting a job or a place at university (Birkeland, et al., 2006; p. 327): “...applicants appear to be distorting their responses on personality dimensions they view as particularly relevant to the specific job for which they are applying. Across most jobs, these dimensions appear to be conscientiousness and emotional stability. In some cases, however, such as sales jobs, the dimension might be extraversion, or even agreeableness.”

The answers of items, which are betrayed as not important are not faked and therefore the answers are either honest or neutral (Ziegler & Buehner, 2009).

However, it has to be mentioned that also the experimental setting (field or labour) and the type of test design (within or between subject design) has an influence on the degree to which answers are distorted: In within subject designs responses are more distorted than in between subject designs (Bye & Holtgraves, 2008; Viswesvaran & Ones, 1999) and responses in field settings with real-life applicants are less distorted than responses in labour settings with simulated applicant settings (Birkeland, et al., 2006).

3.1.5 Do all fake to the same extend?

People differ not only concerning the items they fake, but also in the extent they fake. People who are faking increase their scores by a constant amount (Zickar & Robie, 1999) – but this amount differs from person to person. Thus, two different faking styles were found: Slight fakers and extreme fakers. Whereas honest / regular responders use the full range of response options, slight fakers are more likely to

choose mildly positive options. Extreme fakers choose the most positive options with a high frequency (Zickar, et al., 2004).

3.1.6 Socially desirable responding and individual differences

Are there individual differences between respondents who fake and those who do not? Research revealed that male scored higher on social desirability scales (Ones & Viswesvaran, 1998) and were more often member of an extreme faking class (Ziegler, 2007). Moreover, a positive correlation of cognitive ability and faking in general could be proven by Grubb and McDaniel (2007) as well as by Pauls and Crost (2005a).

Self-monitors tended to rate themselves higher than low self-monitors on socially desirable items (Tunnel, 1980). Therefore, McFarland and Ryan (2000; p. 813) conclude that “high self-monitors are more aware of their social surroundings, are more aware of what is socially appropriate, and are better at manipulating the impressions they make on others. All of this leads to the conclusion that high self-monitors may be more adept at faking to look good.” Furthermore, individuals high in self efficacy of positive self-presentation (Pauls & Crost, 2005a; Ziegler, 2007) and low in integrity (McFarland & Ryan, 2000) were found to fake more.

As far as personality traits are concerned, the following results were found: Faking was found to correlate positively with agreeableness (Grubb & McDaniel, 2007) and conscientiousness (Ones, Viswesvaran, & Reiss, 1996), with those who are high in conscientiousness having a higher probability to be an extreme faker (Ziegler, 2007). In contrast, a significant negative correlation between conscientiousness and magnitude of faking was found by Byle and Holtgraves (2008) as well as by McFarland and Ryan (2000). The results concerning neuroticism are conflicting, too: Whereas a negative relationship between socially desirable responding and neuroticism was found with those who are less neurotic having higher scores on lie scales (Ones, et

al., 1996), other researcher found a positive relationship (McFarland & Ryan, 2000). Why are the results conflicting? One explanation for these conflicting results may be the usage of lie scales in older studies, which are not able to differ between regular responders and fakers clearly – as is known nowadays (Ziegler & Buehner, 2009). Moreover, as socially desirable responding consists of two parts (Paulhus, 1986) – the self-deception and the impression management part – it may be the self-deception component that causes the correlations in the conflicting traits. Whether the conflicting results concerning personality factors are caused by heterogeneous correlations on the dimension of personality facets cannot be answered, because until now, this subject has not been researched. Being young, having higher reasoning scores and having higher scores in the personality factors extraversion, openness, agreeableness and conscientiousness and lower in neuroticism were variables favouring the extreme faking class and so differing between slight and extreme fakers (Ziegler, 2007).

3.1.7 Goals of the present study

Research in the field of socially desirable responding showed that people have different faking styles: Whereas some respondents answer honestly even when asked to distort their responses using the full range of the rating scale, others fake a bit (slight fakers) choosing mildly positive options on rating scales. A third group, the extreme fakers select the most positive option with high frequency. Individual differences between slight and extreme fakers in personality factors, intelligence, gender, and age were seldom investigated. Moreover, individual differences in personality facets between slight and extreme fakers are completely unknown, although personality facets have proven to bring further insights (Costa & McCrae, 1995; Lounsbury, et al., 2002). To close this gap, this study tries to shed light on individual differences between slight and extreme fakers. Therefore, it is hypothesized that respondents fake

when instructed to do so (H1), that they fake to a different amount showing the two different faking styles (H2), and that the faking style is stable across different personality facets (H3). Moreover, it is hypothesized that extreme fakers have significant lower personality scores in neuroticism (H4a) and significant higher scores in extraversion (H4b), openness (H4c), agreeableness (H4d) and conscientiousness (H4e). Slight and extreme fakers should also differ in their intelligence score with extreme fakers having higher values in reasoning (H5). Finally, extreme fakers should be rather male and younger whereas slight fakers should be rather female (H6) and older (H7).

3.2 Method

3.2.1 Procedure

The data were collected in a German university. Participants were randomly assigned to one of two groups: A control group, who had to fill out a test of general personality two times honestly and an experimental group, which had to fill out the personality test honestly and under a faking instruction at time two. Participants of both groups also worked on a cognitive ability test and on several other tests, like a lexical knowledge test, which are not reported in this study. The computer versions of all tests were used.

3.2.2 Participants

Participants in this study were $N = 326$ undergraduate psychology students of the Ludwig-Maximilians-University (LMU) in Munich. Due to technical problems with the computer versions of the tests, data sets were lost so that only the results of 312 participants were examined (304 as far as cognitive abilities are investigated). 247

participants (79%) were female. Mean age was 25 years ($SD = 5.5$) with a range from 20.5 to 53.

The control group consisted of 157 students (121 female) and the experimental group of 155 participants (126 female). All students received study participation credits for their participation in the study.

3.2.3 Measures

3.2.3.1 Instruments

Personality was assessed with the NEO-PI-R (Costa & McCrae, 1992) in a German adaptation (Ostendorf & Angleitner, 2004), which allows a comprehensive assessment of general personality. The NEO-PI-R contains 240 items, measuring the five factor model (FFM): neuroticism, extraversion, openness, agreeableness, and conscientiousness. Respondents with a minimum age of 16 can rate their statements in the questionnaire on a five-point Likert-type scale of endorsement, ranging from *strongly disagree* to *strongly agree*.

Cognitive ability was tested with the help of the basic module of the Intelligence Structure Test 2000 R (Amthauer, et al., 2001). This module consists of 180 items measuring in nine subtests verbal, figural, and numerical reasoning. Combined, the verbal, figural and numerical score build the reasoning score (see Beauducel, et al., 2001, for detail concerning the theoretical basis and factor structure).

3.2.3.2 Instruction

The NEO-PI-R was given to the respondents with different instructions. In the honest condition (control group), respondents were only asked to fill out the questionnaires using the regular instructions, whereas in the faking condition (experimental group), fake-good instructions were used at the second time. Due to the

advice of Rogers (1997) to use a realistic scenario with a warning against too obvious faking and the fact that the participants were students in their first semester, the following university related faking instruction was used:

Universities have to select their students. For this task a number of instruments like the following are being tested right now. Please imagine that you are participating in a student selection procedure. Of course, it is your goal to get an admission as a psychology student. Therefore, you have to fill out the following questionnaire in a way that assures your admission. However, you have to be careful since a test expert will check the results for obvious faking and you do not want to be spotted.

3.2.3.3 *Statistical analyses*

SPSS 17.0 is used to calculate the internal consistency for the personality and the intelligence test.

To calculate personality scores of participants, negative formulated items will be recoded in accordance with the manual of the NEO-PI-R.

To avoid response bias concerning the middle category of the five-point Likert-type scale of the NEO-PI-R, the scale will be collapsed into a four-point Likert-type scale for all IRT analyses as proposed by Rost, Carstensen and von Davier (1999).

To find the faked facets and to make sure that higher ratings in the faking condition are caused by socially desirable responding and not by high “real” personality traits of the respondents, an Analysis of Covariance (ANCOVA) will be calculated for each facet with honest answers as covariate.

Furthermore, the computer program Windows Mixed Rasch Model Analysis (WINMIRA) 2001 (Davier, 2001) will be used to identify distinct subpopulations, like honest / regular responders or fakers. WINMIRA requires that the frequency of the lowest answering category of all items is unequal zero, meaning that for every item at least one participant has to choose the lowest answering possibility of the scale named *strongly disagree*. If this is not the case, all items of the affected facet (eight items)

will be reverse-coded. The different conditions (honest group / experimental group) will be combined into one dataset to determine overlap in classes across conditions in conformance with Zickar, Gibby, and Robie (2004). This dataset will be used to calculate different class analyses for each facet of the NEO-PI-R with the help of WINMIRA. Class solutions will be calculated until all information criteria rise again. In line with other studies using Rasch Models to detect subgroups even in latent classes (Eid & Rauber, 2000; Rost, et al., 1999), the partial credit model (PCM; Masters, 1982) will be used. Profiles of item locations of the two- and three- class solutions will be analyzed to check if the content of items caused different classes (Rost, et al., 1999). Moreover, plots of option thresholds for each facet will be examined, searching for items with estimation problems. In both cases, items causing problems will be eliminated. If less than five items per facet remain, the whole facet will be eliminated, because scale length is an important factor in the accurate identification of classes within WINMIRA (Zickar & Burnfield, 2003). For these corrected facets, a second run of class analyses will be conducted. The class where an information criteria fits and which is clearly interpretable (class size, option thresholds, mean, content), will be chosen to determine the number of classes needed to fit the data (for an overview of information criteria see Bozdogan (1987) or Read and Cressie (1988)). All interpretable facets and classes will be coded with a dichotomous indicator variable (0 = regular responders / slight fakers; 1 = extreme fakers) and a latent class analysis (LCA) will be carried out to check whether the response style is consistently used across facets.

Finally, individual differences between respondents using different faking styles will be explored. ANOVAs will be used to identify differences in cognitive ability and age. To identify individual differences in personality factors and facets, ANCOVAs

with post-hoc tests will be used. Why? Because the “true” personality trait measured might be influenced by the response style of a person. For example, two respondents with the same “true” level of neuroticism might have different scores in a personality test only due to the fact that one disproportionately favours extreme categories of Likert-type scales irrespective of particular item content, whereas the other favours middle categories. To reassure that this response style does not influence empirical differences between respondents of different faking classes in personality traits, the effect of the response style will be controlled for (covariate in the ANCOVA). Thus, differences in personality traits of respondents using different faking styles can be detected regardless of individual response styles (middle / extreme crossing). For the analyses and calculation of the indicator variable describing the response style see last study. Due to the fact that the hypotheses concerning differences in personality scores are directed, no Bonferroni correction will be made. For the post-hoc tests, effect sizes (Hedge’s g) will be calculated. According to the conventions of Cohen (1988) a Hedge’s g of .20 describes a small, .50 a moderate and a Hedge’s g of .80 describes a large effect. Differences in gender between the faking classes will be tested using a χ^2 -test.

3.3 Results

As can be seen in Table 9, the internal consistencies of the NEO-PI-R facets were ranging from $\alpha = .45$ to $\alpha = .87$ in the honest group and from $\alpha = .42$ to $\alpha = .83$ in the faking group. Cronbachs Alphas for factors ranged from $\alpha = .59$ to $\alpha = .88$ and $\alpha = .68$ to $\alpha = .90$, respectively. Hence, not all internal consistencies were acceptable, but comparable to those of the test-handbook (Costa & McCrae, 1992). Cronbachs Alphas for the intelligence test IST-2000-R were between $\alpha = .88$ and $\alpha = .95$. The

experimental group (faking condition) had higher means than the honest group in almost all facets – as far as neuroticism is concerned: lower means.

3.3.1 The occurrence of socially desirable responding

Hypothesis 1 stated that respondents fake under a faking instruction. To identify the faked facets and to verify that higher means in personality facets in the faking condition are caused by socially desirable responding and not by higher personality traits, ANCOVAs were calculated for each facet.

As Table 8 shows, there was a significant effect of socially desirable responding on personality facets in the faking condition after controlling for the “real” personality traits (honest condition). Significant moderate to large effects were found for 22 personality facets ($p < .001$; $\eta^2 > .08$; $1-\beta = 1$) and significant small effects were observed in six facets ($p < .01$; $\eta^2 < .04$; $1-\beta > .52$). Only in two facets (O2, O6) no such significant influence could be found.

Table 8
Testing the occurrence of socially desirably responding by using ANCOVAs

Facet	Facet faked?	<i>F</i> -Value ^a	<i>P</i>	<i>partial eta</i> ²	<i>1-β</i>
A1	yes	44.2	<.001	.125	1.00
A2	yes	28.2	<.001	.084	1.00
A3	yes	84.2	<.001	.214	1.00
A4	yes	102.0	<.001	.248	1.00
A5	yes	7.0	.009	.022	.75
A6	yes	13.2	<.001	.041	.95
C1	yes	317.1	<.001	.506	1.00
C2	yes	180.8	<.001	.369	1.00
C3	yes	298.4	<.001	.491	1.00
C4	yes	234.3	<.001	.431	1.00
C5	yes	412.5	<.001	.572	1.00
C6	yes	219.5	<.001	.415	1.00
E1	yes	60.9	<.001	.165	1.00
E2	yes	31.6	<.001	.093	1.00
E3	yes	177.1	<.001	.364	1.00
E4	yes	38.6	<.001	.111	1.00
E5	yes	4.1	<.001	.013	.52
E6	yes	7.5	<.001	.024	.78
N1	yes	260.0	<.001	.457	1.00
N2	yes	318.0	<.001	.507	1.00
N3	yes	222.6	<.001	.419	1.00
N4	yes	145.8	<.001	.321	1.00
N5	yes	241.3	<.001	.439	1.00
N6	yes	361.8	<.001	.539	1.00
O1	yes	115.8	<.001	.273	1.00
O2	no	0.8	.382	.002	.14
O3	yes	7.1	.008	.022	.75
O4	yes	18.2	<.001	.056	.99
O5	yes	203.1	<.001	.397	1.00
O6	no	3.7	.056	.012	.48

Notes. ^a *df* (1;309); N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-conscientiousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation. All facets in bold letters were not faked.

Table 9
Scale characteristics for NEO-PI-R and IST-2000-R scales

F	α	M	SD	F	α	M	SD	F	α	M	SD	F	α	M	SD	F	α	M	SD
Control Group / Honest Condition ^a																			
N	.88	93.4	26.1	E	.75	117.8	18.5	O	.59	129.4	14.8	A	.80	117.7	17.5	C	.84	119.3	20.3
N1	.86	17.3	6.2	E1	.66	23.8	3.3	O1	.72	21.4	4.3	A1	.75	20.1	4.3	C1	.74	21.9	3.9
N2	.73	14.2	4.8	E2	.76	20.1	4.9	O2	.75	23.6	4.6	A2	.66	18.0	4.5	C2	.75	18.5	4.9
N3	.87	13.8	6.3	E3	.83	16.1	5.7	O3	.81	24.6	4.1	A3	.65	23.8	3.4	C3	.75	22.0	4.4
N4	.81	16.8	5.6	E4	.67	17.8	4.2	O4	.62	17.6	4.2	A4	.65	17.2	4.3	C4	.66	20.6	3.9
N5	.62	17.9	4.3	E5	.60	17.2	4.9	O5	.82	21.5	5.1	A5	.76	16.9	4.7	C5	.83	19.0	5.1
N6	.84	13.6	5.5	E6	.80	22.9	4.6	O6	.45	20.7	3.3	A6	.65	21.7	3.5	C6	.78	17.3	4.8
Experimental Group / Faking Condition ^b																			
N	.90	51.5	18.5	E	.75	132.1	14.3	O	.68	130.3	13.9	A	.74	129.7	13.7	C	.90	154.1	16.5
N1	.77	9.1	4.0	E1	.72	26.3	3.0	O1	.72	16.1	4.4	A1	.79	22.7	3.7	C1	.74	27.2	3.0
N2	.71	7.6	3.6	E2	.75	22.7	3.7	O2	.82	23.7	4.1	A2	.57	20.3	3.6	C2	.55	23.6	3.1
N3	.81	6.7	4.0	E3	.73	22.4	3.7	O3	.80	23.2	3.9	A3	.73	26.3	3.0	C3	.71	27.3	2.9
N4	.66	11.2	3.8	E4	.53	20.2	3.0	O4	.47	19.6	3.0	A4	.47	20.6	3.3	C4	.72	25.8	3.5
N5	.59	11.3	3.6	E5	.66	16.7	4.1	O5	.82	27.3	3.6	A5	.68	16.5	4.0	C5	.83	27.1	3.6
N6	.82	5.7	3.5	E6	.80	23.9	3.8	O6	.42	20.4	3.0	A6	.58	23.3	3.0	C6	.78	23.1	4.0
All participants ^c																			
Reasoning					.95	114.0	23.8												
Verbal Reasoning					.85	38.4	7.8												
Numeric Reasoning					.94	40.8	11.0												
Figural Reasoning					.88	34.8	9.2												

Notes. ^a $n = 157$; ^b $n = 155$; ^c $n = 304$; F = Factor / facet of the NEO-PI-R; α = Cronbachs Alpha; N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-consciousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation.

3.3.2 Searching for Subgroups

All facets of the NEO-PI-R, in which no participant used the lowest category *strongly disagree* to answer a single item, were recoded to enable the usage of WINMIRA. Altogether, 12 facets from 4 factors had to be recoded (see Table 10).

Table 10
Recoded facets of the NEO-PI-R

Neuroticism	Extraversion	Openness to Experience	Agreeableness	Conscientiousness
N1 Anxiety	E1 Warmth	O1 Fantasy	A1 Trust	C1 Competence
N2 Angry hostility	E2 Gregariousness	O2 Aesthetics	A2 Straight- forwardness	C2 Order
N3 Depression	E3 Assertiveness	O3 Feelings	A3 Altruism	C3 Dutifulness
N4 Self- conscientiousness	E4 Activity	O4 Actions	A4 Compliance	C4 Achievement striving
N5 Impulsiveness	E5 Excitement seeking	O5 Ideas	A5 Modesty	C5 Self-discipline
N6 Vulnerability	E6 Positive emotions	O6 Values	A6 Tender- mindedness	C6 Deliberation

Notes. Black coloured facets had to be recoded.

3.3.2.1 *Winmira 1st Run*

Hypothesis 2 stated that respondents fake to a different amount showing different faking styles / classes. To identify these classes, different class analyses were calculated using WINMIRA. Analyzing the results of the class analyses, four facets showed one-class solutions, nine facets three-class solutions and for the remaining 17 facets, two classes were needed to fit the data (see Table 11 for an overview). Plots of the item locations for the two- and three-class solutions showed that participants interpreted some items differently. Accordingly, not all item locations were (almost) parallel and items had to be removed from facets, like in facet O6 (see Figure 9).

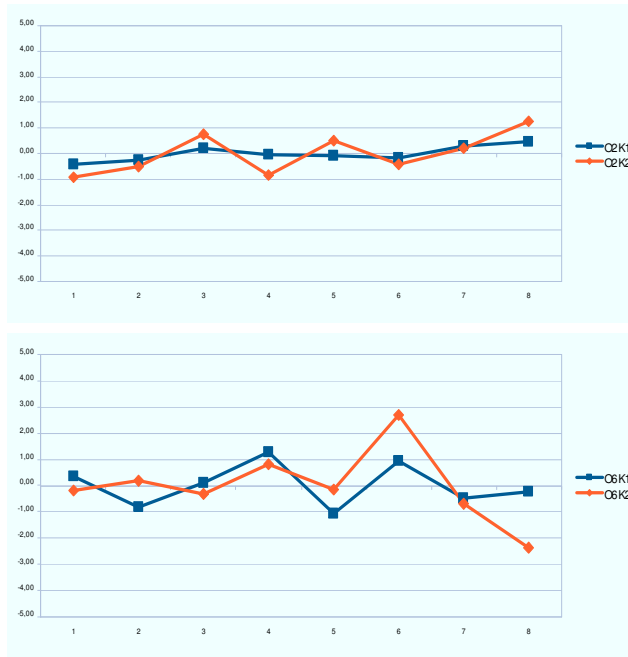


Figure 9
 (Almost) parallel item locations for O2; variations in item locations for O6

Examining the plots of option thresholds for all facets, two generalizations can be made: In eight facets of the two-class solutions threshold estimates were interpretable, meaning that thresholds had (almost) a correct ordering, class sizes were not indicating a small group of outliers and mean differences were interpretable (see Figure 10 for an example). But estimating problems (item locations > |6|) made an interpretation in some facets impossible (see Figure 11 for an example). In such cases, items causing estimating problems were removed and WINMIRA analyses were repeated.

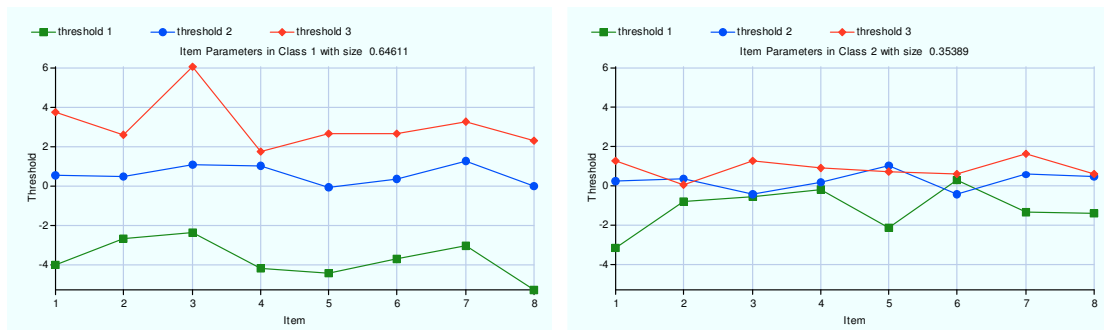


Figure 10
 Threshold estimates for facet E5 class 1 & 2 – an interpretable example

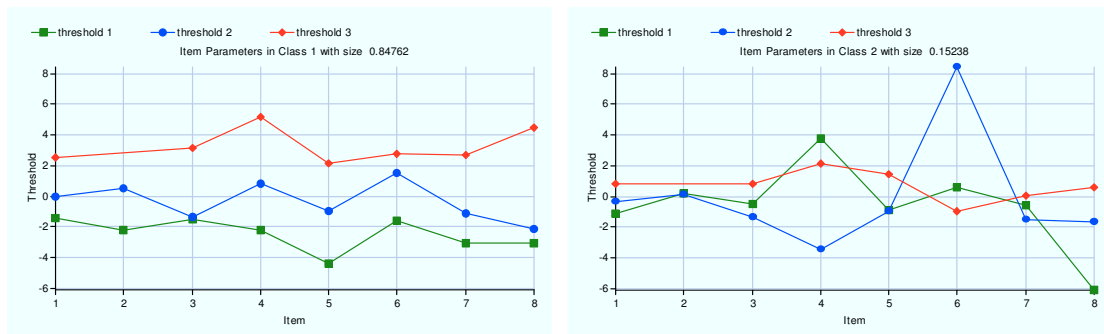


Figure 11

Threshold estimates for facet O6 class 1 & 2 – an example with estimation problems

3.3.2.2 Winmira 2nd Run

After eliminating items due to estimation problems or due to difficulties in item interpretation, one-, two-, and three-class solutions were calculated again using WINMIRA. This time, three facets showed a one-class solution, five facets a three-class solution and 13 facets a two-class solution (see Table 11 and Table 12 for an overview). 5 out of these 13 facets with two class-solutions according to information criteria had classes clearly interpretable as honest / slight faking classes and extreme faking classes. In the other eight facets with two-class solutions estimating problems could not be solved finally. Thus, eight facets from the first run and five facets of the second WINMIRA run were included in the next step – the latent class analysis.

Table 11
Class solutions for NEO-PI-R scales

WINMIRA 1st and 2nd run						
Facet	Nr class 1st run	Item- location 1st run	Thres- holds 1st run	Nr class 2nd run	Item- location 2nd run	Thres- holds 2nd run
A1	2	parallel	not clear	3	parallel	almost clear
A2	3	parallel	clear-cut	3	almost fit	almost clear
A3	1	parallel	clear-cut	1	almost fit	clear-cut
A4	2	parallel	clear-cut	-	-	-
A5	3	variations	not clear	3	parallel	clear-cut
A6	2	parallel	clear-cut	-	-	-
C1	2	almost fit	clear-cut	-	-	-
C2	3	parallel	not clear	2	parallel	clear-cut
C3	2	variations	clear-cut	-	-	-
C4	2	variations	not clear	-	-	-
C5	1	parallel	clear-cut	1	parallel	clear-cut
C6	2	parallel	not clear	2	almost fit	not clear
E1	1	parallel	not clear	2	parallel	clear-cut
E2	2	almost fit	clear-cut	-	-	-
E3	3	variations	not clear	2	parallel	not clear
E4	3	variations	not clear	2	parallel	clear-cut
E5	2	parallel	clear-cut	-	-	-
E6	2	variations	not clear	2	almost fit	not clear
N1	3	parallel	not clear	2	parallel	clear-cut
N2	2	parallel	not clear	3	almost fit	clear-cut
N3	2	parallel	clear-cut	-	-	-
N4	3	parallel	clear-cut	3	almost fit	almost clear
N5	3	variations	clear-cut	2	parallel	not clear
N6	2	variations	not clear	1	almost fit	clear-cut
O1	3	parallel	clear-cut	2	parallel	clear-cut
O2	2	parallel	not clear	2	parallel	not clear
O3	2	variations	clear-cut	-	-	-
O4	2	parallel	not clear	2	parallel	not clear
O5	1	variations	not clear	2	almost fit	not clear
O6	2	variations	not clear	2	almost fit	clear-cut

Notes. N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-conscientiousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation. All facets in bold letters were used in the following latent class analysis; the information criteria AIC was used.

3. SDR and individual differences

Table 12
Information criteria for each facet and each class of WINMIRA 1st and 2nd run

Class	A1			A2			A3			A4			A5			A6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	3534.76	3665.77	3700.77	4480.57	4634.03	4675.03	3900.53	4053.99	4094.99	5010.76	5186.68	5233.68	3442.07	3573.07	3608.07	4594.59	4770.51	4817.51
K2	3527.65	3785.92	3854.92	4391.66	4694.84	4775.84	3911.67	4214.85	4295.85	4956.30	5304.39	5397.39	3426.89	3685.16	3754.16	4543.35	4891.45	4984.45
K3	3524.15	3909.68	4012.68	4375.53	4828.43	4949.43	3945.94	4398.85	4519.85	4991.72	5512.00	5651.00	3406.09	3791.62	3894.62	4575.94	5096.22	5235.22
K4	3570.12	4082.91	4219.91	4451.57	5054.19	5215.19							3477.85	3990.64	4127.64			
	C1			C2			C3			C4			C5			C6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	4412.53	4588.45	4635.45	4360.47	4513.94	4554.94	4415.30	4591.22	4638.22	4842.52	5018.44	5065.44	3819.84	3973.30	4014.30	4016.67	4170.14	4211.14
K2	4374.87	4722.96	4815.96	4309.57	4612.75	4693.75	4375.01	4723.11	4816.11	4744.12	5092.22	5185.22	3831.85	4135.03	4216.03	3986.10	4289.28	4370.28
K3	4381.47	4901.75	5040.75	4326.80	4779.70	4900.70	4386.48	4906.76	5045.76	4767.88	5288.16	5427.16	3848.65	4301.55	4422.55	4009.53	4462.43	4583.43
	E1			E2			E3			E4			E5			E6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	3387.19	3518.19	3553.19	4959.02	5134.94	5181.94	3119.37	3227.92	3256.92	3718.99	3849.99	3884.99	5361.88	5537.80	5584.80	4140.86	4294.32	4335.32
K2	3384.37	3642.64	3711.64	4904.00	5252.10	5345.10	3116.94	3330.29	3387.29	3698.41	3956.68	4025.68	5285.91	5634.01	5727.01	4098.79	4401.97	4482.97
K3	3398.68	3784.21	3887.21	4991.85	5512.13	5651.13	3120.94	3439.10	3524.10	3728.82	4114.35	4217.35	5303.37	5823.65	5962.65	4131.13	4584.04	4705.04
	N1			N2			N3			N4			N5			N6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	3256.10	3387.11	3422.11	3858.56	4012.02	4053.02	4357.81	4533.73	4580.73	3547.63	3678.64	3713.64	3678.22	3809.22	3844.22	2587.80	2696.35	2725.35
K2	3241.51	3499.78	3568.78	3842.13	4145.31	4226.31	4298.19	4646.29	4739.29	3504.76	3763.03	3832.03	3610.37	3868.64	3937.64	2611.57	2735.09	2768.09
K3	3253.80	3639.32	3742.32	3824.32	4277.22	4398.22	4312.78	4833.06	4972.06	3460.82	3846.35	3949.35	3647.87	4033.40	4136.40	2646.78	2964.93	3049.93
K4				3839.79	4442.41	4603.41				3490.59	4003.38	4140.38						
	O1			O2			O3			O4			O5			O6		
	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC	AIC	BIC	CAIC
K1	3746.62	3877.63	3912.63	3140.77	3249.32	3278.32	4520.13	4696.05	4743.05	3740.47	3871.48	3906.48	3024.27	3132.82	3161.82	3922.47	4053.48	4088.48
K2	3654.88	3913.15	3982.15	3116.15	3329.51	3386.51	4500.31	4848.41	4941.41	3715.56	3973.83	4042.83	3017.06	3230.41	3287.41	3905.17	4163.44	4232.44
K3	3668.75	4054.28	4157.28	3157.03	3475.19	3560.19	4535.96	5056.24	5195.24	3728.60	4114.13	4217.13	3127.60	3445.76	3530.76	3932.02	4317.55	4420.55

Notes. K1 = one-class solution; K2 = two-class solution; K3 = three-class solution; K4 = four-class solution; grey facets indicate results from WINMIRA 1st run; black facets indicate results from WINMIRA 2nd run; bold numbers indicate the best solution according to the particular information criteria.

3.3.2.3 Latent Class Analysis

To test whether socially desirable responding is scale specific or a response set throughout all facets (H3), a latent class analysis was executed. As Table 13 shows, three classes were needed to fit the data according to the BIC and CAIC and five according to the AIC.

Table 13
Latent class analysis with 13 facets of the NEO-PI-R

model	class nr.	class size	sum score	Information Criteria		
				AIC	BIC	CAIC
1	1	1.00	5.25	5028.01	5076.67	5089.67
2	1	0.67	4.30	4839.51	4940.57	4967.57
	2	0.33	7.18			
3	1	0.49	4.17	4754.87	4908.34	4949.34
	2	0.29	5.13			
	3	0.22	7.88			
4	1	0.37	4.36	4722.98	4928.84	4983.84
	2	0.23	5.45			
	3	0.23	4.31			
	4	0.16	8.31			
5	1	0.25	5.49	4707.36	4965.63	5034.63
	2	0.24	4.47			
	3	0.21	5.83			
	4	0.20	3.27			
6	5	0.12	8.69	4711.11	5021.78	5104.78
	1	0.25	4.43			
	2	0.23	5.54			
	3	0.18	3.18			
	4	0.16	5.82			
	5	0.12	8.68			
	6	0.07	5.50			

Notes. model = number of classes chosen in the LCA; sum score = sum of the means of the facets included in the LCA; bold class solutions indicate the best solution according to the particular criteria.

Due to the fact that five classes were not interpretable with regard to the content and item locations, the three-class solution will be used. The estimated mean probabilities of assigning a person to the right class were 89.8% for class 1, 81.2% for class 2, and 90.8% for class 3. All probabilities of assigning a person to a wrong class were below 11.4%. So class assignment can be regarded as good. The data used in the LCA were dichotomous variables indicating regular responders / slight fakers (coded with 0) and extreme fakers (coded with 1). Hence, the class with clearly the highest sum score in the

LCA can be described as the class of the extreme fakers. Two classes with sum scores, which are closer, remain. Due to the fact that in the beginning no differentiation could be made between honest responders and slight fakers, the second class with slightly higher means seems to be the class where respondents fake slightly and consistent. The last class with the lowest sum score can therefore be seen as class in which respondents answer honestly or fake – they switch between faking and not faking. Accordingly, the three classes of the LCA can be described as regular responders / fakers (switchers), slight fakers and extreme fakers. This result is confirmed by analyzing the item locations of the three-class solution of the LCA: Whereas the item profiles of two classes (class 2 and 3 in Figure 12) are parallel with some exceptions, the item profile of the first class is either going its own way (regular responding) or identical with one of the other classes (slight and extreme faking).

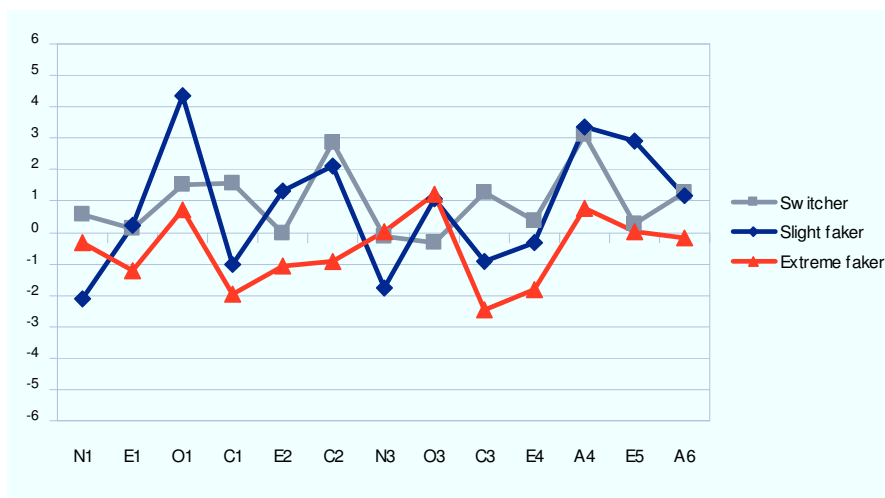


Figure 12

Item locations for the switcher (grey), slight fakers (blue) and extreme fakers (red) of the three-class solution of 13 facets in the LCA

N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; E2 = Gregariousness; C2 = Order; N3 = Depression; O3 = Feelings; C3 = Dutifulness; E4 = Activity; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A6 = Tender-mindedness.

As Table 14 shows, 83% of the respondents of the honest condition are in class one, which can be interpreted as switching class (regular responding / faking), whereas 83%

of the faking condition are in classes two and three – the slight and extreme faking classes. This distribution also supports the interpretation of the classes.

Table 14
Distribution LCA classes across experimental groups

experimental group	class	frequency	percentage	cumulative percentage
honest condition	1	130	82.8	82.8
	2	12	7.6	90.4
	3	15	9.6	100.0
	sum	157	100.0	
faking condition	1	26	16.8	16.8
	2	82	52.9	69.7
	3	47	30.3	100.0
	sum	155	100.0	

Notes. Class 1 = switcher; class 2 = slight faker; class 3 = extreme faker.

3.3.3 Socially desirable responding and individual differences

Before testing for individual differences between the three LCA classes in personality factors and facets (H4), intelligence (H5), gender (H6), and age (H7), a possible correlation of the faking style and the response style will be examined using a χ^2 -test. Results indicate a correlation between the response style (middle / extreme crosser) and the faking style (switcher / slight faker / extreme faker): Half of the middle and extreme crossers were in the switching class. But whereas the other half of the extreme crossers were slightly more in the extreme faking class, more middle crossers were in the slight faking class than in the extreme faking class ($\chi^2(2) = 22.70, p < .001$).

Table 15
Distribution of respondents across response sets and styles

response style	response set			sum
	switcher	slight faker	extreme faker	
middle crosser	111	74	27	212
extreme crosser	45	20	35	100
sum	156	94	62	312

To examine individual differences in personality traits between the faking classes and to control for the response style, ANCOVAs were calculated. The dependent

variables were the scores of the personality test in the honest setting; the fixed factor was the indicator variable for faking and the covariate the indicator variable for the response style.

As Table 18 shows, respondents of the switcher class differ from slight and extreme fakers in the personality factors neuroticism, agreeableness, and conscientiousness significantly and with small to moderate effect sizes – and in at least one personality facet of each personality factor. Main differences were found between switchers and extreme fakers in the personality facets of conscientiousness (e.g. competence: $F(2;308) = 9.38; p < .001; \text{Hedge's } g = .72$).

Extreme fakers have in all personality factors and in nearly all personality facets lower mean scores in neuroticism and higher scores in extraversion, openness, agreeableness, and conscientiousness than slight fakers. The only exceptions are the facets N5, E2, E3, and O6 where the personality scores of the slight and extreme fakers are almost identical. The effect for the personality factor agreeableness is significant and almost moderate ($F(2;308) = 3.87; p < .05; \text{Hedge's } g = .44$). The effects for the personality facets openness for actions (O4; $F(2;308) = 2.69; p > .05; \text{Hedge's } g = .42$), straightforwardness (A2; $F(2;308) = 2.65; p > .05; \text{Hedge's } g = .42$), and altruism (A3; $F(2;308) = 3.04; p < .05; \text{Hedge's } g = .48$) are also significant and almost moderate, whereas the effect of compliance (A4; $F(2;308) = 8.04; p < .001; \text{Hedge's } g = .29$) is rather small. As far as conscientiousness is concerned, differences between slight and extreme fakers are not significant. However, almost moderate effect sizes suggest a meaningful difference in the personality facets competence (C1; $\text{Hedge's } g = .40$) and achievement striving (C4; $\text{Hedge's } g = .43$) as well as in the personality factor itself (C; $\text{Hedge's } g = .37$).

Significant class differences concerning reasoning ($F(1;301) = .83; p > .05; \text{partial } \eta^2 < .01; 1-\beta = .19$), verbal ($F(1;301) = 2.68; p > .05; \text{partial } \eta^2 = .02; 1-\beta = .53$),

numeric ($F(1;301) = 1.64; p > .05; \text{partial } \eta^2 = .01; 1-\beta = .35$) or figural reasoning ($F(1;301) = .48; p > .05; \text{partial } \eta^2 < .01; 1-\beta = .13$), or age ($F(2;309) = 1.70; p > .05; \text{partial } \eta^2 = .01; 1-\beta = .36$) were not found.

As can be seen in Table 16 half of the male and female respondents were in the switching class. The other male respondents were equally distributed across the slight and extreme faking class. More women were gathered in the slight faking class than in the extreme faking class, although this difference was not significant ($\chi^2(2) = .63; p > .05$). Moreover, the same proportion of men and women faked in the honest condition (16% / 17%), and approximately the same proportion switched in the faking condition (14% / 17%).

Table 16
Gender distinguished by LCA class

gender	class according LCA			sum
	switcher	slight faker	extreme faker	
male	34	17	14	65
female	122	77	48	247
sum	156	94	62	312

Table 17
Gender distinguished by LCA class and experimental condition

experimental group	gender	switcher	slight faker	extreme faker
honest	male	84%	8%	8%
	female	83%	7%	10%
fake	male	14%	48%	38%
	female	17%	54%	29%

Table 18

ANCOVAs indicating differences in personality facets and factors between LCA classes (controlled for response style)

	<i>F</i>	<i>partial eta</i> ²	1-β	<i>M</i> sw	<i>SD</i> sw	<i>M</i> sf	<i>SD</i> sf	<i>M</i> ef	<i>SD</i> ef	<i>g</i> sw-sf	<i>g</i> sw-ef	<i>g</i> sf-ef
N	2.55	.02	.51	95.81	24.91	92.56	22.69	86.34	26.65	.13	.37*	.26
E	2.51	.02	.50	117.25	17.88	121.13	20.02	124.92	20.74	-.21	-.41	-.19
O	.60	.00	.15	129.82	13.79	128.43	17.02	133.10	21.91	.09	-.20	-.24
A	3.87*	.02	.70	118.72	17.09	112.73	17.36	120.23	16.37	.35**	-.09	-.44*
C	15.51***	.09	1.00	115.40	48.52	124.19	15.81	130.77	20.95	-.22***	-.36***	-.37
N1	2.19	.01	.45	17.61	6.15	17.55	5.95	15.47	6.12	.01	.35*	.35
N2	1.58	.01	.33	14.27	4.68	14.50	4.94	13.13	5.53	.05	.23	.26
N3	3.27*	.02	.62	14.44	6.11	13.03	5.70	12.08	6.32	-.24	.38*	.16
N4	1.45	.01	.31	17.09	5.21	16.49	4.64	15.50	5.77	.12	.30	.19
N5	.56	.00	.14	18.13	4.34	17.56	3.50	17.73	5.17	.14	.09	-.04
N6	2.18	.01	.44	14.25	5.20	13.43	5.34	12.44	5.16	.16	.35*	.19
E1	.64	.00	.16	23.68	3.34	23.61	4.37	24.71	4.40	.02	-.28	-.25
E2	.50	.00	.13	20.45	4.70	20.68	5.04	20.32	5.45	-.05	.03	.07
E3	3.43*	.02	.64	15.97	5.45	17.64	5.36	17.34	5.50	-.31**	-.25	.06
E4	6.65***	.04	.91	17.37	3.85	18.84	4.28	19.76	5.00	-.37**	-.57**	-.20
E5	.90	.01	.21	17.02	4.77	17.43	4.61	18.53	5.66	-.09	-.30	-.22
E6	.96	.01	.22	22.80	4.21	22.93	5.08	24.26	5.62	-.03	-.31	-.25
O1	.11	.00	.07	21.50	4.09	21.28	4.61	21.32	6.03	.05	.04	-.01
O2	1.19	.01	.26	24.10	4.32	23.16	5.22	24.10	6.53	.20	-.00	-.16
O3	.34	.00	.10	24.58	3.92	24.18	3.87	24.98	4.88	.10	-.09	-.19
O4	2.69	.02	.53	17.54	3.96	17.49	4.30	19.35	4.50	.01	-.44*	-.42*
O5	.38	.00	.11	21.30	4.99	21.27	4.62	22.44	6.36	.01	-.21	-.22
O6	.27	.00	.09	20.81	3.38	21.05	2.99	20.90	3.89	-.07	-.03	.04
A1	1.91	.01	.40	20.17	4.29	18.96	4.32	20.13	4.87	.28*	.01	-.26
A2	2.65	.02	.52	18.20	4.52	17.30	4.46	19.18	4.59	.20	-.22	-.42*
A3	3.04*	.02	.59	23.67	6.36	23.35	3.83	25.21	3.97	.06	-.27*	-.48*
A4	8.04***	.05	.96	17.46	4.14	15.30	4.38	16.52	4.07	.51***	.23	-.29*
A5	.74	.01	.18	17.20	4.67	16.46	4.92	16.87	4.32	.16	.07	-.09
A6	.92	.01	.21	22.03	3.20	21.37	3.71	22.32	3.33	.19	-.09	-.27
C1	9.38***	.06	.98	21.37	3.63	22.55	3.51	23.97	3.58	-.33**	-.72***	-.40
C2	9.85***	.06	.98	17.86	4.78	19.62	4.13	20.81	5.01	-.39**	-.61***	-.26
C3	7.08***	.04	.93	21.45	4.16	22.71	3.62	23.84	4.36	-.32**	-.57***	-.29
C4	8.34***	.05	.96	19.91	3.87	20.88	3.70	22.56	4.21	-.25*	-.67***	-.43
C5	7.85***	.05	.95	18.13	4.57	19.93	4.66	20.94	5.80	-.39**	-.57***	-.20
C6	5.00**	.03	.81	16.77	4.86	18.17	4.63	18.66	5.76	-.29**	-.37**	-.10

Notes. $n_{sw} = 156$; $n_{sl} = 94$; $n_{ef} = 62$; * $p < .05$, ** $p < .01$, *** $p < .001$; sw = switching class; sl = slight faker class; ef = extreme faker class; $F = F$ -value with $df(2;308)$; $g =$ Hedge's g ; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; N2 = Angry hostility; E2 = Gregariousness; O2 = Aesthetics; A2 = Straightforwardness; C2 = Order; N3 = Depression; E3 = Assertiveness; O3 = Feelings; A3 = Altruism; C3 = Dutifulness; N4 = Self-conscientiousness; E4 = Activity; O4 = Actions; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; E6 = Positive emotions; O6 = Values; A6 = Tender-mindedness; C6 = Deliberation.

3.4 Discussion

The goal of this study was to examine whether people using different faking styles differ in personality factors and facets, intelligence, age or gender – regardless of their response style.

3.4.1 The occurrence of socially desirable responding (SDR)

To answer the question whether people using different faking styles differ, socially desirable responding must take place. To reassure that higher means in the faking condition are caused by SDR and not by high personality traits of the person, an ANCOVA was used. Results show that respondents instructed to answer a global measure of personality in a socially desirable way faked all personality facets with the exception of openness for aesthetics (O2) and openness for values (O6). This is in line with earlier studies, which showed that openness is a factor often less faked (Furnham, 1997; McFarland & Ryan, 2000; Ziegler, 2007). Accordingly, SDR does not take place independent of scale content. Due to the fact that 28 out of 30 personality facets were faked hypothesis 1 – people fake when instructed to do so – can be verified.

3.4.2 The occurrence of different faking styles

Using Rasch / Mixed-Rasch Models, different faking styles could be found. However, it has to be mentioned, that different faking styles could not be found for every personality facet – probably for the same reasons as in the last study (see last study for a detailed explanation). Therefore hypothesis 2 – the emergence of different faking classes – could only be partly confirmed.

Did respondents use a constant faking style across those personality facets where different faking styles were found (H3)? Further analysis showed that 83% of the respondents in the faking condition engaged, in slight or extreme faking, whereas the

other 17% did not constantly use the same faking style across all personality facets, but switched between the faking classes or answered honestly. In contrast, in the honest condition, 83% of the respondents answered honestly / switched and only 17% were constantly slight or extreme faking across the personality facets. Zickar et al. (2004) found 7.2% to 22.9% of respondents in the honest condition to be in the faking class and a sizable amount of fakers in the honest class. Hence, the results of this study replicate that not all respondents instructed to fake do so and that not all people asked to answer honestly do so as well. However, due to the fact that 83% of the fakers did not change their faking style across the personality facets, hypothesis 3 – a constant faking style across personality facets – can be verified for the vast majority of respondents.

3.4.3 SDR and individual differences in personality factors / facets

Individual differences between regular responders / fakers (switchers), slight and extreme fakers were investigated (H4). To make sure that “real” personality traits of respondents are not contaminated by their response styles, it was controlled for. What personality traits make a switcher to a switcher? As results indicate, people with lower scores in conscientiousness have a higher probability to be switcher: The scores of switchers and slight fakers as well as the scores of switchers and extreme fakers differ in all facets of conscientiousness significantly and with remarkably effect sizes (moderate to strong for the differences between switchers and extreme fakers; small to moderate for differences between switchers and slight fakers). Moreover, switchers and extreme fakers differ in the personality facets of anxiety (N1), depression (N3), vulnerability (N6), activity (E4), openness for actions (O4), and altruism (A3). To sum up, switchers are less active than extreme fakers, concerning their hunger and openness for adventure (O4), their level of engagement for others (A3) and their activity level

itself (E4). Instead, they are more concerned having a higher level of anxiety (N1, N3, N6).

What is the main difference between people switching between regular responding / faking and (permanent) slight fakers? Switchers have significant lower scores in conscientiousness (C) and higher scores in agreeableness (A) by small to moderate effect sizes, indicating that switchers have a higher probability to compromise in interpersonal conflicts, repressing aggression (A4) and are able to trust others more easily (A1).

How do slight and extreme fakers differ? Slight and extreme fakers differ most significantly and with up to moderate effect sizes in their scores of straightforwardness (A2), altruism (A3), compliance (A4), and openness for actions (O4): Extreme fakers are more active concerning their care for others (A3), their willingness to compromise in interpersonal conflicts (A4), and concerning their hunger for adventure (O4). The nearly moderate effect sizes of competence (C1) and dutifulness (C3) indicate – although not significant – that extreme fakers tend to be a little bit more conscientious than slight fakers. Summing up, the less conscientious, the less active, and the more self-referential a person is, the higher is the probability that he / she switches between regular responding / faking. The more active a person is (in relation to activity, openness for actions or active help for others), the more likely the person is to engage in extreme faking. Due to the fact that the faking classes differ in at least one facet of each personality factor significantly and in the supposed direction, hypothesis 4 – respondents of different faking classes differ in all personality traits – was verified.

3.4.4 SDR and individual differences in fluid intelligence

According to hypothesis 5, slight and extreme fakers should differ in their reasoning scores with extreme fakers having higher scores. Extreme fakers had higher scores than

slight fakers in reasoning, but this difference was not significant. This might be due to the fact that the sample of students was partly selected by intelligence when selected for one of the restricted places at university. Therefore, unfortunately this hypothesis could not be verified, although the correlation of cognitive ability and faking has been shown before (Grubb & McDaniel, 2007; Pauls & Crost, 2005a) and extreme fakers showed already higher reasoning scores than slight fakers in other studies (Ziegler, 2007).

3.4.5 SDR and individual differences in age and gender

Extreme fakers tended to be younger, slight fakers tended to be rather female. The hypothesis of more male respondents being extreme fakers (H6) could not be confirmed and also the results concerning age are in the right direction, but not significant (H7). Due to the fact that this study consisted of a student sample, the variance concerning age might be too small to detect correlations between age and faking style. Moreover, 79% of the participants were female: Only 14 men were in the class of extreme fakers and only 17 in the slight faking class (and 34 were switchers). To achieve representative results concerning gender differences, gender should be more equally distributed, too.

3.4.6 Limitations

Due to the fact that personality facets are only measured with the help of eight items, estimation problems occurred in WINMIRA analyses, because not all rating categories were used to answer an item. For example, no respondent answered the item “I will never manage to bring structure in my life” by using the most extreme category *strongly disagree*. Moreover, items like “I keep myself up to date and normally the decisions I make are intelligent” were interpreted differently by respondents. These two effects yielded in the amount of only 13 facets, which could be used in the LCA to test whether the faking style is consistent across these personality facets.

Using a student sample also limited the results concerning differences in age and intelligence because of the small variance and the high homogeneity the student sample has in this variables. As the high majority of psychology students are female, male respondents were underrepresented, too.

3.4.7 Implications and future directions

This study not only confirms the existence of different faking styles and honest answers from respondents who were instructed to fake, but also shows that the vast majority of slight and extreme fakers fake their responses always in the same way. Moreover, slight and extreme fakers differ in personality facets and factors. Furthermore, extreme fakers tend to be younger and have higher scores in reasoning – while females tend to be rather slight fakers. Former results concerning differences in personality traits between respondents with different faking styles were scanty and conflicting. To some extent this might be due to the different methods used to detect fakers over the time, but also due to the fact that no former study controlled for the response style of the respondents (middle / extreme crossing).

Thus, this study sheds further light on the process of why people cross where they cross on non-cognitive self-report questionnaires with Likert-type scales. However, if these differences in personality traits cause the different faking styles or are only one more symptom has to be investigated further. On the one hand, results clearly indicate that extreme fakers have a higher level of different kinds of activity, indicating that differences in personality traits may be the reason for extreme fakers to distort their answers in a more socially desirable way on Likert-type scales. On the other hand, the fact that the different faking styles could not be found in all facets conflict with this assumption. Therefore, future research is necessary trying to prove whether the reason for not finding the faking styles in all facets is due to this unique combination of

methodical problems (estimation problems), homogenous sample and comparatively small Likert-type scale. If this holds true, differences in personality traits might be (the) one underlying factor for the response set – in combination with a higher level of reasoning. However, there is another point further research should examine: The existence of different faking classes in real applicant settings with a higher amount of male respondents and also controlling for response styles. Thus, the insights this study were able show would be confirmed and extended.

**4. Psychometric quality of a
questionnaire used in personnel assessment**

4 Psychometric quality of a questionnaire used in personnel assessment

4.1 Present Study

To test whether the psychometric quality of non-cognitive self-report questionnaires remain even when response sets and response styles are likely to occur, the psychometric quality of the first German Integrity Test is examined in a neutral situation, where response styles are probable and in a simulated applicant setting, where the situational pressure is likely to cause socially desirable responding.

4.1.1 What is an Integrity Test?

The goal of personnel assessment is to select such employees out of a pool of applicants who have high potential to show favourite behaviours like conscientiousness or teamwork. Therefore, tools are developed. In the USA, the other way round is practiced for years with great success: In the personnel selection process the probability of counterproductive acts like absenteeism, alcoholism, drug abuses or theft are tested with the help of integrity tests (Marcus & Schuler, 2004). What are integrity tests? Integrity tests are self-report questionnaires, which measure the probability of counterproductive working behaviour (CWB). Counterproductive working behaviours are “volitional acts by members of an organization that violate the legitimate interests of the organization or its individual members” (Marcus & Wagner, 2007; p. 161).

In the meantime meta-analyses were able to prove that integrity tests are not only valid instruments to measure counterproductive working behaviour (CWB), but also predict performance criteria. Schmidt and Hunter (1998) illustrated that integrity tests are able to explain more variance of the criteria job success (27%) than variables of conscientiousness (18%). Moreover, Ones, Viswesvaran and Schmidt (1993) were able to show with the help of 68,771 respondents that the corrected predictive validity of all

tested integrity tests is .34 for performance criteria measured with the help of supervisor ratings, awards, and production results.

The job-related attitudes and self-evaluations inventory (Inventar berufsbezogener Einstellungen und Selbsteinschätzungen, IBES; Marcus, 2006) is the first integrity test in German language (Hossiep & Bräutigam, 2007b). However, there is a lack of empirical results, which prove the potential of the IBES beside the manual. Therefore, this paper not only tests the psychometric quality of the IBES with the help of a real life setting, but also examines its value for the explanation of performance criteria variance beyond an intelligence and a personality test.

4.1.2 IBES – field of application

The aim of integrity tests is to identify respondents who have a higher probability to conduct counterproductive working behaviours (Marcus, et al., 2006). Therefore, the IBES is used for selection of applicants older than 16, but also for research in the field of organizational- and personality related psychology. According to Marcus et al. (2006) the test should not be the only selection tool, but used as a first step in stepwise personnel assessment and used before more complex and more expensive selection methods to reduce the pool of applicants.

4.1.3 History of Integrity Tests

In the USA integrity tests have a long and successful tradition: In 1989 more than 40 different integrity tests were already available (O'Bannan, Goldinger, & Appleby, 1989). How did they evolve? The kinds of instruments nowadays used and known as integrity tests began with an attempt by John Reid (Reid Report; Reid, 1957) to detect dishonesty in job applicants without having to use polygraph tests. Nowadays, integrity tests are no longer viewed as surrogates for polygraphs, but the focus typically remains on the prediction of counterproductive work behaviours (Berry, Sackett, & Wiemann,

2007). With the Reid Report (Reid, 1957), John Reid can be seen as the founder of overt integrity tests (Ash, 1991). What are overt integrity tests? In the history of integrity tests two categories of tests were developed, which were later on subsumed under the expression “integrity tests”. The first kind of integrity tests are called overt integrity tests (Sackett, Burris & Callahan, 1989) like the Reid Report (Reid, 1957), the Stanton Survey (Klump, 1964) or the Personnel Selection Inventory (LondonHouse, 1995). This kind of test is characterised by measuring the likelihood of counterproductive working behaviours based on responses to questions, which explicitly ask about counterproductive behaviour, not leaving any doubt regarding the construct the tests assesses (Bye & Holtgraves, 2008). An example for an item of an overt integrity test is: “Have you ever thought of stealing money from your workplace without doing it in reality?” (Marcus, et al., 2006). The second kind of integrity tests are personality based tools, called covert integrity tests (Sackett, Burris, & Callahan, 1989), like the Personnel Reaction Blank (Gough, 1971) or the Employment Inventory (Paajanen, 1986). Items of these kinds of tests are less transparent / overt, like “I am more sensible than adventurous” (Marcus, et al., 2006) and do not necessarily alert the test taker to what is being measured (Bye & Holtgraves, 2008). With the help of personality based items, which have a proven connection to counterproductive behaviour these kind of tests try to extract such applicants who are more likely to show CWB (Horn, Nelson, & Brannick, 2004).

Beside the detection of less favourable applicants the two kinds of integrity tests have another similarity: “...the focus in test construction is on predicting specific target criteria rather than on measuring theoretically founded personality constructs“ (Marcus, Lee, & Ashton, 2007; p. 2). Therefore, they are a member of the “criterion-focused occupational personality scales“ (COPS; Ones & Viswesvaran, 2001), a group of tests, which measures individual differences beyond the domain of cognitive abilities. In

contrast to well-defined traits and structural models of personality, these kind of tests are mainly constructed to predict important work-related criteria with the focus in test construction being on predicting target criteria rather than on measuring theoretically founded personality constructs (Marcus, et al., 2007).

4.1.4 Integrity and counterproductive working criteria

Counterproductive working behaviour (CWB) seems to decrease job performance and increase costs to the company (Sackett & DeVore, 2002). The negative effects of CWBs on organizational effectiveness can be attributed, among other things, to economic losses caused through theft and broken equipment, fraud, legal proceedings, failure to meet production deadlines, and poor quality work (Lanyon & Goodstein, 2004). With the help of 576,460 data points (respondents) Ones, Viswesvaran and Schmidt (1993) were able to show in their meta-analysis that integrity test validities are substantial for predicting broad counterproductive behaviours on the job: Across all integrity tests and all kinds of counterproductive working behaviour, integrity had a corrected predictive validity of .47.

4.1.5 Integrity and job-performance criteria

Although integrity tests are generally designed to predict counterproductive working behaviour, they have also been found to predict job performance. Schmidt and Hunter (1998) presented integrity tests as the selection tool that provided the greatest incremental validity above general mental ability tests, which are the best single predictor of job performance: Integrity tests were able to explain more variance of the criteria job success (27%) than variables of conscientiousness (18%). Moreover, with the help of 68,771 respondents Ones, Viswesvaran and Schmidt (1993) were able to demonstrate that the corrected predictive validity of all tested integrity tests is .34 for performance criteria measured with the help of supervisor ratings, awards, and

production results. According to Berry, Sackett and Wiemann (2007; p. 272), this relationship between integrity and performance “should not be surprising, given that CWBs are related to other performance behaviours such as organizational citizenship behaviours (Dalal, 2005; Sackett, Berry, Wiemann, & Laczó, 2006), and that supervisors’ overall performance ratings reflect judgments regarding CWB (Rotundo & Sackett, 2002)”. So, Marcus, Hoft and Riediger (2006) conclude that the practical value of integrity tests in general seems no longer a matter of debate.

4.1.6 Goals of the present study

Previous evaluations of the job-related attitudes and self-evaluations inventory (IBES) not only criticised a lack of sense and face validity of single items as well as the low reliability and criterion validity of single scales, but also the low amount of studies, which proved the quality of the IBES (Hossiep & Bräutigam, 2007a).

Therefore, not only the psychometric quality of the IBES should be analyzed, but also its potential to explain variance in comparison to a personality test and an intelligence test. In a within subject design the factor validity, the internal consistency, the construct-, and criterion validity of the inventory will be tested in two situations: In an honest situation as well as in a simulated applicant situation (the two situations the IBES is constructed for according to Marcus (2006)).

What are the hypotheses in detail? With the help of confirmatory factor analysis the validity of the factors of the IBES will be tested in an honest situation (H1a) as well as in a simulated applicant setting (H1b). Furthermore, the overall score and the single scales of the IBES should be reliable (H2). This internal consistency should hold true for the honest situation (H2a) and the simulated applicant situation (faking situation; H2b). Due to construct validity, the IBES should not correlate in both situations with construct-far variables like intelligence or with other applicant tests like wire-bending or

paper cut out tests as well as with the personality factors extraversion and openness (divergent validity; H3a/b).

On the other hand, the IBES should correlate with construct-close variables like neuroticism, agreeableness, or conscientiousness (convergent validity). These correlations with similar constructs should be there in the honest condition (H4a) and should remain under the faking condition (H4b). Beside the construct validity the IBES should prove its criterion validity. Therefore, it should correlate with supervisor ratings and school grades – in the honest situation (H5a) as well as under the faking condition (H5b). Due to the fact that supervisor ratings are a broader measure of integrity as school grades (counterproductive behaviour like absenteeism or theft influence supervisor ratings but not school grades), the correlations between the IBES and supervisor ratings should rather be higher than lower as the correlations of the IBES with school grades. Furthermore, the IBES should verify its validity beyond an intelligence test and a personality test (incremental validity; H6a: honest situation; H6b: faking situation).

Why are different situations tested? There is a tendency to adopt answers under situational pressure like in applicant settings. This tendency to fake responses under situational pressure can be seen as systematic measurement error (Ziegler, 2007).

Due to higher values in favourite attitudes in applicant settings (Alliger & Dwight, 2000), the variance and therefore also the correlations of variables can rise under situational pressure. However, because of sealing effects the variance may be often restricted under situational pressure so that correlations are not higher but equal in comparison to honest situations. This effect should also take place in this study: As mentioned in hypothesis 4, the IBES should correlate with neuroticism, agreeableness, or conscientiousness in the honest situation and under situational pressure in the faking condition these correlations should not alter. The same holds true for the IBES itself:

The correlations of the IBES scales with itself should be equal or higher under the faking condition (H7).

4.2 Method

4.2.1 Procedure

The data were collected in the chemical industry. According to Viswesvaran and Ones (1999) a within group design was used. Participants first completed the personality test NEO-PI-R in a German adaption (Ostendorf & Angleitner, 2004) and then the first German integrity test IBES (Marcus, 2006). Both tests were presented at two times: First with a neutral instruction (honest condition) and later with a simulated applicant setting as instruction (faking condition). The order of the tests was not varied, because Byle and Holtgraves (2008) were able to prove that the order (honest – fake or fake – honest) does not make any difference. For test taking, participants got a little present, feedback of their results when wished, and due to the fact that the tests were taking place during working hours, participants did not have to work for the test taking time.

4.2.2 Participants

182 trainees of a German company in the chemical industry were invited to take part. 62 were trainees to skilled chemical workers, 30 to skilled pharmaceutical workers, 38 to biological laboratory workers and 52 to chemical laboratory workers. 134 (74%) trainees took part. Details of the data sample can be seen in Table 19.

Differences between participants and non-participants in the predictors of the original hiring procedure are not significant. Apart from the memory task ($g < .44$) all other predictors have small effect sizes ($g < .25$) according to the convention of Cohen (1988). Hereby the predictors of the original hiring procedure are the education level

and the results of the following tests conducted during the personnel assessment: dictation, calculation test, science-technical test, paper cut out test, wire-bending test, intelligence test, and a memory task. Also the performance in their personnel selection interview, which was rated on the scales job perception, motivation, CV, test result, personal impression, and an overall result were predictors for the original hiring procedure.

Table 19
Statistic of participants grouped by their profession

	PHK part.	CHK part.	CL part.	BL part.	Sum part.	Non part.
<i>N</i>	23	50	35	26	134	48
Age <i>M</i>	19.10	19.30	19.97	19.85	19.54	19.56
Age <i>SD</i>	1.86	2.82	1.89	1.85	2.28	2.55
Gender						
Male %	26.09	76.00	45.71	23.08	49.25	52.08
Female %	73.91	24.00	54.29	76.92	50.75	47.92
School Education						
Gymnasium %	4.35	20.00	48.57	73.08	35.07	39.58
Realschule %	95.65	66.00	51.43	26.92	59.70	60.42
Hauptschule %	0	14.00	0	0	5.22	0
Years traineeship <i>M</i>	1.96	1.54	2.11	1.60	1.77	1.71
Years <i>SD</i>	.83	.73	.80	.81	.81	.80

Notes. PHK = skilled pharmaceutical worker, CHK = skilled chemical worker, CL = chemical laboratory worker, BL = biological laboratory worker, *N* = Number of participants; Age *M* = Mean of age, Age *SD* = Standard deviation of age; Education = Years of scholarship; Gymnasium = 9 years of secondary school resulting in diploma qualifying for university admission, Realschule = 6 years of secondary school resulting in school-leaving certificate, Hauptschule = 5 years of secondary school resulting in lowest level school leaving certificate, years traineeship *M* = mean of fulfilled duration of traineeship, years *SD* = Standard deviation years traineeship; part. = participants.

4.2.3 Measures

4.2.3.1 Instruments

Personality was administered by a German paper-pencil adaption of the NEO-PI-R (Ostendorf & Angleitner, 2004), which allows a comprehensive assessment of general personality. The NEO-PI-R contains 240 items, measuring the five factor model: neuroticism, extroversion, openness, agreeableness, and conscientiousness. Respondents with a minimum age of 16 can rate their statements on a five-point Likert-type scale of endorsement, ranging from *strongly disagree* to *strongly agree*.

The “Inventar berufsbezogener Einstellungen und Selbsteinschätzungen“ (job-related attitudes and self-evaluations inventory; IBES; Marcus, 2006) is the first integrity test in German language. Like the NEO-PI-R, the IBES is also a paper-pencil self-report questionnaire with the same five-point Likert-type scale and norms available. In contrast to the NEO-PI-R, only the total score of the 115 items should be evaluated, although it consists of two parts (an overt and a personality-based part) with the following four / five subscales, respectively: general trust (“Vertrauen”, VT, 14 items), perceived pervasiveness / perceived counter-productivity norms (“geringe Verbreitung unerwünschten Verhaltens”, GV, 9 items), rationalizations of deviant behaviour (“Nicht-Rationalisierung”, NR, 19 items), and behavioural intentions (“Verhaltensabsichten”, VA, 18 items) are the 60 items of the overt part, whereas the 55 items of the personality based part consist of the scales positive self-concept (“Selbstwertgefühl”, GS, 19 items), dependability / reliability (“Zuverlässigkeit / Voraussicht”, ZV, 15 items), manipulativeness / Machiavellianism (“Vorsicht”, VO, 7 items), stimulus seeking (“Zurückhaltung”, ZH, 7 items) and trouble avoidance (“Konfliktmeidung”, KM, 7 items). The higher the score of the respondents, the lower is the probability of counterproductive behaviour.

4.2.3.2 *Instruction*

The self-report questionnaires NEO-PI-R and IBES were given two times to the respondents, but with different instructions. First, in the honest condition, respondents were only asked to fill out the questionnaires, whereas in the faking condition, the following fake-good instruction was used:

Please fill out the personality test with following situation in mind, which you already went through:

You are applying for your current trainee job and you are invited to the personnel selection procedure, which takes place now.

Please remember how you felt in your original assessment. One task of the assessment, which is now going on, is to fill out the following questionnaire.

Of course, your primary goal is to get a trainee job. Please fill out the questionnaire in a way that makes sure that you are the one to get the trainee job. However, you must be careful not to be detected as dishonest. Please begin with the questionnaire.

Before filling out the IBES, respondents were asked to remember the applicant situation, they should imagine being in:

Please imagine being in your original hiring procedure. You really want to get this trainee job, but be careful not to be detected as dishonest.

4.2.3.3 *Hiring procedure*

To examine the correlations of the IBES with other assessment procedures, the respondent's results from their original hiring procedure were included in the analysis. The company's hiring procedure consisted of the following parts: a dictation (DIK, 10 min.), a calculation test (REC, 15 min.), a science-technical test (NTP, 15 min.), a paper cut out test (AUS, 5 min.), and a wire-bending test (DRA, 15 min.). After a break, general mental ability was tested in a one hour lasting test (HIT) measuring the verbal, numeric and figural intelligence. It also contained a memory task (MA).

Being successful in this first step of the stepwise hiring procedure, applicants were invited to a personnel selection interview. After the interview, the job perception, motivation, CV, test results, personal impression, and an overall outcome were rated on a five-point Likert-type scale by a member of the human resources department and a manager of the workspace the applicant is applying for. The Likert-type scale was labelled from *very good* to *none* (1 = *very good*, 2 = *good*, 3 = *satisfying*, 4 = *seldom*, 5 = *none*).

4.2.3.4 *Performance criteria*

Two performance criteria were examined: supervisor ratings and school grades. Important are the supervisor ratings consisting of items concerning skills

(“Fertigkeiten”), knowledge (“Kenntnissen”), transfer (“Transfer”), disposition (“Disposition”), teamwork (“Teamverhalten”), work requirements (“Arbeitsanforderungen”), work quality (“Arbeitsqualität”), will to perform (“Leistungsbereitschaft”), trustiness (“Zuverlässigkeit”), learning transfer (“Erfassen von Lerninhalten”) and an overall rating. The seven-point Likert-type scale of the overall rating was ranging from *very satisfying* to *not satisfying* (*very satisfying, very satisfying to satisfying, satisfying, satisfying to almost satisfying, almost satisfying to not satisfying, not satisfying*). Moreover, school grades from the job related-school, the trainees have to attend, were another performance criteria. Due to the fact that school grades are only available for trainees in their second or third year of traineeship, all statistical analyses concerning these criteria could only be conducted for this subgroup ($n = 69$).

4.2.3.5 Statistical analyses

With the help of t-tests for independent samples differences between participants and non-participants in the variables of the original hiring procedure were analyzed. Not only the significance but also the strength of the effect was examined by calculating Hedge’s g . As convention for the effect size, the advise of Cohen (1988) was used.

The question whether answers under the honest condition differ from answers in the simulated applicant setting (faking condition) was examined by the usage of t-test for paired samples. Not only is the significance of the difference analyzed, but the effect size Hedge’s g also. According to Cohen (1988), effects up to .20 are small effects, effects up to .50 are medium-sized, and effects higher than .80 are large effects. Furthermore, the significance of differences in variance of mean values is examined using the F -test, in the honest condition as well as in the faking condition.

With the help of AMOS 6.0, a computer program to calculate confirmatory factor analyses, the factor validity of the IBES scales was examined. Due to the fact, that the IBES measures a construct consisting of more than one dimension no g-factor model was tested. Instead, as recommended by the author, factor validity was tested on the level of the subscales (Marcus, 2006). With the help of the Mardia test, multivariate normal distribution was tested. Due to the fact that it was violated most of the times, Bollen-Stine Bootstraps ($N = 1000$) were conducted to correct the p -value for the χ^2 -tests. For the assessment of the model fit, χ^2 , df , and p -values as well as the global goodness of fit indices $SRMR$ and $RMSEA$ were used. According to the recommendations of Hu and Bentler (1999) and Beauducel and Wittmann (2005) the $SRMR$ (Standardized Root Mean Square Residual) should be smaller / equal .11 and the $RMSEA$ (Root Mean Squared Error of Approximation) smaller than .08 for less than 250 participants. Because of the violated multivariate normal distribution, the CFI (Comparative Fit Index) is not used to test the model fit as recommended by Cheung and Rensvold (2001).

SPSS 17.0 is used to calculate the internal consistency for the IBES subscales and the IBES overall score.

To determine construct validity, correlations between the IBES overall score and construct-far / construct-close variables like intelligence or personality are examined. Correlations are also used to test criterion validity: In this case, correlations between the IBES overall scores and supervisor ratings as well as school grades are calculated.

Moreover, hierarchical regression analyses were conducted to test the incremental validity of the IBES beyond an intelligence test and a personality test. Intelligence was entered first, followed by personality in the second step and the integrity test scores in a third. Supervisor ratings and school grades served as dependent variables. All calculations were conducted for a neutral and a simulated applicant setting.

4.3 Results

In the present study, respondents were asked to answer the NEO-PI-R and the IBES under an honest condition and under a simulated applicant setting (faking condition). The descriptive statistics of the IBES – of this study as well as the ones published in the test-handbook (Marcus, 2006) – can be seen in Table 20. Only the scale stimulus seeking (ZH), who's mean is nearly identical under the two conditions, all other scales as well as the IBES overall score show significant higher means in the simulated application setting by middle and high effect sizes. Moreover, the standard deviations of the scales general trust (VT), positive self-concept (GS), and stimulus seeking (ZH) are significantly lower in the simulated applicant setting.

Table 20
Descriptive Statistics of the IBES

	Honest Instruction		Faking Instruction		Honest vs. Fake		Marcus lab situation		Marcus SP4 Fake Instr.		Marcus SP5 real application	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>g</i>	<i>F</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
I_all	379.6	44.50	423.95	40.71	1.04***	1.19	378.60	35.8	406.5	37	432.7	40.80
Overt part of the IBES												
VT	45.71	7.45	50.44	5.88	.70***	1.60 ^a	47.10	6.70	48.40	7.10	49.90	6.40
GV	25.31	5.04	30.37	4.97	1.01***	1.03	24.70	4.80	27.20	5.10	29.00	5.00
NR	66.35	10.33	75.29	9.25	.91***	1.25	66.70	9.10	72.40	10.40	77.50	10.00
VA	64.87	11.01	72.04	10.09	.68***	1.19	61.30	10.40	69.20	10.10	75.30	9.00
Personality-based part of the IBES												
GS	62.86	10.03	72.74	8.04	1.08***	1.56 ^a	63.40	9.70	68.30	9.30	74.50	9.00
ZV	54.94	6.59	59.85	5.78	.79***	1.30	54.30	6.70	57.40	6.60	60.90	6.10
VO	17.70	4.19	20.13	3.71	.61***	1.28	19.80	4.50	21.40	6.50	22.20	3.90
ZH	21.13	3.38	21.10	2.84	-.01	1.42 ^a	21.00	3.70	21.90	3.40	22.00	2.90
KM	20.69	3.46	21.99	3.38	.38***	1.04	20.20	4.00	20.30	4.00	21.50	3.70

Notes. $N = 134$. * $p < .05$, ** $p < .01$, *** $p < .001$. I_all = IBES overall score; VT = general trust; GV = perceived pervasiveness; NR = rationalizations; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulativeness; ZH = stimulus seeking; KM = trouble avoidance. g = Hedge's g for t-tests of paired samples with the significance level of the t-test. ^a = $F_{\text{empirisch}} > F_{(133,133;0.95)}$. Marcus lab situation = sample 1-3 according to the handbook (Marcus, 2006); Marcus SP4 = values of a student sample with fake-instruction according to the test-handbook (Marcus, 2006); Marcus SP5 = values of the norm sample in a real applicant setting according to the test-handbook (Marcus, 2006).

4.3.1 Factor Validity

To prove factor validity (H1), each scale of the IBES will be analysed by a confirmatory factor analysis. Both, the exact as well as the approximated fit (Fit-Indices), in the neutral (H1a) and in the simulated applicant setting (H1b) were determined. In the neutral condition the scales perceived pervasiveness (GV) and trouble avoidance (KM) are the only ones in which the multivariate normal distribution was not violated. The scales manipulateness (VO) and rationalizations of deviant behaviour (NR) are the only scales, which have not only a very good approximated fit, but an exact fit also. Apart from these, the scales perceived pervasiveness (GV) and general trust (VT) are the only ones lying in the accepted area of the *RMSEA* and of the *SRMR* as suggested by Hu and Bentler (1999), while all other scales have only acceptable values in the *SRMR*.

Table 21

Results of the CFA, proving the factor validity of the scales in the honest condition

	<i>Kurtosis</i>	<i>C.R.</i>	χ^2	<i>df</i>	<i>p</i>	<i>RMSEA</i>	<i>SRMR</i>
Overt part of the IBES							
VT	17.97	4.91	127.99	77	.033	.07	.07
GV	1.84	.76	50.71	27	.018	.08	.06
NR	50.81	10.41	235.80	152	.085	.06	.07
VA	45.16	9.74	320.29	135	.001	.10	.08
Personality-based part of the IBES							
GS	28.78	5.90	386.80	152	.001	.11	.09
ZV	37.00	9.48	179.05	90	.004	.09	.08
VO	6.67	3.44	18.82	14	.427	.03	.04
ZH	5.48	2.82	58.94	14	.001	.16	.10
KM	2.15	1.11	47.31	14	.001	.13	.09

Notes. $N = 134$; Kurtosis = Multivariate Kurtosis; C.R. = Critical Ratio / Mardia Test; VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulateness; ZH = stimulus seeking; KM = trouble avoidance. Bold written values are values, which are in between expected limits.

In accordance with Marcus (2006) the inter-correlations of the IBES scales were calculated to define the inner structure of the IBES. Due to the fact that scales can correlate higher with each other in applicant settings than in anonymous settings (Marcus, 2006), the inter-correlations are calculated for both settings (H7). Table 22 shows that the correlation in the simulated applicant setting (above the diagonal; $r =$

.45) is in most cases larger as the correlations in the honest situation (values below the diagonal; $r = .39$).

Table 22

Inter-correlations of the IBES scales to determine the internal structure of the IBES

	Overt part of the IBES				Personality-based part of the IBES				
	VT	GV	NR	VA	GS	ZV	VO	ZH	KM
VT	1	.57**	.59**	.42**	.52**	.40**	.12	.13	.31**
GV	.68**	1	.67**	.55**	.59**	.43**	.34**	.22**	.41**
NR	.65**	.71**	1	.74**	.72**	.69**	.34**	.19*	.40**
VA	.53**	.62**	.76**	1	.69**	.70**	.37**	.21*	.41**
GS	.59**	.59**	.64**	.60**	1	.75**	.37**	.15	.38*
ZV	.39**	.38**	.55**	.56**	.69**	1	.34**	.17	.35**
VO	.01	.06	.14	.28**	.01	.19**	1	.46**	.30**
ZH	.05	.02	.08	.12	-.12	.13	.43**	1	.32**
KM	.32**	.27**	.40**	.42**	.23**	.25**	.31**	.41**	1

Notes. $N = 134$; * $p < .05$, ** $p < .01$; below the diagonal = neutral situation; above the diagonal = simulated applicant setting; VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulateness; ZH = stimulus seeking; KM = trouble avoidance.

To prove the factorial validity of the IBES scales in the simulated applicant setting (H1b) as well, confirmatory factor analyses for each scale are calculated. This time, no scale shows multivariate normal distribution. The scales manipulateness (VO), stimulus seeking (ZH) and trouble avoidance (KM) are the only ones, which have an exact fit and a *RMSEA* within the proposed limits of Hu and Bentler (1999). In this setting, all scales also show acceptable values for the *SRMR* (see Table 23).

Table 23

CFA results to prove factorial validity in the simulated applicant setting

	Kurtosis	C.R.	χ^2	df	p	RMSEA	SRMR
Overt part of the IBES							
VT	45.11	12.34	147.89	77	.001	.08	.08
GV	14.79	6.08	81.86	27	.001	.12	.09
NR	132.04	27.05	315.22	152	.001	.09	.08
VA	112.95	24.36	319.87	135	.001	.10	.08
Personality-based part of the IBES							
GS	124.85	25.58	342.59	152	.001	.10	.09
ZV	72.04	18.46	178.81	90	.001	.09	.08
VO	17.56	9.05	22.70	14	.065	.07	.06
ZH	17.76	9.16	22.05	14	.078	.07	.06
KM	9.02	4.65	23.22	14	.057	.07	.07

Notes. $N = 134$; Kurtosis = Multivariate Kurtosis; C.R. = Critical Ratio / Mardia Test; VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulateness; ZH = stimulus seeking; KM = trouble avoidance. Bold written values are the values, which are in between the expected limits.

4.3.2 Reliability

To determine the internal consistency, SPSS 17 was used to calculate Cronbachs Alphas for the IBES overall score and for the subscales. In the neutral / honest condition (H2a), Cronbachs Alpha for the overall scale is $\alpha = .96$ ($\alpha = .95$ overt part; $\alpha = .89$ personality based part). Reliabilities of the subscales range in the honest condition from $\alpha = .58$ to $\alpha = .89$ (see Table 24). The lowest internal consistencies have the scales stimulus seeking (ZH) and trouble avoidance (KM). With a reliability of $\alpha = .62$, or $\alpha = .58$ these scales are not reliable.

Table 24

Reliabilities of this study compared to those of the test-handbook (Marcus, 2006)

	Number of items	Neutral Instruction ^a	Simulated applicant Inst ^a	Marcus SP4 Simulated applicant Inst ^b	Marcus SP5 Real Hiring ^c
IBES overall	115	.96	.96	.93	.94
<i>Overt part</i>	60	.95	.95	.89	.92
VT	14	.85	.81	.69	.77
GV	9	.78	.78	.68	.74
NR	19	.89	.89	.82	.87
VA	18	.89	.90	.75	.78
<i>Personality part</i>	55	.89	.90	.85	.88
GS	19	.88	.85	.83	.86
ZV	15	.82	.82	.78	.75
VO	7	.77	.71	.62	.64
ZH	7	.62	.53	.49	.38
KM	7	.58	.58	.58	.55

Notes. ^a $N = 134$; ^b $N = 175$; ^c $N = 332$; IBES overall = IBES overall score; Overt part = total score of the scales, which belong to the overt part of the IBES; Personality part = total score of the scales, which belong to the personality-based part of the IBES; VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulateness; ZH = stimulus seeking; KM = trouble avoidance. Marcus SP4 = values of a student sample with fake-instruction according to the test-handbook (Marcus, 2006); Marcus SP5 = values of the norm sample in a real applicant setting according to the test-handbook (Marcus, 2006).

In the simulated applicant setting, the reliability of the IBES is $\alpha = .96$ for the total score ($\alpha = .95$ overt part; $\alpha = .90$ personality-based part). For the IBES subscales, Cronbachs Alphas range in the simulated applicant setting from $\alpha = .53$ up to $\alpha = .90$. The reliabilities of the scales stimulus seeking (ZH) and trouble avoidance (KM) are once again below the limits. Overall, the reliabilities found in this study are a little bit higher than the reliabilities found by Marcus (2006).

4.3.3 Construct Validity

Determining the construct validity, bivariate correlations between the IBES and construct-close as well as construct-far variables were calculated. In the honest condition, the correlations between the IBES and construct-far settings (H3) show, that there is nearly no correlation between intelligence and IBES results ($r < .01$ to $r = -.10$). A little bit higher correlations could be found between the IBES overall score and the personal impression in the personnel selection interview ($r = .20$), as well as between the IBES and the personality factors extraversion ($r = .19$) and openness ($r = .21$). To sum up, the correlations with the construct-far variables remain low.

Table 25

Divergent validity in the honest condition

	IBES		Overt part of the IBES				Personality-based part of the IBES					
	all	Oall	VT	GV	NR	VA	Pall	GS	ZV	VO	ZH	KM
Intelligence												
HIT	-.10	-.09	-.10	-.13	-.03	-.07	-.10	-.15	-.04	.01	.05	-.07
HIT Ver	-.10	-.08	-.02	-.13	-.04	-.10	-.11	-.13	-.15	.05	.06	-.05
HIT Num	<.01	-.04	-.05	-.07	-.05	.00	.07	.04	.12	.02	.00	-.03
HIT Fig	-.10	-.09	-.17*	-.09	.00	-.08	-.11	-.17	-.07	.01	.07	-.03
Hiring tests												
DIK	.03	.04	.09	.02	.01	.02	.03	-.04	.06	.10	.03	-.01
NTP	.06	-.03	-.07	.01	.02	-.06	.19*	.15	.14	.11	.13	.05
DRA	.04	.06	-.01	.04	.11	.05	.00	.01	.02	-.04	.08	-.10
AUS	.17	.17*	.16	.10	.14	.17*	.12	.13	.11	-.01	.03	.05
REC	.10	.04	.06	.00	.05	.01	.17*	.05	.09	.22*	.21*	.14
MA	.04	.08	.07	.03	.07	.08	-.02	-.08	.09	-.03	-.02	.01
Hiring interview												
OUT	.14	.13	.12	.09	.13	.10	.14	.05	.10	.18*	.12	.08
MOT	.18	.17	.15	.15	.18*	.12	.15	.05	.06	.13	.21*	.16
CV	.16	.15	.09	.14	.11	.16	.16	.04	.13	.22*	.11	.08
TRE	.01	-.02	-.02	-.05	.01	-.03	.05	-.00	.08	.10	.05	-.04
JOP	.05	.05	.05	.05	.03	.05	.04	.01	-.03	.08	.13	.04
PIM	.20*	.20*	.13	.17*	.23*	.15	.17*	.14	.15	.03	.10	.09
Divergent personality factors												
E	.19*	.26**	.41**	.19*	.22*	.12	.05	.36**	.15	-.50**	-.50**	.05
O	.21*	.21*	.21*	.13	.27**	.10	.17	.19*	.27**	-.19	.04	.02

Notes. $N = 134$. * $p < .05$, ** $p < .01$, IBES all = IBES overall score; Oall = total score of the overt part of the IBES; Pall = total score of the personality-based part of the IBES; VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulativeness; ZH = stimulus seeking; KM = trouble avoidance; HIT = total score of the intelligence test; HIT Ver = value of the verbal intelligence; HIT Num = value of the numerical intelligence; HIT Fig = value of the figural intelligence; NTP = science-technical test; DRA = wire-bending test; AUS = paper cut out test; MA = memory task; OUT = total score of the hiring interview; MOT = motivation; CV = Curriculum vitae; TRE = test results; JOB = job perception; PIM = personal impression; E = Extraversion; O = Openness. All values are recoded, so that higher values indicate a better performance.

Proving the correlations of the IBES in the honest condition with construct-close variables (H4a), it could be shown (see Table 26), that the total score of the IBES correlates moderate to high with those personality factors, which were used for the construction of personality based integrity tests: neuroticism ($r = -.47$), agreeableness ($r = .59$), and conscientiousness ($r = .65$). On the facet level, the neuroticism facets angry hostility ($r = -.58$), depression ($r = -.42$) and impulsiveness ($r = -.45$) correlate high negatively with the IBES total score. As far as agreeableness is concerned, the facets general trust ($r = .47$), straightforwardness ($r = .49$), altruisms ($r = .45$) and compliance ($r = .52$) are the ones correlating high with the IBES overall score. Conscientiousness is the factor, which facets have the highest correlations with the IBES overall score (dutifulness: $r = .55$; self-discipline: $r = .59$; deliberation: $r = .61$). Looking on the subscale level of the IBES, there is a high negative correlation between neuroticism and the IBES subscale positive self-concept ($r = -.77$) and a high positive correlation between conscientiousness and the IBES subscale dependability ($r = .82$).

In the simulated applicant setting the correlations between the IBES and construct-far (H3b) / construct-close variables (H4b) were examined, too. Nearly no correlation between the IBES and construct-far variables, like intelligence, tests of the original hiring procedure or the personality factor extraversion were found ($r = -.14$ to $r = .11$). The only significant correlation on the level of the IBES overall score was the one with the personality factor openness for experiences ($r = .31$, see Table 27).

Table 26
Convergent validity in the honest condition

	IBES		Overt part of the IBES					Personality-based part of the IBES				
	all	Oall	VT	UV	NR	VA	Pall	GS	ZV	VO	ZH	KM
Convergent personality factors												
N	-.47**	-.43**	-.47**	-.43**	-.35**	-.31**	-.45**	-.77**	-.38**	.21*	.29**	.01
A	.59**	.57**	.67**	.43**	.47**	.42**	.51**	.41**	.33**	.07	.33**	.53**
C	.65**	.51**	.34**	.38**	.46**	.52**	.75**	.74**	.82**	.10	-.06	.23**
Convergent personality facets												
N1	-.22*	-.22**	-.24**	-.29**	-.16	-.15	-.17*	-.51**	-.10	.24**	.34**	.12
N2	-.58**	-.50**	-.56**	-.46**	-.39**	-.37**	-.60**	-.76**	-.46**	.07	.01	-.24**
N3	-.42**	-.40**	-.48**	-.36**	-.32**	-.27**	-.38**	-.67**	-.32**	.23**	.26**	-.02
N4	-.21*	-.20*	-.25**	-.25**	-.16	-.11	-.18*	-.52**	-.17	.26**	.38**	.14
N5	-.45**	-.39**	-.31**	-.34**	-.33**	-.37**	-.46**	-.53**	-.35**	-.11	-.01	-.11
N6	-.40**	-.36**	-.39**	-.34**	-.32**	-.24**	-.38**	-.70**	-.42**	.25**	.34**	.14
A1	.47**	.53**	.74**	.49**	.41**	.29**	.30**	.39**	.17	-.04	-.07	.26**
A2	.49**	.46**	.42**	.36**	.45**	.36**	.45**	.29**	.23**	.17	.45**	.48**
A3	.45**	.43**	.55**	.28**	.37**	.29**	.42**	.45**	.32**	-.06	.06	.34**
A4	.52**	.45**	.48**	.35**	.38**	.37**	.52**	.38**	.31**	.18*	.30**	.60**
A5	.23**	.20*	.19*	.17*	.14	.19*	.25**	.03	.17*	.12	.52**	.27**
A6	.28**	.29**	.39**	.13	.24**	.23**	.20*	.18*	.18*	-.10	.10	.24**
C1	.40**	.35**	.38**	.29**	.33**	.23**	.41**	.63**	.50**	-.21*	-.36**	.02
C2	.43**	.32**	.13	.25**	.28**	.40**	.53**	.46**	.64**	.15	-.11	.17*
C3	.55**	.44**	.35**	.34**	.38**	.43**	.62**	.59**	.64**	.02	.14	.23**
C4	.37**	.30**	.23**	.16	.29**	.29**	.42**	.48**	.53**	-.06	-.21*	.12
C5	.59**	.45**	.27**	.29**	.40**	.51**	.70**	.70**	.72**	.18*	-.02	.15
C6	.61**	.46**	.23**	.40**	.44**	.48**	.73**	.53**	.71**	.30**	.28**	.36**

Notes. $N = 134$. * $p < .05$, ** $p < .01$. IBES all = IBES overall score; Oall = total score of the overt part of the IBES; Pall = total score of the personality-based part of the IBES; N = Neuroticism; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; A1 = Trust; C1 = Competence; N2 = Angry hostility; A2 = Straightforwardness; C2 = Order; N3 = Depression; A3 = Altruism; C3 = Dutifulness; N4 = Self-consciousness; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; A6 = Tender-mindedness; C6 = Deliberation, VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulativeness; ZH = stimulus seeking; KM = trouble avoidance. Due to the huge amount of correlations, the heights of the correlations are interpreted rather than the significance. Therefore, no Bonferroni correction was made.

Table 27
Divergent validity in the simulated applicant situation

	IBES	Overt part of the IBES					Personality-based part of the IBES					
	all	Oall	VT	GV	NR	VA	Pall	GS	ZV	VO	ZH	KM
Intelligence												
HIT	-.01	.00	-.06	.05	.05	-.05	-.02	-.07	.03	.02	.05	-.03
HIT Ver	.09	.10	.06	.16	.10	.04	.07	.06	.07	.08	.06	-.03
HIT N	-.02	-.04	-.13	-.03	-.04	.02	.02	.00	.08	-.04	.01	.00
HIT Fig	-.07	-.07	-.07	-.02	.02	-.14	-.07	-.13	-.09	.06	.04	-.02
Hiring tests												
DIK	.06	.06	.10	.00	.04	.06	.05	.01	.02	.09	.01	.07
NTP	.10	.07	.07	.02	.08	.05	.12	.16	.15	.06	-.03	-.06
DRA	.01	.01	-.03	.01	.02	.02	.00	-.04	.03	.04	.09	-.05
AUS	-.01	.02	.03	.08	-.01	.00	-.06	-.02	-.03	-.16	-.03	.01
REC	.05	.02	.01	.07	.03	-.02	.09	.07	.07	.03	.09	.09
MA	.01	.06	.07	.04	.10	-.01	-.06	-.09	.04	-.12	-.03	.00
Hiring interview												
OUT	-.06	-.04	.00	-.03	-.05	-.04	-.09	-.03	-.09	.00	-.16	-.11
MOT	-.10	-.07	-.02	-.12	-.07	-.05	-.13	-.14	-.04	-.03	-.19	-.11
CV	-.13	-.10	-.04	-.07	-.08	-.11	-.17	-.13	-.13	-.12	-.13	-.12
TRE	.09	.10	.14	.06	.04	.11	.06	.10	.00	.03	-.03	.07
JOP	-.14	-.10	-.06	-.04	-.04	-.16	-.17	-.17	-.14	.00	-.08	-.15
PIM	-.04	-.03	.08	-.05	-.03	-.07	-.06	-.01	-.07	.05	-.13	-.09
Divergent personality factors												
E	.11	.17*	.32**	.00	.12	.13	.00	.18*	.14	-.41**	-.35**	.08
O	.31**	.34**	.33**	.16	.30**	.31**	.22*	.30**	.26**	-.14	.06	.10

Notes. $N = 134$. * $p < .05$, ** $p < .01$, IBES all = IBES overall score; Oall = total score of the overt part of the IBES; Pall = total score of the personality-based part of the IBES; VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulateness; ZH = stimulus seeking; KM = trouble avoidance. HIT = total score of the intelligence test; HIT Ver = value of the verbal intelligence; HIT N = value of the numerical intelligence; HIT Fig = value of the figural intelligence; NTP = science-technical test; DRA = wire-bending test; AUS = paper cut out test; MA = memory task; OUT = total score of the hiring interview; MOT = motivation; CV = Curriculum vitae; TRE = test results; JOB = job perception; PIM = personal impression; E = Extraversion; O = Openness. All values are recoded, so that higher values indicate a better performance.

Comparable to the honest condition, the correlation of the IBES with the construct-close personality factors neuroticism ($r = -.48$), conscientiousness ($r = .75$), and agreeableness ($r = .53$) are also moderate to high (see Table 28).

Table 28
Convergent validity in the simulated applicant situation

	IBES		Overt part of the IBES					Personality-based part of the IBES				
	all	Oall	VT	UV	NR	VA	Pall	GS	ZV	VO	ZH	KM
Convergent personality factors												
N	-.48**	-.46**	-.51**	-.36**	-.41**	-.31**	-.44**	-.62**	-.42**	-.07	.14	-.11
A	.53**	.51**	.62**	.39**	.41**	.35**	.49**	.39**	.29**	.21*	.46**	.49**
C	.75**	.68**	.43**	.44**	.62**	.68**	.75**	.74**	.84**	.32**	.04	.33**
Convergent personality facets												
N1	-.17	-.16	-.27**	-.17*	-.14	-.03	-.15	-.32**	-.15	.00	.23**	.04
N2	-.54**	-.49**	-.54**	-.38**	-.47**	-.30**	-.54**	-.65**	-.45**	-.13	-.09	-.26**
N3	-.33**	-.35**	-.44**	-.21*	-.29**	-.25**	-.27**	-.45**	-.27**	.03	.15	-.01
N4	-.21*	-.24**	-.28**	-.24**	-.22*	-.12	-.13	-.32**	-.17*	.11	.28**	.01
N5	-.46**	-.41**	-.32**	-.33**	-.37**	-.35**	-.46**	-.53**	-.39**	-.21*	-.04	-.19*
N6	-.52**	-.49**	-.48**	-.37**	-.39**	-.40**	-.50**	-.62**	-.55**	-.14	.11	-.12
A1	.41**	.43**	.65**	.39**	.32**	.22*	.33**	.32**	.24**	.05	.16	.34**
A2	.32**	.28**	.24**	.19*	.28**	.21*	.35**	.20*	.11	.28**	.58**	.35**
A3	.45**	.45**	.53**	.23**	.37**	.37**	.41**	.47**	.35**	.11	.08	.22*
A4	.56**	.49**	.48**	.42**	.45**	.34**	.60**	.46**	.39**	.34**	.38**	.64**
A5	.06	.06	.12	.10	-.01	.04	.06	-.06	-.06	.05	.43**	.12
A6	.29**	.33**	.45**	.18*	.25**	.26**	.20*	.17*	.14	.00	.18*	.23**
C1	.51**	.51**	.43**	.30**	.46**	.47**	.44**	.52**	.58**	.03	-.20*	.17
C2	.61**	.56**	.36**	.42**	.50**	.54**	.59**	.62**	.63**	.24**	-.05	.29**
C3	.58**	.52**	.29**	.34**	.46**	.56**	.59**	.58**	.69**	.18	.07	.21*
C4	.57**	.51**	.32**	.28**	.49**	.51**	.58**	.58**	.68**	.22*	.00	.24**
C5	.71**	.65**	.41**	.41**	.59**	.66**	.70**	.68**	.80**	.31**	.08	.25**
C6	.70**	.59**	.33**	.43**	.55**	.59**	.77**	.65**	.74**	.51**	.25**	.42**

Notes. $N = 134$. * $p < .05$, ** $p < .01$. IBES all = IBES overall score; Oall = total score of the overt part of the IBES; Pall = total score of the personality-based part of the IBES; N = Neuroticism; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; A1 = Trust; C1 = Competence; N2 = Angry hostility; A2 = Straightforwardness; C2 = Order; N3 = Depression; A3 = Altruism; C3 = Dutifulness; N4 = Self-consciousness; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; A6 = Tender-mindedness; C6 = Deliberation, VT = general trust; GV = perceived pervasiveness; NR = rationalizations of deviant behaviour; VA = behavioural intentions; GS = positive self-concept; ZV = dependability; VO = manipulateness; ZH = stimulus seeking; KM = trouble avoidance. Due to the huge amount of correlations, the heights of the correlations are interpreted rather than the significance. Therefore, no Bonferroni correction was made.

4.3.4 Criterion Validity

To examine the correlations between the IBES and performance criteria, the performance criteria supervisor ratings and school grades were correlated with the IBES overall score of both experimental conditions, with the personality factors of both experimental conditions, and with the tests of the original hiring procedure.

Table 29

Correlations of the predictors with the criteria in both test settings

	Supervisor ratings ^a	School Grades ^b
IBES overall score (honest instruction)	-.24**	-.15
IBES overall score (fake instruction)	-.18*	.01
HIT general intelligence	-.17*	-.20
HIT verbal intelligence	-.20*	-.14
HIT numeric intelligence	-.03	-.02
HIT figural intelligence	-.12	-.22
Dictation (DIK)	-.09	-.01
Science-technical test (NTP)	-.22*	-.33**
Wire-bending test (DRA)	-.21*	-.27*
Paper cut out test (AUS)	-.28**	-.25*
Calculation test (REC)	-.13	-.26*
Memory task (MA)	-.12	-.17
Overall outcome rated in hiring interview (OUT)	.29**	.27*
Motivation rated in hiring interview (MOT)	.14	.23
Curriculum vitae rated in hiring interview (CV)	.32**	.20
Test results rated in hiring interview (TRE)	.23**	.29*
Job perception rated in hiring interview (JOP)	.06	.17
Personal impression rated in hiring interview (PIM)	.23**	.20
Neuroticism (honest instruction)	.04	-.13
Extraversion (honest instruction)	.01	.11
Openness (honest instruction)	-.21*	-.26*
Agreeableness (honest instruction)	-.12	-.05
Conscientiousness (honest instruction)	-.25**	-.13
Neuroticism (fake instruction)	.01	-.20
Extraversion (fake instruction)	.01	.21
Openness (fake instruction)	-.21*	-.20
Agreeableness (fake instruction)	-.09	.11
Conscientiousness (fake instruction)	-.13	.08

Notes. ^a $n = 134$; ^b $n = 69$, * $p < .05$ ** $p < .01$.

Significant correlations between the IBES and the ratings of the supervisors – in the honest ($r = -.24$; H5a) as well as in the simulated applicant setting ($r = -.18$, H5b) – were found. The correlation of the IBES with school grades is in both situations not significant ($r = -.15$; $r = .01$).

4.3.5 Incremental Validity

In addition to the criterion validity, the incremental validity of the IBES (H6) will be examined with the help of hierarchical regression analyses. Results show that the IBES alone is able to explain 6% of the variance of the supervisor ratings in the honest condition (H6a). The factors of the NEO-PI-R explain 9% of the variance of the supervisor ratings. Beyond intelligence and personality factors, the IBES is not able to explain further variance, neither of the supervisor ratings nor of the school grades ($\Delta R^2 = .01$). While conscientiousness is the only significant predictor ($\beta = -.29$) for supervisor ratings, conscientiousness ($\beta = -.37$) and openness to experience ($\beta = -.27$) are significant predictors for school grades. Intelligence alone explains only 3 to 4% of the criteria variance, depending on the criteria, what can be explained due to the fact that respondents are already selected by intelligence, because it was one of the main tests in their original hiring procedure.

Table 30
Hierarchical regressions in the honest condition

	Variable	β	r	r_{sp}	R^2	ΔR^2
Regression of supervisor ratings ^a						
Step 1	General Intelligence	-.17	-.17	-.17	.03	.03*
Step 2	General Intelligence	-.16	-.17	-.15	.15	.12
	Neuroticism	-.06	.04	-.05		
	Extraversion	.13	.01	.11		
	Openness to Experience	-.17	-.21	-.15		
	Agreeableness	-.11	-.12	-.10		
	Conscientiousness	-.29**	-.25	-.24		
Step 3	General Intelligence	-.15	-.17	-.15	.15	<.01
	Neuroticism	-.07	.04	-.06		
	Extraversion	.12	.01	.10		
	Openness to Experience	-.16	-.21	-.14		
	Agreeableness	-.08	-.12	-.06		
	Conscientiousness	-.25*	-.25	-.17		
	IBES overall score	-.07	-.24	-.04		
Regression of school grades ^b						
Step 1	General Intelligence	-.20	-.20	-.20	.04	.04
Step 2	General Intelligence	-.15	-.20	-.14	.21	.17*
	Neuroticism	-.20	-.13	-.15		
	Extraversion	.26	.11	.21		
	Openness to Experience	-.27*	-.26	-.25		
	Agreeableness	-.07	-.05	-.06		
	Conscientiousness	-.37*	-.13	-.28		
Step 3	General Intelligence	-.15	-.20	-.14	.22	.01
	Neuroticism	-.24	-.13	-.17		
	Extraversion	.23	.11	.18		
	Openness to Experience	-.27*	-.26	-.25		
	Agreeableness	.01	-.05	.01		
	Conscientiousness	-.29*	-.13	-.19		
	IBES overall score	-.16	-.15	-.09		

Notes. ^a $n = 134$; ^b $n = 69$; $\Delta R^2 =$ Change in R^2 ; * $p < .05$ ** $p < .01$ *** $p < .001$; $\beta =$ standardised regression coefficient; $r =$ zero order correlation; $r_{sp} =$ semi-partial correlation.

Testing the incremental validity of the IBES in the simulated applicant setting (H6b), results show that the IBES was not able to explain variance beyond intelligence and personality factors, neither for the criteria supervisor rating nor for the criteria school grades ($\Delta R^2 < .01$). Intelligence ($\beta = -.17$) and openness to experience ($\beta = -.21$) were found to be significant predictors for supervisor ratings, whereas openness to experience ($\beta = -.31$) and extraversion ($\beta = .30$) were significant predictors for school grades.

Table 31
Hierarchical regressions in the simulated applicant setting

	Variable	β	r	r_{sp}	R^2	ΔR^2
Regression of supervisor rating ^a						
Step 1	General Intelligence	-.17*	-.17	-.17	.03	.03*
Step 2	General Intelligence	-.14	-.17	-.14	.09	.06
	Neuroticism	-.04	.01	-.03		
	Extraversion	.12	.01	.10		
	Openness to Experience	-.21*	-.21	-.19		
	Agreeableness	-.05	-.09	-.04		
	Conscientiousness	-.12	-.13	-.10		
Step 3	General Intelligence	-.15	-.17	-.15	.09	<.01
	Neuroticism	-.05	.01	-.04		
	Extraversion	.09	.01	.07		
	Openness to Experience	-.19	-.21	-.16		
	Agreeableness	-.01	-.09	-.01		
	Conscientiousness	-.06	-.13	-.33		
	IBES overall score	-.10	-.18	-.05		
Regression of school grades ^b						
Step 1	General Intelligence	-.20	-.20	-.20	.04	.04
Step 2	General Intelligence	-.13	-.20	-.13	.18	.14
	Neuroticism	-.18	-.20	-.14		
	Extraversion	.30*	.22	.24		
	Openness to Experience	-.31*	-.20	-.29		
	Agreeableness	.04	.11	.03		
	Conscientiousness	-.12	.08	-.09		
Step 3	General Intelligence	-.13	-.20	-.13	.18	<.01
	Neuroticism	-.19	-.20	-.14		
	Extraversion	.28	.22	.22		
	Openness to Experience	-.30*	-.20	-.25		
	Agreeableness	.06	.11	.05		
	Conscientiousness	-.07	.08	-.04		
	IBES overall score	-.07	<.01	-.04		

Notes. ^a $n = 134$; ^b $n = 69$, $R^2 =$ corrected R^2 ; $\Delta R^2 =$ change in R^2 ; * $p < .05$ ** $p < .01$ *** $p < .001$; $\beta =$ standardised regression coefficient; $r =$ zero order correlation; $r_{sp} =$ semi-partial correlation.

4.4 Discussion

The aim of this study was the investigation of the psychometric quality of the first German integrity test. Therefore, the factorial validity, the reliability, the construct validity and the criterion validity of the job-related attitudes and self-evaluations inventory (IBES; Marcus, 2006) was examined – in a neutral and a simulated applicant setting. It could be shown, that the IBES is reliable, valid and is able to predict performance criteria. Nevertheless, the advice of the author to use the IBES overall score for interpreting results instead of subscales should be followed.

4.4.1 Factor Validity

Proving the factor validity of the IBES, confirmatory factor analyses were calculated for each subscale in both experimental settings. In the neutral setting (H1a), only 4 out of 9 scales had an acceptable approximated fit. For the other five scales only one fit-index was in between the conventions suggested by Hu and Bentler (1999). In accordance with the IBES test-handbook (Marcus, 2006), the scale stimulus seeking (“Zurückhaltung“) had the worst results.

In the simulated applicant setting (H1b), 3 of 9 scales show an exact as well as an approximated fit. For all other scales not all fit-indices were in between the conventions suggested by Hu und Bentler (1999). To analyse the inner structure of the IBES, the inter-correlations of the IBES scales were determined in the neutral and in the simulated applicant setting, too. As expected, higher inter-correlations ($r = .43$ vs. $r = .36$) were found in the applicant setting (H7). Summing up, the factor validity of the IBES scales could be proven in both tests settings (H1a/b) – although a better result for single subscales would be desirable.

4.4.2 Reliability

To determine reliability, Cronbachs Alphas were calculated. The IBES overall score, the overt and the personality-based part of the IBES as well as most of the subscales are reliable. The reliabilities found in this study obtained values slightly higher as the ones reported in the test-handbook of the IBES. Cronbachs Alphas of the scales stimulus seeking and trouble avoidance are less satisfying. This holds true for both experimental settings and is in accordance with the test-handbook: Marcus (2006) came to the conclusion, that the internal consistency of the subscale stimulus seeking could not satisfy. Therefore, hypothesis 2a and 2b could be more or less confirmed.

4.4.3 Construct Validity

In the neutral setting, the IBES does merely correlate with construct-far variables (H3a), like general intelligence. On the other side, the IBES correlates high with construct-close variables (H4a), like the personality factor conscientiousness. This indicates the independence of the construct being close to the Big 5: Like in meta-analyses of American integrity tests (Ones, 1993), especially the personality factors neuroticism, agreeableness, and conscientiousness are the ones, which correlate high with integrity tests.

To prove the construct validity also in the simulated applicant setting, correlations with construct-far (H3b) and construct-close variables (H4b) were analyzed in this experimental setting, too. On the one hand, low correlations with general intelligence show – as expected – the difference between the constructs, proving the divergent validity in the simulated applicant setting (H3b). On the other hand, the IBES correlates high with personality factors like conscientiousness, proving the convergent validity of the IBES in this setting (H4b), too. Due to the existence of divergent and convergent validity in both experimental settings, construct validity could be confirmed.

4.4.4 Criterion Validity

To determine criterion validity, correlations of the IBES with performance criteria (H5) like supervisor ratings or school grades were analyzed. Significant correlations between the IBES and the supervisor ratings in both settings as well as low correlations between the IBES and school grades in the neutral setting verify the criterion validity of the inventory. Similar results were found by Marcus (2007), who also analyzed the correlation of the IBES with supervisor ratings of trainees in a simulated applicant setting. However, the correlations between school grades and the IBES were lower than those of Marcus (2007) and lower than those of the supervisor ratings. Due to the fact

that counterproductive working behaviours influence supervisor ratings broadly (Rotundo & Sackett, 2002), whereas school grades are less influenced by counterproductive behaviours like theft, the lower correlation of the IBES with school grades is in accordance with the expectations and the construct.

4.4.5 Incremental Validity

To test the incremental validity of the IBES beyond an intelligence and a personality test, hierarchical regression analyses were calculated for both settings and both performance criteria (supervisor rating, school grades). Beyond intelligence and personality, the IBES was not able to explain further variance, neither in the neutral situation nor in the simulated applicant setting (rejecting hypothesis 6a and 6b). Therefore, the proposal of Marcus (2006) to use the IBES as an early instrument in a sequential personnel selection procedure seems to be justified.

4.4.6 Situational Influence

In situations like in a personnel selection process, situational pressure can lead to socially desirable responding in non-cognitive self-report questionnaires, resulting in higher means in favourable attitudes and in lower means in less favourable ones (Alliger & Dwight, 2000). Due to higher variance under situational pressure, correlations often rise when socially desirable responding occurs. Why often and not always? Because of ceiling effects when the additional variance is systematic as assumed by Ziegler and Buehner (2009): Rating scale categories of a questionnaire are often limiting, so that variance and thus correlations cannot rise in those situations any more, but remain more or less unchanged. What was found in this study? The IBES scores in the simulated applicant setting were significant higher and the standard deviations (always but not always significantly) smaller. The correlations of the IBES with construct-close

personality factors and with itself rose in comparison to the neutral situation. However, the IBES proved to be reliable and valid even in the simulated applicant setting.

4.4.7 Limitations

In this study, no real applicant setting was used – only a simulated one: Trainees were asked to answer as in their original hiring procedure. Further research using real applicant settings is needed to replicate these results.

Moreover, analyses using school grades as a criterion were only conducted with trainees in their second and third year of traineeship – the others did not have school grades, yet. Therefore, the number of respondents used to calculate criterion validity sinks to 69. Larger sample sizes were more favourable, too.

4.4.8 Implications and future directions

The IBES is mainly used in situations with situational pressure – when decisions for / against applicants have to take place in the personnel selection process. Research showed that in imagined applicants setting there are higher means in integrity tests than in “straight-take” conditions (Jackson, et al., 2000) and in real-life application settings responses are more distorted than in non-applicant settings (Birkeland, et al., 2006). Already Deinzer et al. (1995) mentioned: „...we always measure persons in situations, not persons; there is no psychological measurement in the situational vacuum“. Whereas some authors have argued that this response distortion can affect the construct and criterion-related validity (Schmit & Ryan, 1992; Zickar & Drasgow, 1996), much research has indicated that the criterion-related validities of self-report personality and integrity measures are not significantly affected (Barrick & Mount, 1996; Smith & Ellingson, 2002; Viswesvaran & Ones, 1999; Viswesvaran, et al., 2001; Ziegler & Buehner, 2009) and that criterion-related validity is mainly caused by differences in personality traits and less by differences in socially desirable responding (Ziegler &

Buehner, 2008). Moreover, former studies were able to prove that construct validity is still given, too (Ellingson, et al., 2001; Smith, Hanges, & Dickson, 2001). Nonetheless, not all concerns and questions are answered properly. For example as Ziegler, Danay, Schoelmerich and Buehner (2010) were able to prove, criterion validity on domain level was not influenced by faking, but criterion validity on facet level: Criterion validities on facet level were affected in size and sometimes in direction, too. Therefore, further research in this field is needed. However, what about the validity of the first integrity test in German language?

In this study with trainees of the chemical industry, which have an heterogeneous scholar background, the job-related attitudes and self-evaluation inventory (IBES; Marcus, 2006) has proven to be a reliable and valid test – in a neutral situation as well as in a simulated applicant setting. However, the reliability of the scales stimulus seeking and trouble avoidance should be improved, as well as the factorial validity of some scales. To sum up, the IBES was not able to explain variance beyond an test of general intelligence and beyond a personality test, but it is a suitable first instrument in a sequential personnel selection process and in research situations, as suggested by Marcus (2006).

**5. Summaries / Abstracts
and Conclusions**

5 Summaries / Abstracts and Conclusions

5.1 Summaries / Abstracts of the studies presented

5.1.1 Study 1

Some respondents tend to prefer extreme categories when answering a self-report questionnaire. This response style might distort questionnaire results – therefore it is a long searched for phenomenon. Unfortunately, results concerning individual differences of middle and extreme crossers are rare and conflicting. So a student sample ($N = 312$) was used to examine individual differences between middle and extreme crossers concerning personality factors and facets, fluid intelligence and its facets, age, and gender. This study shows that the response style is consistent across personality facets. Moreover, middle and extreme crossers differ significantly in personality facets of all five personality factors. Extreme crossers are also significantly younger and have significantly lower scores in verbal reasoning. Differences in gender were not found.

5.1.2 Study 2

Research on SDR has shown that faking styles exist. As research concerning differences is scanty, the objective of this study is to examine whether and how extreme and slight fakers differ in individual variables, controlling for response styles. Therefore, 326 students filled out personality and intelligence tests twice – with a faking instruction in the experimental setting at the second time. Almost all fakers fake their responses always in the same way. Extreme fakers differ significantly in agreeableness, in four personality facets. Moreover, they have lower means in most neuroticism facets and higher in conscientiousness facets. Extreme fakers tend also to be younger and have higher scores in reasoning. Females tend to be rather slight fakers.

5.1.3 Study 3

The psychometric quality of the first German Integrity Test (Marcus, 2006) was explored using a sample ($N = 134$) of applicants in the chemical industry. Together with the data from their original hiring procedure, their school grades and their supervisor ratings, the analyses showed that the test is reliable, although two subscales fail. Moreover, construct validity could be confirmed; only one scale came of badly. The results of this study demonstrate also the factor validity and criterion validity of the IBES, only incremental validity beyond intelligence and a personality test could not be proven. Therefore, results indicate that the psychometric quality of the IBES is given, although some details should be enhanced.

5.2 Conclusions regarding the goals of the present project

To shed further light on the question why respondents cross where they cross on non-cognitive self-report questionnaires, three different goals were stated in the introduction: To identify individual differences between middle and extreme crossers, to identify individual differences between respondents with different faking styles and to analyze the psychometric quality of a test constructed for the usage in the personnel selection process, where the occurrence of response distortion is probable.

To achieve these goals three studies were conducted. The results are reported and discussed above. In the following three sections, each goal will be reviewed with integration of the results. Finally, a short summary and an outlook for future research are given.

5.2.1 Extreme response style and individual differences

Results of non-cognitive self-report questionnaires might be threatened by response styles distorting respondents answers (Baumgartner & Steenkamp, 2001). This was the

reason why response styles are a long and heavily researched topic. Nevertheless, up to now, not all questions concerning response styles could be answered. For example, it was not clear in what personality traits and personality facets respondents with different response styles (extreme vs. middle crossers) might differ. Results concerning differences on the level of personality factors were rare and conflicting, results concerning differences on the level of personality facets did not exist. Moreover, former studies did either not use latent class analyses to identify middle and extreme crossers or they used the same personality test scores to identify the response style as well as the level of the specific personality trait. Therefore, the goal of the first study was to shed light on individual differences between middle and extreme crossers, taking the problems of contamination of personality test scores by response styles into account.

What are the results of the first study? Results show that extreme crossers have significant higher scores (with almost modest effect sizes) in the personality factors extraversion, openness, agreeableness, and conscientiousness, but no significant lower mean in neuroticism was found. What about differences on the level of personality facets? Results show that classes of middle and extreme crossers were not found in all facets, but consistently in facets of all domains. In which personality facets do middle and extreme crosser differ? Middle crossers are less active concerning a lot of different personality traits: They are less active concerning their level of activity (E4), their need for changes (E5), their hunger for adventure (O4), their level of curiosity (O5), their tendency to question social / political norms (O6) or the level of engagement for others (A3). They are less open-hearted (E1) and frolic (E6). Instead, they are more self-referential (A3), more realistic (O1), more sceptical (A1) and unsentimental (A6), being less open for feelings (O3). They are more concerned having a higher level of anxiety (N1, N3). Furthermore, they have a lower faith in their competences (C1) and a lower level of dutifulness (C3), achievement-striving (C4) and self-discipline (C5). In

contrast, extreme crossers are more active in many ways, are more in contact with their surrounding, and have a higher level of assertiveness and conscientiousness (see table 32-34 for an overview of former and present results). However, if these differences in personality traits are the reason or a symptom of middle or extreme crossing has to be investigated further. Moreover, extreme crossers were found to have lower scores in verbal reasoning and they were significant younger. These are the individual differences between respondents engaging in unintentional response distortion (ERS). What are the differences between respondents engaging intentionally in response distortion by adapting their answers in a socially desirable way?

5.2.2 Socially desirable responding and individual differences

The question whether and how people with different levels of socially desirable responding / faking styles have individual differences in personality traits, intelligence, age, and gender is still not answered properly: Results concerning differences on the level of personality factors were rare and conflicting, results concerning differences on the level of personality facets did not exist. Therefore, this study tries to shed light on these topics, controlling the first time for different response styles (middle / extreme crossing).

Results show that respondents instructed to answer a global measure of personality in a socially desirable way faked all personality facets with the exception of openness for aesthetics (O2) and openness for values (O3). This is in line with earlier studies, which showed that openness is a factor often less faked (Furnham, 1997; McFarland & Ryan, 2000; Ziegler, 2007). Therefore, SDR does not take place independent of scale content. Moreover, different faking styles could not be found in all personality facets. Further analysis showed that in those facets where different faking styles were found, people either were constantly slight or extreme faking or switched between honest

responding / faking. Results of this study replicate that not all respondents instructed to fake do so and that not all people asked to answer honestly do so as well.

However, do respondents with different faking styles (slight fakers, extreme fakers, switchers) differ in personality factors and facets? To make sure that “real” personality traits of respondents are not contaminated by their response styles, it was controlled for. What personality traits make a switcher to a switcher? As results indicate, people with lower scores in conscientiousness have a higher probability to be switcher: The scores of switchers and slight fakers as well as the scores of switchers and extreme fakers differ in all facets of conscientiousness significantly and with remarkably effect sizes.

Moreover, switchers are less active than extreme fakers, concerning their hunger and openness for adventure (O4), their level of engagement for others (A3) and their activity level itself (E4). Instead, they are more concerned having a higher level of anxiety (N1, N3, N6).

What is the main difference between people switching between regular responding / faking and (permanent) slight fakers? Switchers have significant higher scores in agreeableness (A) by small to moderate effect sizes, indicating that switchers have a higher probability to compromise in interpersonal conflicts, repressing aggression (A4) and are able to trust others more easily (A1).

How do slight and extreme fakers differ? Slight and extreme fakers differ most significantly and with up to moderate effect sizes in their scores of straightforwardness (A2), altruism (A3), compliance (A4), and openness for actions (O4): Extreme fakers are more active concerning the care for others (A3), their willingness to compromise in interpersonal conflicts (A4), and concerning their hunger for adventure (O4). The nearly moderate effect sizes of competence (C1) and dutifulness (C3) indicate – although not significant – that extreme fakers tend to be a little bit more conscientious than slight fakers.

Summing up, the less conscientious, the less active, and the more self-referential and anxious a person is, the higher is the probability that he / she switches between regular responding / faking. The more active a person is (concerning activity, openness for actions or active help for others), the more likely the person is to engage in extreme faking.

Moreover, extreme fakers tended to be younger, slight fakers tended to be rather female, but no significant differences were found (see Table 32-34 for an overview of former and present results).

5.2.3 Psychometric quality of a questionnaire used in personnel selection

The goal of the third study was to analyze the psychometric quality of a test constructed for the usage in the personnel selection process, where the occurrence of response distortion is probable. Therefore, the factorial validity, the reliability, the construct validity, the criterion validity, and the incremental validity of the job-related attitudes and self-evaluations inventory (IBES; Marcus, 2006) was examined – in a neutral situation, where response sets are likely and in a simulated applicant setting, where socially desirable responding probably takes place.

Proving the factor validity of the IBES, confirmatory factor analyses were calculated for each subscale in both experimental settings. To analyse the inner structure of the IBES, the inter-correlations of the IBES scales were determined in the neutral and in the simulated applicant setting, too. Results prove the factor validity of the IBES scales in the neutral setting as well as in the simulated applicant setting – although a better result for single subscales would be desirable.

To determine reliability, Cronbachs Alphas were calculated. The IBES overall score, the overt and the personality-based part of the IBES as well as most of the subscales are reliable. The reliabilities found in this study obtained values slightly higher as the ones

reported in the test-handbook of the IBES. Cronbachs Alphas of the scales stimulus seeking and trouble avoidance are less satisfying. This holds true for neutral setting as well as for the simulated applicant setting and is in accordance with the test-handbook.

To prove construct validity, correlations with construct-far and construct-close variables were analyzed in both settings. On the one hand, low correlations with general intelligence show – as expected – the difference between the constructs, proving divergent validity. On the other hand, the IBES correlates high with personality factors like conscientiousness, proving the convergent validity of the IBES. This indicates the independence of the construct being close to the Big 5: Like in meta-analyses of American integrity tests (e.g. Ones, 1993), especially the personality factors neuroticism, agreeableness, and conscientiousness are the ones which correlate high with integrity tests. Due to the evidence regarding divergent and convergent validity in both experimental settings, construct validity could be confirmed.

To determine criterion validity, correlations of the IBES with performance criteria like supervisor ratings or school grades were analyzed in the neutral setting as well as in the simulated applicant setting. Significant correlations between the IBES and the supervisor ratings in both settings as well as low correlations between the IBES and school grades in the neutral setting verify the criterion validity of the inventory.

To test the incremental validity of the IBES beyond an intelligence and a personality test, hierarchical regression analyses were calculated for both settings and both performance criteria (supervisor rating, school grades). Beyond intelligence and personality factors, the IBES was not able to explain further variance, neither of supervisor ratings nor of school grades – in both settings. Hence, the proposal of Marcus (2006) to use the IBES as an early instrument in a sequential personnel selection procedure seems to be justified.

Summing up, to test whether the psychometric quality of non-cognitive self-report questionnaires remain even when response sets and response styles are likely to occur, the psychometric quality of the first German Integrity Test was examined in a neutral situation, where response styles are probable and in a simulated applicant setting, where the situational pressure is likely to cause socially desirable responding.

What was found in this study? The IBES scores in the simulated applicant setting were significant higher and the standard deviations (always but not always significantly) smaller. The correlations of the IBES with construct-close personality factors and with itself rose in comparison to the neutral situation. However, the IBES proved to be reliable and valid in the neutral setting as well as in the simulated applicant setting.

Table 32

Former and present results concerning the correlation of personality factors and response sets (SDR) / response styles (ERS)

	response set (socially desirable responding)				responses style (extreme crossing)	
	faking in general	switcher	slight faker	extreme faker	middle crosser	extreme crosser
N	↓ Ones et al., 1996 ↑ McFarland & Ryan, 2000	↑ than ef this study	-	↓ than sw Ziegler, 2007 this study	-	↑ higher anxiety Baumgartner & Steenkamp, 2001 Berg & Collier, 1953 Lewis & Taylor, 1955
E	-	-	-	↑ Ziegler, 2007	↓ this study	↑ Austin, Deary & Egan, 2006 Meiser & Machunsky, 2008 this study
O	-	-	-	↑ Ziegler, 2007	↓ this study	↑ this study
A	↑ Grubb & McDaniel, 2007	↑↓ higher than sf lower than ef this study	↓ than sw / ef this study	↑ than sf Ziegler, 2007 this study	↓ this study	↑ this study
C	↑ Ones, Viswesvaran & Reiss, 1996 ↓ Byle & Holtgraves, 2008 McFarland & Ryan, 2000	↓ than sf & ef this study	↑ than sw this study	↑ than sw Ziegler, 2007	↓ this study	↑ Austin, Deary & Egan, 2006 this study

Notes. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; ↑ indicates positive correlations; ↓ indicates negative correlations.

Table 33

Results concerning the correlation of personality facets and response sets / response styles

	response sets (SDR)						response styles (ERS)	
	switcher		slight faker		extreme faker		middle crosser	extreme crosser
N1	↑	than ef	-	-	↓	than sw	↑	↓
N2		-	-	-		-	-	-
N3	↑	than ef	-	-	↓	than sw	↑	↓
N4		-	-	-		-	-	-
N5		-	-	-		-	-	-
N6	↑	than ef	-	-	↓	than sw	-	-
E1		-	-	-		-	↓	↑
E2		-	-	-		-	-	-
E3	↓	than sf	↑	than sw		-	↓	↑
E4	↓	than sf / ef	↑	than sw	↑	than sw	↓	↑
E5		-	-	-		-	↓	↑
E6		-	-	-		-	↓	↑
O1		-	-	-		-	↓	↑
O2		-	-	-		-	-	-
O3		-	-	-		-	↓	↑
O4	↓	than ef	↓	than ef	↑	than sw / ef	↓	↑
O5		-	-	-		-	↓	↑
O6		-	-	-		-	↓	↑
A1	↑	than sf	↓	than sw		-	↓	↑
A2		-	↓	than ef	↑	than sf	-	-
A3	↓	than ef	↓	than ef	↑	than sw / ef	↓	↑
A4	↑	than sf	↓	than sw / ef	↑	than sf	-	-
A5		-	-	-		-	-	-
A6		-	-	-		-	↓	↑
C1	↓	than sf / ef	↑	than sw	↑	than sw	↓	↑
C2	↓	than sf / ef	↑	than sw	↑	than sw	-	-
C3	↓	than sf / ef	↑	than sw	↑	than sw	↓	↑
C4	↓	than sf / ef	↑	than sw	↑	than sw	↓	↑
C5	↓	than sf / ef	↑	than sw	↑	than sw	↓	↑
C6	↓	than sf / ef	↑	than sw	↑	than sw	-	-

Notes. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; N1 = Anxiety; A1 = Trust; C1 = Competence; N2 = Angry hostility; A2 = Straightforwardness; C2 = Order; N3 = Depression; A3 = Altruism; N3 = Dutifulness; N4 = Self-conscientiousness; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; A5 = Modesty; C5 = Self-discipline; N6 = Vulnerability; A6 = Tender-mindedness; C6 = Deliberation; ↑ indicates positive correlations; ↓ indicates negative correlations.

Table 34

Former and present results concerning the correlation of age, gender, cognitive ability and response sets (SDR) / response styles (ERS)

	response set (socially desirable responding)				responses style (extreme crossing)	
	faking in general	switcher	slight faker	extreme faker	middle crosser	extreme crosser
age	-	-	-	↓ being young Ziegler, 2007	↑ this study	childhood/adolescence high lower /stable middle age rise in old age Das & Dutta, 1969 Hamilton, 1968
gender	male higher on SDR scales Ones & Visweswaran, 1998	-	-	male Ziegler, 2007	-	female Austin, Deary & Egan, 2006 Berg & Collier, 1953 Eid & Rauber,
cognitive ability	↑ Grubb & McDaniel, 2007 Pauls & Crost, 2005a	-	-	↑ Ziegler, 2007	-	↓ simplistic thinking Baumgartner & Steenkamp, 2001 Eid & Rauber, 2000 ↓ lower cognitive ability Brenzelmann, 1960 Das & Dutta, 1969 Light, Zax & Gardiner, 1965
verbal reasoning	-	-	-	-	↑ this study	↓ this study
numeric reasoning	-	-	-	-	-	-
figural reasoning	-	-	-	-	-	-
education	-	-	-	-	-	↓ Greenleaf, 1992 Marin, Gamba & Marin, 1992
leadership level	-	-	-	-	-	↓ Eid & Rauber, 2000

5.2.4 Summary and outlook

The present work aimed at exploring individual differences between middle and extreme crossers as well as between respondents with different faking styles – to shed light on the question why people cross where they cross on non-cognitive self-report questionnaires.

Study 1 is the first proving that middle and extreme crossers differ in a wide range of personality traits, accounting for the contamination of personality scores by response styles. Results show that middle and extreme crossers differ significantly and with up to moderate effect sizes in personality facets of all five personality factors with extreme crossers having higher scores in extraversion, openness, agreeableness, and conscientiousness and lower in neuroticism. Furthermore, the response style has proven to be consistent across personality facets. Hence, this study provides further insight into the process of why people cross where they cross on non-cognitive self-report questionnaires with Likert-type scales, when no situational pressure is given. What does happen when situational pressure is given, like in a personnel selection situation?

When situational pressure is given, most people answer in a socially desirable way. However, people distort their answers to different extents – using different faking styles. In this study, three different faking styles were found: Slight fakers, who are more likely to choose mildly positive options, extreme fakers, who choose the most positive / negative option with a high frequency and switchers, who “switch” between regular responding and faking. What personality traits characterize people with different faking styles, when the response style is controlled for? Switchers can be characterized by significant lower personality traits of conscientiousness, lower activity related facets and higher neuroticism facets. Slight and extreme fakers differ mostly due to the lower agreeableness level of slight fakers. To sum up, the lower the level of neuroticism and

the higher the level of conscientiousness and activity related personality traits, the higher is the probability for respondents to favour extreme categories on Likert-type scales.

Whether the difference in personality traits is the reason for or a symptom of the different response styles and sets has to be explored further. Furthermore, the results for the socially desirable responding study should be examined in a real applicant setting and the results of the ERS study with a more heterogeneous sample in order to replicate these results and to find more evidence for the proposed differences. Former results in this field of research are rare and conflicting, which might be caused by very different methods used to identify response sets, response styles, and individual differences. In this project, methods were used to avoid contamination: Contamination of differences in personality traits between respondents with different response styles (middle / extreme crosser) by the response style itself in the ERS study and contamination of differences in personality traits between respondents with different faking styles by the response style in the SDR study. Hopefully, other researchers will take up the ideas and methods used here to confirm the results and to solve the remaining questions why people cross where they cross. Study 3 showed that the psychometric quality of a test used in the personnel selection process remains despite response sets and styles and confirms the applicability of non-cognitive self-report questionnaires with Likert-type scales in the organizational practice. Due to the fact that self-report questionnaires are such a universal and often used instrument in research, organisational and clinical practices, further efforts should be undertaken to reassure why people cross where they cross on non-cognitive self-report questionnaires with Likert-type scales. This project was not the first, but hopefully a remarkable step towards understanding the issue presented.

References

6 References

- 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*(1), 59-72.
- Allport, G. W., & Odbert, H. S. (1936). Trait-Names a Psycho-Lexical Study. *Psychological Monographs, 47*(1), 1-171.
- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *I-S-T 2000 R; Intelligenz-Struktur-Test 2000 R; Intelligence Structure Test 2000*. Göttingen: Hogrefe.
- Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual Differences in Response Scale Use: Mixed Rasch Modelling of Responses to Neo-Ffi Items. *Personality and Individual Differences, 40*(6), 1235-1245.
- Barrick, M. R., & Mount, M. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology, 44*, 1-26.
- Barrick, M. R., & Mount, M. K. (1996). *Effects of Impression Management and Self-Deception on the Predictive Validity of Personality Constructs*.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next? *International Journal of Selection and Assessment, 9*(1-2), 9-30.

- Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). *Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness*.
- Baumgartner, H., & Steenkamp, J. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2), 143-156.
- Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on Fluid and Crystallized Intelligence: Facets for Verbal, Numerical, and Figural Intelligence. *Personality and Individual Differences*, 30(6), 977-994.
- Beauducel, A., & Wittmann, W. (2005). Simulation Study on Fit Indexes in Cfa Based on Data with Slightly Simple Structure. *Structural Equation Modeling*, 41-75.
- Belli, R. F., Conrad, F. G., & Wright, D. B. (2007). Cognitive Psychology and Survey Methodology: Nurturing the Continuing Dialogue between Disciplines - Introduction. *Applied Cognitive Psychology*, 21(2), 141-144.
- Berg, I. A., & Collier, J. S. (1953). Personality and Group Differences in Extreme Response Sets. *Educational and Psychological Measurement*, 13(2), 164-169.
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal Deviance, Organizational Deviance, and Their Common Correlates: A Review and Meta-Analysis. *Journal of Applied Psychology*, 92(2), 410-424.
- Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A Review of Recent Developments in Integrity Test Research. *Personnel Psychology*, 60(2), 271-301.

- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures. *International Journal of Selection and Assessment, 14*(4), 317-335.
- Blanchard, D. D., McGrath, R. E., Pogge, D. L., & Khadivi, A. (2003). A Comparison of the Pai and Mmpi-2 as Predictors of Faking Bad in College Students. *Journal of Personality Assessment, 80*(2), 197-205.
- Block, J. (1971). Further Conjectures Regarding Acquiescence. *Psychological Bulletin, 76*(3), 205-210.
- Bolt, D. M., & Johnson, T. R. (2009). Addressing Score Bias and Differential Item Functioning Due to Individual Differences in Response Style. *Applied Psychological Measurement, 33*(5), 335-352.
- Borgatta, E. F. (1964). The Structure of Personality Characteristics. *Behavioral Science, 12*, 8-17.
- Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (Aic): The General Theory and Its Analytical Extensions. *Psychometrika, 52*(3), 345-370.
- Brengelmann, J. C. (1960a). Extreme Response Set, Drive Level and Abnormality in Questionnaire Rigidity. *Journal of Mental Science, 106*(442), 171-186.
- Brengelmann, J. C. (1960b). A Note on Questionnaire Rigidity and Extreme Response Set. *Journal of Mental Science, 106*(442), 187-192.

- Byle, K. A., & Holtgraves, T. M. (2008). *Integrity Testing, Personality, and Design: Interpreting the Personnel Reaction Blank*.
- Casillas, A., Robbins, S., McKinniss, T., Postlethwaite, B., & Oh, I. S. (2009). Using Narrow Facets of an Integrity Test to Predict Safety: A Test Validation Study. *International Journal of Selection and Assessment, 17*(1), 119-125.
- Cattell, R. B. (1943a). The Description of Personality. I. Foundations of Trait Measurement. *Psychological Review, 50*, 559-594.
- Cattell, R. B. (1943b). The Description of Personality: Basic Traits Resolved into Clusters. *Journal of Abnormal and Social Psychology, 38*(4), 476-506.
- Cattell, R. B. (1945). The Description of Personality - Principles and Findings in a Factor Analysis. *American Journal of Psychology, 58*(1), 69-90.
- Cheung, G. W., & Rensvold, R. B. (2001). The Effects of Model Parsimony and Sampling Error on the Fit of Structural Equation Models. *Organizational Research Methods, 4*, 236-264.
- Cohen, J. (Ed.). (1988). *Statistical Power Analysis for the Behavioural Sciences* (2 ed.). Hillsdale: Lawrence Erlbaum.
- Colbert, A. E., Mount, M. K., Harter, J. K., Witt, L. A., & Barrick, M. R. (2004). Interactive Effects of Personality and Perceptions of the Work Situation on Workplace Deviance. *Journal of Applied Psychology, 89*(4), 599-609.

- Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on Personality Measures: Implications for Selection Involving Multiple Predictors. *International Journal of Selection and Assessment*, 17(1), 47-60.
- Costa, P. T., & McCrae, R. R. (1985). *The Neo Personality Inventory*. Odessa, FL: Psychol. Assess. Resources.
- Costa, P. T., & McCrae, R. R. (1988). Personality in Adulthood: A Six-Year Longitudinal Study of Self-Reports and Spouse Ratings on the Neo Personality Inventory. *Journal of Personality and Social Psychology*, 54, 853-863.
- Costa, P. T., & McCrae, R. R. (1992). *Revised Neo Personality Inventory (Neo-Pi-R) and Neo Five-Factor Inventory (Neo-Ffi) Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1995). Domains and Facets - Hierarchical Personality-Assessment Using the Revised Neo Personality-Inventory. *Journal of Personality Assessment*, 64(1), 21-50.
- Couch, A. S., & Keniston, K. (1960). Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable. *Journal of Abnormal and Social Psychology*, 60, 151-174.
- Cromwell, R. L., & Caldwell, D. F. (1962). A Comparison of Ratings Based on Personal Constructs of Self and Others. *J. clin. Psychol.*, 18, 43-46.
- Cronbach, L. J. (1946). Response Sets and Test Validity. *Educational and Psychological Measurement*, 6(4), 475-494.

- Cronbach, L. J. (1950). Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement, 10*(1), 3-31.
- Das, J. P., & Dutta, T. (1969). Some Correlates of Extreme Response Set. *Acta Psychologica, 29*(1), 85-92.
- Davier, M. v. (2001). Winmira 2001. Groningen, The Netherlands: ASC-Assessment Systems Corporation USA and Science Plus Group.
- Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F., et al. (1995). Situational Effects in Trait Assessment: The Fpi, Neoffi, and Epi Questionnaires. *European Journal of Personality, 9*, 1-23.
- Digman, J. M. (1990). Personality Structure - Emergence of the 5-Factor Model. *Annual Review of Psychology, 41*, 417-440.
- Digman, J. M., & Shmelyov, A. G. (1996). The Structure of Temperament and Personality in Russian Children. *Journal of Personality and Social Psychology, 71*(2), 341-351.
- Eid, M., & Rauber, M. (2000). Detecting Measurement Invariance in Organizational Surveys. *European Journal of Psychological Assessment, 16*(1), 20-30.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social Desirability Corrections in Personality Measurement: Issues of Applicant Comparison and Construct Validity. *Journal of Applied Psychology, 84*(2), 155-166.

- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the Influence of Social Desirability on Personality Factor Structure. *Journal of Applied Psychology, 86*(1), 122-133.
- Fiske, D. W. (1949). Consistency of the Factorial Structures of Personality Ratings from Different Sources. *Abnorm. Social Psychology, 44*, 329-344.
- Furnham, A. F. (1997). Knowing and Faking One's Five-Factor Personality. *Journal of Personality Assessment, 69*(1), 229-243.
- Galesic, M., & Tourangeau, R. (2007). What Is Sexual Harassment? It Depends on Who Asks! Framing Effects on Survey Responses. *Applied Cognitive Psychology, 21*(2), 189-202.
- Goldberg, L. R. (1981). Language and Individual Differences: The Search for Universals in Personality Lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology*. Beverly Hills, CA: Sage.
- Gough, H. G. (1971). Assessment of Wayward Impulse by Means of Personnel Reaction Blank. *Personnel Psychology, 24*(4), 669-677.
- Greenleaf, E. A. (1992). Measuring Extreme Response Style. *Public Opinion Quarterly, 56*, 328-351.
- Grice, H. P. (1975). Logic and Conversation. In P. Cole & M. J.L. (Eds.), *Syntax and Semantics: Speech Acts* (Vol. 3, pp. 41-58). New York: Academic Press.

- Grubb, W. L., & McDaniel, M. A. (2007). The Fakability of Bar-On's Emotional Quotient Inventory Short Form: Catch Me If You Can. *Human Performance, 20*(1), 43-59.
- Hamilton, D. L. (1968). Personality Attributes Associated with Extreme Response Style. *Psychological Bulletin, 69*(3), 192-203.
- Harzing, A.-W. (2006). Response Styles in Cross-National Survey Research: A 26-Country Study. *International Journal of Cross Cultural Management, 6*(2), 243-266.
- Hesterly, S. O. (1963). Deviant Response Patterns as a Function of Chronological Age. *Journal of Consulting Psychology, 27*(3), 210-234.
- Hirsh, J. B. (2009). Choosing the Right Tools to Find the Right People. *Psychologist, 22*(9), 752-755.
- Horn, J., Nelson, C. E., & Brannick, M. T. (2004). Integrity, Conscientiousness, and Honesty. *Psychological Reports, 95*(1), 27-38.
- Hossiep, R. V., & Bräutigam, S. (2007a). Job-Related Attitudes and Self-Evaluations Inventory. *Zeitschrift Fur Personalpsychologie, 6*(2), 85-90.
- Hossiep, R. V., & Bräutigam, S. (2007b). Tests and Tools. Inventar Berufsbezogener Einstellungen Und Selbsteinschätzungen (Ibes) Von B. Marcus. *Zeitschrift für Personalpsychologie, 6*(2), 85-90.

- Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling, 6*, 1-55.
- Hurtz, G. M., & Donovan, J. J. (2000). *Personality and Job Performance: The Big Five Revisited*.
- Huselid, M. A. (1995). The Impact of Human-Resource Management-Practices on Turnover, Productivity, and Corporate Financial Performance. *Academy of Management Journal, 38*(3), 635-672.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The Impact of Faking on Employment Tests: Does Forced Choice Offer a Solution? *Human Performance, 13*(4), 371-388.
- Johnson, T., Kulesa, P., Llc, I., Cho, Y. I., & Shavitt, S. (2005). The Relation between Culture and Response Styles: Evidence from 19 Countries. *Journal of Cross-Cultural Psychology, 36*(2), 264-277.
- Judge, T. A., & Ilies, R. (2002). Relationship of Personality to Performance Motivation: A Meta-Analytic Review. *Journal of Applied Psychology, 87*(4), 797-807.
- Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian Version of the Neo-Pi-R: An Examination of Universal and Culture-Specific Aspects of the Five-Factor Model. *European Journal of Personality, 14*(3), 265-278.
- Kerrick, J. (1954). The Effects of Intelligence and Manifest Anxiety on Attitude Change

through Communications. *Unpublished doctoral dissertation, University of Illinois.*

Klages, L. (1926). *The Science of Character*. London: Allen & Unwin.

Klump, C. S. (Ed.). (1964). *Stanton Survey*. Charlotte, NC: Stanton Corporation.

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5*, 213-236.

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology, 50*, 537-567.

Landfield, A. W. (1968). The Extremity Rating Revisited within the Context of Personal Construct Theory. *Br. J. soc. clin. Psychol., 7*, 135-139.

Lanyon, R. I., & Goodstein, L. D. (2004). Validity and Reliability of a Pre-Employment Screening Test: The Counterproductive Behavior Index (Cbi). *Journal of Business and Psychology, 18*(4), 533-553.

Lewis, N. A., & Taylor, J. A. (1955). Anxiety and Extreme Response Preferences. *Educational and Psychological Measurement, 15*(2), 111-116.

Liebert, R. M., & Liebert, L. L. (1998). *Personality Strategies & Issues*. Pacific Grove: Brooks / Cole Publishing.

Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of Age, Sex, and Intelligence Level to Extreme Response Style. *J Pers Soc Psychol, 2*(6), 907-909.

- LondonHouse (Ed.). (1995). *Personnel Selection Inventory: Psi-Cs with Math*. Rosemont, IL: McGraw-Hill / London House.
- Lounsbury, J. W., Sundstrom, E., Loveland, J. L., & Gibson, L. W. (2002). Broad Versus Narrow Personality Traits in Predicting Academic Performance of Adolescents. *Learning and Individual Differences, 14*(1), 65-75.
- Marcus, B. (2003). Attitudes Towards Personnel Selection Methods: A Partial Replication and Extension in a German Sample. *Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale, 52*(4), 515-532.
- Marcus, B. (2006). *Inventar Berufsbezogener Einstellungen Und Selbsteinschätzungen [Job-Related Attitudes and Self-Evaluations Inventory]. Testmanual*. Göttingen: Hogrefe.
- Marcus, B. (2007). Recent Evidence on the Job-Related Attitudes and Self-Evaluation Inventory (Ibes). *Zeitschrift Fur Personalpsychologie, 6*(3), 129-132.
- Marcus, B., Hoft, S., & Riediger, M. (2006). Integrity Tests and the Five-Factor Model of Personality: A Review and Empirical Test of Two Alternative Positions. *International Journal of Selection and Assessment, 14*(2), 113-130.
- Marcus, B., Lee, K., & Ashton, M. C. (2007). *Personality Dimensions Explaining Relationships between Integrity Tests and Counterproductive Behavior: Big Five, or One in Addition?*

- Marcus, B., & Schuler, H. (2004). Antecedents of Counterproductive Behavior at Work: A General Perspective. *Journal of Applied Psychology, 89*(4), 647-660.
- Marcus, B., & Wagner, U. (2007). Combining Dispositions and Evaluations of Vocation and Job to Account for Counterproductive Work Behavior in Adolescent Job Apprentices. *Journal of Occupational Health Psychology, 12*(2), 161-176.
- Marin, G., Gamba, R., & Marin, B. (1992). Extreme Response Style and Acquiescence among Hispanics: The Role of Acculturation and Education. *Journal of Cross-Cultural Psychology, 23*, 149-174.
- Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika, 47*(2), 149-174.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in Faking across Noncognitive Measures. *Journal of Applied Psychology, 85*(5), 812-821.
- Meiser, T., & Machunsky, M. (2008). The Personal Structure of Personal Need for Structure - a Mixture-Distribution Rasch Analysis. *European Journal of Psychological Assessment, 24*(1), 27-34.
- Menon, G. (1994). Judgments of Behavioral Frequencies: Memory Search and Retrieval Strategies. In N. Schwarz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (pp. 161-172). New York: Springer Verlag.
- Mitsos, S. B. (1961). Personal Constructs and the Semantic Differential. *Journal abnormal social Psychology, 62*, 433-434.

- Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality Predictors of Extreme Response Style. *Journal of Personality*, 77(1), 261-286.
- Norman, W. T. (1963). Toward an Adequate Taxonomy of Personality Attributes - Replicated Factor Structure in Peer Nomination Personality Ratings. *Journal of Abnormal Psychology*, 66(6), 574-583.
- O'Bannan, R. M., Goldinger, L. A., & Appleby, G. S. (Eds.). (1989). *Honesty and Integrity Testing: A Practical Guide*. Atlanta: Applied Information Resources.
- Office (Ed.). (2009). *Statistical Yearbook 2009 for the Federal Republic of Germany*. Wiesbaden: Statistisches Bundesamt (Federal Statistical Office).
- Ones, Viswesvaran, & Schmidt (1993). Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *Journal of Applied Psychology*, 78, 679-703.
- Ones, D. S. (1993). *The Construct Validity of Integrity Test*. Unpublished doctoral dissertation, University of Iowa City, Iowa.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In Support of Personality Assessment in Organizational Settings. *Personnel Psychology*, 60(4), 995-1027.
- Ones, D. S., & Viswesvaran, C. (1998). *The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection*.

- Ones, D. S., & Viswesvaran, C. (2001). Integrity Tests and Other Criterion-Focused Occupational Personality Scales (Cops) Used in Personnel Selection. *International Journal of Selection and Assessment*, 9(1-2), 31-39.
- Ones, D. S., & Viswesvaran, C. (2003). Job-Specific Applicant Pools and National Norms for Personality Scales: Implications for Range-Restriction Corrections in Validation Research. *Journal of Applied Psychology*, 88(3), 570-577.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of Social Desirability in Personality Testing for Personnel Selection: The Red Herring. *Journal of Applied Psychology*, 81(6), 660-679.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. (1993). Comprehensive Metaanalysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *Journal of Applied Psychology Monograph*, 78, 679-703.
- Ongena, Y. P., & Dijkstra, W. (2007). A Model of Cognitive Processes and Conversational Principles in Survey Interview Interaction. *Applied Cognitive Psychology*, 21(2), 145-163.
- Ostendorf, F., & Angleitner, A. (2004). *Neo-Pi-R: Persönlichkeitsinventar Nach Costa Und McCrae*. Göttingen: Hogrefe.
- Paajanen, G. E. (1986). *Development and Validation of the Pdi Employment Inventory*. Presented at the 94th Annual Convention of the American Psychological Association.

- Paulhus, D. L. (1986). Self-Deception and Impression Management in Test Responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality Assessment Via Questionnaire* (pp. 142-165). New York: Springer.
- Paulhus, D. L. (1991). Measurement and Control of Response Bias. In J. Robinson, P. Shaver & L. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (Vol. 1, pp. 17-59). San Diego: Academic Press.
- Paulhus, D. L. (2002). Socially Desirable Responding: The Evolution of a Construct. In H. I. Braun, D. N. Jackson & D. E. Wiew (Eds.), *The Role of Constructs in Psychological and Educational Measurement* (pp. 49-69). Mahwah NJ: Erlbaum.
- Pauls, C. A., & Crost, N. W. (2005a). Cognitive Ability and Self-Reported Efficacy of Self-Presentation Predict Faking on Personality Measures. *Journal of Individual Differences, 26*(4), 194-206.
- Pauls, C. A., & Crost, N. W. (2005b). Effects of Different Instructional Sets on the Construct Validity of the Neo-Pi-R. *Personality and Individual Differences, 39*(2), 297-308.
- Peterson, M. H., Griffith, R. L., & Converse, P. D. (2009). Examining the Role of Applicant Faking in Hiring Decisions: Percentage of Fakers Hired and Hiring Discrepancies in Single- and Multiple-Predictor Selection. *Journal of Business and Psychology, 24*(4), 373-386.
- Read, T. R. C., & Cressie, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York: Springer-Verlag.

- Reid, R. J. (Ed.). (1957). *The Reid Report*. Chicago, IL: John E. Reid Associates.
- Roberts, B. W., Harms, P. D., Caspi, A., & Moffitt, T. E. (2007). Predicting the Counterproductive Employee in a Child-to-Adult Prospective Study. *Journal of Applied Psychology, 92*(5), 1427-1436.
- Rogers, R. (1997). Researching Dissimulation. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (pp. 309-327). New York: Guilford.
- Ross, C. K., Steward, C. A., & Sinacore, J. M. (1995). A Comparative-Study of 7 Measures of Patient Satisfaction. *Medical Care, 33*(4), 392-406.
- Rost, J., Carstensen, C. H., & von Davier, M. (1999). Are the Big Five Rasch Scalable? A Reanalysis of the Neo-Ffi Norm Data. *Diagnostica, 45*(3), 119-127.
- Rothstein, M., & Goffin, R. (2006). The Use of Personality Measures in Personnel Selection: What Does Current Research Support? *Human Resource Management Review, 16*, 155-180.
- Rotundo, M., & Sackett, P. R. (2002). The Relative Importance of Task, Citizenship, and Counterproductive Performance to Global Ratings of Job Performance: A Policy-Capturing Approach. *Journal of Applied Psychology, 87*(1), 66-80.
- Roussos, L., & Stout, W. (1996). A Multidimensionality-Based Dif Analysis Paradigm. *Applied Psychological Measurement, 20*(4), 355-371.

- Rynes, S. L. (1991). Recruitment, Job Choice, and Post-Hire Consequences: A Call for New Research Directions. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., Vol. 2, pp. 339-444). Palo Alto, CA: Consulting Psychologists Press.
- Rynes, S. L. (1993). Who'S Selecting Whom? Effects of Selection Practices on Applicant Attitudes and Behavior. In N. Schmitt & W. C. Borman (Eds.), *Personnel Selection in Organizations*. San Francisco, CA: Jossey-Bass.
- Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity Testing for Personnel Selection: An Update. *Personnel Psychology*, 42, 491-529.
- Sackett, P. R., & DeVore, C. J. (2002). Counterproductive Behaviors at Work. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of Industrial, Work, and Organizational Psychology* (Vol. 1, pp. 145-164). London: Sage.
- Sackett, P. R., & Wanek, J. E. (1996). New Developments in the Use of Measures of Honesty, Integrity, Conscientiousness, Dependability, Trustworthiness, and Reliability for Personnel Selection. *Personnel Psychology*, 49(4), 787-829.
- Salgado, J. F. (1997). The Five Factor Model of Personality and Job Performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J. F. (2002). The Big Five Personality Dimensions and Counterproductive Behaviors. *International Journal of Selection and Assessment*, 10(1-2), 117-125.

- Schmidt, F. L., & Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, *124*(2), 262-274.
- Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond Alpha: An Empirical Examination of the Effects of Different Sources of Measurement Error on Reliability Estimates for Measures of Individual Differences Constructs. *Psychological Methods*, *8*(2), 206-224.
- Schmit, M. J., & Ryan, A. M. (1992). Test-Taking Dispositions: A Missing Link? *Journal of Applied Psychology*, *77*, 629-637.
- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in Personnel Selection: Factor Structure in Applicant and Nonapplicant Populations. *Journal of Applied Psychology*, *78*(6), 966-974.
- Schroder, H. M., Driver, M., & Streufert, S. (1967). *Human Information Processing*. New York: Holt, Rinehart & Winston.
- Schuman, H., & Presser, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- Schwarz, N. (2007). Cognitive Aspects of Survey Methodology. *Applied Cognitive Psychology*, *21*(2), 277-287.

- Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response Scales: Effects of Category Range on Reported Behavior and Subsequent Judgments. *Public Opinion Quarterly*, 49, 388-395.
- Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2008). Psychological Testing in Personnel Selection, Part I: A Century of Psychological Testing. *Public Personnel Management*, 37(1), 99-109.
- Sharf, J. C., & Jones, D. P. (2000). Employment Risk Management. In J. F. Kehoe (Ed.), *Managing Selection in Changing Organizations* (pp. 271-318). San Francisco: Jossey-Bass.
- Simon, H. (1957). *Models of Man*. New York: Wiley.
- Smith (1967). Explorations in Nonconformity. *Journal of Social Psychology*, 71(1), 133-155.
- Smith, & Conrey, F. R. (2007). Mental Representations Are States Not Things: Implications for Implicit and Explicit Measurement. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit Measures of Attitudes: Progress and Controversies* (pp. 247-264). New York: Guilford.
- Smith, & Ellingson, J. E. (2002). Substance Versus Style: A New Look at Social Desirability in Motivating Contexts. *Journal of Applied Psychology*, 87(2), 211-219.

- Smith, Hanges, P. J., & Dickson, M. W. (2001). Personnel Selection and the Five-Factor Model: Reexamining the Effects of Applicant's Frame of Reference. *Journal of Applied Psychology, 86*, 304-315.
- Strack, F., Schwarz, N., & Wanke, M. (1991). Semantic and Pragmatic Aspects of Context Effects in Social and Psychological-Research. *Social Cognition, 9*(1), 111-125.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Press.
- Tett, R. P., Jackson, D. N., & Rothstein, M. G. (1991). Personality Measures as Predictors of Job Performance: A Meta-Analytic Review. *Personnel Psychology, 44*, 703-742.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin, 103*(3), 299-314.
- Tourangeau, R., Rips, L. C., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tunnel, G. (1980). Intraindividual Consistency in Personality Assessment - the Effect of Self-Monitoring. *Journal of Personality, 48*, 220-232.
- Tupes, E. C., & Christal, R. E. (1961). *Recurrent Personality Factors Based on Trait Ratings: USAF, Lacland Air Force Base, TX*.

- Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the Criterion-Related Validation of Selection Procedures: A Critical Review and Recommendations for Practice. *Personnel Psychology, 61*(4), 871-925.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological Measurement, 59*(2), 197-210.
- Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do Impression Management Scales in Personality Inventories Predict Managerial Job Performance Ratings? *International Journal of Selection and Assessment, 9*(4), 277-289.
- Warr, P. B., & Coffman, T. L. (1970). Personality, Involvement and Extremity of Judgement. *British Journal of Social and Clinical Psychology, 9*(2), 108-121.
- Warwick, D., & Lininger, C. (1975). *The Sample Survey: Theory and Practice*. New York: McGraw-Hill.
- Webster's New International Dictionary* (1925). (2nd unabridged edition ed.). Merriam, Springfield.
- Wright, P. M., & Boswell, W. R. (2002). Desegregating Hrm: A Review and Synthesis of Micro and Macro Human Resource Management Research. *Journal of Management, 28*(3), 247-276.
- Zickar, M. J., & Burnfield, J. L. (2003). A Monte Carlo Investigation of the Polytomous Mixed Rasch Model. *Unpublished manuscript*.

- Zickar, M. J., & Drasgow, F. (1996). Detecting Faking on a Personality Instrument Using Appropriateness Measurement. *Applied Psychological Measurement, 20*(1), 71-87.
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering Faking Samples in Applicant, Incumbent, and Experimental Data Sets: An Application of Mixed-Model Item Response Theory. *Organizational Research Methods, 7*(2), 168-190.
- Zickar, M. J., & Robie, C. (1999). Modeling Faking Good on Personality Items: An Item-Level Analysis. *Journal of Applied Psychology, 84*(4), 551-563.
- Ziegler, M. (2007). *Situational Demand and Its Impact on Construct and Criterion Validity of a Personality Questionnaire: State and Trait, a Couple You Just Can T Study Separately!* Unpublished Dissertation, Ludwig-Maximilians-Universität München.
- Ziegler, M., & Buehner, M. (2008). Higher Order Personality Factors: What Role Do States Have? *International Journal of Psychology, 43*(3-4), 370-370.
- Ziegler, M., & Buehner, M. (2009). Modeling Socially Desirable Responding and Its Effects. *Educational and Psychological Measurement, 69*(4), 548-565.
- Ziegler, M., Danay, E., Schoelmerich, F., & Buehner, M. (2010). Predicting Academic Success with the Big 5 Rated from Different Points of View: Self-Rated, Other Rated and Faked. *European Journal of Personality (in press)*.

- Zuckerman, M., & Norton, J. (1961). Response Set and Content Factors in the California F Scale and the Parental Attitude Research Instrument. *Journal of Social Psychology, 53*, 199-210.

2003 - 2005	LMU München: Lehrstuhl für Allgemeine und Experimentelle Psychologie, Forschungspraktikantin
2000 - 2001	Sun Microsystems GmbH, München Werkstudentin Human Resources

Studium

2007 - 2010	Promotion an der LMU München
2002 - 2007	Psychologie-Studium an der LMU München Abschluss: Dipl. Psychologin
1997 - 2002	BWL-Studium an der FH-Landshut und der Anglia Polytechnik University Chelmsford / England Abschluss: Dipl. Betriebswirtin (FH)

Berufliche Ausbildung und Berufstätigkeit (I)

1995 - 1997	Angestellte der Commerzbank München
1993 - 1995	Ausbildung zur Bankkauffrau bei der Commerzbank München Abschluss: Bankkauffrau

Bildungsweg

1984 - 1993	Gymnasium in Kiel bzw. Landshut Abschluss: Allgemeine Hochschulreife
1980 - 1984	Grundschule in Neunkirchen/Saar bzw. Kiel

Publikationsliste

- Böbel, E., Gerber-Braun, B., Binsler, M. J., Spörrle, M. & Försterling, F. (2005). *Wir bitten um ein differenziertes Urteil: Möglichkeiten der Reduktion von Nondifferentiation bei Ratingskalen* [Abstract]. In K. W. Lange, K.-H. Bäuml, M. W. Greenlee, M. Hammerl & A. Zimmer (Hrsg.), *Experimentelle Psychologie. Beiträge zur 47. Tagung experimentell arbeitender Psychologen* (S. 21). Lengerich: Pabst Science Publishers.
- Gerber-Braun, B., Böbel, E., Ringenberg, I., Spörrle, M. & Försterling, F. (2005). *Sag's durch den Raum – Der Einfluss von freier Raumvorgabe bei der Beantwortung offener Fragestellungen* [Abstract]. In K. W. Lange, K.-H. Bäuml, M. W. Greenlee, M. Hammerl & A. Zimmer (Hrsg.), *Experimentelle Psychologie. Beiträge zur 47. Tagung experimentell arbeitender Psychologen* (S. 57). Lengerich: Pabst Science Publishers.
- Gerber-Braun, B., Hassebrauk, A., Spörrle, M. & Försterling, F. (2004). *Ein Me(e/h)r von Zeilen? Wie "endlos" sollten open-ended questions sein?* [Abstract]. In D. Kerzel, V. Franz & K. Gegenfurtner (Hrsg.), *Beiträge zur 46. Tagung experimentell arbeitender Psychologen* (S. 86). Lengerich: Pabst Science Publishers.
- Gerber-Braun, B., Spörrle, M., Binsler, M. J. & Försterling, F. (2004). *Sag's durch die Zeile – Frageformateffekte bei offenen Antwortformaten.* [Abstract]. In T. Rammsayer, S. Grabianowski & S. Troche (Hrsg.), *44. Kongress der Deutschen Gesellschaft für Psychologie: 100 Jahre Deutsche Gesellschaft für Psychologie* (S. 31). Lengerich: Pabst Science Publishers.
- Gerber-Braun, B., Zielger, M. & Bühner, M. (2007). *Criterion validity of personality measures: Faking or trait?* 13th Biennial Meeting of the International Society for the Study of Individual Differences (ISSID). in press.
- Kau, H.-P., Gerber-Braun, B., Böbel, E. & Spörrle, M. (2005). *Using a multi-level tailored design process to develop a customer satisfaction survey for university evaluation* [Abstract]. In K. W. Lange, K.-H. Bäuml, M. W. Greenlee, M. Hammerl & A. Zimmer (Hrsg.), *Experimentelle Psychologie. Beiträge zur 47. Tagung experimentell arbeitender Psychologen* (S. 102). Lengerich: Pabst Science Publishers.
- Spörrle, M., Barth, K., Gerber-Braun, B., Stich, J. & Försterling, F. (2004). *Mit welchem Wort soll ich's benennen? Eine lexikalische und korpusbasierte Analyse deutscher Verben* [Abstract]. In D. Kerzel, V. Franz & K. Gegenfurtner (Hrsg.), *Beiträge zur 46. Tagung experimentell arbeitender Psychologen* (S. 251). Lengerich: Pabst Science Publishers.
- Spörrle, M., Gerber-Braun, B. & Försterling, F. (2007). *The Influence or Response Lines on Response Behavior in the Context of Open-Question Formats.* *Swiss Journal of Psychologie*, 66 (2), 103-107.
- Wimmer, B.M., Breugst, N., Gerber-Braun, B., & Woschée, R. (2009). *The EI-cing on the cake: Emotional intelligence and performance in contexts of emotional labour* [Poster]. 11th European Congress of Psychology (ECP), Oslo, July 7-10, 2009.