



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Differentialpsychologische Aspekte der Fakeability“

Verfasserin

Christa Brunflicker

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, im Oktober 2008

Studienkennzahl lt. Studienblatt: A 298

Matrikelnummer: 9701013

Studienrichtung lt. Studienblatt: Psychologie

Betreuer: Ao. Univ.-Prof. Mag. Dr. Martin E. Arendasy

Mein Dank gilt

Herrn Ao. Univ.-Prof. Mag. Dr. Martin Arendasy
für die fachliche Unterstützung und die Geduld, die er mit mir hatte,

den DirektorInnen, LehrerInnen, SchülerInnen
sowie meinem Bekanntenkreis
für ihre Hilfsbereitschaft bei der Durchführung der Studie,

meiner Freundin Frau Mag. Alexandra Stummer,
für Ihre fachliche und emotionale Unterstützung,
sie schaffte es immer wieder mich zu motivieren,

meiner Familie,
besonders erwähnen möchte ich meine Mutter Maria
und meine Schwester Michaela,
sie haben immer an mich geglaubt und mir Mut gemacht,
dankbar bin ich auch meiner verstorbenen Großmutter Maria Neumann,
von der ich so viel über und für das Leben lernen durfte.

Besonders dankbar bin ich meinem Lebensgefährten Wolfgang Tisch,
der auch in schweren Zeiten immer für mich da war
und mir dieses Studium ermöglicht hat.

Inhaltsverzeichnis

1.	<i>Einleitung</i>	1
2.	<i>Theoretische Ausgangslage</i>	4
2.1.	Differentielle Psychologie und Persönlichkeitspsychologie	4
2.1.1.	Einführung und Begriffsdefinition	4
2.1.2.	Inhaltliche Konzepte	6
2.1.3.	Gütekriterien	7
2.1.4.	Die Gewinnung empirischer Daten	8
2.2.	Tests im Persönlichkeitsbereich	10
2.2.1.	Subjektive Tests	10
2.2.2.	Response-Sets	11
2.2.3.	Kontrolle der Fehlerfaktoren	13
2.3.	Fakeability	15
2.3.1.	Untersuchungsmethoden	15
2.3.2.	Praxisbezogene Forschung	18
2.4.	Faktorentheorie der Persönlichkeit nach R. B. Cattell	21
2.4.1.	Extraversion-Introversion	25
2.4.2.	Emotionale Stabilität vs. Emotionale Labilität	26
2.5.	Das Repression-Sensitization-Konstrukt	28
2.5.1.	Beschreibung des Konstruktes	28
2.5.2.	Die Entwicklung des R-S-Konstruktes	29
2.5.3.	Unterschiede zwischen Repressern und Sensitizern	30
2.5.4.	Weiterentwicklung des R-S-Konstruktes	33
2.5.4.1.	Eindimensionale Erfassung	33
2.5.4.2.	Zweidimensionale Erfassung	34
2.6.	Selbstdarstellung	36
2.6.1.	Begriffsdefinitionen	36
2.6.2.	Selbstdarstellungstechniken	37
2.7.	Theoretische Grundlagen des Moralischen Urteils	40
2.7.1.	Die Stadien der Moral nach Piaget	40
2.7.1.1.	Die Regeln des Marmelspiels	41
2.7.1.2.	Heteronome und autonome Moral	41
2.7.2.	Die kognitive Entwicklungstheorie von L. Kohlberg	42
2.7.2.1.	Das Stufenkonzept	43

2.7.2.2.	Die Stufen der Entwicklung des moralischen Urteils	45
2.7.2.3.	Affekt und Kognition (Inhalt und Struktur)	48
2.7.2.4.	Moralische Dilemmata	50
2.7.3.	Die Stufen der Entwicklung des logischen Denkens nach Piaget	50
2.8.	Empirische Befunde	52
2.8.1.	Fakeability	52
2.8.2.	Differentialpsychologische Aspekte der Fakeability	55
3.	Empirische Untersuchung	56
3.1.	Fragestellung und Operationalisierung	56
3.2.	Instruktionsbedingungen	56
3.2.1.	Instruktion der Faking-bad-Bedingung	57
3.2.2.	Instruktion der Faking-good-Bedingung	57
3.3.	Ziele und Hypothesen der Untersuchung	58
3.3.1.	Hypothesen zur Faking-bad-Bedingung	58
3.3.2.	Hypothesen zur Faking-good-Bedingung	60
3.3.3.	Vermutete differentialpsychologische Unterschiede	61
3.4.	Auswahl und Begründung der Methoden	64
3.4.1.	Das Testmaterial	64
3.4.1.1.	Sozialdatenerhebung	64
3.4.1.2.	16-Persönlichkeits-Faktoren-Test Revidierte Fassung (Schneewind & Graf, 1998)	65
3.4.1.3.	Repression-Sensitization-Coping-Inventar (Huwe, Hennig & Netter, 1996)	69
3.4.1.4.	Moralisches Urteil Test (Lind, 1985)	71
3.4.1.5.	Impression Management Skala (Mummendey & Eifler, 1994)	74
3.5.	Versuchsanordnung und Durchführung	77
3.5.1.	Versuchsanordnung	77
3.5.2.	Durchführung	77
3.5.3.	Untersuchungsablauf	78
4.	Auswertung und Interpretation der Ergebnisse	79
4.1.	Statistische Deskription der Strichprobe	79
4.1.1.	Filtersetzung	79
4.1.2.	Gesamtstichprobe nach Filter	79
4.2.	Stichprobe der Faking-bad-Bedingung	79
4.2.1.	Verteilung von Geschlecht, Alter und Bildung	80
4.2.2.	Ergebnisse der Persönlichkeitsfragebogen	84
4.2.2.1.	Ergebnisse des 16 PF-R	84
4.2.2.2.	Ergebnisse des RSCI	85

4.2.2.3.	Ergebnisse der IMS	86
4.2.2.4.	Ergebnisse des MUT	87
4.3.	Stichprobe der Faking-good-Bedingung	89
4.3.1.	Verteilung von Geschlecht, Alter und Bildung	89
4.3.2.	Ergebnisse der Persönlichkeitsfragebogen	91
4.3.2.1.	Ergebnisse des 16 PF-R	91
4.3.2.2.	Ergebnisse des RSCI	93
4.3.2.3.	Ergebnisse der IMS	94
4.3.2.4.	Ergebnisse des MUT	95
4.4.	Prüfung der Hypothesen	96
4.4.1.	Geschlecht	98
4.4.2.	Alter	100
4.4.3.	Bildung	104
4.4.4.	Extraversion versus Introversion	107
4.4.5.	Emotionale Stabilität vs. Emotionale Instabilität	109
4.4.6.	Repression vs. Sensitization	110
4.4.7.	Moralische Urteilsfähigkeit	112
4.4.8.	Positive Selbstdarstellung	116
5.	Diskussion	118
5.1.	Untersuchungsergebnisse	118
5.1.1.	Differentialpsychologische Aspekte	118
5.1.2.	Allgemeine Fakeability	122
5.2.	Abschließende Bemerkungen	124
5.2.1.	Kritikpunkte	124
5.2.2.	Anregungen für zukünftige Untersuchungen	124
6.	Zusammenfassung	125
	Literaturverzeichnis	130

Abbildungsverzeichnis

Abb. 1:	Model of Applicant Faking (Snell, Sydell & Lueke, 1999, S. 222)	19
Abb. 2:	Diagramm-Darstellung vom hierarchischen Aufbau der Persönlichkeit (Häcker & Stapf, 1998, S. 267)	24
Abb. 3:	Unabhängigkeit von „Extraversion vs. Introversion“ und „Emotionaler Stabilität vs. Emotionaler Labilität“ (Häcker & Stapf, 1998, S. 267)	27
Abb. 4:	Prozentuelle Aufteilung der Probanden auf die Versuchsgruppen	80
Abb. 5:	Geschlechterverteilung in der Faking-bad-Bedingung	81
Abb. 6:	Prozentuelle Verteilung der Altersklassen in der Faking-bad-Stichprobe	82
Abb. 7:	Prozentueller Anteil der Bildungsklassen in der Faking-bad-Bedingung	83
Abb. 8:	Prozentueller Anteil der Intro- und Extravertierten in der Faking-bad-Bedingung	84
Abb. 9:	Prozentueller Anteil der emotional stabilen vs. emotional instabilen Personen in der Faking-bad-Bedingung	85
Abb. 10:	Prozentueller Anteil von Repressern und Sensitizern in der Faking-bad-Bedingung	86
Abb. 11:	Prozentueller Anteil der Probanden mit geringer bzw. mittlerer Tendenz zur positiven Selbstdarstellung	87
Abb. 12:	Prozentuelle Verteilung der Gruppen zur moralischen Urteilsfähigkeit	88
Abb. 13:	Prozentuelle Aufteilung der Probanden auf die Versuchsgruppen	89
Abb. 14:	Geschlechterverteilung in der Faking-good-Stichprobe	90
Abb. 15:	Prozentuelle Verteilung der Altersklassen in der Faking-good-Stichprobe	90
Abb. 16:	Prozentueller Anteil der Bildungsklassen in der Faking-good-Stichprobe	91
Abb. 17:	Prozentueller Anteil der Intro- und Extravertierten in der Faking-good-Stichprobe	92
Abb. 18:	Prozentuelle Verteilung von emotional stabilen und emotional instabilen Probanden in der Faking-good-Stichprobe	93

Abb. 19:	Prozentueller Anteil von Repressern und Sensitizern in der Faking-good-Stichprobe	94
Abb. 20:	Prozentueller Anteil der Probanden mit geringen bzw. mittleren IM-Tendenzen_____	94
Abb. 21:	Prozentuelle Verteilung der moralischen Urteilsfähigkeit_____	95

Tabellenverzeichnis

Tab. 1:	Cattells Beschreibung der bivariaten, klinischen und multivariaten Forschungsmethoden (Quelle: Pervin, 2000, S. 241)	23
Tab. 2:	Das Fünf-Faktoren-Modell der Persönlichkeit (Quelle: Amelang & Bartussek, 2001)	25
Tab. 3:	Angstabwehrmechanismen von Repressern und Sensitizern (Krohne, 1975, S. 58)	29
Tab. 4:	Ängstlichkeit und Angstleugnung bei Repressern und Sensitizern	32
Tab. 5:	Einteilung nach Crowne und Marlowe (Quelle: Amelang & Bartussek, 2001, S. 437)	35
Tab. 6:	Die vier Typen der Stressbewältigung	35
Tab. 7:	Assertive und Defensive IM-Strategien (Quelle: Mummendey, 1990)	38
Tab. 8:	Assertive und Defensive IM-Taktiken (Quelle: Mummendey, 1990)	39
Tab. 9:	Die vier verschiedenen Stadien der Anwendung der Regeln des Murmelspiels	41
Tab. 10:	Gegenüberstellung der Phasen nach Piaget mit den Stufen nach Kohlberg (Lind, 2000, S. 53)	45
Tab. 11:	Die drei Ebenen der Moralentwicklung (Lind, 2000, S. 50)	45
Tab. 12:	Der affektive Aspekt: Stufenmodell moralischer Motive nach Kohlberg (Lind, 2002a, S. 53)	46
Tab. 13:	Klassifikation der moralischen Urteile nach Entwicklungsebenen und –stufen (Quelle: Kohlberg, 1974, S. 60-61, Tab. 2)	47
Tab. 14:	Überblick über die Primärskalen des 16 PF-R	66
Tab. 15:	Die fünf Globalskalen des 16 PF-R	67
Tab. 16:	Primär- und Globalfaktorenstruktur des 16 PF-R	67
Tab. 17:	Normierungswerte männlich für die Skalen des 16 PF-R (vgl. Manual, S. 105)	68
Tab. 18:	Normierungswerte weiblich für die Skalen des 16 PF-R (vgl. Manual, S. 105)	68

Tab. 19:	Die vier resultierenden Extremgruppen von Stressbewältigungstypen des RSCI	70
Tab. 20:	Maße für den kognitiv-strukturellen Aspekt des MUT (C-Werte)	74
Tab. 21:	Rohwerte der IMS	76
Tab. 22:	Aufteilung der Probanden auf die Versuchsgruppen	80
Tab. 23:	Altersklassen in der Faking-bad-Stichprobe	81
Tab. 24:	Bildungsverteilung in der Faking-bad-Stichprobe	83
Tab. 25:	Anteil der Intro- und Extravertierten in der Faking-bad-Stichprobe	84
Tab. 26:	Anteil der emotional stabilen und emotional instabilen Probanden in der Faking-bad-Stichprobe	85
Tab. 27:	Anteil von Repressern und Sensitizern in der Faking-bad-Stichprobe	85
Tab. 28:	Tendenz zur positiven Selbstdarstellung in der Faking-bad-Bedingung	86
Tab. 29:	Verteilung der moralischen Urteilsfähigkeit in der Faking-bad-Stichprobe	87
Tab. 30:	Aufteilung der Probanden auf die Versuchsgruppen	89
Tab. 31:	Altersklassen in der Faking-good-Stichprobe	90
Tab. 32:	Bildungsverteilung in der Faking-good-Stichprobe	91
Tab. 33:	Anteil der Intro- und Extravertierten in der Faking-good-Stichprobe	92
Tab. 34:	Anteil der Emotional Stabilen und Instabilen in der Faking-good-Stichprobe	92
Tab. 35:	Represser und Sensitizer in der Faking-good-Stichprobe	93
Tab. 36:	Personen mit geringer bzw. mittlerer IM-Tendenz in der Faking-good-Stichprobe	94
Tab. 37:	Verteilung der Gruppen zum moralischen Urteil	95
Tab. 38:	Prüfung der H1A₁ innerhalb der Altersklassen	101
Tab. 39:	Prüfung der H1B₁ innerhalb der Altersklassen	101
Tab. 40:	Prüfung der H1C₁ innerhalb der Altersklassen	101
Tab. 41:	Prüfung der H1D₁ innerhalb der Altersklassen	102

Tab. 42:	Prüfung der H1E₁ innerhalb der Altersklassen	102
Tab. 43:	Prüfung der H2A₁ innerhalb der Altersklassen	102
Tab. 44:	Prüfung der H2B₁ innerhalb der Altersklassen	103
Tab. 45:	Prüfung der H2C₁ innerhalb der Altersklassen	103
Tab. 46:	Prüfung der H2D₁ innerhalb der Altersklassen	103
Tab. 47:	Prüfung der H2E₁ innerhalb der Altersklassen	104
Tab. 48:	Prüfung der H1A₁ innerhalb der Bildungsklassen	104
Tab. 49:	Prüfung der H1B₁ innerhalb der Bildungsklassen	104
Tab. 50:	Prüfung der H1C₁ innerhalb der Bildungsklassen	105
Tab. 51:	Prüfung der H1D₁ innerhalb der Bildungsklassen	105
Tab. 52:	Prüfung der H1E₁ innerhalb der Bildungsklassen	105
Tab. 53:	Prüfung der H2A₁ innerhalb der Bildungsklassen	106
Tab. 54:	Prüfung der H2B₁ innerhalb der Bildungsklassen	106
Tab. 55:	Prüfung der H2C₁ innerhalb der Bildungsklassen	106
Tab. 56:	Prüfung der H2D₁ innerhalb der Bildungsklassen	107
Tab. 57:	Prüfung der H2E₁ innerhalb der Bildungsklassen	107
Tab. 58:	Ergebnisse der Introvertierten in der Faking-bad-Stichprobe	108
Tab. 59:	Ergebnisse der Extravertierten in der Faking-bad-Stichprobe	108
Tab. 60:	Ergebnisse der Introvertierten in der Faking-good-Stichprobe	108
Tab. 61:	Ergebnisse der Extravertierten in der Faking-good-Stichprobe	109
Tab. 62:	Ergebnisse der Emotional Instabilen in der Faking-bad-Stichprobe	109
Tab. 63:	Ergebnisse der Emotional Stabilen in der Faking-bad-Stichprobe	110
Tab. 64:	Ergebnisse der Emotional Instabilen in der Faking-good-Stichprobe	110
Tab. 65:	Ergebnisse der Emotional Stabilen in der Faking-good-Stichprobe	110

Tab. 66:	Ergebnisse der Represser in der Faking-bad-Stichprobe	111
Tab. 67:	Ergebnisse der Sensitizer in der Faking-bad-Stichprobe	111
Tab. 68:	Ergebnisse der Represser in der Faking-good-Stichprobe	112
Tab. 69:	Ergebnisse der Sensitizer in der Faking-good-Stichprobe	112
Tab. 70:	Prüfung der H1A₁ innerhalb der MU-Klassen	112
Tab. 71:	Prüfung der H1B₁ innerhalb der MU-Klassen	113
Tab. 72:	Prüfung der H1C₁ innerhalb der MU-Klassen	113
Tab. 73:	Prüfung der H1D₁ innerhalb der MU-Klassen	113
Tab. 74:	Prüfung der H1E₁ innerhalb der MU-Klassen	114
Tab. 75:	Prüfung der H2A₁ innerhalb der MU-Klassen	114
Tab. 76:	Prüfung der H2B₁ innerhalb der MU-Klassen	115
Tab. 77:	Prüfung der H2C₁ innerhalb der MU-Klassen	115
Tab. 78:	Prüfung der H2D₁ innerhalb der MU-Klassen	115
Tab. 79:	Prüfung der H2E₁ innerhalb der MU-Klassen	116
Tab. 80:	Ergebnisse der Probanden mit geringer IM-Tendenz (FB)	116
Tab. 81:	Ergebnisse der Probanden mit mittlerer IM-Tendenz (FB)	117
Tab. 82:	Ergebnisse der Probanden mit geringer IM-Tendenz (FG)	117
Tab. 83:	Ergebnisse der Probanden mit mittlerer IM-Tendenz (FG)	117

1. Einleitung

Seit jeher war die Erfassung und Messung der Persönlichkeit von großem Interesse für die Psychologie. Die Differentielle Psychologie befasst sich mit der Art des Erlebens und Verhaltens des Individuums sowie den Ursachen und Bedingungen von Verhaltensunterschieden zwischen den Menschen (Amelang & Bartussek, 2001).

Die Erfassung von Persönlichkeitseigenschaften erfolgt zumeist mittels Persönlichkeitsfragebogen. Die Frage, ob Individuen ihre Antworten in einem Persönlichkeitsfragebogen verfälschen können, beschäftigt Psychologen seit vielen Jahren. Man hat festgestellt, dass Persönlichkeitsfragebogen sensitiv sind gegenüber einer absichtlichen Verfälschung der Antworten durch die befragten Personen. Die Verfälschung kann sowohl in eine positive Richtung (*fake good*) als auch in eine negative Richtung (*fake bad*) gehen (Amelang & Bartussek, 2001). Dieses Phänomen ist unter einer Vielzahl von Namen einschließlich *Response distortion*, *Impression Management*, *soziale Erwünschtheit* sowie *Selbstdarstellung* untersucht worden (Rosse, Stechner, Levin & Miller, 1998; Viswesvaran & Ones, 1999).

Eine Möglichkeit zur Erfassung der Fakeability beinhaltet das Vergleichen der Antworten von ein und derselben Person unter unterschiedlichen Instruktionsbedingungen. Hier wird auf die Bestimmung der maximalen Grenzen der Fakeability von verschiedenen Persönlichkeitsskalen gezielt (*Within-Subjects-Design*). Ein anderer Ansatz vergleicht die von verschiedenen Gruppen (Stellenbewerber, Gefangene, Studenten, ...) erhaltenen Antworten miteinander. Man spricht dann von einem *Between-Subjects-Design*. Hierzu gehört auch die Erforschung der Fakeability in tatsächlichen Bewerbungssituationen. Die vorliegende Arbeit konzentriert sich nur auf die Effekte von Instruktionsbedingungen in einem *Within-Subjects-Design* (Viswesvaran & Ones, 1999).

Wie bereits erwähnt, sind in bisherigen Untersuchungen verschiedene Instruktionsbedingungen benutzt worden, um festzustellen, ob Personen ihre Antworten auf einen Persönlichkeitsfragebogen verfälschen können. Die drei am häufigsten verglichenen Fragebogen-Instruktionen sind: *a) respond honestly* (ehrlich antworten), *b) fake good* bzw. so zu antworten, dass man einen möglichst *vorteilhaften Eindruck* hinterlässt *c) fake bad* bzw. so zu antworten, dass man um einen *ungünstigen Eindruck* macht (Viswesvaran & Ones, 1999).

Die zentrale Fragestellung der Untersuchung lautet: Besteht ein Zusammenhang zwischen der Fähigkeit einer Person zur Verfälschung von Persönlichkeitsfragebogen und spezifischen Merkmalen dieser Person? Die Untersuchung soll klären inwieweit demographische Daten wie Alter, Geschlecht und Bildung sowie Persönlichkeitseigenschaften wie Extra- und Introversion, Emotionale Stabilität vs. Emotionale Instabilität, Repression vs. Sensitization, das Ausmaß der moralischen

Urteilsfähigkeit und die Tendenz zur positiven Selbstdarstellung mit der Fakeability in Zusammenhang stehen.

Die vorliegende Arbeit befasst sich unter anderem mit der Persönlichkeitstheorie von R. B. Cattell. Cattells Theorie ist die bei weitem ausführlichste auf Faktorenanalyse basierende Theorie der Persönlichkeit. Spearman führte die grundlegenden Ideen der Faktorenanalyse ein, indem er 1904 einen Generalfaktor (z.B. allgemeine Intelligenz) sowie einen spezifischen Faktor fand (z.B. Gedächtnis). Spearman isolierte bei seiner Methode einzelne Faktoren. Von Thurstone wurde diese Methode aufgegriffen und 1931 auf die Analyse multipler Faktoren erweitert (Hall, Lindzey & Campbell, 1997).

Cattell knüpft an die Vorarbeiten von Allport und Odbert an. Diese hatten 1936 „im Zuge einer psycholexikalischen Studie aus einem Wörterbuch 17953 Begriffe zur Kennzeichnung von Eigenschaften herausgesucht“ (Amelang & Bartussek, 2001, S. 315). In einem mehrstufigen Verfahren reduzierte Cattell diesen großen Variablensatz. Letztendlich erhielt er sechzehn Faktoren (Pervin, 2000).

Es folgt eine kurze Beschreibung der Globalskala *Extraversion* sowie der Primärskala *Emotionale Stabilität* von Cattells *16-Persönlichkeits-Faktoren-Test Revidierte Fassung* (16 PF-R).

Anschließend folgt eine Einführung in die Theorie über das Persönlichkeitskonstrukt *Repression-Sensitization*, welche in der Wahrnehmungsforschung ihren Ausgangspunkt hat. Represser und Sensitizer unterscheiden sich in der Art und Weise, wie sie eine bedrohliche Situation wahrnehmen bzw. wie sie mit dieser umgehen. Am repressiven Pol des eindimensionalen Kontinuums sind die so genannten „Verleugner“ lokalisiert, während am sensitiven Pol Personen lokalisiert sind, welche man als „Sensibilisierer“ bezeichnen könnte (Amelang & Bartussek, 2001). Weitere bestätigte bzw. vermutete Unterschiede zwischen Personen am repressiven und am sensitiven Pol werden dargestellt. Daraufhin folgt ein kurzer Überblick über die Weiterentwicklung des Repression-Sensitization-Konstruktes vom eindimensionalen Konzept hin zur zweidimensionalen Erfassung unter Einbeziehung von *Niedrig-Ängstlichen* und *Defensiv-Hochängstlichen* (Huwe, Hennig & Netter, 1996).

Nachfolgend werden kurz der Begriff Impression Management sowie eine Beschreibung des von Mummendey und Eifler (1994) eingeführten Konstrukts der *positiven Selbstdarstellung* erläutert. Man spricht von *Selbstdarstellungs-* oder *Impression-Management-Theorien*, wenn „die Beeinflussung der Personenwahrnehmung von Interaktionspartnern als Hauptzweck selbstbezogener Kognitionen oder bestimmendes Moment sozialer Interaktion gesehen wird“ (Mummendey, 1990, S. 127). Die Tendenz, sich selbst als positiv darzustellen, ist wohl die üblichste Art und Weise der Selbstpräsentation. Positive Selbstdarstellung meint „die Tendenz eines Individuums ..., andere Personen zur Zuschreibung von als günstig bewerteten Merkmalen an dieses Individuum zu veranlassen“ (Mummendey & Eifler, 1994, S. 9).

Die Darstellung der Theorie zur *moralischen Urteilsfähigkeit* ist der nächste Punkt. Es wird ein kurzer Überblick gegeben über die Arbeiten von Jean Piaget, dessen Erkenntnisse über die heteronome und autonome Moral wesentlich zu einer Theorie des moralischen Urteils beigetragen haben. Lawrence

Kohlberg hat seine Kognitive Theorie des moralischen Urteils auf die Arbeiten von Piaget aufgebaut und den Ansatz von Piaget differenziert und erweitert. Es wird das Kohlbergsche Stufensystem mit den sechs moralischen Stufen beschrieben, welche den *affektiven Aspekt* des moralischen Urteilens repräsentieren. Neben dem affektiven Aspekt existiert auch noch ein kognitiver Aspekt des moralischen Urteils. Der *kognitive Aspekt* meint die eigentliche moralische Urteilsfähigkeit, d. h. das Beurteilen der Qualität von Argumenten unabhängig von der eigenen Einstellung zu diesen (Heidbrink, 1991).

In der vorliegenden Untersuchung wird zunächst ein Abriss über das Thema Differentielle Psychologie und Persönlichkeitspsychologie im Allgemeinen präsentiert. Es erfolgt eine Beschreibung der Fehlerfaktoren von Persönlichkeitsfragebogen. Für die vorliegende Arbeit von besonderem Interesse ist der Fehlerfaktor *Verfälschbarkeit*. Verschiedene Ansätze, um eine Antwortverzerrung zu verhindern bzw. absichtlich herbei zu führen, werden dargestellt.

Diese Annahmen und Daten bilden die Grundlage der eigenen Untersuchung, in welcher die Fähigkeit zur Verfälschung von Persönlichkeitsfragebogen aus differentialpsychologischer Sicht untersucht wird. Die zugrunde liegenden Hypothesen, das Untersuchungsdesign sowie die Durchführung werden in den darauf folgenden Kapiteln näher erläutert. Den Abschluss bilden die statistische Auswertung der Daten anhand der aufgestellten Hypothesen sowie die Diskussion der daraus resultierenden Ergebnisse.

In dieser Arbeit werden für bestimmte Personengruppen (Forscher, Probanden, etc.) nur die männlichen Bezeichnungen verwendet. Dies dient lediglich zur besseren Lesbarkeit des Textes, die Autorin meint selbstverständlich beide Geschlechter.

2. Theoretische Ausgangslage

2.1. Differentielle Psychologie und Persönlichkeitspsychologie

2.1.1. Einführung und Begriffsdefinition

Dieses Kapitel dient der kurzen Definition und Erläuterung grundlegender Begriffe wie *Differentielle Psychologie*, *Person*, *Persönlichkeit*, *Persönlichkeitspsychologie*, *Persönlichkeitstheorie*, Messung der Persönlichkeit mittels *Fragebogen*, *Tests* etc. Zunächst folgen einige Begriffsdefinitionen zum besseren Verständnis.

Durch den Terminus **Differentielle Psychologie** wird ziemlich eindeutig jenes Teilgebiet der Psychologie bezeichnet, welches sich mit Unterschieden im Erleben und Verhalten zwischen den Menschen beschäftigt. Dieser Begriff scheint jedoch immer mehr durch den Ausdruck „Persönlichkeitspsychologie“ ersetzt zu werden, dieser wäre dann synonym zu Differenzieller Psychologie zu verstehen. Unter Persönlichkeitspsychologie versteht man aber sehr oft ein Teilgebiet der Differentiellen Psychologie. Es handelt sich hierbei um jenes Gebiet, welches sich nicht mit Unterschieden im Leistungsbereich beschäftigt, sondern mit Unterschieden im Bereich des Temperaments (Amelang & Bartussek, 2001).

Die Autorin hält es nach Amelang und Bartussek. In der vorliegenden Arbeit wird Differentielle Psychologie als Oberbegriff aufgefasst. Innerhalb dieses Gebietes erfolgt eine Unterscheidung in den Leistungs- und den Persönlichkeitsbereich, wobei für die vorliegende Arbeit allein der Persönlichkeitsbereich von Belang ist. Amelang und Bartussek halten eine Abgrenzung der Persönlichkeitspsychologie von der Differentiellen Psychologie jedoch für unerheblich (2001).

Die differentialpsychologische Betrachtung von Unterschieden bezieht sich nicht nur auf Unterschiede zwischen Personen zu einem gegebenen Zeitpunkt (Querschnittsbetrachtung interindividueller Unterschiede), obgleich dies ihr Hauptgegenstand ist. Auch Unterschiede innerhalb je einer Person in Abhängigkeit von bestimmten Situationen oder Zeitpunkten (Längsschnittbetrachtung intraindividueller Unterschiede) können Gegenstand der Differentiellen Psychologie sein, allerdings nur dann, wenn diese Abhängigkeit bestimmter Verhaltensweisen von situativen oder zeitlichen Gegebenheiten selbst wieder zwischen Individuen variiert ... (Amelang & Bartussek, 2001, S. 4)

Differentialpsychologische Fragestellungen befassen sich mit

- der Beschaffenheit von Merkmalen, in denen es interindividuelle Differenzen gibt
- dem Ausmaß dieser Differenzen
- der wechselseitigen Abhängigkeit solcher Merkmale
- den Ursachen der Differenzen sowie
- ihrer Beeinflussbarkeit durch Training, Umweltveränderungen, Medikamente etc. (Anastasi, 1966; zitiert nach Amelang & Bartussek, 2001, S. 4)

Die differentialpsychologische Betrachtungsweise beinhaltet nicht bloß Unterschiede zwischen einzelnen Personen, sondern auch Unterschiede zwischen Gruppen von Individuen. Der Terminus „Gruppe“ bedeutet hierbei, dass Personen, die ein oder mehrere gemeinsame Merkmale aufweisen, zu einer Gruppe zusammengefasst werden. Diese Gruppe wird dann mit Personen verglichen, bei denen dieses Merkmal bzw. diese Merkmale anders ausgeprägt sind. Gebräuchliche Klassifikationen in diesem Sinne wären bspw. das Geschlecht, das Alter, die Rasse, der Status etc. „Auch Unterteilungen nach dem Ausprägungsgrad in bestimmten Variablen wie Intelligenz, Leistungsmotiviertheit oder der emotionalen Stabilität, etwa in den Kategorien ‚hoch‘, ‚mittel‘ und ‚niedrig‘, können für eine entsprechende gruppenweise Zusammenfassung von Personen herangezogen werden“ (Amelang & Bartussek, 2001, S. 33).

Die **Persönlichkeitspsychologie** hat sich aus der Differentiellen Psychologie und aus der Charakterologie heraus entwickelt. Die **Person** (der einzelne Mensch, das Individuum) bildet den Ausgangspunkt der Persönlichkeitspsychologie (Häcker & Stapf, 1998).

Häcker und Stapf wiederum verstehen die Persönlichkeitspsychologie sowohl als Teil der Allgemeinen Psychologie als auch als eigenständige Grunddisziplin der Psychologie. Die Allgemeine Psychologie ist an der „Erstellung allgemeiner Gesetzmäßigkeiten“ interessiert, während die Persönlichkeitspsychologie „nach den Bedingungen für die individuellen Besonderheiten“ fragt (1998, S. 627).

Der Gegenstand der Persönlichkeitspsychologie ist die **Persönlichkeit**. Bislang konnte keine Einigung über eine allgemein gültige Definition der Persönlichkeit erzielt werden (Häcker & Stapf, 1998). Herrmann (1972) vertritt die Auffassung, dass es sich bei Persönlichkeit um ein „theoretisches bzw. hypothetisches Konstrukt“ handelt und „inhaltlich als *ein bei jedem Menschen eigenartiges, relativ stabiles und den Zeitlauf überdauerndes Verhaltenskorrelat*“ (zitiert nach Häcker & Stapf, 1998, S. 627).

Cattell versteht unter Persönlichkeit (zitiert nach Amelang & Bartussek, 2001, S. 314) „die Gesamtheit nichtsituativer Verhaltensbedingungen oder dasjenige, *which permits a prediction of what a person will do in a given situation*“.

Nach Mummendey (1990) wird unter *Persönlichkeit*

allgemein das Insgesamt der interindividuellen Unterschiede verstanden, wobei für eine psychologische Erforschung von Persönlichkeitsmerkmalen vor allem die einigermaßen situationsübergreifenden Unterschiede und Besonderheiten im Verhalten von Interesse sind. Die erfaßbaren Aspekte des Verhaltens, hinsichtlich deren Individuen sich unterscheiden und die nicht nur vollständig aktueller, vorübergehender Art sind, werden als *Persönlichkeitseigenschaften* bezeichnet. (S. 205)

Häcker & Stapf (1998) sehen das *Ziel jeder Persönlichkeitstheorie* darin

ihren Gegenstand nicht nur adäquat zu beschreiben, sondern auch zu erklären, d. h. auf seine Bedingungszusammenhänge zurückzuführen. Entsprechend soll eine P.theorie [Persönlichkeitstheorie; Anm. der Verfasserin] jegliches individuelle Erleben und Verhalten als ihren Spezialfall umfassen und – mit Kenntnis der jeweils wirkenden Bedingungen – vorhersagbar machen (S. 628).

Um dieses Ziel zu erreichen gibt es verschiedene Vorgehensweisen. **Nomothetische Ansätze** („allgemeine Gesetze aufstellende *erklärende Psychologie*“) werden von **idiographischen Ansätzen** („einmalige Abläufe beschreibende, von Sinnbezügen ausgehende *verstehende Psychologie*“) unterschieden (Häcker & Stapf, 1998, S. 628). Faktorentheorien der Persönlichkeit (z. B. von Cattell, Eysenck, Guilford) sind Beispiele für nomothetische Ansätze, geisteswissenschaftliche Typologien (z. B. von Allport) für idiographische Ansätze (Häcker & Stapf, 1998). Die idiographische Richtung betont die *Einzigartigkeit* des Individuums, während die nomothetische Richtung die *Gleichheit* der Individuen voraussetzt. Die *moderne Persönlichkeitstheorie* steht in gewisser Weise zwischen diesen beiden Richtungen. Dazu wäre zu bemerken: Menschen unterscheiden sich natürlich voneinander, doch sie tun dies entlang bestimmten quantifizierbaren und messbaren Dimensionen (Eysenck & Eysenck, 1987).

Die Vielfalt der Persönlichkeitstheorien lässt sich laut Häcker (1998, S. 628) „anhand der zu ihrer Erstellung angewandten Methoden (z.B. philosophische, psychoanalytische, faktorenanalytische), der in ihr als analytischer Einheit vorherrschenden Konstrukte (z.B. Eigenschaften, Faktoren, Einstellungen, Selbst) sowie der dominierenden Aspekte der Persönlichkeitsbetrachtung“ unterscheiden.

Bislang ist eine „übergreifende, allgem. [allgemein; Anm. d. Verf.] anerkannte Theorie“ der Persönlichkeitspsychologie noch ausständig. Die Anwendungsbereiche sind sehr breit gestreut (Häcker & Stapf, 1998, S. 629).

2.1.2. Inhaltliche Konzepte

In der Persönlichkeitspsychologie unterscheidet man zwei Klassen von Merkmalen: **traits** (Eigenschaften, Wesenszüge) und **states** (Zustände). Die *Eigenschaft (trait)* ist eine „konstante,

überdauernde und spezifische Art des Verhaltens. Sie kann entlang eines Kontinuums gemessen werden und wird sowohl zur Beschreibung von einzelnen Personen als auch zur Vorhersage ihres zukünftigen Verhaltens benutzt“ (Zimbardo, 1995, S. 476). Ein *Zustand (state)* hingegen wird als einzelnes Ereignis oder Vorkommnis verstanden, z. B. kann ein an und für sich geselliger Mensch sich gelegentlich auch ungesellig verhalten (Eysenck & Eysenck, 1987).

Die Interaktion zwischen Situationen und Eigenschaften erzeugt vorübergehend innere Bedingungen, welche als Zustände (states) bekannt sind. Sowohl Eigenschaften als auch Zustände der Persönlichkeit sind durch Fremd- oder Selbstbeurteilungen (Fragebogen) messbar (Eysenck & Eysenck, 1987).

Nach Eysenck und Eysenck (1987) sind Eigenschaften und Zustände

intervenierende oder vermittelnde Variablen, die zur Erklärung der individuellen Verhaltensunterschiede in dem Maße von Nutzen sind, wie sie sich in ein angemessenes theoretisches System einordnen lassen. Die Beziehung zwischen Eigenschaften oder Zuständen und dem Verhalten ist gewöhnlich eine indirekte, insofern sie durch die zwischen jenen und anderen bedeutenden Faktoren existierende Wechselwirkung beeinflusst oder gemäßigt wird. (S. 35-36)

2.1.3. Gütekriterien

Die Erfassung von Persönlichkeitseigenschaften erfolgt im Optimalfall mit Instrumenten, welche „hinsichtlich wichtiger Gütekriterien überprüft, also nach testtheoretischen Prinzipien konstruiert sind“ (Mummendey, 1990, S. 205).

Soll ein Fragebogen bzw. Test wirklich brauchbar sein, muss er bestimmten **Gütekriterien** entsprechen. Diese Gütekriterien sind Objektivität, Reliabilität und Validität (Amelang & Bartussek, 2001):

- **Objektivität** eines Tests bedeutet, dass die Testergebnisse von der Person des Auswerterers unabhängig sind.
- **Reliabilität** (Zuverlässigkeit) spricht man einem Test dann zu, wenn er dasjenige Persönlichkeitsmerkmal, das er vorgibt zu messen, exakt misst, d. h. bei wiederholter Anwendung unter gleichen Bedingungen zu gleichen Ergebnissen führt.
- **Validität** (Gültigkeit) des Tests besagt, dass er dasjenige Persönlichkeitsmerkmal, das er zu messen vorgibt, auch tatsächlich zuverlässig misst und nicht irgendein mit ihm verwandtes. So soll ein Fragebogen zur Erfassung der Angst nicht Aggressivität messen.

2.1.4. Die Gewinnung empirischer Daten

Um Informationen über die Persönlichkeit zu gewinnen können **unterschiedliche Verfahren bzw. Vorgehensweisen** angewandt werden (Amelang & Bartussek, 2001):

- Verhaltensbeobachtungen
- Selbst- vs. Fremdratings (Beurteilungen durch sich selbst oder durch andere Personen)
- Fragebogen
- Leistungstests
- Experimente
- Physiologische Messungen
- etc.

Bei den in der Persönlichkeitspsychologie angewandten Forschungsmethoden wird unterschieden zwischen **qualitativen** (Wesensschau, Deutung, Hermeneutik) und **quantitativen Methoden** (Beobachtung, Messen und Test, Experiment, Modellbildung, Simulation). Die Unterscheidung der beiden Methoden verliert laut Häcker und Stapf (1998) jedoch zunehmend an Bedeutung.

Mittlerweile gibt es eine Vielzahl von Persönlichkeitstheorien, auf deren Basis Messmethoden zur Erfassung der Persönlichkeit entwickelt bzw. erstellt wurden. Die Methoden, welche hauptsächlich Anwendung finden, sind Persönlichkeitsfragebogen und Persönlichkeitstests.

Ein **Persönlichkeitsfragebogen** (engl. *questionnaire, self-inventory*) ist laut Häcker und Stapf (1998) ein besonders konstruierter Fragebogen zur Erfassung der Ausprägung von Persönlichkeitseigenschaften. Die Elemente des Fragebogens werden als Testaufgabe oder Item bezeichnet. Mit diesen Items oder Feststellungen werden keine Tatsachen erfragt, sondern der Pb [Proband; Anm. d. Verf.] soll eine Art Selbstbeurteilung vornehmen, welche dann Auskünfte über sein eigenes Verhalten, seine Gefühle, seine Gepflogenheiten usw. gibt. Im Gegensatz zu den Befragungsbogen sind die Fragen bzw. Antworten auf die Fragen validiert, d. h. die Antworten stehen in einem geprüften Zusammenhang zu einer ps. [psychologisch; Anm. d. Verf.] bedeutsamen Aussage. (S. 626)

Bei Persönlichkeitsfragebogen unterscheidet man zwischen empirischen und rationalen Skalen. Unter **empirischen Skalen** versteht man Persönlichkeitsfragebogen, „welche ohne theoretische Orientierung konzipiert und anschließend einer empirischen Validierung unterzogen wurden“. Als **rationale Skalen** bezeichnet man die „mit persönlichkeits-theoretischen Vorstellungen verbundenen Persönlichkeitsfragebogen“ (Häcker & Stapf, 1998, S. 626).

Der entscheidende Vorteil von Selbstberichtsdaten liegt darin, dass viele Verhaltensbereiche erfasst werden können. Diese Art der Informationsgewinnung ist noch dazu äußerst ökonomisch (Mummendey, 1990).

Persönlichkeitstests definieren Häcker und Stapf (1998) als

Testverfahren (auf der Basis unterschiedlicher Testkonzeptionen konstruiert: z. B. projektiv oder psychometrisch), mit welchen empirisch abgrenzbare Dimensionen quantitativ und/oder qualitativ erfaßt werden können. Persönlichkeitsdimensionen (in diesem engeren Sinne) stellen solche Bereiche der Person dar, welche in Abgrenzung zu allgemeinen oder spezifischen Fähigkeits- oder Leistungsdimensionen diejenigen Dimensionen erfassen, welche die Art und Weise des Vollzugs von Verhalten bestimmen. Solche Dimensionen sind je nach Persönlichkeitssystem enger oder weiter gefaßte Faktoren wie z. B. Extraversion / Introversion, ps. Labilität / ps. Stabilität, ... (S. 629).

Beim **Experiment** werden mehrere unabhängige Variablen so in einem Forschungsdesign kombiniert (orthogonales Design), dass man ihren Einfluss auf eine bestimmte abhängige Variable unabhängig voneinander schätzen kann (Lind, 1993).

2.2. Tests im Persönlichkeitsbereich

2.2.1. Subjektive Tests

Sowohl Persönlichkeitsfragebogen wie auch Persönlichkeitstests werden zu den subjektiven Tests gezählt.

Häcker und Stapf (1998) definieren subjektive Tests als

eine Gruppe von Tests – meist Persönlichkeitstests -, bei denen im Unterschied zu Intelligenz- und anderen Fähigkeitstests gefordert wird, daß die Vp [Versuchsperson; Anm. d. Verfasserin] in der Lage ist, Auskunft über das eigene Erleben und Verhalten zu geben. S. T. [Subjektive Tests; Anm. d. Verfasserin] werden in der Regel als Fragebogen konzipiert und für die Persönlichkeitsdiagnostik und die Interessenmessung verwendet. (S. 845)

Persönlichkeitsfragebogen sind die bekanntesten und populärsten Verfahren zur Messung der Persönlichkeit. Sie sind jedoch oft transparent, d. h. es ist für Probanden oft klar, welche Konstrukte die Tests messen. Da die Probanden dann dazu in der Lage wären, Schlüsse zu ziehen, welche Konstrukte die Items messen könnten, könnte sie ihre Antworten in eine positive Richtung verfälschen (Khorramdel & Kubinger, 2006). Die *Transparenz* von Persönlichkeitsfragebogen ist ein häufig angeführter Grund, diese nicht für Entscheidungen bei der Personalauswahl einzusetzen. Es wurde nachgewiesen, dass Probanden ihre Antworten dahingehend verfälschen können, um einen vorteilhaften (fake good) oder ungünstigen (fake bad) Eindruck von sich zu erzeugen (Furnham, 1986; 1990). Die Mittelwerte zwischen Gruppen von Personen, die einen Persönlichkeitsfragebogen ehrlich beantworten, unterscheiden sich signifikant von denen, die durch verfälschte Antworten einen vorteilhaften oder ungünstigen Eindruck erzeugen (Topping & Gorman, 1997).

Persönlichkeitsfragebogen beinhalten eine Reihe von kurzen Statements über Verhaltensweisen. Die Probanden sollen angeben, ob die jeweilige Verhaltensweise für sie zutreffend ist oder nicht. Seit der Einführung von Persönlichkeitsfragebogen Anfang des 20. Jahrhunderts werden sie eingesetzt für die Personalauswahl, in klinischen Settings sowie in gerichtlichen Entscheidungen. Hinsichtlich der Effektivität von Persönlichkeitsfragebogen gibt es jedoch Kontroversen. Befürworter der Tests argumentieren, dass sie zufrieden stellende Reliabilität und Validität haben, während Kritiker auf ihre Oberflächlichkeit und auf ihre Anfälligkeit zur vorsätzlichen Verzerrung verweisen (Topping & Gorman, 1997).

Mummendey (1995) charakterisiert das Problem der Persönlichkeitsmessung mittels Fragebogen folgendermaßen:

Fragebogen werden gewöhnlich als „*reaktive*“ Meßinstrumente bezeichnet, weil der Gegenstand der wissenschaftlichen Untersuchung, nämlich die antwortende bzw. reagierende Person, das Ergebnis der Messung selbst wissentlich oder unabsichtlich *beeinflussen* kann. Je nach dem Grad an Absichtlichkeit und je nach dem Ausmaß der Veränderung des Antwortverhaltens aufgrund des Wissens um die eigene Rolle als Versuchsperson kann man von *Beschönigungstendenzen* bis zu *Verfälschungstendenzen* sprechen. (S. 159)

2.2.2. Response-Sets

Häcker & Stapf (1998) beschreiben das **Response-set** als

Antworttendenz, wird die Antwort eines Pbn [Probanden; Anm. d. Verf.] auf die Testaufgabe eher durch die Form der Testaufgabe hervorgerufen, ist diese Antworttendenz konsistent und verfälscht das Testergebnis, so spricht man nach Cronbach (1946) von r.s. ... Tendiert die Vp eher dazu eine Ja-Antwort zu bevorzugen, so spricht man von Ja-Sage-Tendenzen, Zustimmungstendenzen oder acquiescence. Läßt die Testantwort eine unentschiedene Antwort zu, so konnte man feststellen, daß bestimmte Vpn diese Antwortkategorie bevorzugen. Auch die Wahl extremer Antwortkategorien ist als eine solche Antworttendenz entdeckt worden. Ein sehr systematisch untersuchter r.s. ist die soziale Erwünschtheit. Sie ist dadurch gekennzeichnet, daß der Pb die seiner Meinung nach sozial erwünschte Antwort gibt.... Da die r.s. die Meßintention der Tests maßgeblich beeinflussen, die Fehlervarianz also erhöhen, ist man bemüht, über die Konstruktion von Testverfahren solche r.s. auszuschalten oder durch entsprechende Korrekturmaßnahmen nach der Testdurchführung zu korrigieren. (S. 737)

Die **Fehlerfaktoren** im Persönlichkeitsfragebogen (Response-Sets) werden von Amelang und Bartussek (2001) unterteilt in:

1. *Absichtliche Verstellung (Verfälschung)*
2. *Soziale Erwünschtheit*
3. *Akquieszenz*
4. *Weitere Fehlerfaktoren* (bspw. Bevorzugung extremer Antwortkategorien)

Das Testkuratorium (1986, S. 359) definiert das Gütekriterium **Verfälschbarkeit** als das „Ausmaß, in dem ein Test die individuelle Kontrolle über Art und Inhalt der verlangten bzw. gelieferten Informationen ermöglicht.“

Persönlichkeitsfragebogen sind wie eingangs schon erwähnt also „sensitiv gegenüber einer absichtlichen Verfälschung in *jeder* Richtung. Instruktionen zum *faking good* bzw. *faking bad* haben Resultate zur

Folge, die sich nicht nur voneinander, sondern auch jeweils von den unter Normalinstruktion erhaltenen Mittelwerten unterscheiden“ (Amelang & Bartussek, 2001, S. 171).

Persönlichkeitsfragebogen sind häufig für Verfälschungsabsichten der Probanden nicht stabil, wie Fragebogenkonstrukteure selbst nachweisen konnten. Das Ausmaß der Verfälschung hängt jedoch vom Anlass der Testdurchführung ab (Häcker & Stapf, 1998). Karner (2002) sieht Probleme bei der Verwendung von Persönlichkeitsfragebogen, wenn die dabei erhaltenen Scores wichtige Konsequenzen für die Beantwortenden haben.

Darüber hinaus stellt anscheinend auch die instruktionsgemäße Übernahme verschiedener Rollen für die Probanden keinerlei Probleme dar z. B. wenn die Probanden dazu aufgefordert werden, „die Fragen so zu beantworten, wie dieses typische Angehörige einzelner Berufsgruppen ... vermutlich täten, und es zweckmäßig wäre, um eine in diesen Bereichen ausgeschriebene Stelle zu erhalten“ (Amelang & Bartussek, 2001, S. 171).

Testsituationen, welche negative Konsequenzen für die Testpersonen implizieren, fördern das absichtliche Verfälschen von Persönlichkeitsfragebögen. Dies konnte von Kubinger (2002) in drei Experimenten mit jeweils verschiedenen Persönlichkeitsinventaren sowie jeweils unterschiedlichen Antwortformaten (vier-kategoriell, dichotom, analog) nachgewiesen werden. Jedes Experiment erbrachte zumindest eine Skala, die verfälscht wurde (fake good), trotz der Tatsache, dass die erwarteten negativen Konsequenzen ziemlich klein waren. Vor allem das vier-kategorielle und das dichotome Antwortformat begünstigten die absichtliche Verfälschung.

Seiwald (2002) berichtet über vier Studien, bei denen Verfälschung von Persönlichkeitsinventaren nicht nur bei Erwachsenen sondern auch bei Kindern sowie bei verschiedenen Antwortformaten (dichotom, analog, Q-sort) auftritt. Gut ausgeklügelte Antwortformate wie das Q-sort können zwar die Akquieszenz reduzieren aber nicht notwendigerweise faking verhindern.

Khorramdel und Kubinger haben 2006 den Einfluss von drei Variablen auf die Scores von Persönlichkeitsskalen gemessen: time limits bei der Beantwortung, Antwortformat (dichotom versus analog) und Warnungen vor verfälschten Antworten. Das besondere an dieser Studie ist, dass es sich bei den Probanden tatsächlich um Bewerber für einen Job gehandelt hat. Khorramdel und Kubinger konnten zeigen, dass verschiedene Faktoren die Antworten auf Items von Persönlichkeitstests beeinflussen. Die Warnung vor unehrlichen Antworten brachte keinen nennenswerten Effekt. Das Antwortformat und die Interaktion zwischen Antwortformat und Zeitbegrenzung zeigten bei einigen der Skalen einen signifikanten Effekt. Ein analoges Antwortformat scheint gegenüber einem dichotomen Antwortformat vorzuziehen zu sein, wenn man Verfälschung verhindern will, es gibt jedoch keine Garantie dafür.

Die Verfälschung in Richtung **soziale Erwünschtheit** und die **Akquieszenz** stellen ebenso zwei wichtige *Response-Sets* dar. Bei der sozialen Erwünschtheit handelt es sich um „die universell vorhandene

Neigung, sozial erwünschte Verhaltensweisen und Einstellungen zu bevorzugen, bzw. für die eigenen auszugeben“ (Mittenecker, 1982, S. 100). Die Tendenz mit „Ja“ zu antworten, also einer Aussage zuzustimmen, wird als Akquieszenz bezeichnet.

Hofmann & Kubinger üben Kritik an der Vorgehensweise bei der Testkonstruktion (2001, S. 298): „Trotz verschiedenster Bemühungen um Lügen- bzw. so genannte „Validitäts“-Skalen in herkömmlichen Persönlichkeitsfragebogen ..., scheint sowohl in der Praxis als auch in der „praktizierten“ Vorgehensweise bei der Konstruktion psychologisch-diagnostischer Persönlichkeitsverfahren das Phänomen der Verfälschbarkeit ignoriert.“

Lügenskalen bestehen aus Items, welche Ansichten und Verhaltensweisen zum Inhalt haben, die entweder sozial wünschenswert sind, aber selten ausgeübt werden oder häufig ausgeübt aber sozial nicht erwünscht sind (Jackson & Francis, 1999).

2.2.3. Kontrolle der Fehlerfaktoren

Zur Kontrolle der Verfälschbarkeit durch Fragebogen-Instruktion sagt Mummendey (1995) etwa folgendes: Gewöhnlich beinhalten die üblichen Instruktionen für die Beantwortung von Fragebogen irgendeinen Zusatz. Der Proband wird darum gebeten, er möge möglichst „*offen und ehrlich*“ antworten.

Eine weitere Möglichkeit, absichtlichen Verfälschungstendenzen entgegen zu wirken, besteht in der Vorgabe einer ausführlichen **Zusatzinstruktion**. Darunter versteht Mittenecker (1982), dass

1. der Proband aufgefordert wird, in seinem eigenen Interesse möglichst spontane und aufrichtige Antworten zu geben, selbst wenn er bei einigen Fragen nicht verstehen sollte, inwiefern diese von Bedeutung sind,
2. die Bereitschaft des Versuchsleiters erklärt wird, die jeweiligen Ergebnisse mit jedem Probanden ausführlich zu besprechen und den Probanden zu beraten,
3. versichert wird, man irre sich, wenn man meine, durchschauen zu können, welche Antwort die günstigere sei,
4. man behauptet, dass selbst ein leichtes „Frisieren“ der Antworten erkennbar sei.

Das Beantworten von Fragebogen unter **Zeitdruck** ist eine eher indirekte Art, Verfälschungstendenzen durch Instruktion zu verringern zu versuchen. Die Annahme dahinter ist, wenn Probanden zur Eile angetrieben werden, antworten sie eher ehrlich, da beschönigende Antworten längere Zeit in Anspruch nehmen würden. Mittlerweile gibt es jedoch empirische Ergebnisse, die genau das Gegenteil vermuten lassen (Mummendey, 1995; Schneider & Hübner, 1980; vgl. Kap. 2.3.1.)

Schneider und Hübner berichten über die Kontrolle von „social desirability“ durch

Verwendung von Zusatzinstruktionen. Obwohl es zum Teil ökonomische Überlegungen sein mögen, die Autoren vieler Fragebogen bewogen haben, die Aufforderung zur schnellen Beantwortung in die Instruktion aufzunehmen, vermuten wir dahinter die mehr oder minder explizite Annahme, daß eine längere Testbearbeitungszeit wohl vor allem zur Verfälschung der Antworten in eine sozial erwünschte Richtung bzw. zur Darstellung eines guten Eindrucks verwendet wird Unter diesem Blickwinkel kann der „Zeitdruck“, ..., ebenfalls als eine Maßnahme gegen Verfälschungstendenzen verstanden werden. (S. 566)

Wenn Persönlichkeitsbewertungen solcherart verwendet werden, dass die Befragten beruhend auf ihren Testwerten gerangreicht werden und diese Rangordnung dazu verwendet wird, um Personalentscheidungen zu treffen, so konnten individuelle Unterschiede in der sozialen Erwünschtheit die Rangordnung verdrehen und folglich Einstellungsentscheidungen beeinflussen (Dilchert, Ones, Viswesvaran & Deller, 2006).

In den heute verwendeten Persönlichkeitsinventaren schenkt man den beiden Fehlerfaktoren „soziale Erwünschtheit“ und „Akquieszenz“ Beachtung. Als Beispiel sei hier der **16 PF-R** zu nennen, dieser enthält sowohl eine **Impression Management Skala** zur Erfassung der sozialen Erwünschtheit wie auch eine **Akquieszenz-Skala**. Außerdem beinhaltet der 16 PF-R auch eine **Infrequenz-Skala**, mit deren Hilfe wird die Häufigkeit der Antwortalternative „?“ (bzw. „weiß nicht“) erfasst. Diese Verfälschungstendenzen werden zwar erhoben, eine nachfolgende Korrektur ist jedoch nicht möglich. Sofern die Testergebnisse auffällige Werte in diesen Skalen aufweisen, sind sie nur mit Vorsicht interpretierbar bzw. sollten nicht in eine Gesamtauswertung mit einbezogen werden. Laut Schneewind und Graf (1998) ist es wichtig mit dem jeweiligen Probanden abzuklären, warum er dementsprechend geantwortet hat, „gegebenenfalls ist eine erneute Testung erforderlich“ (S. 78).

2.3. Fakeability

Begriffe wie soziale Erwünschtheit, Impression Management, Verfälschung, absichtliche Verzerrung und self enhancement finden in der Literatur Verwendung, um response distortion zu beschreiben (Ones, Viswesvaran & Reiss, 1996; Viswesvaran & Ones, 1999; Graham, McDaniel, Douglas & Snell, 2002).

„Frankness, social desirability, claiming unlikely virtues, denying common faults and unpopular attitudes, exaggerating personal strengths, good impression, self-enhancement, and faking are all terms used to describe *response distortion*“ (Ones, Viswesvaran & Reiss, 1996, S. 660).

2.3.1. Untersuchungsmethoden

Es existieren zwei unterschiedliche Zugänge zur Erfassung der Fakeability: **Between-Subjects-Designs** und **Within-Subjects-Designs**. Beim Vergleich zwischen Individuen (*Between-Subjects*) beantwortet eine Gruppe der Probanden wahrheitsgetreu, während die andere Gruppe entweder zum faking good oder zum faking bad instruiert wird. Das *Within-Subjects-Design* ist so angelegt, dass ein und dieselbe Versuchsperson denselben Fragebogen mehrmals beantworten muss, einmal wahrheitsgemäß sowie unter einer faking good und/oder faking bad-Instruktion. Hierbei können jedoch Testeffekte auftreten (Viswesvaran & Ones, 1999).

In der Literatur zum Thema Verfälschung können mehrere Forschungsrichtungen unterschieden werden (Viswesvaran & Ones, 1999). Eine Forschungsrichtung beinhaltet *experimentelle Studien* (Martin et al., 2002), bei denen die, unter **verschiedenen Instruktionsbedingungen** (honest, fake good, fake bad) erhaltenen, Antworten verglichen werden. Diese Studien werden in *Laboruntersuchungen* durchgeführt. Sie haben zum Ziel die maximalen Grenzen der Fakeability aufzuzeigen. Testwerte, die unter verschiedenen Instruktionsbedingungen erreicht wurden, werden miteinander verglichen. Das liefert den Forschern Daten über die maximale Verfälschbarkeit von Persönlichkeitsfragebogen und –tests (Viswesvaran & Ones, 1999). Diese experimentellen Studien verwenden entweder Within-Subjects-Designs oder Between-Subjects-Designs. Mittlerweile besteht für die Forscher kein Zweifel mehr daran, dass Personen dazu in der Lage sind ihre Antworten auf Persönlichkeitsfragebogen zu verfälschen und somit ihre Testwerte zu erhöhen, wenn sie dazu aufgefordert werden. In einer Meta-Analyse über die Verfälschung von Persönlichkeitsfragebogen fand man Testwerte, welche durch Verfälschung um fast eine halbe Standardabweichung erhöht wurden. Jedoch konnte hiermit nicht bewiesen werden, dass Bewerber in realen Einstellungs-Situationen auch tatsächlich faken (Rosse, Stechner, Levin & Miller, 1998).

Eine zweite Forschungsrichtung bedient sich so genannter *Felduntersuchungen*. Hier werden Antworten von verschiedenen Personengruppen (Studenten, Bewerber, im jeweiligen Unternehmen eingestellte Personen) verglichen. Diese Studien haben zum Ziel, das Niveau der Antwortverzerrung in realen

Situationen zu bestimmen. Vertreter dieser Forschungsrichtung sehen die Ergebnisse der Laboruntersuchungen im Vergleich zur tatsächlichen Antwortverzerrung in realen Auswahl-Situationen als „worst-case-scenario“ (Rosse et al., 1998; Peeters & Lievens, 2005).

Um die Frage zu beantworten, ob Personen ihre Antworten in einer sozial erwünschten Weise verfälschen können, werden in Untersuchungen häufig die Richtungen, in welche die Antworten verfälscht werden sollen, vorgegeben (fake good, fake bad, antworten wie ein Jobbewerber). Am häufigsten werden unterschiedliche Instruktionsbedingungen zur Untersuchung der Fakeability verwendet, da die Forscher der Annahme sind, bei dieser Methode die maximalen Grenzwerte von Antwortverzerrungen zu erhalten (Dilchert et al., 2006).

Personen sind dazu in der Lage ihre Testscores bei Persönlichkeitsfragebogen zu ändern, wenn sie instruiert wurden bestimmte Eindrücke zu erzeugen z.B. fake good, fake bad, ehrlich antworten oder ein Bewerber für einen Job zu sein (Schmit & Ryan, 1993).

Bei herkömmlichen Persönlichkeitsfragebogen erkennen die Probanden oft die Eigenschaften, welche gemessen werden. Hofmann und Kubinger führten hierzu eine Untersuchung mit dem NEO-FFI durch. Sie stellten signifikante Mittelwertsunterschiede fest „zwischen Testwerten von Personen, die den Fragebogen praktisch komplett durchschauten, und solchen, die höchstens einzelne von ihm gemessene Eigenschaften identifizieren“ (2001, S. 303). Die grundsätzliche Verfälschbarkeit von Persönlichkeitsfragebogen konnte mit den Ergebnissen deutlich gemacht werden. Hofmann und Kubinger geben jedoch zu bedenken, „dass die Beantwortung eines Persönlichkeitsfragebogens stets im Kontext der damit verbundenen Konsequenzen gesehen werden muss“ (2001, S. 303).

Schneider und Hübner (1980) untersuchten den Einfluss von Verfälschungsinstruktionen auf die Bearbeitungsdauer eines Persönlichkeitsfragebogens anhand des Freiburger Persönlichkeitsinventars (FPI-k). Sie konnten signifikante Unterschiede in den Bearbeitungszeiten bezüglich der drei Instruktionsbedingungen feststellen. Ihnen gelang somit der Nachweis, dass die mittleren Bearbeitungszeiten die Relation „Guter Eindruck“ < „Kontrollbedingung“ < „Schlechter Eindruck“ ergeben. Dieses Ergebnis führten sie darauf zurück, dass jeder Mensch im Alltag versucht einen möglichst guten Eindruck von sich bei anderen zu hinterlassen.

Die durchschnittliche Bearbeitungszeit war bei der Aufforderung einen „guten Eindruck“ zu hinterlassen am kürzesten. Schneider und Hübner (1980) vermuteten deshalb, dass Zusatzinstruktionen wie Tempozusätze die Beantwortung in eine positive Richtung verfälschen. Die lange Bearbeitungszeit bei der Instruktion „schlechter Eindruck“ wird dadurch erklärt, dass im Menschen Abwehrreaktionen und innere Konflikte ausgelöst werden, wenn man ihn auffordert bei der Beantwortung eines Persönlichkeitsfragebogens ein schlechtes Bild von sich zu erzeugen.

Tempozusätze sollten eigentlich dazu dienen die Verfälschung der Antworten zu verhindern, und somit eine wahrheitsgemäße Beantwortung gewährleisten. Diese Maßnahme erscheint im Hinblick auf die theoretische Herleitung und empirisch gesicherte Wirksamkeit allerdings fragwürdig (Schneider & Hübner, 1980).

Die benötigte Zeit für die Beantwortung der Items kann scheinbar nicht bei allen Assessment-Werkzeugen herangezogen werden, um Faking zu ermitteln (Kluger, Reilly & Russell, 1991).

Wenn sich ein Proband dazu entschließt nicht wahrheitsgemäß zu antworten, kann zusätzlicher Zeitaufwand für die erfolgreiche Bearbeitung des Prozesses der Verfälschung erforderlich sein. Bei nicht wahrheitsgemäßer Beantwortung eines Items kommt zusätzlich zu der Zeit für wahrheitsgemäße Beantwortung noch die Zeit dazu, die jemand benötigt, um das Item verfälschend zu beantworten, behaupten wiederum Griffith und McDaniel (2007).

Marcus (2006) weist darauf hin, dass die Höhe des verfälschten Scores abhängig ist vom True-Score. Ein True-Score, der ohnehin schon hoch ist, schränkt die Möglichkeit einer Score-Erhöhung ein.

Marcus (2006) übt Kritik an den herkömmlichen Untersuchungen zur Verfälschbarkeit von Persönlichkeitsfragebogen. Seine Kritik bezieht sich darauf, dass die jeweiligen Versuchspersonen mit einer Situation konfrontiert werden, in welcher sie durch Verfälschung weder etwas zu gewinnen noch zu verlieren haben.

Lässt man unsystematische Fehlerkomponenten und systematische Fehler beiseite, so setzt sich der Testscore eines Bewerbers auf einem verfälschbaren Messinstrument wie einem Persönlichkeitstest zusammen aus dem wahren Wert des Bewerbers auf dem gemessenen Trait und einem zusätzlichen „faking Score“. Theoretisch kann der verfälschte Testscore auch negative Werte (bekannt als faking bad) annehmen, obwohl es eher unwahrscheinlich ist, dass das Phänomen des faking bad im Kontext der Personalauswahl auftritt (Marcus, 2006).

Marcus (2006) bedauert, dass die Forschung bis dato sehr wenig über die Varianz der absichtlichen Antwortverzerrung in realen Bewerbungs-Situationen weiß. Aus verschiedenen Gründen erlauben die wenigen Studien, die sich direkt mit der Varianz der Verfälschung befassten, keine Überprüfung der Abweichung des True Score vom verfälschten Score in realen Bewerbungs-Settings.

Das experimentell induzierte Verfälschen von Persönlichkeitstests führt zu einem Sammeln von Testwerten am oberen Ende der Verteilung, da die Probanden instruiert werden, so hohe Testergebnisse wie möglich zu erzielen. Würde man die beiden Werte vergleichen, würde dies wahrscheinlich zu sehr hohen negativen Korrelationen zwischen den True Scores und den verfälschten Scores führen (Marcus, 2006).

In der Forschung über das Faking wurden oftmals die ermittelten Scores auf Skalen zur **sozialen Erwünschtheit** herangezogen, um Verfälschung zu definieren. Die Forscher gingen dabei davon aus, dass Personen mit hohen Werten auf Skalen zur Sozialen Erwünschtheit ihre Antworten verfälschen. Bei Personen, die niedrige Scores auf Skalen zur Sozialen Erwünschtheit haben, waren die Forscher der Ansicht, diese Personen hätten ehrlich geantwortet (Graham et al., 2002). Als wichtiger Kritikpunkt wäre hier zu nennen, dass soziale Erwünschtheit nicht mit Verfälschung gleichzusetzen ist.

2.3.2. Praxisbezogene Forschung

Laut Griffith und McDaniel (2007) hat sich der Gebrauch von Persönlichkeitsfragebogen als Instrumente für die Personalselektion in den letzten 20 Jahren massiv erhöht. Selbstratings werden von Organisationen aus mehreren Gründen bevorzugt. Zunächst haben sich viele Persönlichkeitsmessungen als nützlich hinsichtlich personalverwandter Entscheidungen herausgestellt. Ein weiteres Argument für die Verwendung von Persönlichkeitstests liegt in deren ökonomischem Aspekt. Sie sind einfach zu handhaben und können auch einer großen Gruppe von Bewerbern gleichzeitig vorgelegt werden. Selbstratings wie Persönlichkeitstests haben weniger nachteiligen Einfluss als alternative Methoden der Personalauswahl wie beispielsweise kognitive Eignungsprüfungen.

Obgleich Persönlichkeitsfragebogen im Allgemeinen als wirkungsvolle Werkzeuge für die Personalauswahl angesehen werden wurde Kritik an ihnen geübt, da sie von den Jobbewerbern verfälscht werden können. Dieses Phänomen ist unter einer Vielzahl von Namen einschließlich *Response distortion*, *Impression Management*, *soziale Erwünschtheit* sowie *Selbstdarstellung* untersucht worden (Rosse, Stechner, Levin & Miller, 1998; Viswesvaran & Ones, 1999). Obwohl diese Bezeichnungen begrifflich verschieden sind, betreffen sie in der Praxis alle die Verfälschung von Testwerten auf einem Persönlichkeitsinventar. Im Zuge einer Stellenbewerbung erhöhen Antwortende häufig ihre Testscores auf positiven Attributen und senken ihre Testscores auf nicht wünschenswerten Merkmalen. Zum gegenwärtigen Zeitpunkt wissen die Forscher wenig über die Auswirkungen, die mit verfälschendem Antwortverhalten verbunden sind. So existieren bislang wenige Antworten über die Natur des verfälschenden Antwortverhaltens und dem tatsächlichen Ausmaß der Verfälschung. Deshalb schlugen Forscher vor, man solle in der Praxis auch Stellenbewerber berücksichtigen, die verfälschende Antworten geben. Dennoch solle man bei der Personalauswahl darum bemüht sein Verfälschungen zu reduzieren (Griffith & McDaniel, 2007).

Schließlich kann es sein, so geben Griffith und McDaniel (2007) zu bedenken, dass sich Fakers als Angestellte kontraproduktiv zur Organisation verhalten werden. Aber bislang gibt es weder Beweise, die diese Position unterstützen, noch konnte das Gegenteil bewiesen werden. Zurzeit ist wenig über die Eigenschaften von Personen bekannt, die beschließen, ihre Testscores durch Verfälschung zu erhöhen. Bis heute ist die Wissenschaft beim Aufbau eines grundlegenden Verständnisses des Phänomens der Fakeability ineffizient gewesen, so lautet die Kritik von Griffith und McDaniel.

Martin, Bowen und Hunt (2002) definieren Verfälschung als vorsätzlichen Versuch das eigene Persönlichkeitsprofil dahingehend zu verfälschen, damit es der Vorstellung einer idealen Persönlichkeit [für einen spezifischen Job; Anm. der Verf.] entspricht. Diese Technik der Verfälschung hat ihren Fokus auf der Fähigkeit der Probanden ihre Antworten zu ändern, aber nicht notwendigerweise auf der Effektivität dieser Veränderungen.

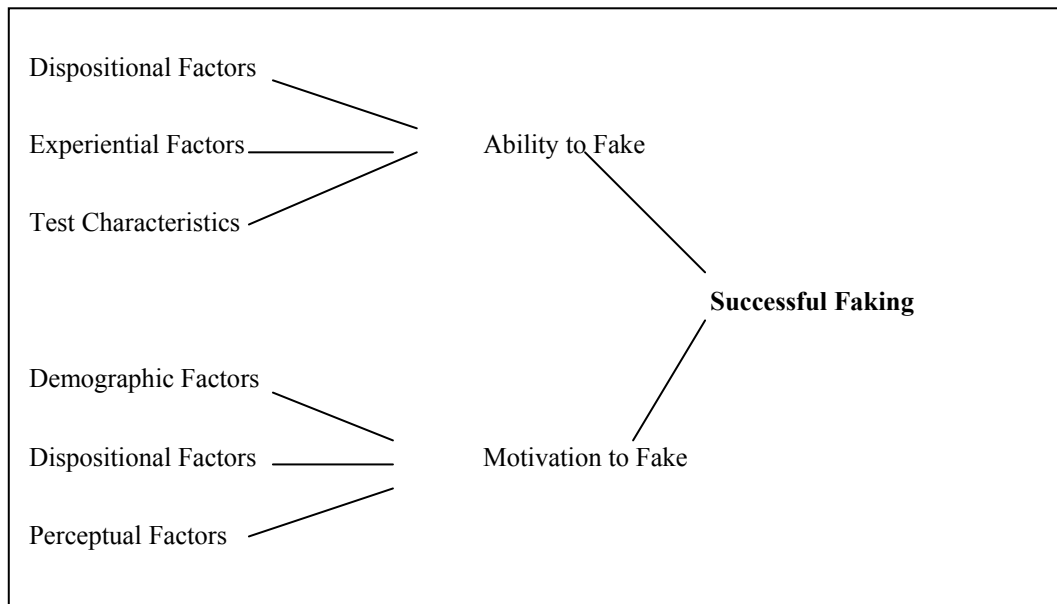


Abb. 1: Model of Applicant Faking (Snell, Sydell & Lueke, 1999, S. 222)

Die Forschung über die Verfälschung bzw. absichtliche Antwortverzerrung, hat in der Personalauswahl eine lange Tradition (Ones et al., 1996). Das Verfälschen eines Auswahl-Messinstrumentes kann als eine bewusste Antwortverzerrung einer Person definiert werden, um günstig dazustehen (Peeters & Lievens, 2005). Das Faking ist besonders unter Verwendung von nicht-kognitiven Messinstrumenten wie Persönlichkeitstests, biographischen Inventaren (Lebenslaufdaten) und Integritätstests untersucht worden (Martin et al., 2002).

Die Debatte über das Ausmaß, in dem sich Bewerber tatsächlich mit der absichtlichen Antwortverzerrung bei Einstellungstests beschäftigen, muss noch geklärt werden. Einige Studien haben berichtet, dass Bewerber Persönlichkeitstests nicht verfälschen, und selbst wenn sie tun, ist die Validität der Tests davon nicht betroffen (Ones et al., 1996). Andere Studien wiederum bestätigen, dass Verfälschung in Auswahl-Situationen vorkommt und die Kriteriumsvalidität von Persönlichkeitstests (Kluger et al., 1991; Rosse et al., 1998) dadurch gemindert wird. Zudem haben Forscher ihre Aufmerksamkeit auf praxisnahe Ergebnisse in Zusammenhang mit der Verfälschung gerichtet. Sie entdeckten, dass absichtliche Antwortverzerrung einen bedeutenden Einfluss darauf haben kann, welcher Bewerber eingestellt wird (Rosse et al., 1998).

Faking kann durch Situationen hervorgerufen werden. Es ist zum Beispiel gut dokumentiert, dass Personen, welche vom Versuchsleiter zum Verfälschen angewiesen wurden, diese Anweisung befolgen. Ausgehend von diesen Daten, könnte man nun auch annehmen, dass Personen auch effektiv verfälschen können, wenn sie durch Eigeninteresse dazu motiviert sind. Dies wäre beispielsweise der Fall, wenn von Personen in eine positive Richtung verfälschte Antworten als notwendig erachtet werden, um eine Arbeitsstelle zu erhalten (Graham et al., 2002).

Zickar, Gibby und Robie (2004) liefern Ergebnisse, die darauf hindeuten, dass vorherige Annahmen über die Natur der absichtlichen Antwortverzerrung auf Persönlichkeitsinventaren zu einschränkend gewesen sind. In den letzten paar Jahren zeigte die Forschung großes Interesse daran, die Natur und die Effekte der Verfälschung in Bewerber-Einstellungen zu untersuchen. Fast alle Faking-Studien haben sich auf die Unterschiede zwischen Befragten konzentriert, die hoch motiviert waren verfälschende Antworten zu geben (Stellenbewerber oder Personen, die aufgefordert wurden zu verfälschen) und solchen, die wenig dazu motiviert waren (bereits eingestellte Personen oder Personen, die angewiesen wurden aufrichtig zu antworten). In einer Meta-Analyse über die Fakeability von Persönlichkeitsfragebogen fanden Viswesvaran und Ones (1999) heraus, dass die Mittelwerte von Bewerber-Stichproben um 0,48 bis 0,65 Standardabweichungen über den Mittelwerten der Angestellten-Stichproben lagen.

Eine dichotome Unterscheidung zwischen Personen, die verfälschen, und solchen, die dies nicht tun, vereinfacht die kontinuierliche Verteilung dieser Variable grob.¹

Mehrere Meta-Analysen haben durchwegs bestätigt, dass durch die Durchführung von Persönlichkeitstests und anderen verfälschbaren Auswahl-Instrumenten in Bewerbungs-Situationen die Kriteriumsvalidität nicht verringert wird. Dies hatte eine Debatte darüber zufolge, ob der absichtlichen Verfälschung möglicherweise verdeckte Effekte bei Ergebnissen der Personalauswahl, wie Einstellungsentscheidungen und Entscheidungsgenauigkeit, zugesprochen werden können (Marcus, 2006).

Die Bedeutsamkeit von Studien über die Verfälschung bei Jobbewerbern ist abhängig vom Einfluss der Verfälschung auf die Testscores. In einer Bewerbungssituation scheint es so zu sein, dass die Mehrheit der Bewerber den höchstmöglichen Score erreichen will. Ist sowohl dieser Wunsch als auch die Fakeability bei allen Bewerbern gleich ausgeprägt, so wird durch das Faken bloß eine Konstante zu jedermanns Testwert hinzugefügt. Die Bedeutung des Fakens als Faktor im Auswahlverfahren wäre in diesem Falle nur minimal (Nguyen, Biderman, McDaniel, 2005).

¹ vgl. Zickar et al., 2004, die einen genaueren Gruppierungsansatz verfolgen, der jedoch Übergänge zwischen den Gruppen nicht ausschließt.

2.4. Faktorentheorie der Persönlichkeit nach R. B. Cattell

Nach Häcker und Stapf (1998) versteht man unter den Faktorentheorien der Persönlichkeit „diejenigen Persönlichkeitstheorien, deren Beschreibungsdimensionen (Persönlichkeitsfaktor) mittels der Methode der Faktorenanalyse die untersuchten Persönlichkeitsvariablen kategorisieren. Die gewonnenen Persönlichkeitsfaktoren hängen stark von den unterschiedlichen Ausgangsdaten (...) ab“ (S. 264).

Die grundlegenden Ideen der Faktorenanalyse wurden eingeführt von Spearman, der 1904 einen Generalfaktor (z. B. allgemeine Intelligenz) sowie einen spezifischen Faktor fand (z. B. Gedächtnis). Spearman isolierte bei seiner Methode einzelne Faktoren. Thurstone hat diese Methode aufgegriffen und 1931 auf die Analyse multipler Faktoren erweitert (Hall et al., 1997). Auch wenn Thurstone und andere Pionierarbeit leisteten, war es in Amerika „vor allem J. P. Guilford, der das faktorenanalytische Studium der Persönlichkeit auf die Beine brachte“ (Eysenck & Eysenck, 1987, S. 65). J. P. Guilford, H. J. Eysenck und R. B. Cattell verfolgten bei ihren Studien der Persönlichkeit mittels Faktorenanalyse das hochgesteckte Ziel, „der Persönlichkeit in ihrer Gesamtheit gerecht zu werden“ (Amelang & Bartussek, 2001, S. 308).

Cattell ist dem *nomothetischen Traitansatz* verpflichtet, als wichtigstes Mittel des Erkenntnisgewinns benutzt er die Faktorenanalyse (Amelang & Bartussek, 2001). Cattells Theorie ist bei weitem die ausführlichste auf Faktorenanalyse basierende Theorie der Persönlichkeit. Seine Suche nach den Hauptfaktoren der Persönlichkeit hat sich über Jahrzehnte erstreckt (Eysenck & Eysenck, 1987). Seine Arbeiten hierzu sind derart umfangreich, dass sich die Autorin mit der Erläuterung seines - für die vorliegende Arbeit relevanten - Persönlichkeitsfragebogens (16 PF-R; 16-Persönlichkeits-Faktoren-Test Revidierte Fassung) begnügt.

Allport und Odbert haben 1936 nahezu 18.000 persönlichkeitsrelevante Begriffe des „Webster’s New International Dictionary“ (Ausgabe von 1925) gesammelt. Als Grundlage für Cattells Studien zur Entwicklung eines Modells der Persönlichkeit dienten die rund 4.500 allgemeinen und stabilen Eigenschaftswörter, die so genannten personal traits (Hall et al., 1997) sowie die circa 100 Begriffe umfassende Kategorie „passing activities and temporary states“ (Amelang & Bartussek, 2001, S. 365). In einem mehrstufigen Verfahren reduzierte Cattell diesen großen Variablensatz und hat letztlich einen Pool von 171 Variablen angelegt. Diese waren mehrheitlich in Form von Gegensatzpaaren angeordnet. Cattell nahm weitere Reduktionen vor, schließlich erhielt er

35 Cluster von Variablen mit jeweils 6 bis 12 Elementen ... Diese Cluster dienten ihm als empirische Variablen zur Fremdbeurteilung (L-Daten) und führten schließlich zur Entwicklung der 12 Cattell’schen Faktoren. Genau diese Cattell-Variablen stellen auch die Grundlage vieler nachfolgender Untersuchungen anderer Wissenschaftler dar, die die ursprünglichen 35 Cluster jedoch häufig etwas modifizierten. (Amelang & Bartussek, 2001, S. 365)

Dies lieferte die Basis für Cattell und andere Forscher formale Taxonomien zu entwickeln. (Hall et al., 1997).

Cattell bezeichnet diese 12 Faktoren als

Source-Traits, d. h. eine Quelle von generalisierten fundamentalen Einflüssen, die zusammen variieren und speziellere Traits organisieren ... Konkret beobachtbar sind nur Verhaltensweisen, die insoweit, als zwischen ihnen Korrelationen bestehen, „*Surface Traits*“ ausmachen und die Grundlage für die Erschließung der dispositionellen Source-Traits liefern. (Amelang & Bartussek, 2001, S. 316)

Die 35 verschiedenen Variablen wurden von Cattell **surface-traits** (Oberflächeneigenschaften) genannt. **Source-traits** (Grundeigenschaften, fundamentale Persönlichkeitsdimension) nannte er die 16 bipolaren Faktoren (Häcker & Stapf, 1998).

Persönlichkeit bedeutet für Cattell nach Häcker und Stapf (1998)

die Gesamtheit der nicht situativen Verhaltensbedingungen, die eine Prognose darüber ermöglichen, wie sich jemand in einer gegebenen Situation verhalten wird. Die Reaktion R eines Individuums ist nicht nur eine Funktion f seiner Persönlichkeit P, sondern auch der bestimmten Situation S. Die *Verhaltensgleichung* von Cattell berücksichtigt deshalb beide Größen

$$R = f(S, P). \text{ (S. 264)}$$

Die Intention von Cattell war es, dass „Vorhersagen des Verhaltens stets auf der Grundlage von faktorenanalytisch identifizierbaren Eigenschaften erfolgen sollten und Einzeltestscores als Prediktorvariablen von geringerer Bedeutung sind“ (Häcker, Schwenkmezger & Utz, 1979, S. 17).

„*Wesenszüge* [traits; Anm. d. Verf.] sind jene Persönlichkeitsmerkmale, die hinsichtlich *Zeit und Situation stabil* bleiben ... als weitreichende innere Dispositionen, sich auf eine ganz bestimmte Art und Weise zu verhalten, definiert ...“ (Pervin, 2000, S. 224).

Die Theorie der *Persönlichkeitswesenszüge* geht von der Hauptannahme aus, dass Menschen ein weites Spektrum von Eigenschaften besitzen, um auf jeweils besondere Weise zu reagieren. Diese Dispositionen werden Persönlichkeitswesenszüge genannt. Diese Wesenszüge repräsentieren die Grundbausteine der menschlichen Persönlichkeit (Pervin, 2000).

Viele Theoretiker der Wesenszüge wenden für die Klassifikation von Wesenszügen das statistische Verfahren der Faktorenanalyse an. Durch diese Technik wird eine Gruppe von Items oder Antworten (Faktoren) gebildet, wobei die Items der einen Gruppe (Faktor) untereinander eng verwandt sind und sich von denen einer anderen Gruppe (Faktor) unterscheiden. (Pervin, 2000, S. 249)

Pervin (2000) und Hall et al. (1997) beschreiben die Vorgehensweise von Cattell folgendermaßen: R.B. Cattell bezog seine Daten aus drei Datenquellen - *Lebensprotokolldaten (L-Daten)*, *Fragebogen-Daten (F-Daten)* sowie *Daten aus objektiven Tests (OT-Daten)*. Zuerst hat Cattell die L-Daten einer Faktorenanalyse unterzogen, dabei fand er fünfzehn Faktoren, welche scheinbar auf die meisten Persönlichkeiten zutreffen. Anschließend untersuchte er die F-Daten auf vergleichbare Faktoren. Er entwickelte daraufhin den *16 Persönlichkeits-Faktoren-Test*, dieser beinhaltet zwölf Wesenszüge, die auch mit den L-Daten übereinstimmen sowie vier Faktoren (questionnaire specifics), welche nur mit der Fragebogen-Methode zu gewinnen sind. Diese Ergebnisse verwendete Cattell für die Forschung und Entwicklung von objektiven Tests. Die OT-Daten wurden schließlich auch einer Faktorenanalyse unterzogen, daraus resultierten 21 Faktoren.

Laut Pervin (2000) hat Cattell zwischen bivariaten, multivariaten und klinischen Ansätzen in der Persönlichkeitsforschung unterschieden. Cattell bevorzugte die multivariate Untersuchung der Beziehungen zwischen vielen Variablen. Er hat auch zwischen Fähigkeits-, Temperament- und dynamischen Wesenszügen unterschieden sowie zwischen Oberflächen- und Grundwesenszügen. Unter Grundwesenszügen versteht man eine Verbindung von Verhaltensweisen. Mittels Faktorenanalyse wurden sie entdeckt. Sie sind die Eckpfeiler der Persönlichkeit. Zu Cattells Forschungsbemühungen zählt die Entwicklung von Fragebogen. Dennoch wollte er beweisen, dass dieselben Faktoren auch anhand von Ratings und objektiven Tests nachweisbar sind. Cattell ging davon aus, dass das Verhalten in einer spezifischen Situation Motivationsvariablen (Ergs und Sentiments) sowie temporäre Einflüsse (Zustände und Rollen) widerspiegelt.

Bivariate Methode	Klinische Methode	Multivariate Methode
Wissenschaftliche Strenge, kontrollierte Experimente Beachtung von wenigen Variablen Vernachlässigung von wichtigen Phänomenen Vereinfacht, stückweise	Intuition Beachtung vieler Variablen Studium wichtiger Phänomene Interesse an allgemeinen Ereignissen und komplexen Mustern von Verhaltensweisen (Gesamtpersönlichkeit)	Wissenschaftliche Strenge, objektive und quantitative Analysen Beachtung vieler Variablen Studium wichtiger Phänomene Interesse an allgemeinen Ereignissen und komplexen Mustern von Verhaltensweisen (Gesamtpersönlichkeit)

Tab. 1: Cattells Beschreibung der bivariaten, klinischen und multivariaten Forschungsmethoden (Quelle: Pervin, 2000, S. 241)

Das entscheidende Merkmal von Cattells Persönlichkeitsfaktoren ist, dass diese nicht unabhängig voneinander sind, sondern miteinander korrelieren. Dadurch lassen sich aus ihnen Faktoren zweiter Ordnung extrahieren (Eysenck & Eysenck, 1987).

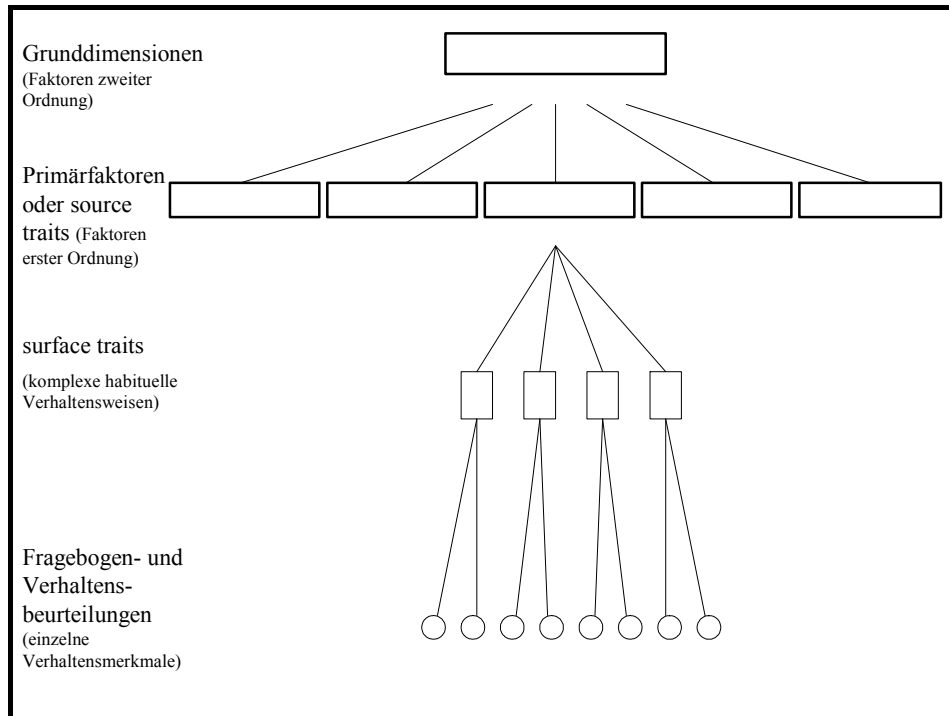


Abb. 2: Diagramm-Darstellung vom hierarchischen Aufbau der Persönlichkeit (Häcker & Stapf, 1998, S. 267)

Bei zahlreichen Reanalysen von Cattells Daten sowie später auch bei anderen Datensätzen resultierten immer wieder fünf Faktoren – die **Big Five**², welche die Persönlichkeit recht gut beschreiben. Die Autorin möchte hier einen kurzen Überblick über die Big Five geben, da sie auch im Zusammenhang mit den für die vorliegende Arbeit relevanten Untersuchungen zum Thema Faking Verwendung fanden (Amelang & Bartussek, 2001).

² Für eine detailliertere Beschreibung der Big Five Persönlichkeitsfaktoren siehe Amelang & Bartussek, 2001 (S. 364 ff).

Faktor I	Extraversion (Extraversion)
Faktor II	Verträglichkeit (Agreeableness)
Faktor III	Gewissenhaftigkeit (Conscientiousness)
Faktor IV	Emotionale Labilität (Neuroticism)
Faktor V	Offenheit für Erfahrungen (Openness to Experience)

Tab. 2: Das Fünf-Faktoren-Modell der Persönlichkeit (Quelle: Amelang & Bartussek, 2001)

2.4.1. Extraversion-Introversion

Bereits in den Anfängen der Persönlichkeitspsychologie wurde Extraversion als Konzept beschrieben. Das Konstrukt Extraversion vs. Introversion geht hauptsächlich auf C.G. Jung zurück. Es war allerdings auch Gegenstand späterer Untersuchungen, wie etwa von Cattell oder H.J. Eysenck. Extraversion-Introversion hat im „Big-Five“-Modell der Persönlichkeit einen festen Platz, es zählt somit als eine der fünf „großen“ Persönlichkeitsdimensionen (Schneewind & Graf, 1998).

Der Extravertierte ist nach Jung ein Individuum, dessen Denken, Fühlen, Interesse und Handeln auf die Umwelt und auf Objekte gerichtet sind. Er richtet seine psychische Energie auf die Außenwelt, hat guten und leichten Kontakt zur Umwelt und versteht es, sich neuen Situationen ohne Schwierigkeiten anzupassen (Hall & Campbell, 1998).

Den Introvertierten beschreibt Jung als Individuum, dessen psychische Energie nach innen gerichtet ist. Seine Gedanken, Interessen und selbst das Handeln sind – im Gegensatz zum Extravertierten – auf das eigene Ich gerichtet. Infolgedessen weist er eine Neigung zur Nachdenklichkeit, zur Verschlossenheit und zur Analyse der eigenen psychischen Zustände auf. Weiters fällt es ihm schwerer, Kontakt zur Umwelt herzustellen und sich anzupassen (Hall & Campbell, 1998).

Schneewind und Graf (1998) beschreiben Extravertierte folgendermaßen:

Extravertierte Personen sind kontaktorientiert und suchen aktiv die Beziehung zu anderen Menschen. Sie sind einerseits an den Bedürfnissen und Gefühlen anderer interessiert und gleichzeitig bereit, sich selbst anderen zu öffnen. Es fällt ihnen leicht, Kontakte zu anderen zu knüpfen. Sie schließen sich gerne zu gemeinsamen Aktivitäten der Gesellschaft anderer an und genießen es, im Mittelpunkt der allgemeinen Beachtung zu stehen. Dabei wirken sie lebhaft, spontan und begeisterungsfähig. (S. 62)

Im Gegensatz dazu sieht die Charakterisierung der Introvertierten laut Schneewind und Graf (1998) wie folgt aus: „Introvertierte Personen dagegen sind eher einzelgängerisch veranlagt und verbringen ihre Zeit lieber allein als in Gesellschaft anderer. Sie gehen weniger aus sich heraus, wirken reserviert, zurückhaltend und eher sozial gehemmt. Sie öffnen sich anderen gegenüber nur schwer“ (S. 62).

Die Skala Extraversion des 16 PF-R setzt sich aus unterschiedlichen Komponenten zusammen, die in den Ladungen von sechs Primärskalen abgebildet werden. Die sechs Primärskalen sind *Wärme*, *Dominanz*, *Lebhaftigkeit*, *Soziale Kompetenz*, *Privatheit* und *Selbstgenügsamkeit* (Schneewind & Graf, 1998).

2.4.2. Emotionale Stabilität vs. Emotionale Labilität

Bei der Skala *Emotionale Stabilität vs. Emotionale Labilität* handelt es sich um eine Primärskala des 16 PF-R. Mit dieser Skala wird das emotionale Wohlbefinden erfasst sowie das subjektive Gefühl einer Person, mit den Herausforderungen des Alltagslebens zurechtzukommen. Auf der einen Seite des Kontinuums befinden sich Personen, welche als emotional stabil und ausgeglichen bezeichnet werden können. Die Personen am anderen Ende sind als stimmungslabil zu bezeichnen (Schneewind & Graf, 1998).

Personen mit hohen Werten ... sind ausgeglichen und meistern ihr Leben ohne größere Einbrüche. Beispielsweise geben sie an, im Alltag kaum Schwierigkeiten zu begegnen, mit denen sie nicht fertig werden würden. Abends beim Zubettgehen sind sie gewöhnlich damit zufrieden, wie der Tag verlaufen ist. Extrem hohe Werte können jedoch auch ein Hinweis darauf sein, daß der Proband wenig geneigt ist, sogenannte „negative Gefühle“ zu berichten (...).

Personen mit niedrigen Werten leiden mehr unter Stimmungsschwankungen als andere Personen. Sie haben das Gefühl, nicht ausreichend Kontrolle über ihr Leben zu haben. Sie reagieren eher, während Personen mit hohen Werten aktiv handeln. Ihre gefühlsmäßigen Bedürfnisse erleben sie als nicht sehr gut befriedigt. Wenn Kleinigkeiten schiefgehen, haben sie das Gefühl, nicht damit zurechtzukommen. (Schneewind & Graf, 1998, S. 68)

Einschub zu H.J. Eysenck: In Eysencks Theorie der Persönlichkeit existiert neben dem Konstrukt „Extraversion/Introversion“ auch ein Persönlichkeitskonstrukt „Emotionale Stabilität vs. Emotionale Labilität“. Dieses wurde von Eysenck als Neurotizismus spezifiziert. Die Persönlichkeitsdimensionen „Extraversion vs. Introversion“ und „Neurotizismus“ sind unabhängig voneinander, wie folgende Darstellung veranschaulicht.³

³ Für die vorliegende Arbeit besitzt die Persönlichkeitsdimension „Neurotizismus“ von H.J. Eysenck keine Relevanz, deshalb verzichtet die Autorin auf eine detailliertere Darstellung (vgl. Amelang & Bartussek, 2001, S. 327 ff)

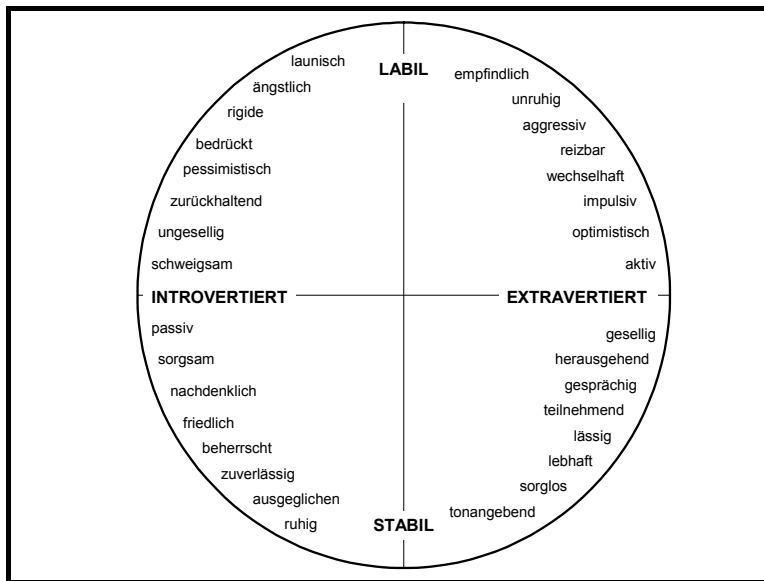


Abb. 3: Unabhängigkeit von „Extraversion vs. Introversion“ und „Emotionaler Stabilität vs. Emotionaler Labilität“ (Häcker & Stapf, 1998, S. 267)

2.5. Das Repression-Sensitization-Konstrukt

2.5.1. Beschreibung des Konstruktes

„Im Rahmen der Darstellung neuerer kognitionspsychologisch orientierter Ansätze zur Erforschung von Angst und Angstverarbeitung wurde das Konstrukt „Repression-Sensitization“ als ein zentrales Konzept der Angstabwehr vorgestellt“ (Krohne, 1975, S. 66).

Das Persönlichkeitsmerkmal „*Repression versus Sensitization*“ wird als kontinuierlich variierendes Merkmal aufgefasst, welches durch die Bezeichnung der beiden Extreme Repression und Sensitization umschrieben wird. Die beiden Pole des Persönlichkeitsmerkmals Repression versus Sensitization werden als zwei einander entgegen gesetzte Arten aufgefasst, wie Personen mit angstausslösenden Reizen oder Situationen fertig werden (Amelang & Bartussek, 2001).

Von **Sensitization** spricht man, wenn die Person ihre Aufmerksamkeit einem potentiell angstausslösenden Sachverhalt zuwendet. Wendet eine Person ihre Aufmerksamkeit von einem potentiell angstausslösenden Sachverhalt ab, so spricht man von **Repression** (Krohne, 1996; Amelang & Bartussek, 2001). Laut Krohne (1996) tendieren Represser dazu, „den Bedrohungsgehalt von Situationen bzw. Reizen zu verleugnen oder herunterzuspielen“ (S. 95), während Sensitizer den Bedrohungsreizen erhöhte Aufmerksamkeit schenken.

Unter Repression versteht man die Vermeidung von angstausslösenden Situationen. **Represser** (Abwehler, Vermeider) wehren angstausslösende Situationen ab. Sie verleugnen und vermeiden solche Situationen. **Sensitizer** (Intellektualisierer, Zuwender) hingegen wenden ihre Aufmerksamkeit verstärkt angstausslösenden Situationen zu. Sie verhalten sich aversiven Ereignissen gegenüber besonders empfänglich und sensibilisiert (Krohne, 1975; Amelang & Bartussek, 2001). So genannte Sensitizer suchen verstärkt gefahrrelevante Informationen, während sich Represser eher auf Strategien stützen, die darauf abzielen gefahrbezogene Hinweisreize zu meiden (Krohne, 1985a).

Beim R-S-Konstrukt handelt es sich um ein dispositionelles Konstrukt der Angstbewältigung, welches eindimensional ausgerichtet und großteils in der psychodynamischen Tradition begründet ist (Krohne, 1996). Die Theorie der beiden einander entgegen gesetzten Arten der Angstverarbeitung leitet sich aus der Psychoanalyse ab. In der Psychoanalyse bezieht sich diese Theorie auf die als Abwehrmechanismen bezeichneten Ich-Funktionen zur Angstbewältigung. Die Mechanismen der Repression und Sensitization umfassen verschiedene Freud'sche Abwehrmechanismen (Amelang & Bartussek, 2001). Das R-S-Konstrukt bezieht die Vielfalt der Angstabwehrmechanismen auf eine einzige bipolare Dimension mit den beiden Polen Repression und Sensitization (Krohne, 1996).

Krohne (1975) ordnet Repressern und Sensitizern die Angstabwehrmechanismen nach Freud etwa folgendermaßen zu:

<i>Repression</i>	<i>Sensitization</i>
Verdrängung	Isolierung
Verleugnung	Intellectualisierung
Reaktionsbildung	Kompensation
Verschiebung	Depression (Selbstaggression)
Sublimierung	Projektion
Identifikation (Reaktion im Sinne sozialer Erwünschtheit)	Fantasien und Tagträume
Rationalisierung	Zwangsneurotische Reaktionen
Psychosomatische Störungen	

Tab. 3: Angstabwehrmechanismen von Repressern und Sensitizern (Krohne, 1975, S. 58)

2.5.2. Die Entwicklung des R-S-Konstruktes

Die Forschung zum Repression-Sensitization-Konstrukt hat sich aus der Forschung zur *Wahrnehmungsabwehr* (*perceptual-defense*) entwickelt. Die Theorie der Wahrnehmungsabwehr besagt, dass emotional besetzte Wörter wie Tabuwörter oder angstassoziierte Wörter im tachistoskopischen Versuch höhere Erkennungsschwellen aufweisen als neutrale Wörter. Nun fanden Bruner und Postman in den Vierziger-Jahren aber bei Untersuchungen zur Wahrnehmungsforschung heraus, dass es eine Personengruppe gibt, bei denen sich die Wahrnehmungsprozesse entgegengesetzt zur Theorie verhalten. Für diese Gruppe von Personen gilt, dass sie für emotional besetzte Wörter niedrigere Erkennungsschwellen aufwiesen als für neutrale Wörter. Man spricht hier von *perceptual-vigilance* (Amelang & Bartussek, 2001).

Aus der differentialpsychologischen Betrachtung dieses Phänomens entwickelte sich die über viele Jahre aktuelle Forschung zum Konstrukt „Repression-Sensitization“ (Amelang & Bartussek, 2001).

2.5.3. Unterschiede zwischen Repressern und Sensitizern

Die grundlegenden Unterschiede zwischen Personen am repressiven und sensitiven Pol werden von Krohne und Rogner (1985) zusammengefasst:

1. in der Richtung der Aufmerksamkeit in Streß-Situationen (Sensitizer richten ihre Aufmerksamkeit verstärkt auf streßrelevante Information, Represser ziehen sie hiervon ab),
2. im Ausmaß der Antizipation künftiger Gefahr (bei Sensitizern erhöht),
3. in der Art der ausgeübten Kontrolle in Streß-Situationen (Sensitizer kontrollieren verstärkt Information und eigenes Verhalten). (S. 47-48)

Eine Fülle von Verhaltensbereichen wurde untersucht, für die aus dem Repression-Sensitization-Konstrukt Erwartungen über Unterschiede zwischen Repressern und Sensitizern abgeleitet wurden (Krohne, 1975; Amelang & Bartussek, 2001).

Wahrnehmung: Represser erinnern sich eher an Erfolge in Leistungssituationen (Höfer, Wallbott & Scherer, 1985). Für das Erkennen mißerfolgsbezogener Wörter benötigen sie im Tachistoskop längere Expositionszeiten als Sensitizer (Krohne, 1975; 1996). Die Wahrnehmungsschwelle steigt für Represser mit zunehmender emotionaler Bedeutsamkeit der Wörter steiler an als für Sensitizer, während jedoch die gleichzeitig erhobene elektrodermale Aktivität bei Sensitizern stärker ansteigt (Krohne, 1996). Außerdem schenken Sensitizer sexuellen und physisch bedrohlichen Bildern mehr Beachtung (Krohne, 1996). Krohne (1978) berichtet über eine Untersuchung von Haley aus dem Jahre 1974, bei dem den Probanden ein Film gezeigt wurde, der sowohl stress-induzierende als auch neutrale Szenen enthielt. Während Sensitizer ihre Aufmerksamkeit mehr auf die stressrelevanten Szenen richteten, wurde den neutralen Szenen von Repressern mehr Beachtung geschenkt.

Informationsverarbeitung und selektive Aufmerksamkeit: Während Sensitizer bei Problemlösungen nach mehr Information suchen als Represser, sind sie bei der Integration dieser Information zu Konzepten z. B. durch Stresseinwirkung störbare. Represser neigen zur Stereotypisierung, Vereinheitlichung und Reduktion von Mehrdeutigkeit sowie geringerer Offenheit für neue Erfahrungen und Informationen (Krohne, 1975). Bei der kognitiven Verarbeitung mehrdeutiger Vorlagen geben Represser weniger aggressive bzw. sexuelle Inhalte an als Sensitizer, während Sensitizer bedrohliche Assoziationen mehr beachten (Krohne, 1996). Represser schenken ihren eigenen Stresssymptomen weniger Beachtung, sie machen sich auch weniger Gedanken um die Begleitumstände ihrer Symptome (Furnham, Petrides & Spencer-Bowdage, 2002).

Personenwahrnehmung und –beurteilung: Represser geben positivere Selbst- und Fremdbeurteilungen ab. Dies muss jedoch im Zusammenhang gesehen werden mit der Tendenz der Represser zum Ausklammern von Konflikten und gefahrrelevanten Aspekten, ebenso wie der Tendenz zum Verhalten

im Sinne sozialer Erwünschtheit. Dies hat eine Unterschätzung des Unterschiedes zwischen sich und anderen Personen sowie eine undifferenziertere Personenbeurteilung zur Folge (Krohne, 1975). In einigen Untersuchungen wurden Represser auch von Dritten positiver eingeschätzt als Sensitizer, dies ist „offenbar auf geschicktere Strategien der Selbstdarstellung zurückzuführen“ (Krohne, 1996, S. 128).

Konformität und soziale Erwünschtheit: Represser tendieren stärker zum Verhalten im Sinne sozialer Erwünschtheit als Sensitizer. Außerdem sind Sensitizer eher dazu gewillt moralisch schlechte Verhaltensweisen zuzugeben. Zudem berichten Sensitizer auch von erhöhter Aggressionsbereitschaft sowie Selbstaggression. In sozialen Interaktionen sind es jedoch Represser, welche tatsächlich als aggressiver eingeschätzt werden (Krohne, 1975; 1996). Furnham et al. (2002) konnten in einer Untersuchung an Studenten belegen, dass Represser signifikant mehr sozial erwünschte Antworten in Persönlichkeitsfragebogen geben.

Emotionale Anpassung: Represser neigen dazu Konflikte und gefahrrelevante Situationsinterpretationen auszuklammern. Wie schon erwähnt tendieren Represser zu Reaktionen im Sinne sozialer Erwünschtheit. Sie leugnen eigene Schwächen und schildern sich selbst als kontaktfreudig, gut gelaunt, ruhig, selbstbewusst, aktiv, ausgeglichen sowie frei von körperlichen Beschwerden. Zudem berichten Represser über eine leistungsfördernde Wirkung von Angst und Erregung. Die Tendenz Ängstlichkeit, leistungsmindernde Angst, emotionale Labilität und Aggressionsakte zuzugeben sinkt mit erhöhter repressiver Angstabwehr. Im Gegensatz dazu tendieren Sensitizer vermehrt dazu Tatbestände für sich zuzugeben, welche im Allgemeinen als Schwäche angesehen werden (Krohne, 1975). Sensitizer neigen zu emotionaler Fehlangepasstheit (Amelang & Bartussek, 2001).

Vergleich verbaler und physiologischer Reaktion: „Emotionale Reaktionen auf bedrohliche Reize gehen einher mit messbaren physiologischen Erregungsprozessen wie gesteigerter Herz- und Atmungsfrequenz, erhöhtem Blutdruck und gesteigerter Hautleitfähigkeit“ (Amelang & Bartussek, 2001, S. 431). Einige Untersuchungen ergaben, dass sich Represser trotz höherer physiologischer Reaktionen durch bedrohliche Situationen subjektiv weniger betroffen und erregt einschätzen, während Sensitizer einen hohen Grad an Erregung berichten trotz relativ niedriger physiologischer Erregung. Der Grad der Erregung dürfte somit offenbar nicht nur vom Ausmaß der wahrgenommenen Gefahr abhängen. Beim Produzieren freier und evtl. selbst enthüllender Assoziationen sind Represser dann weniger erregt als Sensitizer, wenn man sie darauf vorbereitet hat. Werden sie unvorbereitet in eine derartige Situation gebracht, so zeigen Represser jedoch einen höheren Grad von Erregung als Sensitizer (Krohne, 1975; Amelang & Bartussek, 2001).

Entscheidungsverhalten: In Konfliktsituationen haben Represser kürzere Entscheidungszeiten, weil sie scheinbar die belastende Situation nicht gut ertragen können und deshalb schneller hinter sich bringen wollen. Während Represser den Zustand der Dissonanz allgemein als unangenehm erleben sollen, beachten Sensitizer die Dissonanz in verstärktem Ausmaß. Represser setzen drastische Formen der Dissonanzreduktion ein, während Sensitizer die Dissonanz aufrechterhalten. Represser äußern zudem

einen „höheren Grad von subjektiver Sicherheit hinsichtlich der Richtigkeit eigener Einschätzungen“ (Krohne, 1996, S. 128).

Körperliche Symptome: Sensitizer suchen häufiger Ärzte und Psychotherapeuten auf, erhalten mehr medikamentöse Hilfe und zeigen mehr körperliche Symptome. Sie berichten auch signifikant mehr Krankheiten als Represser. Sensitizer sind öfter wegen psychosomatischer Krankheiten in Behandlung, während Represser öfter wegen organischen Krankheiten behandelt werden (Krohne, 1975; Amelang & Bartussek, 2001). Während bei Repressern, die einen Arzt konsultieren, fast ausschließlich eine organische Erkrankung vorliegt, wurde für Sensitizer bei Erkrankungen oft die Beteiligung einer psychischen Komponente diagnostiziert (Krohne, 1996). Zunächst vermögen Represser Schmerzen besser zu ertragen. Wenn die Schmerzen jedoch anhalten, sinkt bei Repressern die Schmerztoleranz, bei Sensitizern hingegen bleibt sie relativ konstant (Krohne, 1975; 1996).

Gesundheitsverhalten: Unter Gesundheitsverhalten wird wesentlich die Auseinandersetzung mit Informationen über Krankheit und Gesundheit verstanden. Während sich Sensitizer verstärkt mit ihren Krankheitssymptomen auseinander setzen, gehen Represser dieser Auseinandersetzung eher aus dem Weg (Krohne, 1996).

Unterschiede im Elternhaus: Hierzu liegen nur wenige Untersuchungen vor. Diese legen nahe, dass Represser in einem warmen, freundlichen Familienklima aufgewachsen sind, während Sensitizer die eigene Kindheit als eher unglücklich schildern. Es ist jedoch unklar, ob die berichteten Unterschiede im Elternhaus tatsächlich bestanden haben oder von Repressern und Sensitizern im Sinne unterschiedlicher Abwehrmechanismen nur verschieden erinnert werden (Amelang & Bartussek, 2001).

Durch die Berücksichtigung subjektiver Streßbewältigungsreaktionen auf verschiedenen Ebenen lassen sich mehrere neue Einsichten in das R-S-Konstrukt gewinnen: Zunächst ..., daß es nicht möglich ist, auf der Ebene subjektiver Angaben zwischen kognitiven, emotionalen, körperlichen und behavioralen Streß(bewältigungs)reaktionen zu unterscheiden. Der Grund hierfür könnte ... darin gesehen werden, daß die subjektiven Angaben von Repressern und Sensitizern von deren kognitiven Selbstkonzepten überformt werden (Huwe, Hennig & Netter, 1996, S. 169).

	Ängstlichkeit	Angstleugnung
Sensitizer	hoch	niedrig
Represser	niedrig	hoch

Tab. 4: Ängstlichkeit und Angstleugnung bei Repressern und Sensitizern

Represser sind charakterisiert durch ihre Angaben über psychische und physische Nicht-Erregtheit. Sie setzen sich mit der Situation nicht auseinander. Represser geben vor, dass sie Stressoren nicht als solche wahrnehmen. Man könnte sagen, Represser betreiben *nicht* Stressbewältigung, sondern *Nicht-Stressbewältigung*. Aus der Literatur ist die starke physiologische Auslenkung von Repressern in bedrohlichen Situationen jedoch hinreichend bekannt. Emotionale Erfahrungen, insbesondere wenn diese mit einem negativen Gefühl verbunden sind, können von Repressern nur schlecht erinnert werden,

da sie dazu neigen, die Verarbeitung negativen Materials zu vermeiden und es hierdurch zu Enkodierungsdefiziten kommt ... Möglicherweise sind Represser aufgrund dessen auch nicht so gut wie Sensitizer in der Lage, sich in bereits erlebte streßreiche selbstwertrelevante Situationen hineinzusetzen und ihre erlebten Streß(bewältigungs)reaktionen auf die imaginäre Fragebogensituation zu beziehen. (Huwe et al., 1996, S. 170)

„Sensitizer geben ... eine Reihe von Streß(bewältigungs)reaktionen an. Ihre subjektiven Angaben ... sind primär durch Kognitionen, aber auch durch Coping-Verhalten charakterisiert. ... Folgerichtig könnte man in Zukunft auch zwischen eher aktiv und eher passiv bewältigenden Sensitizern differenzieren“ (Huwe et al., 1996, S. 170-171).

2.5.4. Weiterentwicklung des R-S-Konstruktes

2.5.4.1. Eindimensionale Erfassung

Die Forschung zum R-S-Konstrukt hat sich aus frühen „Arbeiten zu Wahrnehmungsschwellen bei neutralem und belastendem Material“ (Höfer et al., 1985, S. 95) heraus entwickelt. Die unterschiedlichen Stile der Angstbewältigung wurden zwei Verhaltensklassen zugeordnet (Repression vs. Sensitization). Diese zwei Klassen sind gekennzeichnet durch einander weitest gehend entgegen gesetzten Formen der Angstbewältigung, deshalb wurde von vielen Autoren die Annahme einer *eindimensionalen, bipolaren Persönlichkeitsdisposition* vorgeschlagen (Krohne, 1985b). Höfer et al. (1985, S. 95) bringen es auf den Punkt: „Dem *klassischen* R-S-Konstrukt liegt damit ein *eindimensionales* Modell des habituellen Umgangs mit angst- oder allgemeiner erregungsauslösenden Situationen zugrunde“.

Formulierungen, die diesem Konzept entsprechen sind neben anderen z. B. „Monitoring/Blunting“ von Miller, „Attention/Rejection“ von Mullen und Suls, „Attention/Avoidance“ von Krohne (Krohne, 1993a) sowie „Distress/Restraint“ von Weinberger und Schwartz (Krohne, Schmukle, Burns, Egloff & Spielberger, 2001).

Krohne hat das R-S-Konstrukt zu einem *Modell der Bewältigungsmodi* weiter entwickelt. Dieses Modell geht davon aus, dass sich Personen grundsätzlich bei der Konfrontation mit aversiven Situationen unterscheiden. Personen zeigen demnach Intoleranz gegenüber einer mit der Situation verbundenen Unsicherheit (Unsicherheitsintoleranz) und/oder Erregung (Erregungsintoleranz). Sie setzen dementsprechende Strategien ein. Als *Vigilanz (VIG)* werden Strategien bezeichnet, die auf den Abbau

von Unsicherheit abzielen, während Strategien zur Reduktion der Erregung zusammenfassend als *kognitive Vermeidung (KOV)* bezeichnet werden (Krohne, 1975; Huwe et al., 1996). Es geht hier vor allem um den intrapsychischen Prozess der Aufmerksamkeitsveränderung hinsichtlich einer Gefahrenquelle, die Aufmerksamkeit soll dabei zwischen den Polen Vigilanz und Kognitive Vermeidung variieren (Krohne, 1976).

Vigilanz entspricht in der traditionellen Terminologie der Angstbewältigungsforschung dem Pol Sensitization, während Kognitive Vermeidung mit Repression gleich gesetzt werden kann (Krohne, 1996; Amelang & Bartussek, 2001). Krohne (1993b) spricht bei Sensitizern von „hoher Vigilanz/niedriger Vermeidung“ und umgekehrt bei Repressern von „niedriger Vigilanz/hoher Vermeidung“.

Die Formen der Angstbewältigung (*Coping*) wurden operational bestimmt durch spezifische Ausprägungen von Ängstlichkeit und Angstleugnung (Krohne, Wigand & Kiehl, 1985). Represser und Sensitizer setzen laut Krohne und Rogner (1985) Coping-Strategien vergleichsweise rigide ein, während Personen im mittleren Bereich dieses Kontinuums Bedrohungen flexibel und eher der Situation angemessen bewältigen. Die Personen im Mittelbereich dieses Kontinuums werden als so genannte „Nichtdefensive“ bezeichnet (Krohne, 1985b).

In intrapsychischen Coping-Mechanismen wird entweder eine sensitive oder eine repressive Strategie realisiert. Unter einer *sensitiven Strategie* verstehen Krohne, Wigand & Kiehl „z. B. Akzentuierung der Gefahr, Betonung der Notwendigkeit von Informations- und Verhaltenskontrolle bei Bedrohung“ (1985, S. 65). „Herunterspielen der Bedrohung, Vermeidung der Beschäftigung mit der Gefahr“ (Krohne, Wigand & Kiehl, 1985, S. 65) wird der *repressiven Strategie* zugeordnet.

2.5.4.2. Zweidimensionale Erfassung

Lange Zeit war die *Repression-Sensitization-Skala* von Byrne aus dem Jahre 1961 das am meisten verwendete Instrument zur Erfassung von kognitiven Bewältigungsstilen (Bijttebier, Vertommen & Steene, 2001). Krohne wies schon 1978 auf die Problematik der eindimensionalen Erfassung von Repression-Sensitization hin, da laut Krohne Repression-Sensitization immer mit *Ängstlichkeit* konfundiert gemessen worden ist.

	<i>Angstleugnung hoch</i>	<i>Angstleugnung niedrig</i>
<i>Ängstlichkeit niedrig</i>	Repressive Angstbewältigung	Nichtdefensive Angstbewältigung
<i>Ängstlichkeit hoch</i>	Erfolgreiche Angstbewältigung	Sensitive Angstbewältigung

Tab. 5: Einteilung nach Crowne und Marlowe (Quelle: Amelang & Bartussek, 2001, S. 437)

Am Repression-Pol von Byrne's Repression-Sensitization-Skala sind auch *Niedrig-Ängstliche* lokalisiert und am Sensitization-Pol auch *Hochängstliche*. Es kann also mit dieser Skala nicht zwischen Repressern und niedrigängstlichen Personen einerseits sowie zwischen Sensitizern und hochängstlichen Personen andererseits differenziert werden (Krohne, 1996). In weiterer Folge waren Testautoren um eine zweidimensionale Erfassung bemüht. Es wurde, wie oben bereits erwähnt, eine zusätzliche Unterscheidung getroffen in *Defensiv-Hochängstliche* und *Niedrigängstliche* (Huwe et al., 1996; Amelang & Bartussek, 2001).

Huwe et al. haben 1996 bei der Konstruktion ihres Fragebogens (siehe Kap. 3.4.1.3.) folgende Einteilung getroffen:

Repressor	Defensivität hoch	Sensitivität gering
Sensitizer	Defensivität gering	Sensitivität hoch
Niedrig-Ängstliche	Defensivität hoch	Sensitivität hoch
Defensiv-Hochängstliche	Defensivität gering	Sensitivität gering

Tab. 6: Die vier Typen der Stressbewältigung

2.6. Selbstdarstellung

2.6.1. Begriffsdefinitionen

Bislang gibt es noch keine umfassende Theorie der Selbstdarstellung, es gibt jedoch mehrere Ansätze, in denen die vorteilhafte Selbstdarstellung (self-presentation bzw. impression-management) eine wichtige Rolle spielt (Herkner, 1991).

Selbstdarstellungsprozesse sind vorwiegend allgemein- und sozialpsychologisch erforscht worden, aber bislang nur selten differentiell- und persönlichkeitspsychologisch (Mummendey & Eifler, 1994).

Hier wäre zum Beispiel die Theorie zum *Self-Monitoring* von Snyder aus dem Jahre 1974 nennen. Auch die *Theorie der objektiven Selbstaufmerksamkeit* von Duval und Wicklund von 1972 beschäftigt sich mit der Selbstdarstellung. Erwähnenswert wäre auch noch die *Theorie der kognitiven Dissonanz* von Festinger (Herkner, 1991; Mummendey & Eifler, 1994).⁴

Paulhus (1984) etwa unterscheidet bei der Beantwortung von Fragen im Sinne sozialer Erwünschtheit zwei Komponenten: self-deception und impression management. Bei *Self-deception* handelt es sich um eine unbewusste Selbsterhöhung. Die jeweilige Person ist von der Richtigkeit ihrer - in eine positive Richtung verfälschten - Antworten überzeugt. *Impression Management* hingegen meint die absichtliche Verfälschung von Antworten in eine positive Richtung (response distortion).

Die Impression-Management-Theorie bezieht sich auf die Beobachtungen von Goffman aus dem Jahre 1959. Man spricht von *Selbstdarstellungs-* oder *Impression-Management-Theorien*, wenn „die Beeinflussung der Personenwahrnehmung von Interaktionspartnern als Hauptzweck selbstbezogener Kognitionen oder bestimmendes Moment sozialer Interaktion gesehen wird“ (Mummendey, 1990, S. 127).

Mummendey und Eifler definieren *Selbstdarstellung* als die „Tendenz eines Individuums ..., andere Personen zu bestimmten Zuschreibungen von Merkmalen an dieses Individuum zu veranlassen“ (1994, S. 3).

⁴ Die Autorin wird an dieser Stelle nicht näher auf die Theorien, welche sich mit der Selbstdarstellung im weitesten Sinne beschäftigen, eingehen. Für die vorliegende Arbeit ist eine detaillierte Ausführung der unterschiedlichen Theorien nicht vonnöten (vgl. Herkner, 1991; Mummendey, 1990).

Unter Selbstdarstellung versteht man Mummendey (1990) zufolge:

1. Die Präsentation von Beurteilungen und Bewertungen der eigenen Person
2. Die Kontrolle von Urteilen und Bewertungen anderer Personen über die eigene Person (im Sinne der Impression-Management-Theorie)

Die Tendenz, sich selbst als positiv darzustellen, ist wohl die üblichste Art und Weise der Selbstpräsentation (Mummendey & Eifler, 1994). **Positive Selbstdarstellung** meint „die Tendenz eines Individuums ..., andere Personen zur Zuschreibung von als günstig bewerteten Merkmalen an dieses Individuum zu veranlassen“ (Mummendey & Eifler, 1994, S. 9).

Bei jeglichem zwischenmenschlichen Verhalten sind Selbstdarstellungsprozesse sowie die dabei beteiligten Kognitionen zu berücksichtigen (Mummendey, 1990).

Soziale Verstärkungsprozesse scheinen für das Selbstdarstellungsverhalten eine maßgebliche Rolle zu spielen. Es wird angenommen, dass Individuen sich weitgehend damit beschäftigen, Zustimmung von anderen zu erhalten und die Ablehnung durch Andere zu vermeiden. Um dies zu erreichen, präsentieren sie sich anderen Personen gegenüber auf positive Art und Weise (Mummendey, 1990).

2.6.2. Selbstdarstellungstechniken

Mummendey und Eifler (1994) zählen eine Reihe von positiven Selbstdarstellungstechniken auf:

- Eigenwerbung betreiben (self-promotion)
- Beispielhaft und vorbildlich erscheinen
- Sich im positiven Licht anderer sonnen
- Vertrauenswürdig und glaubwürdig erscheinen
- Hohen Status und Prestige herauskehren
- Hohe Ansprüche stellen
- Kompetenz und Expertentum signalisieren
- Übertreiben (overstatement)

Mummendey und Eifler haben 1994 auch als **negativ** anzusehende **Selbstdarstellungstechniken** beschrieben. „Die entsprechenden Konzepte suggerieren die Zuschreibung sozial eher unerwünschter Merkmale an die sich selbst darstellende Person“ (S. 5):

- Abwerten von Personen und Gruppen (blasting)
- Sich-Einschmeicheln (ingratiation)
- Einschüchtern (intimidation)
- Hinweisen auf eigene Unzulänglichkeiten (self-handicapping)
- Untertreiben (under-statement)

Diese positiven und negativen Selbstdarstellungstechniken überschneiden sich in vielerlei Hinsicht. Mummendey und Eifler (1994) haben sie deshalb für die Zwecke der Skalenkonstruktion für die Impression-Management-Skala (vgl. Kap. 4.3.1.5.) zu einem globalen Konstrukt „Positive Selbstdarstellung“ zusammengefasst.

Die genaueste Unterteilung trafen Tedeschi, Lindskold und Rosenfeld. Sie haben 1985 eine große Anzahl von Selbstdarstellungstechniken beschrieben. Sie trafen eine Unterscheidung der Impression-Management-Techniken in **Strategien** und in **Taktiken**. Strategien sind eher langfristig und situationsübergreifend, während Taktiken eher kurzfristig und situationsbezogen sind. Außerdem trafen sie noch eine Unterscheidung in assertive und defensive Techniken. Durch **assertive Techniken** soll ein positiver Eindruck erzeugt werden, **defensive Techniken** sollen einen negativen Eindruck verhindern (Herkner, 1991; Mummendey, 1990).

Assertive Impression-Management-Strategien	Defensive Impression-Management-Strategien
<i>Der Akteur ist bemüht, eine langfristig gültige, gute Reputation zu erwerben, die über verschiedene und unterschiedliche Situationen hinweg wirksam ist</i>	<i>Der Akteur ist bemüht, ein Bild von sich zu vermitteln, wonach er über verschiedene und unterschiedliche Situationen hinweg nicht in vollem Maße für sein Handeln verantwortlich zu machen ist</i>
Competence or Expertise (Kompetenz und Expertentum) Attractiveness (sich als attraktiv und liebenswert darstellen) Status and Prestige Credibility & Trustworthiness (Glaubwürdigkeit/Vertrauenswürdigkeit erwecken) Self-disclosure (sich anderen gegenüber öffnen)	Pathologischer Gebrauch von Entschuldigungen Self-handicapping-Strategien (Alkoholismus- und Drogenabhängigkeit; Hilflosigkeit und Ängstlichkeit) Präsentation von Symptomen von Geisteskrankheit

Tab. 7: Assertive und Defensive IM-Strategien (Quelle: Mummendey, 1990)

Assertive Impression-Management-Taktiken	Defensive Impression-Management-Taktiken
<i>Der Akteur versucht, über positive Selbstdarstellung das Publikum so zu beeindrucken, dass seine soziale Macht vergrößert wird und eigene Interessen und Wünsche vom Publikum erfüllt werden</i>	<i>Der Akteur versucht zu verhindern oder zumindest einzuschränken, dass er beim Publikum an Ansehen verliert</i>
Ingratiation (sich beliebt machen, sich einschmeicheln) self-enhancing communication (Ziel: sich oder andere zu erhöhen) Opinion conformity (Meinungskonformität) Favor-doing (Gunst erweisen) Intimidation (Einschüchterung) Supplication (hilfsbedürftig erscheinen) Self-Promotion (Eigenwerbung) Entitlements (verbal Leistung herausstellen) Enhancements (eigene Leistung überbewerten) Social Identity (sich mit einer anerkannten Gruppe identifizieren) Blasting (rivalisierende Gruppen/Personen abwehren) Exemplification (sich als beispielhaft, moralisch integer darstellen)	Predictaments and Face-work (sich aus einer misslichen Lage herauswinden) Excuses (Abstreiten von Verantwortlichkeit) Justifications (Rechtfertigungen) Disclaimers and self-handicapping (andere über eventuelle spätere eigene Misserfolge informieren) Apologies (Entschuldigungen)

Tab. 8: Assertive und Defensive IM-Taktiken (Quelle: Mummendey, 1990)

2.7. Theoretische Grundlagen des Moralischen Urteils

Im Persönlichkeitssystem entsprechen der moralischen Urteilsfähigkeit Verhaltenskontrollen bzw. Über-Ich-Strukturen (Habermas, 1996).

Lind (2000) definiert moralische Urteilskompetenz folgendermaßen:

Hohe moralische Urteilskompetenz soll insbesondere die Fähigkeit meinen, Argumente auch dann moralisch differenziert zu beurteilen, wenn sie gegen die eigene Meinung gerichtet sind. Akzeptiert und differenziert eine Person nur Argumente, die ihre anfängliche Meinung stützen, dann sprechen wir eher von (nachträglicher) Rationalisierung statt von einer Rationalität des Diskurses. (S. 72)

Moralische Urteilsfähigkeit wird definiert als „das Vermögen, moralische Entscheidungen auf der Grundlage von (selbstgewählten) moralischen Prinzipien zu treffen (statt z.B. auf Vorurteilen oder Konformität), und in bezug auf diese Prinzipien konsistent und parteiisch zu handeln“ (Lind, 1993, S. 83).

2.7.1. Die Stadien der Moral nach Piaget

Weit weniger bekannt als seine Theorie der kognitiven Entwicklung ist Jean Piagets *Theorie der Entwicklung des moralischen Urteils bei Kindern*. Dies liegt wohl darin, dass sich Piaget nach der Veröffentlichung seines diesbezüglichen Forschungsberichtes im Jahre 1932 anderen Entwicklungsbereichen zuwandte. Die frühe Arbeit Piagets hat dennoch den gegenwärtigen Stand der moralischen Urteilsforschung wesentlich inspiriert (Heidbrink, 1991).

Piaget sah vor allem folgende Problemstellungen als zentral an:

1. Den Einfluß von Zwang durch die Erwachsenen auf das Kind.
2. Die Auswirkung der sozialen Kooperation auf das moralische Urteil.
3. Den Einfluß der intellektuellen Entwicklung auf die Prozesse des moralischen Denkens.
4. Die Wechselwirkung dieser drei Faktoren. (Heidbrink, 1991, S. 11)

2.7.1.1. Die Regeln des Marmelspiels

Piagets Arbeit umfasst *Interviews* und *Verhaltensbeobachtungen* von ungefähr hundert Kindern, von denen die meisten im Vor- und Grundschulalter waren. Er analysierte die Praxis des Marmelspiels und traf folgende Unterscheidung der Anwendung der Regeln (Heidbrink, 1991):

Stadien	Alter	Anwendung der Regeln
<i>Rein motorisches und individuelles Stadium</i>	0-3	Kind spielt allein nach eigenen Wünschen; motorische Regeln
<i>Egozentrisches Stadium</i>	2-6	Kind versucht Spielregeln nachzuahmen; beim Spiel mit anderen kann allerdings jeder nach der eigenen „Interpretation“ der Regeln spielen
<i>Beginnende Zusammenarbeit</i>	7-10	Kind versucht die Mitspieler zu besiegen; gegenseitige Kontrolle; Vereinheitlichung der Spielregeln; jedoch noch widersprüchliche Auffassungen über die einzelnen Regeln
<i>Kodifizierung der Regeln</i>	ab 11	Spieldpartien sind genau geregelt; die Regeln sind bekannt

Tab. 9: Die vier verschiedenen Stadien der Anwendung der Regeln des Marmelspiels

Beim Regelbewusstsein werden von Piaget drei Stadien unterschieden, diese stehen in gewisser zeitlicher Verzögerung zu den oben genannten Stadien der Spielpraxis (Heidbrink, 1991):

1. Stadium: Motorische Schemata (Individuelle Riten)
2. Stadium: Heteronomie (Regeln sind heilig und unantastbar)
3. Stadium: Autonomes Regelverständnis

2.7.1.2. Heteronome und autonome Moral

„Die *heteronome Moral* beruht auf dem moralischen Zwang der Erwachsenen auf die Kinder“, während die *autonome Moral* „auf Zusammenarbeit und Kooperation der Kinder untereinander“ beruht. „Zwischen heteronomer und autonomer Moral existiert noch ein *Zwischenstadium*, in dem das Kind nicht mehr ausschließlich an der erwachsenen Autorität orientiert ist“ (Heidbrink, 1991, S. 19).

Auch in bezug auf die Entwicklung des Gerechtigkeitsbegriffs kommt Piaget zu einer Unterscheidung von drei großen „Perioden“. Die erste Periode dauert etwa bis zum 7. bzw. 8. Lebensjahr und ist noch ganz durch die Autorität der Erwachsenen geprägt: Gerecht ist das, was die Eltern erwarten. Die zweite Periode liegt nach Piaget etwa zwischen dem 8. und 11. Lebensjahr und ist durch den Primat der Gleichheit über die Autorität bestimmt. Ab dem 11. bzw. 12. Lebensjahr wird der strikte Gleichheitsgedanke relativiert. Piaget bezeichnet dies als „Billigkeitsgefühl“, das die

besondere Lage jedes einzelnen mit einbezieht. Obwohl er sich hier nicht sehr deutlich äußert, kann man davon ausgehen, daß die erste Periode der Heteronomie, die dritte der Autonomie und die zweite dem oben erwähnten „Zwischenstadium“ entspricht. (Heidbrink, 1991, S. 19-20)

Die Verfasserin möchte darauf hinweisen, dass den Altersangaben lediglich eine Orientierungsfunktion zukommt, da die Entwicklung von Kindern natürlich unterschiedlich schnell verläuft (Heidbrink, 1991).

2.7.2. Die kognitive Entwicklungstheorie von L. Kohlberg

Es werden drei Arten von Theorien des Moralerwerbs unterschieden (Kohlberg, 1995):

- Kognitivistische Theorien – kognitiv-entwicklungsorientierter Ansatz
- Sozialisationstheorie – Theorie des sozialen Lernens
- Psychoanalytische Theorie: Aneignung von Moral als Prozess der Internalisierung

Von den zahlreichen Theorien im Bereich der Moral- und Persönlichkeitsentwicklung ist die Kognitive Entwicklungstheorie von Lawrence Kohlberg die am stärksten beachtete (Lind, 2000). Eine in sich geschlossene, vollständig ausgearbeitete und empirisch gesicherte Kognitive Theorie der Moralentwicklung existiert bislang nicht. Der Theoriekern wurde vor allem von Piaget ausgearbeitet. Kohlberg hat ihn auf der Basis von kreativen und umfangreichen Studien erweitert und mehrfach abgewandelt. Kohlberg spricht von einem Ansatz („approach“),

um zu verdeutlichen, dass es sich hier nicht nur um neue Hypothesen handelt, sondern um eine gegenüber alten Ansätzen erweiterte Sichtweise, die beansprucht, dem Gegenstand der Moralphyschologie adäquater zu sein und sich in der Lösung moralbezogener Probleme fruchtbarer zu erweisen. (Lind, 2000, S. 47)

Kohlberg hat seine Theorie der moralischen Urteilsfähigkeit über fast 30 Jahre hinweg erarbeitet, überarbeitet und erweitert. Bezüglich der psychologischen Theoriebildung steht Kohlberg „in der Tradition Piagets, wobei er dessen Ansatz ... differenziert, teilweise revidiert und ... besser begründet“ (Heidbrink, 1991, S. 20).

Vier Thesen sind für die Kognitive Theorie der Moralentwicklung von zentraler Bedeutung (Lind, 2000):

1. Invariante Abfolge der Entwicklungsstufen
2. Strukturelle Ganzheit oder Organisation des moralischen Urteilens
3. Hierarchische Integration der Urteilsstufen
4. Parallelität zwischen der Entwicklung der kognitiven und der affektiven Komponente des moralischen Urteilsverhaltens

Zur besseren Veranschaulichung erfolgt nun eine kurze Beschreibung der vier Thesen. In den nächsten Unterkapiteln wird ausführlicher darauf eingegangen:

1. Invariante Sequenz

„Die qualitativ verschiedenen Stufen der moralischen Entwicklung bilden eine invariante Sequenz. Soziale Faktoren können die Entwicklung beschleunigen oder verlangsamen, aber nicht ihre Reihenfolge verändern“ (Lind, 2000, S. 54).

2. Strukturelle Ganzheit

Darunter versteht man, „dass jede der Urteilsstufen eine strukturelle Ganzheit formt, die verschiedene Verhaltensweisen vereinigt“ (Lind, 2000, S. 56).

3. Hierarchische Integration

Diese Annahme besagt, „dass neue Stufen der Moralität die strukturgenetisch früheren nicht ersetzen, sondern in Form einer hierarchischen Integration in sich aufnehmen“ (Lind, 2000, S. 58).

4. Moralisch-kognitive Parallelität

Die „Entwicklung des Inhalts und der Struktur, d.h. der inhaltlich-affektiven und der strukturell-kognitiven Komponente des moralischen Urteils“ (Lind, 2000, S. 59) verlaufen parallel. (siehe Kap. 2.7.2.3)

2.7.2.1. Das Stufenkonzept

Laut Heidbrink (1991) besteht eine der Grundannahmen kognitiver Entwicklungstheorien darin, dass „kognitive Strukturen Resultate der Interaktion zwischen Organismus und Umwelt sind“ (S. 24). Die kognitive Struktur ist das Produkt der Interaktion zwischen dem Subjekt und den Objekten. In Anlehnung an Piaget bezeichnet Kohlberg qualitative Unterschiede zwischen kognitiven Strukturen als *kognitive Stufen*.

Um Moralstufen zu begreifen, ist es hilfreich, „sie in eine Abfolge der Entwicklung der Gesamtpersönlichkeit einzuordnen“ (Kohlberg, 1995, S. 123).

Nach Kohlberg (1995) hängt moralisches Denken vom logischen Denken ab. Die Entwicklung *formal-operationaler* kognitiver Fähigkeiten ist die Voraussetzung für ein bewusstes moralisches Urteil in eigener oder fremder Sache. Gewöhnlich bilden sich diese Fähigkeiten etwa ab dem zwölften Lebensjahr (Lind, 1989).

Kohlberg und seine Mitarbeiter entwickelten die Theorie der Entwicklung des moralischen Bewusstseins. Dieser Theorie zufolge entwickelt sich die moralische Urteilsfähigkeit nach einem invarianten Muster. Für die Beschreibung der kognitiven Strukturen benützt Kohlberg Ergebnisse der philosophischen Ethik.

„Die Stufen des moralischen Urteils bilden eine invariante, unumkehrbare und konsekutive Reihenfolge diskreter Strukturen“ (Habermas, 1996, S. 138).

Den Übergang von einer Stufe zur nächsten begreift Kohlberg als Lernen. „Die Stufen des moralischen Urteils bilden eine Hierarchie in dem Sinne, daß die kognitiven Strukturen einer höheren Stufe diejenigen der jeweils niedrigeren Stufe aufheben, d. h. sowohl ersetzen wie in reorganisierter und ausdifferenzierter Form aufbewahren“ (Habermas, 1996, S. 138).

„Mit der Vorstellung eines Entwicklungspfades, der sich durch eine *hierarchisch geordnete Folge von Strukturen* beschreiben läßt, steht und fällt das Modell der Entwicklungsstufen“ (Habermas, 1996, S. 138).

Die Hypothese der *invarianten Sequenz* konnte empirisch gut bestätigt werden. Kohlberg und seine Mitarbeiter führten hierzu eine zwanzig Jahre umfassende Längsschnittstudie durch. 58 Personen wurden im Abstand von ungefähr vier Jahren wiederholt interviewt. Die erwarteten Entwicklungsabfolgen haben sich, abgesehen von wenigen Ausnahmen, ergeben (Lind, 2000).

Beim Vergleich von Schülern, Studienanfängern und Absolventen zeigte sich, dass mit höherem Alter und größeren Bildungserfahrungen die Rationalisierung festgefasster Meinungen abnimmt, während die moralische Reflexion zunimmt (Lind, 2000).

Der bisherigen Annahme, dass das moralische Urteil einer invarianten Abfolge von Entwicklungsschritten folgt, stehen allerdings auch widersprüchliche Ergebnisse entgegen. Habermas (1996) konnte im dritten Lebensjahrzehnt auftretende Fälle von Regression in der moralischen Urteilsfähigkeit beobachten. Die Erklärung für dieses Phänomen liefert J. R. Rest, der in seinen Untersuchungen feststellen konnte, „dass die moralische Urteilsfähigkeit von Menschen nicht nur zunehmen, sondern auch abnehmen kann, nämlich dann, wenn Bildungserfahrungen ausbleiben“ (Lind, 2002a, S. 113). Lind kommt zu der Schlussfolgerung, dass eine Stimulation der moralischen Entwicklung durch Bildungsprozesse essentiell ist: „Unter Lebens- und Arbeitsbedingungen die keine solche Stimulation bieten, oder den Menschen gar die Möglichkeit der moralischen Reflexion und des ethischen Diskurses entziehen, müssen wir mit einem schleichenden Verlust der moralischen Urteilsfähigkeit rechnen“ (2002a, S. 118).

Studien mit dem MUT haben ergeben, dass Bildung und Erziehung einen sehr wichtigen Faktor für die Entwicklung der moralischen Urteilsfähigkeit darstellen. Es handelt sich hierbei weder um Effekte der Anpassung an soziale Normen noch um das Produkt genetisch gesteuerter Reifung. Sicherlich haben soziale Normen und biologische Reifungsprozesse in bestimmten Phasen der Moralentwicklung auch wichtige Funktionen, aber ihr Einfluß wurde ... bislang weit überschätzt. (Lind, 2002a, S. 268)

2.7.2.2. Die Stufen der Entwicklung des moralischen Urteils

PIAGET-Phasen	KOHLBERG-Stufen			
A. Imitation	I./II.	III.	IV.	V./VI.
B. Heteronomie	Ich-Andere Dichotomie	Primärgruppe	Sekundärgruppe	Ideale Gesellschaft
C. Autonomie				

Tab. 10: Gegenüberstellung der Phasen nach Piaget mit den Stufen nach Kohlberg (Lind, 2000, S. 53)

Zunächst unterscheidet Kohlberg drei Ebenen der Moralentwicklung, diese sind durch die Beziehung zwischen dem Individuum und der sozialen Umwelt (soziale Perspektive) charakterisiert:

<i>Moralisches Urteil</i>	<i>Soziale Perspektive</i>
A. Präkonventionell	Individuum versus Andere
B. Konventionell	Mitglied der konkreten Gesellschaft
C. Postkonventionell	Mitglied einer idealen Gesellschaft

Tab. 11: Die drei Ebenen der Moralentwicklung (Lind, 2000, S. 50)

Kohlberg (1981) trifft weiter eine Einteilung in sechs moralische Stufen. Diese werden den drei Hauptebenen zugeteilt: der präkonventionellen (Stufen 1 und 2), der konventionellen (Stufen 3 und 4) und der postkonventionellen Ebene (Stufen 5 und 6).

Nach Kohlberg werden die sechs Stufen des moralischen Urteils folgendermaßen unterschieden:

Stufe 1	Das Motiv, eigene körperliche Schäden und Verletzungen (Strafe) zu vermeiden.
Stufe 2	Das Motiv, Vorteile und Vergünstigungen (Belohnungen) zu erlangen, auch wenn dabei gelegentlich Nachteile in Kauf genommen werden müssen.
Stufe 3	Das Motiv, Mißbilligung durch andere zu vermeiden bzw. ihre Anerkennung zu erreichen.
Stufe 4	Das Motiv, die Gesetze und Ordnung der Gesellschaft zu achten und zu ihrer Aufrechterhaltung beizutragen. Affektiv-kognitive Parallelität: Korrelationen zwischen der moralischen Urteilsfähigkeit (MUT) und den Einstellungen zu den sechs Moralstufen.
Stufe 5	Das Motiv, die Regeln freier, demokratischer Willensbildung (Mehrheitsprinzip etc.) zu beachten und die auf ihnen gründenden Verträge einzuhalten.
Stufe 6	Das Motiv, die Würde jedes einzelnen Menschen zu achten und sein Handeln an den universellen Prinzipien der Gerechtigkeit, Vernunft und Logik zu messen.

Tab. 12: Der affektive Aspekt: Stufenmodell moralischer Motive nach Kohlberg (Lind, 2002a, S. 53)

Ebene	Basis des moralischen Urteils	Entwicklungsstufen
I	Moralische Wertung beruht auf äußeren, quasi-physischen Geschehnissen, schlechten Handlungen, oder auf quasi-physischen Bedürfnissen statt auf Personen und Normen	<i>Stufe 1:</i> Orientierung an Bestrafung und Gehorsam. Egozentrischer Respekt vor überlegener Macht oder Prestigestellung bzw. Vermeidung von Schwierigkeiten. Objektive Verantwortlichkeit. <i>Stufe 2:</i> Naiv egoistische Orientierung. Richtiges Handeln ist jenes, das die Bedürfnisse des Ich und gelegentlich die der anderen instrumentell befriedigt. Bewußtsein für die Relativität des Wertes der Bedürfnisse und der Perspektive aller Beteiligten. Naiver Egalitarismus und Orientierung an Austausch und Reziprozität.
II	Moralische Wertung beruht auf der Übernahme guter und richtiger Rollen, der Einhaltung der konventionellen Ordnung und den Erwartungen anderer.	<i>Stufe 3:</i> Orientierung am Ideal des „Guten Jungen“. Bemüht, Beifall zu erhalten und anderen zu gefallen und ihnen zu helfen. Konformität mit stereotypischen Vorstellungen vom natürlichen oder Mehrheits-Verhalten, Beurteilung aufgrund von Intentionen. <i>Stufe 4:</i> Orientierung an Aufrechterhaltung von Autorität und sozialer Ordnung. Bestrebt, „seine Pflicht zu tun“, Respekt vor der Autorität zu zeigen und die soziale Ordnung um ihrer selbst willen einzuhalten. Rücksicht auf die Erwartungen anderer.
III	Moralische Wertung beruht auf Konformität des Ich mit gemeinsamen (oder potentiell gemeinsamen) Normen, Rechten oder Pflichten.	<i>Stufe 5:</i> Legalistische Vertrags-Orientierung. Anerkennung einer willkürlichen Komponente oder Basis von Regeln und Erwartungen als Ausgangspunkt der Übereinstimmung. Pflicht definiert als Vertrag, allgemein Vermeidung der Verletzung von Absichten oder Rechten anderer sowie Wille und Wohl der Mehrheit. <i>Stufe 6:</i> Orientierung an Gewissen oder Prinzipien. Orientierung nicht nur an zugewiesenen sozialen Rollen, sondern auch an Prinzipien der Entscheidung, die an logische Universalität und Konsistenz appellieren. Orientierung am Gewissen als leitendes Agens und an gegenseitigem Respekt und Vertrauen.

Tab. 13: Klassifikation der moralischen Urteile nach Entwicklungsebenen und –stufen (Quelle: Kohlberg, 1974, S. 60-61, Tab. 2)

Kohlberg (1995) beschreibt die Moralstufen wie folgt:

Man versteht die einzelnen Stufen besser, wenn man zuvor eine Vorstellung von drei moralischen Niveaus gewonnen hat. Die präkonventionelle Ebene ist die moralische Denkebene der meisten Kinder bis zum 9. Lebensjahr, einiger Jugendlicher und vieler jugendlicher und erwachsener Straftäter. Der konventionellen Ebene sind in unserer und in anderen Gesellschaften die meisten Jugendlichen und Erwachsenen zuzurechnen. Die postkonventionelle Ebene wird von einer Minorität von Erwachsenen erreicht – und das in der Regel erst nach dem 20. Lebensjahr. (S. 126-127).

Wie Lind (2000) berichtet, konnten Studien belegen, dass die moralische Urteilsfähigkeit mit der Höhe des Schulabschlusses bzw. bei Studenten mit der Anzahl der absolvierten Semester positiv variiert.

Kohlberg hat in weiterer Folge noch eine Unterteilung der sechs Stufen in A- und B-Substufen vorgenommen. Da dies jedoch für die vorliegende Arbeit nicht von Relevanz ist, wird die Autorin auch nicht näher darauf eingehen (vgl. Heidbrink, 1991; Kohlberg, Levine & Hower, 1983).

2.7.2.3. Affekt und Kognition (Inhalt und Struktur)

Zwei Aspekte der moralischen Urteilkraft werden von Kohlberg hervorgehoben:

1. *Inhaltlich-affektiver Aspekt*: Akzeptanz moralischer Prinzipien
2. *Kognitiv-struktureller Aspekt*: Der Grad, „mit dem das Urteilsverhalten eines Individuums den Kriterien der Moral nahe kommt“ (Kohlberg, 1958, zitiert nach Lind, 2000, S. 49)

Lind (1993, S. 83) versteht unter den *affektiven Aspekten* des moralischen Urteilens „die individuellen moralischen Einstellungen, das heißt die moralischen Prinzipien, an denen der einzelne sein Urteilsverhalten ausrichtet“.

Orientiert sich eine Person bei der Bewertung von Argumenten *konsistent* an der moralischen Qualität dieser Argumente, dann zeigt sie hohe moralische Urteilsfähigkeit. Der Begriff der moralischen Urteilsfähigkeit ist allein für den *kognitiven Aspekt* reserviert, „da es wohl ohne Anstrengung oder Erfahrung möglich ist, hohe Stufen des moralischen Urteilens zu präferieren oder zu benutzen (weshalb der affektive Aspekt *keine* Fähigkeit darstellt), aber wohl nicht, sich konsistent an moralischen Prinzipien zu orientieren“ (Lind, 1993, S. 84).

Dabei gilt zu beachten, dass beide Aspekte Eigenschaften desselben Verhaltens sind, sie können also nur analytisch unterschieden, nicht aber dinglich getrennt werden (Lind, 2000).

Piagets entscheidende psychologische Entdeckung war, dass moralisches Verhalten nicht nur eine Frage der bloßen Übereinstimmung mit äußeren Normen und auch nicht nur eine Frage einer starken affektiven Bindung an soziale Werte ... ist, sondern dass beim moralischen Handeln auch kognitive (Urteils-)Prozesse beteiligt sind. Kohlberg hat in seinen frühen Arbeiten die Art der Beziehung mit der Annahme einer affektiv-kognitiven Parallelität zu präzisieren und ... zu überprüfen versucht. (Lind, 2000, S. 149)

Die Annahme einer affektiv-kognitiven Parallelität beinhaltet folgende Punkte (Lind, 2000):

1. Beide Aspekte bilden eine Einheit und können nicht dinglich voneinander getrennt werden.
2. Der affektive und der kognitive Aspekt korrelieren in ihren Ausprägungen empirisch positiv miteinander.
3. Die beiden Aspekte stehen auch hinsichtlich der individuellen moralisch-kognitiven Entwicklung in einem engen Zusammenhang.

Kohlbergs Methode der Erfassung des moralischen Urteilens ist nicht in der Lage, die kognitive Struktur und den affektiven Inhalt theoriegemäß „als nicht trennbare, aber unterscheidbare Aspekte zu operationalisieren und die Hypothese der Parallelität zu untersuchen“ (Lind, 2000, S. 154). Gegenüber herkömmlichen Methoden weist Kohlbergs strukturelle Methode des Klinischen Interviews deutliche Verbesserungen hinsichtlich der Erfassung der *affektiven* Komponente auf. Es gelang ihm jedoch nicht das Problem der adäquaten Operationalisierung der *kognitiven* Komponente zu lösen (Lind, 2000).

Gemäß der Kognitiven Entwicklungstheorie ... sind Affekt und Kognition nicht nur als Einheit zu betrachten, sondern auch so zu operationalisieren, dass beide als Aspekte unterschieden, aber nicht dinglich separiert werden, und dass formale Eigenschaften der kognitiv-strukturellen Entwicklung wie Integration und Differenzierung eindeutig definierbar sind. Diese Forderung lässt sich offenbar nur einlösen, indem Affekt und Kognition als Aspekte oder Eigenschaften ein und desselben Verhaltens einer Person erfasst werden. Der kognitive Aspekt des Urteilsverhaltens existiert nicht abgetrennt von inhaltlich definierten moralischen Affekten; er ist keine Ursache des Urteilsverhaltens, vielmehr ist dessen Struktur selbst der kognitive Aspekt. Und auch die affektiven Dispositionen des moralischen Urteilens sind nicht dinglich getrennt von kognitiven Strukturen zu sehen; vielmehr sind es die Elemente selbst, die eine Struktur konstituieren – und auch erst vollständig definieren. Denn anders als in physikalischen Strukturen reicht im Falle von funktional gerichteten Handlungsstrukturen eine *rein* formale Charakterisierung nicht aus. Affekte sind die *charakteristischen Inhalte* einer kognitiven Struktur. (Lind, 2000, S. 161)

Lind (2000) hat mit dem Moralischen-Urteil-Test (MUT) die *Parallelitätshypothese* „methodisch präzisiert und den affektiven Aspekt als die präferierte moralische Regel, den kognitiven Aspekt als den Grad der strukturellen Verankerung im individuellen Urteil definiert“ (vgl. Kap. 3.4.1.4.).

In den Untersuchungen mit dem MUT zeigte sich zwischen Inhalt und Struktur eine deutliche Parallelität: „Je höher die Konsistenz in Bezug auf moralische Kategorien ist, desto häufiger werden die höheren Stufen der moralischen Argumentation akzeptiert und die niedrigen Stufen abgelehnt“ (Lind, 2000, S. 59). Bei internationalen Vergleichen von Studenten konnte nachgewiesen werden, dass Argumente auf Stufe 6 allen anderen Argumentationsstufen vorgezogen wurden, während die Argumente auf Stufe 1 die stärkste Ablehnung hervorriefen (Lind, 2000).

Linds (2000, S. 167) „Untersuchung hat gezeigt, dass eine Unterscheidung von affektiven und kognitiven Aspekten des moralischen Urteilens nicht nur theoretisch, sondern auch operational möglich ist, und dass beide Aspekte in der Tat, wie die Parallelitätshypothese impliziert, empirisch eng miteinander korrelieren.“ Die Ergebnisse von Linds

empirischen Analysen haben damit bestätigt, dass eine integrale Sichtweise, in der Kognition und Affekt als Aspekte desselben Verhaltens, aber dennoch analytisch unabhängig voneinander erfasst werden können, auch methodisch realisierbar ist; und dass zwischen beiden logisch unabhängigen

Aspekten dennoch ein enger empirischer Zusammenhang besteht, der auf eine entwicklungsmäßige Parallelität hinweist. (2000, S. 167)

Lind & Wakenhut (1985) charakterisieren die kognitiv-moralische Entwicklung nicht nur durch eine kognitive Verankerung von moralischen Grundsätzen, sondern auch durch den notwendigen Prozess der Aufgaben von Verhaltensweisen (Differenzierung) und der zunehmenden Einbeziehung von situativen Umständen sowie konkurrierenden Werten (hierarchische Integration).

2.7.2.4. Moralische Dilemmata

In der Theorie Kohlbergs nehmen moralische Dilemmata eine zentrale Stellung ein. Sie stellen einen der Ausgangspunkte für die Entwicklung der Theorie dar. Außerdem sind sie „zentraler Bestandteil des Kohlbergschen Verfahrens zur Erhebung der jeweiligen individuellen moralischen Stufe“ (Heidbrink, 1991, S. 37).

Für die Analyse der Struktur eines moralischen Dilemmas erscheint es sinnvoll sich an der von Dörner entworfenen Problemlösungstheorie zu orientieren. Probleme lassen sich nach Dörner durch drei Komponenten charakterisieren:

1. *ein unerwünschter bzw. unbefriedigender Anfangszustand;*
 2. *ein erwünschter Endzustand;*
 3. *eine Barriere, die die Transformation des Anfangs- in den Endzustand im Moment verhindert.*
- (Dörner, 1976, zitiert nach Heidbrink, 1991, S. 37).

Alle drei Problemkomponenten – Anfangszustand, Barriere, mögliche Endzustände – müssen bei einem moralischen Dilemma genau bekannt sein. Die Barriere besitzt bei einem moralischen Dilemma eine konstruktive Besonderheit, denn ein Dilemma beinhaltet ein Problem, welches entweder nicht oder zumindest nicht optimal lösbar ist (vgl. Heidbrink, 1991, S. 39).

Die bevorzugte Handlungsalternative gibt keinen Aufschluss über das moralische Urteilsniveau. Es ist nicht die Wahl interessant, sondern die *Begründung* der Wahl (Heidbrink, 1991).

2.7.3. Die Stufen der Entwicklung des logischen Denkens nach Piaget

Obwohl Kohlberg thematisch Piagets Theorie zum moralischen Urteil aufgegriffen hat, orientiert er sich konzeptionell stärker an Piagets Stufen der Entwicklung des logischen Denkens (Heidbrink, 1991). Zum besseren Verständnis skizziert die Autorin hier im Folgenden kurz Piagets Stufenmodell des logischen Denkens.

Jean Piaget unterscheidet fünf Stufen der Intelligenzentwicklung (Rollett, 1997, S. 61):

1. Die sensumotorische Intelligenz (0 bis 1,6 Jahre)
2. Das vorbegrifflich-symbolische Denken (1 bis 4 Jahre)
3. Das anschauliche Denken (4 bis 7 Jahre)
4. Die konkreten Operationen (8 bis 11 Jahre)
5. Die formalen Operationen (ab etwa 11 Jahren)

Die drei logischen Stufen lassen sich folgendermaßen unterteilen (Heidbrink, 1991)⁵:

1. Symbolisches, intuitives Denken
2. Stufe der konkreten Operationen
3. Stufe der formalen Operationen

⁵ Für die vorliegende Arbeit ist eine genauere Darstellung nicht von Belang. Für eine ausführlichere Beschreibung siehe Heidbrink, 1991, S. 78.

2.8. Empirische Befunde

Die empirische Forschung konzentrierte sich vor allem auf die Frage, ob Persönlichkeitsfragebogen und – tests prinzipiell verfälschbar sind und wie man dies verhindern könnte. Es liegen bis dato allerdings nur vereinzelt Studien vor, die sich der Frage widmen, ob bestimmte Personengruppen fähiger im Verfälschen sind als andere.

2.8.1. Fakeability

Zahlreiche Untersuchungen haben gezeigt, dass Testwerte willkürlich in eine vorher festgelegte Richtung verfälscht werden können (Schneider & Hübner, 1980; Martin et al., 2002). Viswesvaran und Ones (1999) kamen in ihrer Metaanalyse allerdings zu dem Schluss, dass Persönlichkeitsfragebogen anfälliger für Verfälschungen in negative Richtung sind. Wenn Probanden also instruiert werden verfälschende Antworten zu geben, so sind sie erfolgreicher im faking bad als im faking good.

Viswesvaran und Ones (1999) haben 51 Studien über bewusste Verfälschung einer Metaanalyse unterzogen. Sie fanden heraus, dass alle Big-Five-Faktoren in gleichem Ausmaß verfälschbar sind. Die Persönlichkeitsfragebogen waren, wie bereits erwähnt, für fake-bad-Instruktionen anfälliger als für fake-good-Instruktionen. Am anfälligsten für Antwortverzerrungen unter faking-Instruktionen erwiesen sich Skalen zur sozialen Erwünschtheit.

Wie schon Häcker et al. bestätigt haben, führt die Methode der Instruktionsvariation zu extremen Verfälschungseffekten. Sie identifizierten zumindest zwei Faktoren, bei denen „Motivationseinflüsse in Richtung konformer und sozial erwünschter Testbeantwortung zu erwarten“ (1979, S. 20) sind, nämlich *Angst* und *kulturelle Konformität*.

Scores der Testskalen mit dem höchsten Grad von sozialer Erwünschtheit zeigten erhöhte Mittelwerte unter „fake good“-Bedingungen. Es scheint also ein Zusammenhang zwischen der Beantwortung im Sinne sozialer Erwünschtheit und dem faking good zu bestehen (Amelang, Schäfer & Yousfi, 2002).

Erfolgreiche Verfälschung erfordert eine genaue Kenntnis vom Persönlichkeitsprofil der Berufszielgruppe, wenn sich eine Person für eine Stelle bewirbt bzw. unter der Instruktionsbedingung, so zu antworten, als bewerbe man sich für eine Arbeitsstelle (Martin et al., 2002).

Karner (2002) hat mit dem Rasch-Modell auf Homogenität geprüft. Unter den Probanden befanden sich sowohl Freiwillige als auch Personen in einer Auswahl-situation. Bei den Probanden, welche die Fragebogen im Zuge eines Auswahlverfahrens beantworteten, wurde eine große Anzahl von Items entfernt, da sie die Vorannahme des Modells nach Homogenität verletzten. Dies deutete Karner als Hinweis auf Verfälschung.

Durch Within-Subjects-Designs erhält man genauere Schätzungen. Dies könnte auch daran liegen, dass es sich bei der Fakeability um eine individuelle Fähigkeit handelt. Manche Menschen sind mehr geneigt oder fähig im Verfälschen von Persönlichkeitsfragebogen als andere. Between-Subjects-Designs könnten hinsichtlich der Interaktionen von Subjekt x Behandlung verfälscht sein und reduzieren die statistische Macht (Viswesvaran & Ones, 1999).

Ein Within-Subjects-Design wurde verwendet, in dem Teilnehmer sowohl einen Situational Judgment Test (SJT)⁶ als auch einen Persönlichkeitstest unter einer "honest"-Instruktion sowie unter einer "fake good"-Instruktion zu beantworten hatten (Nguyen et al., 2005). Die verwendeten Antwortformate waren Knowledge-Format vs. Behavioral Tendency-Format. Beim *Knowledge-Format* geht es darum, die „richtigen“ Antworten zu finden. Es ist resistenter gegenüber einer Verfälschung. Das *Behavioral Tendency-Format* hingegen beinhaltet Selbstberichte über Verhaltenstendenzen. Es ist somit einem Persönlichkeitstest ähnlich und deshalb auch leichter zu verfälschen.

Peeters und Lievens (2005) haben eine Studie an Studenten durchgeführt. Sie kamen zum Ergebnis, dass auch *Situational Judgment Tests* verfälscht werden können. Die Studenten in der Faking-Bedingung erzielten signifikant höhere Testscores als Studenten in der Honest-Bedingung. Das Ergebnis legt nahe, dass Studenten auch einen *SJT über Erfolge von College-Studenten* verfälschen können, wenn sie dazu instruiert werden. SJTs sind verfälschbarer, wenn sie stark mit Persönlichkeits-Traits korreliert sind. Der in der Untersuchung verwendete SJT korrelierte signifikant mit den Big-Five-Persönlichkeitsdimensionen.

Es spricht einiges dafür, dass sich Stellenbewerber im Ausmaß, in dem sie ihre Antworten auf Persönlichkeitsfragebogen verfälschen, unterscheiden. Rosse et al. (1998) fanden heraus, dass der Umfang der Verfälschung (operationalisiert durch die Impression-Management-Skala von Paulhus vom Jahre 1984) zwischen den Stellenbewerbern variierte. Einige Bewerber übertrieben kaum ihre vorhandenen, positiven Persönlichkeitseigenschaften, wohingegen andere ihre Antworten auf einem Niveau verzerrten, welches schon als Betrug zu bezeichnen war. Nur eine kleine Anzahl von bereits eingestellten Personen wies höhere Impression-Management-Werte auf als der durchschnittliche Bewerber.

Die Untersuchung von Rosse et al. (1998) konnte zeigen, dass einige Bewerber ihre Antworten verfälschen, während andere dies nicht tun. Das Ausmaß der Verfälschung variierte zwischen den Bewerbern. Sogar einige Personen, die bereits fix eingestellt waren, antworteten so (aus welchem Grund auch immer), als ob sie die Antworten in eine positive Richtung verzerren wollten.

⁶ Situationsurteil-Tests (SJTs) wurden entwickelt, um die Antworten eines Stellenbewerbers auf eine Reihe von arbeitsverwandten Szenarien zu beurteilen. Ihr Einsatz als Instrumente in der Personalauswahl wird immer populärer (Nguyen et al., 2005).

Größtenteils galt in der Forschung über das Verfälschen von Persönlichkeitsinventaren die Annahme, dass Personen entweder verfälschen oder eben ehrlich antworten. Eine Unterscheidung innerhalb dieser beiden Gruppen wurde im Allgemeinen nicht getroffen. Zickar, Gibby und Robie (2004) verwendeten eine kürzlich entwickelte statistische Technik, die *Mixed-Model Item Response Theory*, um weitere Untergruppen zu identifizieren. Die Versuchspersonen hatten zwei unterschiedliche Persönlichkeitsinventare unter verschiedenen Instruktionsbedingungen zu bearbeiten. Für einen der Persönlichkeitstests erhielten die Autoren eine Stichprobe aus Stellenbewerbern und bereits eingestellten Personen. Für den zweiten Test wurde ein Teil der Personen dazu angehalten ehrlich zu antworten. Zwei weitere Gruppen von Personen hatten diesen unter der Instruktion verfälschende Antworten zu geben, auszufüllen. Eine Unterteilung in drei Kategorien war letztlich erforderlich, um allen Antwortmustern gerecht zu werden. Insgesamt weisen diese Ergebnisse darauf hin, dass vorherige Annahmen über die Natur des Verfälschens auf Persönlichkeitsinventaren zu einschränkend gewesen sind. Die Jobbewerber wiesen auf allen Skalen höhere Mittelwerte auf als die Angestellten. In der Honest-Bedingung wurden durchwegs niedrigere Mittelwerte gefunden als in der Faking-Bedingung.

Es gab bei dieser Untersuchung jedoch auch Personen, die ehrlich antworteten, obwohl sie der Faking-Bedingung zugeteilt waren. Diese Ergebnisse sprechen dafür, dass unterschiedliche Antwortstile sogar innerhalb von kontrollierten experimentellen Bedingungen zu finden sind. Diese Unterschiede können einer Vielfalt von Faktoren zugeschrieben werden: *Fakeability*, *Missverständnis der Instruktionen*, *Ausmaß der Selbsterkenntnis*. Die Unterschiede innerhalb der Gruppe von Bewerbern könnten auf unterschiedliche Motivation (einige könnten diesen Job wirklich wollen und sich demnach mehr dazu gezwungen fühlen zu verfälschen), Unterschiede im Gerechtigkeitsempfinden und nicht zuletzt auf die Fakeability zurückzuführen sein. Personen, die bereits eingestellt waren, könnten sich auch im Ausmaß unterscheiden haben, in dem sie glaubten, dass die Geschäftsführung die Testergebnisse für Personalentscheidungen verwenden würde. Die Ergebnisse der Untersuchung unterstützen indirekt die Annahme, dass Personen, wenn sie verfälschen, ihre Testwerte um einen konstanten Betrag erhöhen (Zickar, et al., 2004).

Die bisherigen empirischen Ergebnisse weisen nicht auf eine systematische Abnahme der Kriteriumsvalidität von Faking-Messungen in tatsächlichen Bewerbungs-Situationen hin. Hingegen wurde durchwegs eine wesentliche Abnahme der Kriteriumsvalidität in Untersuchungen mit induziertem Verfälschen berichtet. Dies erweckt den Anschein, dass die Prozesse des Faking bzw. der Selbstpräsentation sowie deren Ergebnisse in der realen Welt beträchtlich komplizierter sein dürften als diejenigen von simulierten Studien zur Verfälschbarkeit (Marcus, 2006).

Viswesvaran und Ones (1999) haben mit ihrer Metaanalyse über faking in Bewerber-Situationen gezeigt, dass Stellenbewerber ihre Testwerte auf einem Persönlichkeitsinventar um eine halbe Standardabweichung erhöhen können, unter Verwendung eines Between-Subjects-Designs. Bei Untersuchungen mit Within-Subjects-Designs waren sie in der Lage ihre Testscores in der Faking-

Bedingung um ungefähr eine Standardabweichung zu erhöhen. Bei faking-good-Instruktionen waren die Probanden in der Lage ihre Antworten um etwa eine halbe Standardabweichung zu verändern.

Viswesvaran und Ones (1999) Meta-Analyse zeigte, dass das Ausmaß der Verfälschbarkeit für die Skala *Emotionale Stabilität* am größten und für die Skala *Agreeableness* (Verträglichkeit) am geringsten ausfiel.

In der Analyse von Nguyen et al. (2005) spiegelte sich die Hauptwirkung der *Versuchsanordnung* wider. Das Ausmaß der Verfälschung war im Allgemeinen in der Versuchsanordnung Honest -> Faking größer als bei Faking -> Honest. Es scheint so, dass Teilnehmer, die die Tests schon einmal zuvor beantwortet hatten, besser dazu imstande waren die Antworten zu verfälschen.

2.8.2. Differentialpsychologische Aspekte der Fakeability

Die Autorin möchte an dieser Stelle vermerken, dass in der Forschung über die Verfälschbarkeit oftmals die ermittelten Scores auf Skalen zur sozialen Erwünschtheit herangezogen wurden, um Verfälschung zu definieren (Graham et al., 2002).

Mahar, Cologon und Duck (1995) konnten in ihrer Untersuchung *keine geschlechtsspezifischen Unterschiede* in Bezug auf die Fähigkeit zur Verfälschung feststellen.

Ones et al. (1996) fanden heraus, dass Stellenbewerber, welche hohe Werte für *soziale Erwünschtheit* hatten, dazu tendierten auch hohe Werte für *emotionale Stabilität* und *Gewissenhaftigkeit* zu haben. Dies könnte aber auch daran liegen, dass diese Personen es für sozial erwünschter hielten emotional stabil und gewissenhaft zu sein und deswegen ihre Antworten in diese Richtung verfälschten.

Personen mit weniger Jahren an schulischer Bildung erzielten höhere Scores in Skalen zur *sozialen Erwünschtheit* als Personen, deren Ausbildung längere Zeit in Anspruch genommen hatte (Ones et al., 1996).

Bei einer Untersuchung an Studenten stellten Peeters und Lievens (2005) fest, dass *Frauen* besser dazu in der Lage waren den verwendeten *Situational Judgment Test* zu verfälschen als Männer.

McFarland und Ryan (2000; zitiert nach Nguyen et al., 2005) zeigten, dass es individuelle Unterschiede in der Fakeability gibt. Sie fanden auch, dass Personen, die hohe Scores auf den Big Five Persönlichkeitsdimensionen der *Gewissenhaftigkeit* und *Emotionalen Stabilität* erzielten, weniger verfälschten als diejenigen, die niedrige Werte auf denselben Dimensionen erreichten.

3. Empirische Untersuchung

Anhand der Fachliteratur wurden im theoretischen Teil die Fehlerfaktoren von Persönlichkeitsfragebogen erläutert. Verschiedene Ansätze in Bezug auf die absichtliche Herbeiführung bzw. die Verhinderung von Antwortverzerrung sind dargestellt worden.

Die Persönlichkeitstheorie von R. B. Cattell und die Impression-Management-Theorie wurden beschrieben. Historisch aufgebaute Denkansätze speziell in den Bereichen Repression vs. Sensitization und moralischer Urteilsfähigkeit wurden dargestellt.

Im Kapitel empirische Befunde wurde ein Überblick über die Ergebnisse von bisherigen Untersuchungen zur Verfälschung von Persönlichkeitsfragebogen gegeben.

3.1. Fragestellung und Operationalisierung

Die zentrale Fragestellung der Untersuchung lautet: Besteht ein Zusammenhang zwischen der Fähigkeit einer Person zur Verfälschung von Persönlichkeitsfragebogen und spezifischen Merkmalen dieser Person? Die Untersuchung soll klären inwieweit demographische Daten wie Alter, Geschlecht und Bildung sowie Persönlichkeitseigenschaften wie Extra- und Introversion, Emotionale Stabilität vs. Emotionale Instabilität, Repression vs. Sensitization, moralische Urteilsfähigkeit und die Tendenz zur positiven Selbstdarstellung mit der Fakeability in Zusammenhang stehen. Da die bisher vorliegenden Forschungsergebnisse dafür sprechen, dass Personen dazu in der Lage sind einen Persönlichkeitsfragebogen zu verfälschen, sind die Hypothesen gerichtet formuliert.

3.2. Instruktionsbedingungen

Die vorliegende Untersuchung stützt sich auf ein *Within-Subjects-Design*. Sie ist so angelegt, dass ein und dieselbe Versuchsperson denselben Fragebogen mehrmals beantworten muss, einmal wahrheitsgemäß sowie unter einer faking-good und/oder faking-bad-Instruktion. Die Honest-Instruktion ist die gewöhnliche Standartinstruktion des Fragebogens. Sie wird bei dieser Untersuchung jedoch noch um einen Hinweis erweitert: „Hinweis: Es handelt sich hier um eine wissenschaftliche Untersuchung. Um zu aussagekräftigen Ergebnissen zu gelangen, ist es äußerst wichtig, dass Sie die Fragen so ehrlich als möglich beantworten. Die Auswertung des Fragebogens erfolgt anonym. Ihre Daten werden selbstverständlich vertraulich behandelt.“

3.2.1. Instruktion der Faking-bad-Bedingung

Um den Probanden das Verfälschen in eine negative Richtung zu erleichtern, wurde folgende Anforderung an sie gestellt: *„Stellen Sie sich vor, Sie sind auf Arbeitsuche. Vom AMS wird Ihnen ein Vorstellungsgespräch für die Stelle eines Küchengehilfen/einer Küchengehilfin in der Zeit der Hochsaison vermittelt. Sie möchten jedoch auf gar keinen Fall zur Hochsaison im Gastgewerbe tätig sein. Bei Ihrem Vorstellungsgespräch wird Ihnen dieser Fragebogen vorgelegt. Da Sie den Job nicht annehmen wollen, sollten Sie darum bemüht sein, einen besonders schlechten Eindruck zu hinterlassen.“*

Damit sich die Probanden besser orientieren können steht auf dem Deckblatt auch noch ein fiktives Anforderungsprofil für den Beruf des Küchengehilfen/der Küchengehilfin:

„Bei der Beantwortung der Fragen denken Sie bitte daran, dass man als Küchengehilfe/Küchengehilfin bestimmte Voraussetzungen zu erfüllen hat. Nachfolgendes Anforderungsprofil dient Ihnen als Hilfestellung beim Beantworten des Fragebogens.“

Persönliche Anforderungen:

- *Belastbarkeit*
- *hohe Selbständigkeit*
- *Verantwortungsgefühl*
- *Kommunikationsfähigkeit*
- *Kontaktfreude*
- *Lebhaftigkeit*
- *emotionale Wärme“*

3.2.2. Instruktion der Faking-good-Bedingung

Die Instruktion zur Verfälschung in eine positive Richtung lautete wie folgt: *„Stellen Sie sich vor, Sie bewerben sich um einen Posten als Geschäftsführer/in. Es liegt Ihnen viel daran diesen Job zu bekommen. Nun haben Sie es geschafft, zu einem Bewerbungsgespräch eingeladen zu werden. Bei diesem wird Ihnen dieser Fragebogen vorgelegt. Da Sie unbedingt als Geschäftsführer/in tätig sein möchten, sollten Sie nun sehr darum bemüht sein, die Fragen so zu beantworten, dass Sie den bestmöglichen Eindruck hinterlassen.“*

Mahar et al. (1995) fanden heraus, dass sich Probanden an Ihren Stereotypen in Bezug auf die relevante Berufsgruppe orientieren, wenn sie ihr Persönlichkeitsprofil dahingehend ändern sollen, dass sie eine bestimmte Anstellung erhalten bzw. nicht erhalten. Zur besseren Orientierung der Probanden steht deshalb auf dem Deckblatt auch noch das Anforderungsprofil für den Beruf des Geschäftsführers (AMS-Berufsinformationssystem, 2001):

„Bei der Beantwortung der Fragen denken Sie bitte daran, dass man als Geschäftsführer/in bestimmte Voraussetzungen zu erfüllen hat. Nachfolgendes Anforderungsprofil dient Ihnen als Hilfestellung beim Beantworten des Fragebogens.“

Persönliche Anforderungen:

- *Durchsetzungsvermögen*
- *analytisches Denkvermögen*
- *Belastbarkeit*
- *Führungsqualitäten*
- *gutes Auftreten*
- *hohe Selbständigkeit*
- *hohes Verantwortungsgefühl*
- *Kommunikationsfähigkeit*
- *Kontaktfreude*
- *Organisationstalent*
- *unternehmerisches Denken*
- *Verhandlungsgeschick“*

3.3. Ziele und Hypothesen der Untersuchung

Ziel dieser Untersuchung ist die differentialpsychologische Überprüfung der Fakeability. Als Grundlage zur Erhebung der Persönlichkeit dienen die Messung des Extra- und Introversionsgrades mittels 16 PF-R, die Skalenwerte der Probanden auf der Primärskala „Emotionale Stabilität“ des 16 PF-R, die Messung von Repression versus Sensitization mittels RSCI, die Erfassung der moralischen Urteilsfähigkeit mittels MUT sowie die Messung der Tendenz zur positiven Selbstdarstellung durch die IMS (siehe Kap. 3.4.).

Zur besseren Verständlichkeit folgt eine Übersicht über die verwendeten Zeichen bei der Darstellung der Hypothesen: Extraversion (E), Ängstlichkeit (AE), Selbstkontrolle (SK), Unabhängigkeit (UA), Unnachgiebigkeit (UN), honest (HO), faking bad (FB), faking good (FG).

Bei der Hypothesenbildung wurden die Beschreibungen der Globalskalen des 16 PF-R als Anhaltspunkt (Schneewind & Graf, 1998) zur Hilfe genommen. Die Autorin ist danach mit folgenden Hypothesen an die Untersuchung herangetreten:

3.3.1 Hypothesen zur Faking-bad-Bedingung

Die Probanden mussten in der Faking-bad-Bedingung den 16 PF-R so ausfüllen, als ob sie vom AMS zu einer Bewerbung für eine Stelle als Küchengehilfin müssen und diesen Job nicht haben wollen (siehe

Kap. 3.2.1.). Die Autorin erwartet, dass sich die Probanden in der Faking-bad-Bedingung demnach introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher darstellen.

Die *unabhängigen Variablen* sind demographische Daten wie das Alter (jung versus alt), das Geschlecht (männlich versus weiblich) sowie der Bildungsgrad (niedrige versus hohe Bildung) der Probanden. Die weiteren unabhängigen Variablen stellen die Persönlichkeitseigenschaften wie Extra- versus Introversion, Emotionale Stabilität versus Emotionale Instabilität, Repression versus Sensitization, niedrige versus hohe moralische Urteilsfähigkeit sowie niedrige versus hohe Tendenz zur positiven Selbstdarstellung dar. Als *abhängige Variablen* fungieren die Stenwerte auf den Globalskalen des 16 PF-R. Hier stellt der honest-Wert die Baseline dar und lässt somit einen Vergleich mit den gefakten Werten zu.

Bei der *Globalskala Extraversion* ist davon auszugehen, dass die Personen ihr Profil dahingehend ändern, dass sie sich in der Faking-bad-Situation als introvertierter darstellen. Da sie dazu aufgefordert sind, dem fiktiven Anforderungsprofil nicht zu entsprechen, sollten sie sich als unkommunikativ, kontaktscheu, emotional kalt und nicht lebhaft darstellen.

$$H1A_1: \mu_{E\ HO} > \mu_{E\ FB}$$

$$H1A_0: \mu_{E\ HO} = \mu_{E\ FB}$$

Bei der *Globalskala Ängstlichkeit* sollten die Probanden darum bemüht sein sich in der Faking-bad-Situation als ängstlicher darzustellen. Sie sollten das Profil eines emotional instabilen, angespannten und besorgten Menschen anstreben.

$$H1B_1: \mu_{AE\ HO} < \mu_{AE\ FB}$$

$$H1B_0: \mu_{AE\ HO} = \mu_{AE\ FB}$$

Bei der *Globalskala Selbstkontrolle* sollten die Probanden ihre Skalenwerte dahingehend ändern, dass sie in der Faking-bad-Situation den Eindruck erwecken unbeherrscht zu sein. Sie sollten sich als unangepasste Personen darstellen, denen eine perfektionistische Arbeitsweise nicht liegt.

$$H1C_1: \mu_{SK\ HO} > \mu_{SK\ FB}$$

$$H1C_0: \mu_{SK\ HO} = \mu_{SK\ FB}$$

Die Autorin vermutet, dass sich die Probanden bei der *Globalskala Unabhängigkeit* in der Faking-bad-Situation als unabhängiger darstellen. Unabhängige Menschen sind eigensinnig und dominant, dies ist nicht wünschenswert, wenn man als Küchengehilfe in einem Team arbeiten sollte.

$$H1D_1: \mu_{UA\ HO} < \mu_{UA\ FB}$$

$$H1D_0: \mu_{UA\ HO} = \mu_{UA\ FB}$$

Bei der *Globalskala Unnachgiebigkeit* ist davon auszugehen, dass die Personen ihr Profil dahingehend ändern, dass sie sich in der Faking-bad-Situation als empfänglich darstellen. Die Probanden sollten bemüht sein, sich als möglichst abgehoben und offen für Veränderungen darzustellen.

$$H1E_1: \mu_{UN\ HO} > \mu_{UN\ FB}$$

$$H1E_0: \mu_{UN\ HO} = \mu_{UN\ FB}$$

3.3.2. Hypothesen zur Faking-good-Bedingung

Die Versuchspersonen hatten in der Faking-good-Bedingung den 16 PF-R so auszufüllen, als ob sie sich für eine Stelle als Geschäftsführerin bewerben (siehe Kap. 3.2.2.). Die Probanden sollten sich demnach in der Faking-good-Bedingung extravertierter, gelassener, selbstkontrollierter, unabhängiger sowie unnachgiebiger darstellen.

Die *unabhängigen Variablen* sind ebenso wie in der Faking-bad-Bedingung demographische Daten wie das Alter (jung versus alt), das Geschlecht (männlich versus weiblich) sowie der Bildungsgrad (niedrige versus hohe Bildung) der Untersuchungsteilnehmer. Persönlichkeitseigenschaften wie Extra- versus Introversion, Emotionale Stabilität versus Emotionale Instabilität, Repression versus Sensitization, niedrige versus hohe moralische Urteilsfähigkeit sowie niedrige versus hohe Tendenz zur positiven Selbstdarstellung stellen die weiteren unabhängigen Variablen dar. Als *abhängige Variablen* fungieren wieder die Stenwerte auf den Globalskalen des 16 PF-R. Hier stellt ebenfalls der honest-Wert die Baseline dar und lässt somit einen Vergleich mit den gefakten Werten zu.

Bei der *Globalskala Extraversion* ist anzunehmen, dass die Personen ihr Profil dahingehend ändern, dass sie sich in der Faking-good-Situation als extravertierter darstellen. Sie sind dazu aufgefordert dem fiktiven Anforderungsprofil zu entsprechen, deshalb sollten sie sich als kommunikativ, kontaktfreudig, emotional warm und durchsetzungsfähig darstellen.

$$H2A_1: \mu_E\ HO < \mu_E\ FG$$

$$H2A_0: \mu_E\ HO = \mu_E\ FG$$

Die Probanden sollten sich bei der *Globalskala Ängstlichkeit* in der Faking-good-Situation als gelassener darstellen. Sie sollten den Eindruck erwecken, dass es sich bei ihnen um einen emotional stabilen, entspannten und selbstsicheren Menschen handelt.

$$H2B1: \quad \mu_{AE\ HO} > \mu_{AE\ FG}$$

$$H2B0: \quad \mu_{AE\ HO} = \mu_{AE\ FG}$$

Bei der *Globalskala Selbstkontrolle* erwartet die Autorin, dass die Probanden ihre Skalenwerte in der Faking-good-Situation erhöhen. Damit würden sie den Eindruck eines regelbewussten, perfektionistischen Menschen erwecken.

$$H2C1: \quad \mu_{SK\ HO} < \mu_{SK\ FG}$$

$$H2C0: \quad \mu_{SK\ HO} = \mu_{SK\ FG}$$

Bei der *Globalskala Unabhängigkeit* geht die Autorin davon aus, dass sich die Probanden in der Faking-good-Situation als unabhängiger darstellen. Unabhängige Menschen gelten als durchsetzungsfähig und willensstark. Für den Posten als Geschäftsführer sind dies wichtige Persönlichkeitseigenschaften.

$$H2D1: \quad \mu_{UA\ HO} < \mu_{UA\ FG}$$

$$H2D0: \quad \mu_{UA\ HO} = \mu_{UA\ FG}$$

Bei der *Globalskala Unnachgiebigkeit* vermutet die Autorin, dass sich die Personen in der Faking-good-Situation als unnachgiebig darstellen. Die Probanden sollten bemüht sein sich als möglichst auf dem Boden der Tatsachen stehend und resolut darzustellen.

$$H2E1: \quad \mu_{UN\ HO} < \mu_{UN\ FG}$$

$$H2E0: \quad \mu_{UN\ HO} = \mu_{UN\ FG}$$

3.3.3. Vermutete differentialpsychologische Unterschiede

Die Autorin hat zusätzlich noch Überlegungen angestellt, welche Personengruppen wohl fähiger im Verfälschen von Persönlichkeitsfragebogen agieren werden.

Es werden wohl *altersspezifische Unterschiede* bezüglich der Fakeability resultieren. Diese Unterschiede werden vermutlich in die Richtung gehen, dass ältere Personen aufgrund Ihrer größeren Lebenserfahrung besser verfälschen können als jüngere Personen.

Ebenso werden *geschlechtsspezifische Unterschiede* bezüglich der Fakeability angenommen. Frauen werden wohl eine höhere Fakeability aufweisen als Männer. Mahar, Cologon und Duck (1995) konnten in

ihrer Untersuchung keine geschlechtsspezifischen Unterschiede in Bezug auf die Fähigkeit zur Verfälschung feststellen. Intuitiv geht die Autorin dennoch davon aus, dass es einen Unterschied zwischen den Geschlechtern gibt. Die Autorin nimmt an, dass Frauen besser im Verfälschen von Persönlichkeitsfragebogen sind, da sie eher dazu tendieren sozial erwünscht zu handeln.

Zudem werden vermutlich auch *bildungsspezifische Unterschiede* bezüglich der Fakeability resultieren. Die Autorin ist der Meinung, dass gebildete Personen besser dazu imstande sein werden Persönlichkeitsfragebogen zu verfälschen. Ones et al. (1996) fanden heraus, dass Personen mit weniger Jahren schulischer Bildung höhere Scores in Skalen zur sozialen Erwünschtheit erzielten als Personen, deren Ausbildung längere Zeit in Anspruch genommen hatte. Die Autorin geht jedoch intuitiv davon aus, dass bei dieser Untersuchung ein gegenteiliges Ergebnis resultieren wird, da Personen mit höherer Bildung sich durch größere kognitive Differenziertheit auszeichnen.

Extravertierte und introvertierte Menschen werden sich wohl ebenfalls unterscheiden bezüglich ihrer Fakeability. Es wird davon ausgegangen, dass extravertierte Personen besser verfälschen können als introvertierte Personen. Diese Annahme wird dadurch begründet, dass extravertierte Personen in höherem Maße sozial erwünscht handeln. Sie sind mehr darum bemüht als Introvertierte bei ihren Mitmenschen einen guten Eindruck zu hinterlassen.

Es wird ein Unterschied in der Fakeability erwartet zwischen *emotional stabilen und emotional instabilen* Personen. Emotional instabile Personen können besser verfälschen als emotional stabile Personen. Diese Annahme wird dadurch begründet, dass emotional instabile Menschen in ihrer Persönlichkeit nicht so gefestigt sind. Sie sind aus diesem Grund mehr bemüht sozial erwünscht zu handeln. Ones et al. (1996) fanden gegenteiliges heraus, nämlich dass Stellenbewerber, welche hohe Werte für soziale Erwünschtheit hatten, dazu tendierten auch hohe Werte für emotionale Stabilität (und Gewissenhaftigkeit) zu haben. Dies könnte allerdings auch daran liegen, dass diese Personen es für sozial erwünschter hielten emotional stabil zu sein und deswegen ihre Antworten in diese Richtung verfälschten.

Sensitizer und Represser werden sich vermutlich ebenfalls unterscheiden bezüglich ihrer Fakeability. Die Autorin nimmt an, dass Represser besser verfälschen können, da sie stärker zum Verhalten im Sinne sozialer Erwünschtheit tendieren (Krohne, 1975; 1996). In einer Untersuchung an Studenten konnten Furnham et al. (2002) belegen, dass Represser signifikant mehr sozial erwünschte Antworten in Persönlichkeitsfragebogen geben.

Personen, die hohe Werte bezüglich ihrer *Tendenz zu positiver Selbstdarstellung* erzielen, werden auch fähiger beim Verfälschen von Persönlichkeitsfragebogen sein. Die Autorin begründet diese Annahme dadurch, dass „Impression Management“ und „faking“ sehr ähnliche Selbstdarstellungsformen bezeichnen.

Personen, welche über eine hohe *moralische Urteilsfähigkeit* verfügen, sollten auch eine höhere Fakeability aufweisen. Institutionalisierte Bildung führt zu einer höheren moralischen Urteilsfähigkeit, wie bereits von Lind (1993) nachgewiesen werden konnte. Hohe moralische Urteilsfähigkeit benötigt eine kognitive Flexibilität und Differenziertheit. Personen mit einer hohen moralischen Urteilsfähigkeit setzen sich auch mit Argumenten auseinander, die entgegengesetzt ihrer eigenen Meinung sind. Die Autorin geht intuitiv davon aus, dass Personen mit einer höheren Bildung die besseren Faker sind, demnach sollten auch Personen mit hoher moralischer Urteilsfähigkeit besser verfälschen können.

3.4. Auswahl und Begründung der Methoden

Um die dargestellten Fragestellungen zu überprüfen wurden ein Sozialdatenblatt sowie vier Persönlichkeitsfragebogen vorgegeben. Im Anschluss an die Testung erfolgte ein Dankeschön an die Probanden sowie deren Aufklärung über Sinn und Zweck der Untersuchung.

3.4.1. Das Testmaterial

Im Folgenden wird das verwendete Testmaterial im Einzelnen dargestellt. Der Übersicht wegen wird es in der den Probanden dargebotenen Reihenfolge beschrieben.

Um eventuelle Versuchsleitereffekte auszuschalten, erfolgte die Durchführung vorwiegend in Anwesenheit der Autorin. Die Befragung wurde unter Wahrung der Anonymität der Probanden durchgeführt. Zudem erfolgte die Teilnahme auf freiwilliger Basis und ohne Bezahlung.

Die Instruktionen der Fragebogen wurden von der Autorin um folgenden Zusatz erweitert: *Es handelt sich hier um eine wissenschaftliche Untersuchung. Um zu aussagekräftigen Ergebnissen zu gelangen, ist es äußerst wichtig, dass Sie die Fragen so ehrlich als möglich beantworten. Die Auswertung des Fragebogens erfolgt anonym. Ihre Daten werden selbstverständlich vertraulich behandelt.*

Die Auswertung des 16 PF-R erfolgte unter Zuhilfenahme des *Testauswerteprogramm 16 PF-R* für Windows. Die übrigen Persönlichkeitsfragebogen wurden anhand des Auswertungsschemas der jeweiligen Testautoren ausgewertet.

3.4.1.1. Sozialdatenerhebung

Durch das Sozialdatenblatt werden das Alter, das Geschlecht sowie die höchste abgeschlossene Bildung der Probanden erhoben. Die Erfassung dieser Faktoren ist für die Untersuchung wesentlich. Außerdem erhält jede Person einen persönlichen Code, um die Anonymität gewährleisten zu können. Der persönliche Code besteht aus sechs Buchstaben, nämlich den letzten beiden Buchstaben des Vornamens des Probanden, dem ersten und dem letzten Buchstaben des Vornamens der Mutter des Probanden sowie dem zweiten und dem dritten Buchstaben des Vornamens des Vaters des Probanden. Durch den persönlichen Code wird trotz der anonymisierten Auswertung ermöglicht, dass jeder Proband sein persönliches Ergebnis erhält.

Zusätzlich werden die Probanden noch zu Ihrer Einstellung gegenüber Persönlichkeitsfragebogen befragt. Hierzu müssen die Probanden angeben, ob sie gerne Fragebogen ausfüllen. Die Autorin möchte an dieser Stelle jedoch kritisch anmerken, dass die Möglichkeit besteht, dass viele Probanden diese Frage sozial erwünscht mit „Ja“ beantworten, selbst wenn dies ihrer tatsächlichen Einstellung widerspricht.

3.4.1.2. 16-Persönlichkeits-Faktoren-Test Revidierte Fassung (Schneewind & Graf, 1998)

Zur Erfassung relevanter Persönlichkeitseigenschaften und der Fakeability wird der *16 PF-R* (16-Persönlichkeits-Faktoren-Test Revidierte Fassung) von Schneewind und Graf (1998) vorgegeben.

Der 16 PF-R ist ein objektiver Test zur mehrdimensionalen Persönlichkeitsdiagnostik im Erwachsenenalter. Er kann bei Personen im Alter von 17 Jahren bis ins hohe Alter eingesetzt werden. Der Test besteht aus 184 Items. Diese Items können 16 Primärdimensionen, einer Skala zur Erfassung der Tendenz zur sozialen Erwünschtheit sowie fünf Globaldimensionen der Persönlichkeit zugeordnet werden. Die fünf Globaldimensionen sind zusammengesetzt aus Extraversion, Ängstlichkeit, Selbstkontrolle, Unabhängigkeit sowie Unnachgiebigkeit.

Der 16 Persönlichkeits-Faktoren-Test von Raymond B. Cattell ist 1949 in seiner ersten Auflage erschienen. Die erste deutschsprachige Version wurde im Jahre 1983 von Schneewind, Schröder und Cattell publiziert. Die Grundlagen des Tests wurden immer wieder verbessert und so entstand schließlich für den deutschsprachigen Raum der 16 PF-R von Schneewind und Graf, der in dieser Untersuchung Verwendung fand.

Hinsichtlich der Primär- als auch der Globalfaktoren ist die Reliabilität zufrieden stellend. Durchführungs-, Auswertungs- und Interpretationsobjektivität sind durch die Art der Vorgabe und Auswertung gegeben. Laut Schneewind und Graf kann der 16 PF-R als valides Testinstrument angesehen werden. Der 16 PF-R ist sowohl in Einzel- wie auch in Gruppentestung durchführbar, somit erfüllt er auch einen ökonomischen Aspekt. Die Bearbeitungszeit des 16 PF-R nimmt in etwa 30 bis 45 Minuten in Anspruch.

Die Autorin hat sich für die Vorgabe des 16 PF-R entschieden wegen der Erfassung der Globaldimension Extraversion sowie der Primärdimension Emotionale Stabilität. Außerdem sprachen für die Verwendung dieses Tests folgende drei Antwortstil-Indizes: die Impression Management Skala (IM), die Akquieszenzskala (AK) und die Infrequenzskala (IF).

Die IM-Skala beinhaltet zehn Items, mit denen die Antworttendenz zur Sozialen Erwünschtheit erfasst wird. Diese zehn Items gehören nicht gleichzeitig einer anderen Skala an. Personen mit besonders hohen Werten auf dieser Skala neigen dazu, sozial erwünscht zu antworten und sich in einem positiven Licht darzustellen. Sehr niedrige sowie sehr hohe IM-Testwerte können die inhaltliche Aussagekraft der anderen Testwerte mindern. Übliche Cut-off-Werte liegen bei 5 % bzw. 95 %. Die Gründe für sehr hohe IM-Werte sind im Einzelfall zu eruieren, gegebenenfalls muss die Testung wiederholt werden.

Unter Akquieszenz wird die Tendenz verstanden, Antworten unabhängig vom spezifischen Iteminhalt zuzustimmen. Der 16 PF-R beinhaltet 100 Richtig-Falsch-Antworten. Diese 100 Items wurden zur Akquieszenzskala zusammengefasst. Alle Richtig-Antworten (Antwortalternative a) werden als eins

gewertet, die anderen Antworten (b, c) werden als null gewertet. Erreicht eine Person hohe AK-Werte, so spricht dies für eine deutlich ausgeprägte Tendenz zur Zustimmung. Bei einem Rohwert über 70 haben nur 5 % oder weniger der Probanden häufiger zugestimmt. Die Interpretation der AK-Werte bleibt dem professionellen Testanwender überlassen. Es empfiehlt sich Rücksprache mit dem Probanden zu halten bzw. andere Testergebnisse hinzuzuziehen.

Die Infrequenzskala setzt sich aus 51 Items zusammen und misst die Tendenz zufällige Antworten zu geben. Für diese Skala wurden jene Items ausgewählt, deren Zustimmungsrate bei einer der drei Antwortalternativen bei weniger als 5 % lag. Bei allen Items war ausschließlich die Antwortalternative „b“ (? , weiß nicht) betroffen. Somit wurde bei diesen 51 Items die Antwortalternative „b“ mit eins gewertet, die anderen beiden mit null. Bei hohen Werten auf der Infrequenzskala liegt die Vermutung nahe, dass ein Proband andere Antwortalternativen als die überwiegende Mehrheit der Personen gewählt hat. Bei einem IF-Wert im 95 %-Perzentil, sollte geklärt werden, warum die Antwortalternative „b“ so häufig gewählt wurde. Als mögliche Gründe neben der Tendenz, zufällige Antworten zu geben, kommt laut Testautoren folgendes in Frage: die Unfähigkeit, sich zu entscheiden; der Versuch, einen falschen Eindruck zu erwecken; Leseschwierigkeiten.

Nachfolgend wird nun ein Überblick über die Primär- und Globalskalen des 16 PF-R gegeben.

16 PF-R-Primärskalen		
A	Wärme	warmherzig, aufmerksam für die Gefühle und Bedürfnisse anderer vs. reserviert, unpersönlich, distanziert
B	Logisches Schlussfolgern	hoch vs. niedrig
C	Emotionale Stabilität	emotional stabil, ausgeglichen vs. stimmungslabil
E	Dominanz	dominant, durchsetzungsfähig, sich selbst behauptend vs. nachgiebig, kooperativ, konfliktvermeidend
F	Lebhaftigkeit	lebhaft, spontan, gesellig vs. ernst, zurückhaltend, bedacht
G	Regelbewusstsein	regelmäßig, pflichtbewusst vs. unangepasst, nonkonformistisch
H	Soziale Kompetenz	sozial kompetent, kontaktstark vs. scheu, schüchtern
I	Empfindsamkeit	empfindsam, gefühlvoll, sentimental vs. sachlich, unsentimental, robust
L	Wachsamkeit	wachsam, misstrauisch, skeptisch vs. vertrauensvoll, arglos
M	Abgehobenheit	abgehoben, träumerisch, phantasievoll, ideenreich vs. lösungsorientiert, praktisch, auf dem Boden der Tatsachen stehend
N	Privatheit	geschlossen, diskret vs. offen, geradeheraus, natürlich
O	Besorgtheit	besorgt, selbstzweifelnd, verletzlich vs. selbstsicher, selbstzufrieden, selbstbejahend
Q1	Offenheit	offen für Veränderung, experimentierfreudig, aufgeschlossen für Neues vs. am Gewohnten haftend, traditionalistisch
Q2	Selbstgenügsamkeit	selbstgenügsam, einzelgängerisch, zurückgezogen vs. sozial orientiert, anschlussfreudig
Q3	Perfektionismus	perfektionistisch, planvoll, selbstdiszipliniert, ordentlich vs. flexibel, wenig Wert auf Ordnung/Perfektion/Disziplin legend
Q4	Anspannung	angespannt, reizbar, nervös, getrieben vs. entspannt, ruhig, gelassen, geduldig

Tab. 14: Überblick über die Primärskalen des 16 PF-R

16 PF-R-Globalskalen	
Extraversion	extravertiert vs. introvertiert
Ängstlichkeit	ängstlich vs. gelassen
Selbstkontrolle	selbstkontrolliert vs. unbeherrscht
Unabhängigkeit	unabhängig vs. anpassungsbereit
Unnachgiebigkeit	unnachgiebig vs. empfänglich

Tab. 15: Die fünf Globalskalen des 16 PF-R

Die einzelnen Globalskalen setzen sich aus mehreren Primärfaktoren zusammen (Schneewind & Graf, 1998):

Globalskala	Primärskalen
SELBSTKONTROLLE	G Regelbewusstsein Q3 Perfektionismus
EXTRAVERSION	A Wärme E Dominanz F Lebhaftigkeit H Soziale Kompetenz N Privatheit Q2 Selbstgenügsamkeit
UNABHÄNGIGKEIT	A Wärme E Dominanz I Empfindsamkeit O Besorgtheit
ÄNGSTLICHKEIT	C Emotionale Stabilität O Besorgtheit Q4 Anspannung
UNNACHGIEBIGKEIT	M Abgehobenheit Q1 Offenheit für Veränderung

Tab. 16: Primär- und Globalfaktorenstruktur des 16 PF-R

Die Normierung der Skalen des 16 PF-R erfolgte nach Geschlecht und Alter aufgegliedert laut Manual (S. 105) anhand den Werten der beiden folgenden Tabellen:

Mittelwerte und Standardabweichungen der Skalen des 16 PF-R (männlich)								
Skala	Gesamt		18 bis 29 Jahre		30 bis 49 Jahre		über 50 Jahre	
	(n = 532)		(n = 112)		(n = 213)		(n = 207)	
	M	s	M	s	M	s	M	s
A	19,6	4,4	20,1	4,6	19,6	4,4	19,3	4,4
B	7,8	3,2	9,6	2,7	7,9	3,2	6,9	3,0
C	25,3	4,8	24,8	4,8	24,9	5,0	26,0	4,6
E	24,1	4,8	24,5	4,6	23,9	4,7	23,9	4,9
F	15,8	4,4	18,5	4,4	15,9	4,3	14,4	3,8
G	22,2	4,9	20,0	4,7	21,4	4,8	24,1	4,3
H	25,9	7,6	25,8	7,3	25,7	8,0	26,3	7,3
I	20,1	4,8	20,1	5,0	19,8	4,8	20,3	4,7
IM	18,8	4,0	17,5	3,6	18,1	3,6	20,2	4,2
L	23,9	4,1	23,4	4,4	23,7	4,2	24,4	3,9
M	18,5	4,4	20,0	4,9	18,3	4,3	17,9	4,1
N	24,4	5,0	23,7	4,9	24,3	5,1	25,0	5,0
O	22,7	5,6	22,4	5,9	22,7	5,9	22,9	5,3
Q1	22,0	5,3	24,7	4,9	22,2	5,2	20,3	4,8
Q2	21,4	5,7	20,1	5,8	21,6	5,7	21,9	5,6
Q3	23,9	5,6	21,6	5,6	23,8	5,6	25,4	5,1
Q4	22,1	5,7	21,9	5,5	22,2	5,7	22,1	5,9

Tab. 17: Normierungswerte männlich für die Skalen des 16 PF-R (vgl. Manual, S. 105)

Mittelwerte und Standardabweichungen der Skalen des 16 PF-R (weiblich)								
Skala	Gesamt		18 bis 29 Jahre		30 bis 49 Jahre		über 50 Jahre	
	(n = 677)		(n = 130)		(n = 278)		(n = 269)	
	M	s	M	s	M	s	M	s
A	22,1	3,6	22,9	3,1	22,2	3,6	21,6	3,8
B	7,4	3,2	8,6	2,9	7,9	3,1	6,3	3,1
C	24,4	5,0	23,7	4,9	24,5	5,2	24,5	5,0
E	22,6	5,0	23,9	4,6	22,5	5,0	22,1	5,0
F	16,9	4,4	19,3	4,4	17,4	4,3	15,3	4,0
G	22,7	4,3	20,7	4,0	22,4	4,3	24,1	4,0
H	25,5	7,5	26,9	7,3	25,0	7,8	25,4	7,4
I	26,7	4,0	27,0	4,0	26,5	4,0	26,8	3,9
IM	19,6	4,4	17,7	3,6	19,1	4,3	21,1	4,3
L	23,5	4,2	23,0	4,6	23,1	4,5	24,1	3,7
M	17,4	4,3	18,1	4,4	17,1	4,6	17,3	3,9
N	23,4	5,3	22,1	5,3	22,8	5,4	24,6	5,0
O	25,7	5,4	26,3	5,2	25,5	5,6	25,6	5,3
Q1	21,6	5,0	23,5	4,7	22,1	5,0	20,1	4,8
Q2	20,5	5,3	19,0	5,4	19,9	5,2	21,8	5,2
Q3	24,2	5,5	22,2	5,3	23,4	5,7	25,9	4,8
Q4	23,8	5,6	24,8	5,4	23,9	5,8	23,3	5,5

Tab. 18: Normierungswerte weiblich für die Skalen des 16 PF-R (vgl. Manual, S. 105)

3.4.1.3. Repression-Sensitization-Coping-Inventar (Huwe, Hennig & Netter, 1996)

Dieses Inventar von Huwe, Hennig und Netter (1996) dient der „Erfassung der dispositionellen Angstbewältigungsdimensionen Repression-Sensitization in der selbstwertrelevanten Situation des öffentlichen Sprechens“ (S. 157). Die Testkonstrukteure bezeichnen das Repression-Sensitization-Coping-Inventar (RSCI) als ein Verfahren „zur Erfassung repressiver vs. sensibler Streßbewältigung in einer Public-Speaking-Situation“ (Huwe et al., 1996, S. 159).

Die Konstruktion dieses Testverfahrens hatte folgendes zum Ziel:

- 1) Die Konstruktion zweier unabhängiger Skalen zur Erfassung von Defensivität bzw. Sensitivität im Umgang mit einer Belastungssituation, wobei
- 2) subjektive Streß(bewältigungs)reaktionen auf verschiedenen Ebenen Berücksichtigung finden sollten: der Ebene der Kognitionen, der Emotionen, der körperlichen Reaktionen und des Verhaltens.
- 3) Es sollte die Konfundierung von R-S und Ängstlichkeit bzw. Neurotizismus, wenn möglich, weiter reduziert werden, die sich in hohen Korrelationen zwischen entsprechenden Indikatoren niederschlägt.
- 4) Hierfür sollte die selbstwertrelevante Situation des „Öffentlichen Sprechens“ herangezogen werden, da diese in vielen Untersuchungen zuverlässig zu gesteigerter peripher-physiologischer Aktivierung, erhöhter Erregtheit und Ängstlichkeit und zu vermehrtem Auftreten von Streßverarbeitungskognitionen führte, und daher ... auch als Modell-Streßsituation bezeichnet wurde. (Huwe et al., 1996, S. 159-160)

In einer Voruntersuchung wurde einer Stichprobe von 48 Studenten das Angstbewältigungsinventar für selbstwertrelevante Situationen (ABI-E-R) zusammen mit dem Stressverarbeitungsfragebogen (SVF 84) vorgelegt. Insgesamt erhielten die Autoren einen Itempool von 52 Items, welche mit den Skalen Vigilanz bzw. kognitive Vermeidung des ABI-E-R sowie mit den Skalen des SVF 84 korrelierten. Zusätzlich wurden jeweils zehn Items aus den Bereichen emotionaler, verhaltensmäßiger und vegetativer Stressbewältigungsreaktionen gebildet. Der Itempool umfasste schließlich 82 Items. Eine Stichprobe von 112 Personen (70 Männer, 42 Frauen) hatten nun den Fragebogen im Abstand von vier Wochen zwei Mal zu bearbeiten, damit auch die Retest-Reliabilität bestimmt werden konnte. Nur jene Items wurden einer Itemanalyse unterzogen, welche mit der Skala VIG bzw. der Skala KOV des ABI-E-R signifikant ($p < 0.05$) korrelierten. Die 26 trennschärfsten Items wurden schließlich zu den Skalen Sensitivität bzw. Defensivität zusammengefasst (Huwe et al., 1996).

Sowohl die Reliabilität als auch die Retest-Reliabilität der Defensivitäts- und Sensitivitätsskala sind zufrieden stellend. Dem Testverfahren konnte ebenso Durchführungs- und Auswertungsobjektivität bescheinigt werden. Die Konstruktvalidität ist ebenfalls gegeben.

Faktorenanalytisch konnte laut Huwe et al. (1996) die Unabhängigkeit der Tendenz zu defensiven und sensitiven Reaktionen nachgewiesen werden, obwohl die Skalensummenwerte miteinander korrelieren ($r = -.26$). Die Items der *Skala Defensivität* spiegeln Nicht-Erregtheit sowohl in kognitiver, emotionaler als auch körperlicher Hinsicht wieder. Der Skala Defensivität liegt somit ein Generalfaktor zugrunde. Anders verhält es sich bei den Items der *Sensitivitäts-Skala*. Diese laden auf zwei Faktoren, dem aktiven Situations-/Reaktions-Faktor und dem passiven Hilflosigkeits-Faktor. Die Sensitivitätsskala kann somit unterteilt werden in die Skalen aktive Bewältigung und passive Bewältigung, wobei nicht eindeutig ladende Items ausgeschlossen werden (Huwe et al., 1996).

Es konnten durch den RSCI zwei weitere Extremgruppen von Stressbewältigungstypen klassifiziert werden, nämlich Niedrig-Ängstliche und Defensiv-Hochängstliche.

Vier-Felder-Schema der vier Typen des RSCI:

	<i>Defensivität gering</i>	<i>Defensivität hoch</i>
<i>Sensitivität gering</i>	Defensiv Hochängstliche	Represser
<i>Sensitivität hoch</i>	Sensitizer	Niedrig-Ängstliche

Tab. 19: Die vier resultierenden Extremgruppen von Stressbewältigungstypen des RSCI

Die Instruktion des Fragebogens besteht aus der Schilderung einer fiktiven Bewerbungssituation:

Stellen Sie sich vor, dass Sie – in einem Job, der Sie sehr interessiert – zu einem Bewerbungsgespräch eingeladen worden sind und in Kürze (d.h. in 10 Minuten) etwa 5 Minuten Zeit haben, ein Bild von sich und Ihren besonderen Fähigkeiten und Qualitäten zu vermitteln. Sie sollen eine Gruppe von Beurteilern überzeugen, dass Sie der/die geeignete Bewerber/in für die ausgeschriebene Position sind.

Für die aktuelle Untersuchung wurde die Instruktion noch erweitert: *Kreuzen Sie bitte die jeweils für Sie am ehesten zutreffende Antwort an, wobei „0 = auf keinen Fall“ und „3 = mit Sicherheit“ bedeutet.*

Danach folgen 26 Items, welche alle eine emotionale, kognitive, physiologische oder behaviorale Bewältigungsreaktion beinhalten. Die Eintrittswahrscheinlichkeit der jeweiligen Reaktionen ist nun von den Probanden auf einer vierstufigen Skala einzuschätzen. Die Items sind zwei Subskalen zugeordnet, der Subskala „Defensivität“ mit zwölf Items und der Subskala „Sensitivität“ mit 14 Items. Außerdem wird die Subskala „Sensitivität“ noch unterteilt in die Skalen „Aktive Sensitivität“ und „Passive Sensitivität“.

Durch die Berechnung der Summenscores ist es nun möglich zwischen Repressern und Sensitizern sowie Niedrigängstlichen und Defensiv-hochängstlichen Personen zu unterscheiden. Eine Normierung wurde nicht durchgeführt, die Autoren geben lediglich Mittelwerte und Standardabweichungen für die beiden

Subskalen an. Die Testautoren vergeben für jede Antwort Werte von 0 bis 3 bei der Klassifikation von Repression versus Sensitization. Die Autorin hat dies für diese Untersuchung umgeändert und Werte von 1 bis 4 je Antwort vergeben (0 = 1, 1 = 2, 2 = 3, 3 = 4). Dadurch hat sich natürlich auch der Median geändert. Die nachfolgend angegebenen Mediane beziehen sich dementsprechend nur auf diese Untersuchung.

Die Skala *Passive Sensitivität* setzt sich zusammen aus den Items 2, 13, 18, 22 (Md = 8). Die Items 1, 8, 11, 14, 25, 26 (Md = 12) laden auf der Skala *Aktive Sensitivität* (Huwe et al., 1996). Die Sensitivitätsskala beinhaltet 14 Items (Md = 28), während sich die Defensivitätsskala aus 12 Items zusammensetzt (Md = 24). *Represser* haben einen Median über 24 auf der Defensivitätsskala und einen Median unter 28 auf der Sensitivitätsskala. Die Werte der *Sensitizer* liegen bei einem Median unter 24 auf der Defensivitätsskala sowie einem Median über 28 auf der Sensitivitätsskala.

Die Autorin hat sich zur Vorgabe dieses Fragebogens entschieden, da dieses Testverfahren nicht nur die Unterscheidung zwischen Repressern und Sensitizern ermöglicht, sondern auch deren Unterscheidung von Niedrig-Ängstlichen und Defensiv-Hochängstlichen. Ein weiterer Grund für die Entscheidung zur Vorgabe dieses Fragebogens lag in dessen relativ kurzer Bearbeitungszeit.

3.4.1.4. Moralisches Urteil Test (Lind, 1985)

Beim *Moralisches Urteil Test (MUT)* handelt es sich um einen *Experimentellen Fragebogen* von Georg Lind. Der MUT orientiert sich an der *Kognitiven Entwicklungstheorie* des moralischen Urteilens und somit an den Stufen der Moralentwicklung von Lawrence Kohlberg (Heidbrink, 1991). Lind selbst beschreibt den MUT so: „Er ist sowohl ein Fragebogen wie auch ein objektives, experimentelles Testverfahren“ (2000, S. 209). Der MUT wurde im Rahmen des Forschungsprojekts *Hochschulsozialisation* entwickelt. Dieses Projekt umfasst im Jahre 1976 begonnene Längsschnittuntersuchungen und hat Voraussetzungen, Verläufe und Folgen der universitären Sozialisation zum Gegenstand. Es wurde der Versuch unternommen, mit „der Erfassung von inhaltlichen und strukturellen Komponenten des moralischen Urteilsverhaltens ..., zentrale Momente des Sozialisationsprozesses sichtbar zu machen“ (Lind, 2000, S. 106).

Der MUT wurde als „Instrument zur Überprüfung von theoretischen Annahmen in der Forschung“ (Lind, 2000, S. 71) entwickelt. Er ist nicht für individual-diagnostische Auswertungen geeignet. Der MUT ist nach Angaben des Testautors der einzige Test, mit dem unabhängig von der moralischen Einstellung des Probanden, ein Kennwert für die moralische Urteilsfähigkeit erfasst werden kann (G. Lind, schriftl. Mitteilung, 01.10.2002). Vor allem dies sprach für die Vorgabe des MUTs in dieser Untersuchung, da hier lediglich die moralische Urteilsfähigkeit von Interesse ist.

Den Probanden werden zwei moralische Dilemmata zur Bearbeitung vorgelegt – die Geschichte „Arbeiter/Diebstahl“ sowie die Geschichte „Arzt/Euthanasie“. Für die Formulierung von Argumenten im Arbeiter-Dilemma wurde teilweise das Buch und gleichnamige Fernsehspiel *Stellenweise Glatteis* aus dem Jahre 1975 von Max von der Grün zu Hilfe genommen. Das Auswertungshandbuch für das *Kohlberg-Interview* diente im Falle des Euthanasie-Dilemmas als Quelle für konkrete Argumentbeispiele. Bevor der MUT empirisch zum Einsatz kam, wurden die Argumente nochmals von Experten kritisch geprüft. Es war nötig einige Argumente daraufhin zu revidieren. Ziel dieses Tests ist es die affektiven und die kognitiven Aspekte des moralischen Urteils von Personen zu erfassen (Lind, 1978; 2000).

Lind definiert die *affektiven Aspekte* folgendermaßen: „Die affektiven Aspekte des moralischen Urteils sind die individuellen moralischen Einstellungen, d.h. die moralischen Prinzipien, an denen der einzelne sein Urteilsverhalten ausrichtet“ (2002a, S. 87). Die Messintention orientiert sich hierbei an den sechs Stufen der moralischen Entwicklung nach Kohlberg.

Die *kognitiven Aspekte* des moralischen Urteils hingegen meinen die moralische Urteilsfähigkeit. Die Anlage und die Auswertung des MUT sind so konzipiert worden, „dass nur diejenige Person einen hohen Wert auf der Skala für moralische Urteilsfähigkeit erhält, die Argumente auch dann konsistent nach ihrer moralischen Qualität beurteilt, wenn diese der eigenen Meinung zur Lösung eines Dilemmas widersprechen“ (Lind, 2002a, S. 90). Der am häufigsten verwendete Index hierfür ist der *C-Wert*, wobei C für das englische Wort *competence* steht (G. Lind, schriftl. Mitteilung, 01.10.2002).

Während es zahlreiche Tests gibt, welche die moralischen Einstellungen messen, scheint mit Ausnahme vom MUT kein Test zu existieren, mit dem man die moralische Urteilsfähigkeit theoretisch valide erfassen kann (Lind & Wakenhut, 1985; Lind, 2002b).

Für unsere Untersuchung haben wir die Standard-Fassung des MUT (Version 2) verwendet. Sie beinhaltet zwei Dilemmas mit je sechs Pro- und sechs Contra-Argumenten. Jedes Argument ist einer der sechs moralischen Stufen zugeordnet, der MUT beinhaltet demnach für jede Stufe ein Pro- und ein Contra-Argument. Die Probanden müssen moralische Begründungen für das Handeln der Personen in der Geschichte auf einer neunstufigen Skala beurteilen: -4 bedeutet *völlig unakzeptabel*, 4 bedeutet *völlig akzeptabel* (Heidbrink, 1991; Lind, 2000; Lind, 2002a).

Die Instruktion des MUT lautet folgendermaßen: *Auf den nächsten Seiten stehen zwei Geschichten, die jeweils ein Problem beinhalten. Wie stehen Sie zur Lösung dieser Probleme? Wie stehen Sie zu verschiedenen Argumenten, die für und gegen die jeweilige Lösung vorgebracht wurden? Geben Sie bitte eine Antwort, indem Sie das entsprechende Kästchen mit einer Zahl ankreuzen.*

Die Methode des *Experimentellen Fragebogens* wurde der Vorgehensweise beim Experiment entlehnt, in welchem mehrere unabhängige Variablen so in einem Forschungsdesign kombiniert werden

(orthogonales Design), dass man ihren Einfluss auf eine bestimmte abhängige Variable unabhängig voneinander schätzen kann (Lind, 1993).

Durch die experimentelle Anlage des MUT können drei mögliche Einflussgrößen für das moralische Urteilsverhalten systematisch variiert werden: a) *Dilemma-Kontext*, b) *Meinungskonformität*, c) *Stufen der moralischen Orientierung* (Lind, 2000, S. 75ff).

Diese drei Faktoren bilden im Design des MUT die *unabhängigen Variablen* (Lind, 2000; Lind, 2002a):

- Moralische Qualität der vorgegebenen Argumente (*Faktor Stufe*)
- Übereinstimmung oder Nicht-Übereinstimmung der eigenen Meinung mit den vorgegebenen Argumenten (*Faktor Pro-Contra* oder *Meinungskonformität*)
- Situativer Kontext der Argumentation (*Faktor Dilemma*)

Die *abhängige Variable* im MUT ist die „subjektive Beurteilung der Akzeptabilität der vorgegebenen Argumente auf einer Bewertungs-Skala“ (Lind, 2002a, S. 93 ff) durch die Probanden. In der hier verwendeten Standardform des MUT ist diese Skala neun-stufig.

Der MUT ist als ein „dreifaktorielles, vollständig gekreuztes 2 x 2 x 6 Fragebogenexperiment angelegt ... Dabei sind der Faktor *Dilemma* und der Faktor *Meinungskonformität* (oder Pro-Con) als 2-stufige Variable und der Faktor *Stufe* als 6-stufige Variable realisiert“ (Lind, 2000, S. 81).

Untersuchungsergebnisse haben gezeigt, dass Personen einen Diskurs auf der Stufe 6 beim Arztdilemma bevorzugen und ein Argumentieren auf der Stufe 5 beim Arbeiterdilemma (Lind, 2002b).

„Die Determination des Urteilsverhaltens durch die Qualität der Argumente und durch ihre Meinungskonformität sind sicherlich die wichtigsten Indikatoren für die kognitiv-strukturelle Komponente des moralischen Urteilens“ (Lind, 2000, S. 132).

Die Autorin hat als Auswertungsmethode, die von G. Lind (schriftl. Mitteilung, 01.10.2002) vorgeschlagene *intra-individuelle Varianz-Komponenten-Zerlegung* (VKZ) gewählt. Hier wird kein einzelner Summenscore ermittelt, es gilt die Bedeutsamkeit der einzelnen Faktoren zu quantifizieren. Für jede Person werden die auf die einzelnen Faktoren entfallenen Varianzanteile berechnet. Das Antwortmuster jeder einzelnen Person wird analysiert. Es würden sich auch noch andere Auswertungsmethoden anbieten (Heidbrink, 1991; Lind, 2000).

Je konsistenter ein Proband Argumente nach ihrer moralischen Qualität (statt nach ihrer Meinungskonformität) beurteilt, desto höher ist sein C-Wert. Die Bevorzugung einer bestimmten Moral (im Sinne der sechs Stufen Kohlbergs) spielt für den C-Wert keine Rolle. Der C-Wert ist ein

reines Maß der Urteilsfähigkeit, nicht der moralischen Einstellungen. (G. Lind, schriftl. Mitteilung, 01.10.2002)

Beurteilt eine Person die Pro- und Contra-Argumente danach, ob diese mit der eigenen Meinung übereinstimmen, so erhält diese Person niedrige Testwerte für moralische Urteilsfähigkeit im MUT (Lind, 2002c). Der ermittelte C-Wert kann Werte von 0 bis 100 erreichen. Ein C-Wert von 0 bedeutet, dass sich die Person beim Beurteilen von Argumenten überhaupt nicht an moralischen Prinzipien orientiert. Erreicht eine Person den maximalen C-Wert von 100, so orientiert sie sich ausschließlich an moralischen Gesichtspunkten (Lind, 1993).

Höhe der moralischen Urteilskompetenz	C-Score
Niedrige moralische Urteilskompetenz	0-9
Mittlere moralische Urteilskompetenz	10-29
Hohe moralische Urteilskompetenz	30-49
Sehr hohe moralische Urteilskompetenz	50-100

Tab. 20: Maße für den kognitiv-strukturellen Aspekt des MUT (C-Werte)

Da es sich beim MUT um einen Experimentellen Fragebogen handelt, gelten für ihn andere (strengere) Kriterien der Testvalidität und Zuverlässigkeit (Lind, 2000). Durch den MUT konnten Kohlbergs Stufenkonzept sowie seine Ansichten über die kognitiven Aspekte bestätigt werden. Das und die Möglichkeit der simultanen Erfassung der Aspekte des moralischen Urteils erlauben es, dass man von einer validen Messung des Konstrukts spricht (Lind, 1978). Er wurde mittlerweile bei mehr als 200.000 Probanden unterschiedlicher Kulturen, unterschiedlichen Alters, Geschlechts, sozioökonomischen Status und Bildungsniveau eingesetzt. Der MUT hat sich in zahlreichen Quer- und Längsschnittstudien sowie experimentellen Untersuchungen bewährt (Lind, 2000; Lind, 2002c).

Der MUT ist *theorie-kompatibel* (valide), *objektiv* (replizierbar) sowie *leistungsabfordernd* (Lind, 1993). Die Durchführungsdauer des MUT beträgt ungefähr zehn bis zwanzig Minuten. Hinsichtlich seiner Anwendung und Auswertung ist der MUT sehr ökonomisch, zumal auch Gruppentestungen möglich sind (Lind, 1978; Lind, 2002b).

3.4.1.5. Impression Management Skala (Mummendey & Eifler, 1994)

Hierbei handelt es sich um einen Fragebogen zur Erfassung der „positiven“ Selbstpräsentation von Mummendey und Eifler (1994). Die Impression Management Skala (IMS) dient zur Prüfung der Tendenz die Fragen in einer Art und Weise zu beantworten, dass man in einem möglichst guten Licht erscheint. Mit der *IMS* wird die Tendenz eines Individuums erfasst, andere Menschen zu bestimmten Zuschreibungen von Merkmalen an diesem Individuum zu veranlassen (Mummendey & Eifler, 1994).

Nach Mummendey und Eifler (1994, S. 3) soll die antwortende Person „darüber Auskunft geben, wie sehr sie dazu neigt, sich in sozialen Situationen in einer bestimmten Weise zu präsentieren.“

Für die Itemsammlung wurden fünfzehn Studenten gebeten, sich Situationen des Alltags vorzustellen, in denen sie Impression Management betreiben. Für diese Sammlung von Verhaltensweisen wurden ihnen als Leitlinie folgende Bereiche vorgegeben: Freizeitaktivitäten, Familienkontakte, Kontakte mit dem anderen Geschlecht, Studium, Beruf und Öffentlichkeit. Die so gesammelten Verhaltensbeschreibungen wurden in der Ich-Form als Feststellungen formuliert. Jedem Statement wurde ein Statement gegenüber gestellt, welches den Verzicht auf Impression Management ausdrückte. Die Testautoren erhielten einen Pool von mehr als 200 Verhaltensbeschreibungen. Daraus wurden 30 Itempaare, also insgesamt 60 Items gebildet (Mummendey & Eifler, 1994).

Diese 60 Items wurden einer Stichprobe von 146 Personen vorgelegt. Von den 146 Probanden waren 76 männlich und 69 weiblich. Das Durchschnittsalter war 24 Jahre. 87,5 % der Stichprobe waren Studenten. Es wurde eine Schwierigkeitsanalyse sowie eine Trennschärfeanalyse vorgenommen. Von den 60 Items verblieben 17 Items, die den endgültigen Fragebogen bilden (Mummendey & Eifler, 1994).

Die Itemanalyse ergab einen Konsistenzkoeffizienten (Cronbach) von 0,82 sowie eine Split-Half-Reliabilität von 0,83. Bei den vorwiegend studentischen Probanden ist der Fragebogen demnach gut geeignet das Ausmaß positiver Selbstdarstellung zu erfassen. Zudem ist der Fragebogen objektiv in der Durchführung und Auswertung (Mummendey & Eifler, 1994).

Die Testautoren üben Kritik an der IMS: „Selbstverständlich liegt mit dieser Skala nur ein erster Ansatz vor, immerhin aber einer, der unmittelbar *Verhaltenstendenzen*, und nicht nur selbstbezogene Kognitionen und Emotionen behandelt“ (Mummendey & Eifler, 1994, S. 9).

Außerdem sehen Mummendey und Eifler eine wesentliche Einschränkung des Geltungsbereichs der IMS darin, dass ihr

Impression-Management-Konstrukt „‘Positive’ Selbstdarstellung“ zum einen *die Bejahung der Zuschreibung „positiver“ Merkmale*, zum anderen *die Bejahung des Verzichts auf die Zuschreibung „positiver“ Merkmale* thematisiert: Die Bejahung von derart „negativ“ formulierten Selbstdarstellungs-Aussagen im vorliegenden Fragebogen (...) könnte mit Social-Desirability-Tendenzen vermischt sein. Im Gegensatz dazu dürfte die Bejahung „positiv“ formulierter Selbstdarstellungs-Aussagen (...) sozial eher ungünstig wirken. Ein Understatement scheint allgemein eher sozial akzeptiert zu sein als ein Overstatement. (1994, S. 9).

Zur Vermeidung von mittleren Antworttendenzen werden die 17 Items der IMS auf einer vierstufigen Skala vorgegeben. Die Instruktion wurde von der Autorin umgeändert.

Die Instruktion in der vorliegenden Untersuchung lautet: *Unter jedem Satz sind zwischen den Antworten „stimmt nicht“ und „stimmt“ vier Kreise angegeben. Kreuzen Sie bitte immer nur denjenigen von ihnen an, der am ehesten auf Sie zutrifft. Lassen Sie keine Frage aus.*

<i>Rohwert</i>	<i>Tendenz zu positiver Selbstdarstellung</i>
–34 bis –18	wenig Tendenz
–17 bis 17	normale Tendenz
18 bis 34	hohe Tendenz

Tab. 21: Rohwerte der IMS

An dieser Stelle sei angemerkt, dass keine Normierung vorgenommen wurde. Die Autoren machen zudem keine Angaben über Mittelwerte und Standardabweichungen.

Die Autorin hat sich wegen seiner kurzen Bearbeitungszeit für die Vorgabe dieses Fragebogens entschieden. Da die IMS bei Studenten das Ausmaß der positiven Selbstpräsentation gut erfasst und sich die vorliegende Stichprobe vorwiegend aus jüngeren Probanden bzw. Schülern zusammen setzt, schien der Autorin dieses Verfahren geeignet.

3.5. Versuchsanordnung und Durchführung

3.5.1. Versuchsanordnung

Ein Teil der Stichprobe setzte sich vorwiegend aus Bekannten der Autorin zusammen. Der andere Teil der Stichprobe waren Schüler, deren Lehrer zwei Schulstunden zur Verfügung stellten. Jeder Proband der Bekanntenstichprobe muss den 16 PF-R dreimal ausfüllen, jeder Proband der Schülerstichprobe zumindest zweimal, unter jeweils verschiedenen schriftlichen Instruktionsbedingungen. Die drei Instruktionsbedingungen befinden sich auf dem Deckblatt des Fragebogens und werden unterteilt in *faking-good*-, *honest*- und *faking-bad-Bedingung*. Die Reihenfolge der drei Instruktionsbedingungen wird nach dem Zufallsverfahren bestimmt, wodurch sich ein ausbalanciertes Design ergibt. Ein ausbalanciertes Design ist erforderlich, damit Transfereffekte weitgehend ausgeschaltet werden können (Mersman & Shultz, 1998).

Folgende Reihenfolgen (vereinfacht Versuchsgruppen genannt) ergeben sich daraus:

- Versuchsgruppe 1: honest – faking good – faking bad
- Versuchsgruppe 2: honest – faking bad – faking good
- Versuchsgruppe 3: faking good – honest – faking bad
- Versuchsgruppe 4: faking good – faking bad – honest
- Versuchsgruppe 5: faking bad – honest – faking good
- Versuchsgruppe 6: faking bad – faking good – honest
- Versuchsgruppe 7: honest – faking good
- Versuchsgruppe 8: honest – faking bad
- Versuchsgruppe 9: faking good – honest
- Versuchsgruppe 10: faking bad – honest

3.5.2. Durchführung

Die angestrebte Teilnehmeranzahl der Untersuchung lag bei 150 Personen im Alter von siebzehn bis 60 Jahren. Bei den Untersuchungsteilnehmern handelte es sich zum Teil um Personen aus dem Familien-, Bekannten- und Freundeskreis der Autorin. Hierbei ist der so genannte „snowballeffect“ zum Tragen gekommen, das heißt dass diese Personen wiederum Personen kannten, die sich für die Untersuchung als Teilnehmer zur Verfügung stellten. Bei dieser Stichprobe fungierte die Autorin als Versuchsleiterin. Die Untersuchung wurde vorwiegend bei den Bekannten zu Hause durchgeführt, hiermit war dafür gesorgt, dass sich die Probanden wohl fühlen.

Der andere Teil der Stichprobe setzt sich aus Schülern der Maturaklassen der Handelsakademie (HAK) in Ybbs sowie des Stiftsgymnasiums in Melk zusammen. Frau Mag. Alexandra Stummer und die Autorin nahmen hierbei gemeinsam die Funktion der Versuchsleiterinnen ein. Schließt ein Schüler bereits vor

Ende der im Durchschnitt zwei benötigten Unterrichtseinheiten die Untersuchung ab, kann er sich in der restlichen Zeit selbständig still beschäftigen.

3.5.3. Untersuchungsablauf

Aus Gründen der Zeitersparnis wurden zumeist Gruppentestungen durchgeführt. Diese Gruppen setzten sich aus zwei bis achtzehn Personen zusammen. Die Autorin nahm dabei meistens selbst die Funktion der Testleiterin ein. Sie war während der gesamten Untersuchung anwesend. Bei Fragen von Seiten der Untersuchungsteilnehmer konnte sie dadurch auch Hilfestellung leisten, soweit es zulässig ist.

Vorerst wurde den Probanden nur mitgeteilt, dass es sich um eine Untersuchung im Rahmen einer Diplomarbeit und somit um eine wissenschaftliche Untersuchung handelt. Die Teilnehmer wurden darauf hingewiesen, dass es enorm wichtig ist, dass Sie die Anweisungen auf den Deckblättern der jeweiligen Fragebogen genau durchlesen und befolgen.

Die Vorgabenreihenfolge bei der Bekanntenstichprobe sah folgendermaßen aus:

Sozialdatenblatt – 16 PF-R – RSCI– 16 PF-R — MUT – 16 PF-R – IMS

Die Vorgabenreihenfolge bei der Schülerstichprobe hingegen gestaltete sich so:

Sozialdatenblatt – RSCI – 16 PF-R – MUT – 16 PF-R – IMS

Sämtliche Fragebogen und Tests wurden in Papier-Bleistift-Form vorgegeben. Die ausgefüllten Fragebogen jeder Versuchsperson wurden mit einer Büroklammer zusammengesteckt. Dadurch wurde gewährleistet, dass beim Auswerten der Daten keine Verwechslungen entstehen. Erst im Nachhinein wurden die Probanden über die wahre Absicht der Untersuchung informiert.

Bei den Probanden wird wohl im Laufe der Untersuchung, aufgrund der langen Vorgabedauer der Fragebogen, die Motivation zur Mitarbeit nachlassen. Dem wahrscheinlich auftretenden Motivationsverlust von Seiten der Probanden kann man entgegenwirken, indem man den Teilnehmern mitteilt, die Untersuchung wäre äußerst wichtig. Ein zusätzlicher Anreiz für die Probanden bestand darin, dass sie Ihre Ergebnisse, sofern dies gewünscht wurde, schriftlich in Form eines Berichtes erhielten.

Außerdem war die Versuchsleiterin darum bemüht dafür Sorge zu tragen, dass die Untersuchung in einer gemütlichen Atmosphäre stattfand.

4. Auswertung und Interpretation der Ergebnisse

4.1. Statistische Deskription der Strichprobe

Insgesamt nahmen 145 Personen im Alter von 17 bis 56 Jahren an dieser in den Jahren 2003 und 2004 durchgeführten Untersuchung teil. Einige Datensätze mussten ausgeschieden werden, da im Falle der Schülerstichprobe die Zeit zu knapp war. Es gingen 121 vollständige Datensätze in die Untersuchung ein. 51 Personen davon waren männlich und 70 weiblich.

4.1.1. Filtersetzung

Es wurde ein entsprechender Filter über die Daten gelegt, um jene Probanden auszuschneiden, deren Datensätze hinsichtlich der Skalen Impression Management (IM), Infrequenz (IN) und Akquieszenz (AK) im 16 PF-R auffällig sind.

Für die 16 PF-R Skalen zur Akquieszenz, zur Infrequenz und zum Impression Management wurde also ein Filter gesetzt. Datensätze mit auffälligen Werten auf einer dieser drei Skalen gingen nicht in die Untersuchung ein. Bei der honest-Bedingung sind folgende Werte kritisch: IM ab einem Rohwert von 27, AK ab einem Rohwert über 70 (PR ab 96), IF ab einem Rohwert von 9 (PR ab 95). Das Testauswerteprogramm berechnet jedoch keine Rohwerte für diese Skalen, deshalb wurden bei der Skala IM jene Probanden ausgeschieden, die einen Sten-Wert von 10 erhielten. Hierfür wurden zur zusätzlichen Absicherung die Sten-Werte mit den Rohwerten anhand des Test-Manuals verglichen.

4.1.2. Gesamtstichprobe nach Filter

Insgesamt liegen für die Auswertung der Fragestellung 112 vollständige Datensätze vor. Da sich jedoch nur 58 Personen darunter befinden, welche den 16 PF-R unter allen drei Instruktionsbedingungen bearbeitet hatten, erfolgte eine Trennung der Stichprobe in die beiden Faking-Bedingungen. Die Stichprobe wurde somit unterteilt in eine Faking-bad-Stichprobe sowie eine Faking-good-Stichprobe. Die oben genannten 58 Probanden sind in beiden Stichproben anzutreffen. Nachfolgende Berechnungen erfolgten für beide Stichproben getrennt.

4.2. Stichprobe der Faking-bad-Bedingung

Die Stichprobe der Faking-bad-Bedingung umfasst insgesamt 96 Personen. Nach der Filtersetzung setzt sich die Stichprobe aus 89 Personen zusammen. Die 89 Probanden teilen sich wie folgt auf die verschiedenen Reihenfolgen der Vorgabe (vereinfacht Versuchsgruppen genannt) auf:

Versuchsgruppe				
		Häufigkeit	Prozent	Gültige Prozente
Gültig	1	17	19,1	19,1
	2	14	15,7	15,7
	3	7	7,9	7,9
	4	9	10,1	10,1
	5	4	4,5	4,5
	6	7	7,9	7,9
	8	16	18,0	18,0
	10	15	16,9	16,9
	Gesamt	89	100,0	100,0

Tab. 22: Aufteilung der Probanden auf die Versuchsgruppen

Das ergibt folgende prozentuelle Aufteilung:

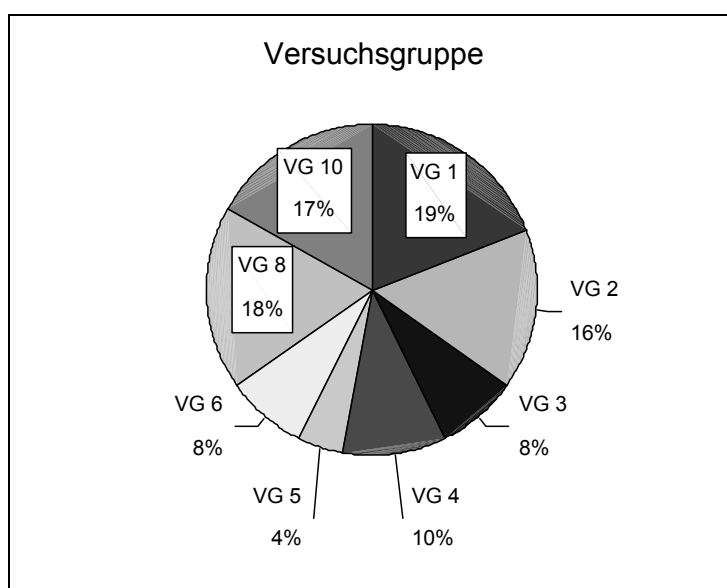


Abb. 4: Prozentuelle Aufteilung der Probanden auf die Versuchsgruppen

4.2.1. Verteilung von Geschlecht, Alter und Bildung

Die Stichprobe setzt sich zusammen aus 41 Männern und 48 Frauen. Die männlichen Probanden machen somit 46 % der Stichprobe aus, während 54 % der Stichprobe weiblich sind. Man kann hier von einem ausgewogenen Geschlechterverhältnis sprechen.

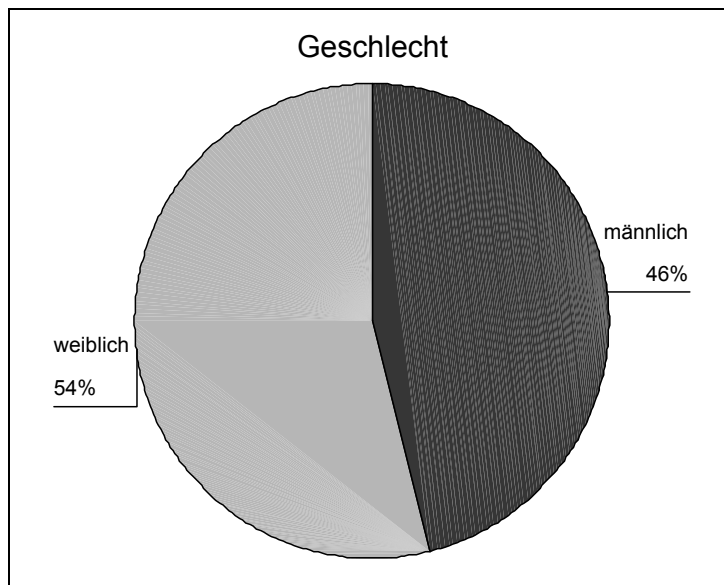


Abb. 5: Geschlechterverteilung in der Faking-bad-Bedingung

Aufgrund der hohen Anzahl an Schülern ist die Stichprobe bezüglich des Alters schief verteilt. Die Gruppe der 17 bis 19 Jahre alten Probanden überwiegt deutlich. In Bezug auf den jeweiligen Lebensabschnitt wurden Altersklassen gebildet.

Altersklassen					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	17 bis 19	38	42,7	42,7	42,7
	20 bis 29	22	24,7	24,7	67,4
	30 bis 39	13	14,6	14,6	82,0
	40 bis 56	16	18,0	18,0	100,0
	Gesamt	89	100,0	100,0	

Tab. 23: Altersklassen in der Faking-bad-Stichprobe

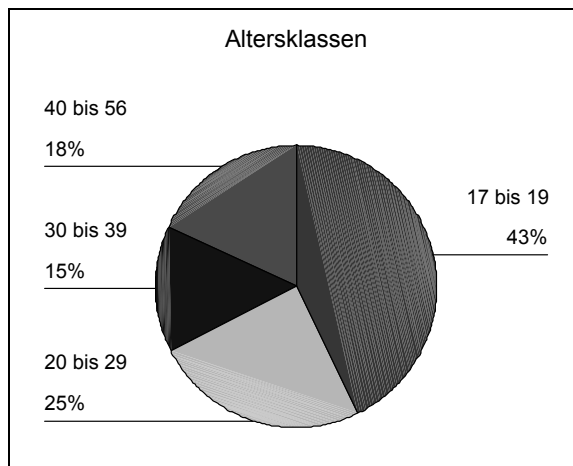


Abb. 6: Prozentuelle Verteilung der Altersklassen in der Faking-bad-Stichprobe

Zur Erhebung der höchsten abgeschlossenen Bildung der Versuchspersonen wurde folgende Rangreihung festgelegt (Kammer für Arbeiter und Angestellte Niederösterreich, mündliche Mitteilung, November 2002):

- *Volksschule*
- *Hauptschule, AHS-Unterstufe, Polytechnische Schule*
- *Berufslehre*
- *Berufsbildende mittlere Schule*
- *Krankenpflegeschule*
- *Meister*
- *Werkmeister*
- *Reifeprüfung: Allgemeinbildende höhere Schule*
- *Reife- und Diplomprüfung: Berufsbildende höhere Schule, Kolleg*
- *Akademie*
- *Fachhochschule*
- *Universität*

Aufgrund der vielen Schüler, die an dieser Untersuchung teilnahmen, ergibt sich hier folgendes Bild:

Ausbildung					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Hauptschule, AHS-Unterstufe, Polytechnische Schule	42	47,2	47,2	47,2
	Berufslehre	18	20,2	20,2	67,4
	Berufsbildende mittlere Schule	6	6,7	6,7	74,2
	Krankenpflegeschule	4	4,5	4,5	78,7
	Meister, Werkmeister	2	2,2	2,2	80,9
	Reifeprüfung: AHS	5	5,6	5,6	86,5
	Reife- u. Diplomprüfung: BHS, Kolleg	8	9,0	9,0	95,5
	Akademie	1	1,1	1,1	96,6
	Universität	3	3,4	3,4	100,0
	Gesamt	89	100,0	100,0	

Tab. 24: Bildungsverteilung in der Faking-bad-Stichprobe

Zur weiteren Berechnung wird die Stichprobe in drei Bildungsklassen unterteilt. Die Bildungsklasse „Pflichtschule“ setzt sich zusammen aus insgesamt 42 Personen, welche die Hauptschule, die AHS-Unterstufe bzw. die Polytechnische Schule absolviert hatten. In der Bildungsklasse „Berufsausbildung“ wurden jene 30 Probanden zusammengefasst, welche zum Zeitpunkt der Untersuchung eine Berufslehre, eine Berufsbildende mittlere Schule, die Krankenpflegeschule sowie die Prüfung zum Meister bzw. Werkmeister abgeschlossen hatten. Die Bildungsklasse „Reifeprüfung“ schließlich umfasst jene 17 Personen, deren höchste abgeschlossene Bildung die Reifeprüfung bzw. der Abschluss an einer Akademie oder Universität ist.

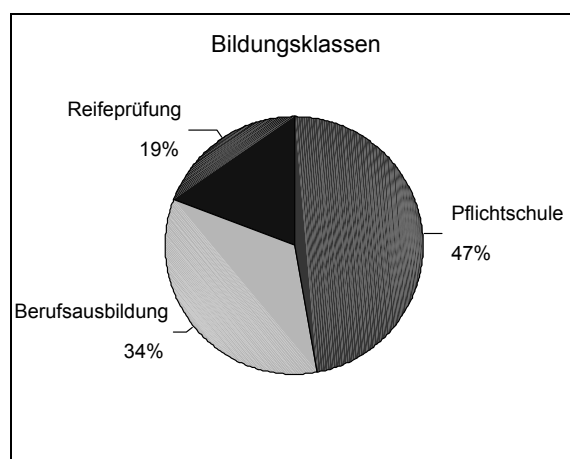


Abb. 7: Prozentueller Anteil der Bildungsklassen in der Faking-bad-Bedingung

4.2.2. Ergebnisse der Persönlichkeitsfragebogen

4.2.2.1. Ergebnisse des 16 PF-R

Die Eingabe der Daten erfolgte im *Testauswerteprogramm 16 PF-R* (Windowsversion). Anhand der geschlechts- und altersspezifischen Normwerte erfolgte die Normierung der Daten. Für die vorliegende Untersuchung sind die Globalskala Extraversion sowie die Primärskala Emotionale Stabilität von Belang.

Zunächst war die Verteilung der Intro- versus Extravertierten in der Stichprobe von Interesse.

Extra-/Introversion					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	introvertiert	53	59,6	59,6	59,6
	extravertiert	36	40,4	40,4	100,0
	Gesamt	89	100,0	100,0	

Tab. 25: Anteil der Intro- und Extravertierten in der Faking-bad-Stichprobe

Laut 16 PF-R sind 60 % der Untersuchungsteilnehmer introvertiert und 40 % extravertiert. Hier kann man von einer ausgewogenen Verteilung sprechen.

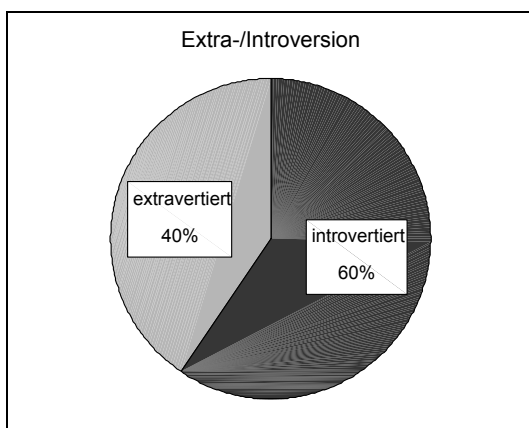


Abb. 8: Prozentueller Anteil der Intro- und Extravertierten in der Faking-bad-Bedingung

Für den Anteil der emotional stabilen und instabilen Personen ergibt sich folgende Verteilung:

Emotionale Stabilität vs. Emotionale Instabilität					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	emotional instabil	40	44,9	44,9	44,9
	emotional stabil	49	55,1	55,1	100,0
Gesamt		89	100,0	100,0	

Tab. 26: Anteil der emotional stabilen und emotional instabilen Probanden in der Faking-bad-Stichprobe

Demnach sind 45 % der Stichprobe als emotional instabil und 55 % emotional stabil zu bezeichnen. Emotional Stabile und Emotional Instabile sind in der Stichprobe annähernd gleich verteilt.

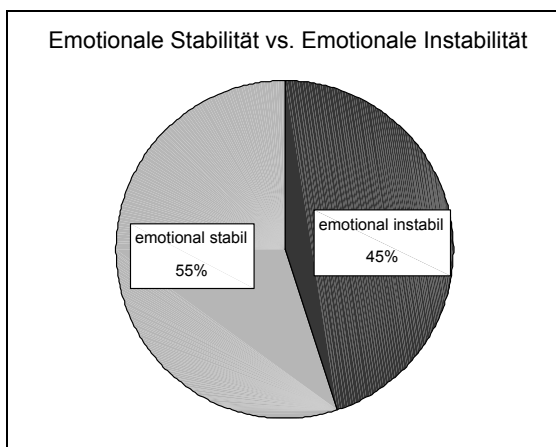


Abb. 9: Prozentueller Anteil der emotional stabilen vs. emotional instabilen Personen in der Faking-bad-Bedingung

4.2.2.2. Ergebnisse des RSCI

Die Ergebnisse des RSCI zeigen folgendes Bild:

Repression vs. Sensitization					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Defensiv-Hochängstliche	1	1,1	1,1	1,1
	Represser	10	11,2	11,2	12,4
	Sensitizer	14	15,7	15,7	28,1
	Niedrig-Ängstliche	64	71,9	71,9	100,0
	Gesamt	89	100,0	100,0	

Tab. 27: Anteil von Repressern und Sensitizern in der Faking-bad-Stichprobe

Somit besteht die Stichprobe aus 11 % Repressern und 16 % Sensitizern. Der Anteil der Niedrig-Ängstlichen und Defensiv-Hochängstlichen ist für die vorliegende Arbeit nicht relevant, deshalb wird

auch nicht näher darauf eingegangen. Für die Berechnungen bezüglich Repression versus Sensitization sind lediglich jene 24 Personen von Interesse, welche man als Represser (10 Personen) bzw. als Sensitizer (14 Personen), ihren Testergebnissen entsprechend, bezeichnen kann.

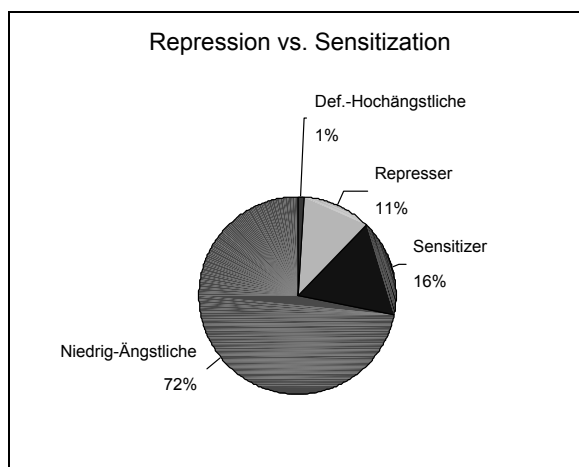


Abb. 10: Prozentueller Anteil von Repressern und Sensitizern in der Faking-bad-Bedingung

4.2.2.3. Ergebnisse der IMS

In der Gesamtstichprobe weist laut IMS nur eine einzige Person eine hohe Tendenz zur positiven Selbstdarstellung auf. Der erreichte Wert liegt jedoch knapp beim Wertebereich der mittleren Tendenz zur positiven Selbstdarstellung. Die Autorin entschied sich deshalb zur Bildung von zwei Gruppen: Geringe Tendenz (-34 bis -1 Punkte) sowie mittlere Tendenz zur positiven Selbstdarstellung (0 bis 34 Punkte).

Positive Selbstdarstellung					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gering	52	58,4	58,4	58,4
	mittel	37	41,6	41,6	100,0
	Gesamt	89	100,0	100,0	

Tab. 28: Tendenz zur positiven Selbstdarstellung in der Faking-bad-Bedingung

Das ergibt folgende prozentuelle Aufteilung:

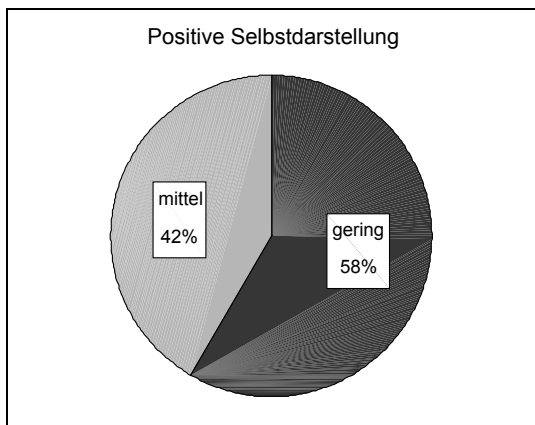


Abb. 11: Prozentueller Anteil der Probanden mit geringer bzw. mittlerer Tendenz zur positiven Selbstdarstellung

4.2.2.4. Ergebnisse des MUT

Die Ergebnisse des MUT zeigen folgendes Bild bezüglich der Höhe der moralischen Urteilsfähigkeit:

Moralisches Urteil					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	niedrig	18	20,2	20,2	20,2
	mittel	46	51,7	51,7	71,9
	hoch	16	18,0	18,0	89,9
	sehr hoch	9	10,1	10,1	100,0
Gesamt		89	100,0	100,0	

Tab. 29: Verteilung der moralischen Urteilsfähigkeit in der Faking-bad-Stichprobe

Die Gruppe der Personen mit mittlerer moralischer Urteilsfähigkeit ($n = 46$) überwiegt deutlich, sie macht rund 52 Prozent der Stichprobe aus.

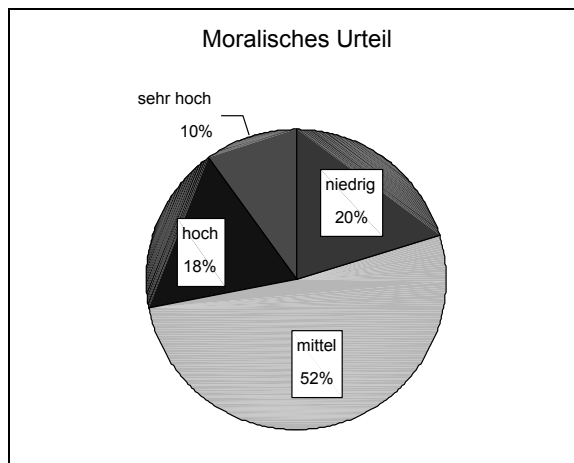


Abb. 12: Prozentuelle Verteilung der Gruppen zur moralischen Urteilsfähigkeit

4.3. Stichprobe der Faking-good-Bedingung

Die Stichprobe der Faking-good-Bedingung setzt sich aus insgesamt 88 Personen zusammen. Nach der Filtersetzung verbleiben 81 Personen. Diese 81 Probanden teilen sich wie folgt auf die verschiedenen Reihenfolgen der Vorgabe (vereinfacht Versuchsgruppen genannt) auf:

Versuchsgruppe				
		Häufigkeit	Prozent	Gültige Prozente
Gültig	1	17	21,0	21,0
	2	14	17,3	17,3
	3	7	8,6	8,6
	4	9	11,1	11,1
	5	4	4,9	4,9
	6	7	8,6	8,6
	7	15	18,5	18,5
	9	8	9,9	9,9
	Gesamt	81	100,0	100,0

Tab. 30: Aufteilung der Probanden auf die Versuchsgruppen

Die Aufteilung der Probanden sieht prozentuell folgendermaßen aus:

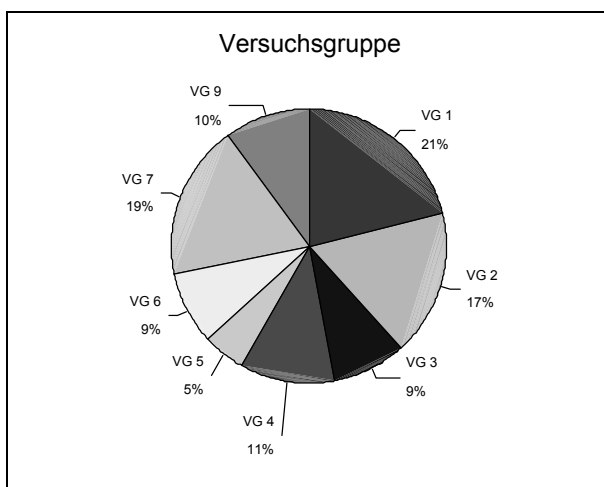


Abb. 13: Prozentuelle Aufteilung der Probanden auf die Versuchsgruppen

4.3.1. Verteilung von Geschlecht, Alter und Bildung

Die Stichprobe besteht aus 37 Männern und 44 Frauen. Die männlichen Probanden machen demnach 46 % der Stichprobe aus, während 54 % der Stichprobe weiblich sind. Der prozentuelle Anteil von männlichen und weiblichen Probanden entspricht jenem der Faking-bad-Bedingung und kann somit als ausgewogen bezeichnet werden.

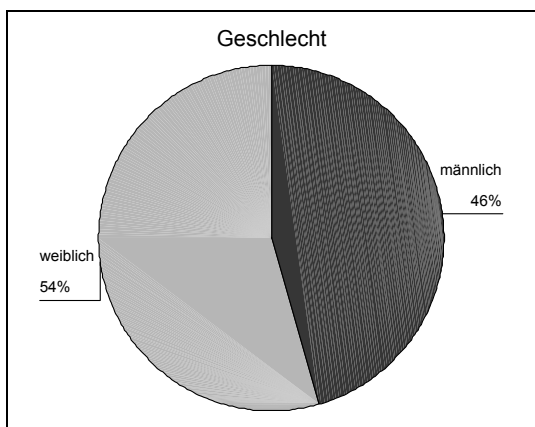


Abb. 14: Geschlechterverteilung in der Faking-good-Stichprobe

Bezüglich des Alters ist auch hier aufgrund der hohen Schüleranzahl die Stichprobe schief verteilt. Es wurden wieder in Bezug auf den jeweiligen Lebensabschnitt Altersklassen gebildet.

Altersklassen					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	17 bis 19	31	38,3	38,3	38,3
	20 bis 29	21	25,9	25,9	64,2
	30 bis 39	13	16,0	16,0	80,2
	40 bis 56	16	19,8	19,8	100,0
	Gesamt	81	100,0	100,0	

Tab. 31: Altersklassen in der Faking-good-Stichprobe

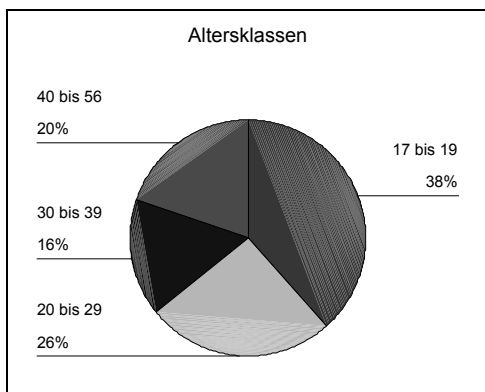


Abb. 15: Prozentuelle Verteilung der Altersklassen in der Faking-good-Stichprobe

Aufgrund der vielen Schüler, die an dieser Untersuchung teilnahmen, sieht die Verteilung der höchsten abgeschlossenen Bildung folgendermaßen aus:

Ausbildung					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Hauptschule, AHS-Unterstufe, Polytechnische Schule	33	40,7	40,7	40,7
	Berufslehre	18	22,2	22,2	63,0
	Berufsbildende mittlere Schule	7	8,6	8,6	71,6
	Krankenschule	4	4,9	4,9	76,5
	Meister, Werkmeister	2	2,5	2,5	79,0
	Reifeprüfung: AHS	5	6,2	6,2	85,2
	Reife- u. Diplomprüfung: BHS, Kolleg	8	9,9	9,9	95,1
	Akademie	1	1,2	1,2	96,3
	Universität	3	3,7	3,7	100,0
	Gesamt	81	100,0	100,0	

Tab. 32: Bildungsverteilung in der Faking-good-Stichprobe

Zur weiteren Berechnung wird die Stichprobe in drei Bildungsklassen unterteilt. Die Bildungsklasse „Pflichtschule“ setzt sich in der Faking-good-Bedingung zusammen aus insgesamt 33 Personen. In der Bildungsklasse „Berufsausbildung“ sind 31 Probanden vertreten. Die Bildungsklasse „Reifeprüfung“ schließlich umfasst 17 Personen.

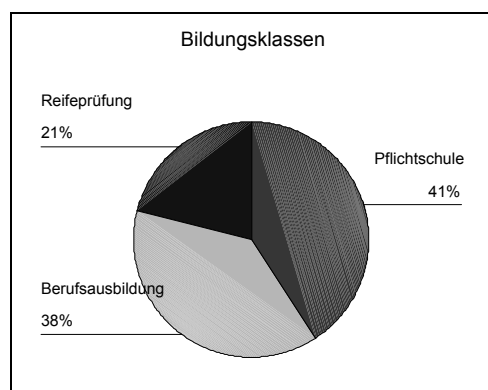


Abb. 16: Prozentueller Anteil der Bildungsklassen in der Faking-good-Stichprobe

4.3.2. Ergebnisse der Persönlichkeitsfragebogen

4.3.2.1. Ergebnisse des 16 PF-R

Die Eingabe und Normierung der Daten erfolgte mittels *Testauswerteprogramm 16 PF-R* (Windowsversion) anhand der geschlechts- und altersspezifischen Normwerte. Von Interesse für die vorliegende Arbeit sind die Globalskala Extraversion sowie die Primärskala Emotionale Stabilität.

Die Werte von 49 Probanden auf der Globalskala Extraversion waren als introvertiert zu bezeichnen, die Werte der übrigen 32 Probanden als extravertiert.

Extra-/Introversion					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	introvertiert	49	60,5	60,5	60,5
	extravertiert	32	39,5	39,5	100,0
	Gesamt	81	100,0	100,0	

Tab. 33: Anteil der Intro- und Extravertierten in der Faking-good-Stichprobe

Laut 16 PF-R sind 60 % der Untersuchungsteilnehmer introvertiert und 40 % extravertiert, das Verhältnis von extra- und introvertierten Personen kann somit als ziemlich ausgewogen angesehen werden.

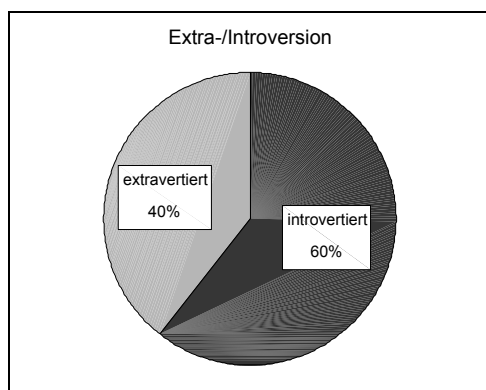


Abb. 17: Prozentueller Anteil der Intro- und Extravertierten in der Faking-good-Stichprobe

Die Verteilung der emotional stabilen und emotional instabilen Probanden gestaltet sich so:

Emotionale Stabilität vs. Emotionale Instabilität					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	emotional instabil	36	44,4	44,4	44,4
	emotional stabil	45	55,6	55,6	100,0
	Gesamt	81	100,0	100,0	

Tab. 34: Anteil der Emotional Stablen und Instabilen in der Faking-good-Stichprobe

Die emotional stabilen Personen machen somit 56 Prozent der Stichprobe aus, die übrigen 44 Prozent sind als emotional instabil zu bezeichnen. Das Verhältnis innerhalb der Stichprobe ist relativ ausgewogen.

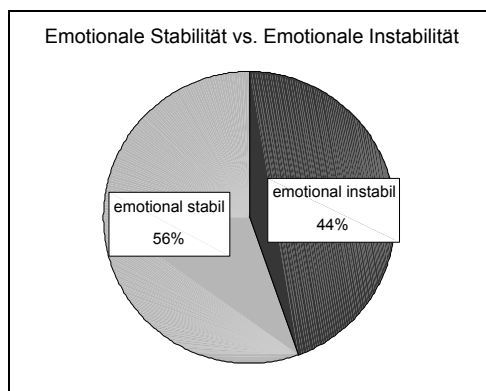


Abb. 18: Prozentuelle Verteilung von emotional stabilen und emotional instabilen Probanden in der Faking-good-Stichprobe

4.3.2.2. Ergebnisse des RSCI

Auch in der Faking-good-Stichprobe ist der Anteil an Repressern und Sensitizern sehr gering.

Repression vs. Sensitization					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Defensiv-Hochängstliche	1	1,2	1,2	1,2
	Represser	7	8,6	8,6	9,9
	Sensitizer	10	12,3	12,3	22,2
	Niedrig-Ängstliche	63	77,8	77,8	100,0
	Gesamt	81	100,0	100,0	

Tab. 35: Represser und Sensitizer in der Faking-good-Stichprobe

Die Stichprobe setzt sich zusammen aus 9 % Repressern (7 Personen) und 12 % Sensitizern (10 Personen). Für die weiteren Berechnungen waren nur diese 17 Personen von Interesse. Der Anteil der Niedrig-Ängstlichen und Defensiv-Hochängstlichen ist für die vorliegende Arbeit nicht relevant, deshalb wird auch nicht näher darauf eingegangen.

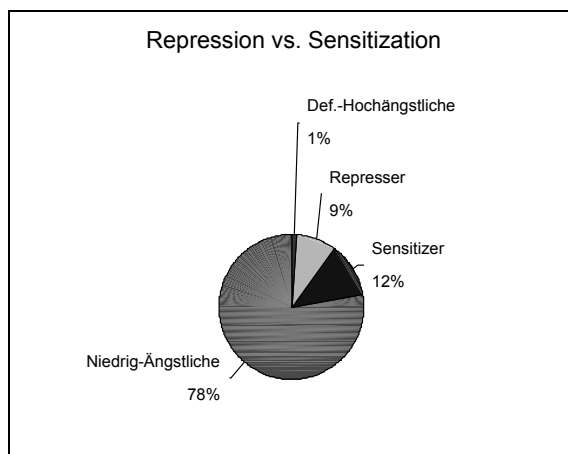


Abb. 19: Prozentueller Anteil von Repressern und Sensitizern in der Faking-good-Stichprobe

4.3.2.3. Ergebnisse der IMS

Personen mit geringer bzw. mittlerer Tendenz zur positiven Selbstdarstellung teilen sich wie folgt auf die Stichprobe auf:

Positive Selbstdarstellung					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gering	49	60,5	60,5	60,5
	mittel	32	39,5	39,5	100,0
	Gesamt	81	100,0	100,0	

Tab. 36: Personen mit geringer bzw. mittlerer IM-Tendenz in der Faking-good-Stichprobe

Auch hier kann man von einem relativ ausgewogenen Verhältnis in der Stichprobe sprechen.

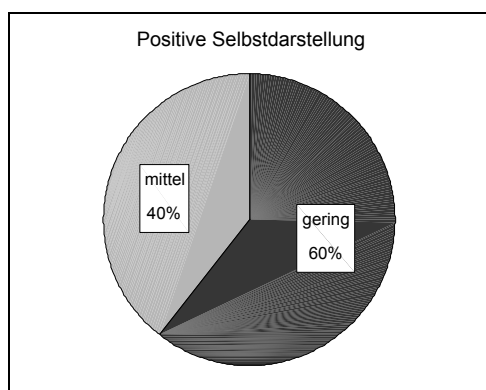


Abb. 20: Prozentueller Anteil der Probanden mit geringen bzw. mittleren IM-Tendenzen

4.3.2.4. Ergebnisse des MUT

Wie schon in der Faking-bad-Stichprobe überwiegt auch hier der Anteil an Personen mit mittlerer moralischer Urteilsfähigkeit.

Moralisches Urteil					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	niedrig	13	16,0	16,0	16,0
	mittel	44	54,3	54,3	70,4
	hoch	15	18,5	18,5	88,9
	sehr hoch	9	11,1	11,1	100,0
	Gesamt	81	100,0	100,0	

Tab. 37: Verteilung der Gruppen zum moralischen Urteil

Prozentuell macht die Gruppe mit mittlerer moralischer Urteilsfähigkeit hier über die Hälfte der Gesamtstichprobe aus.

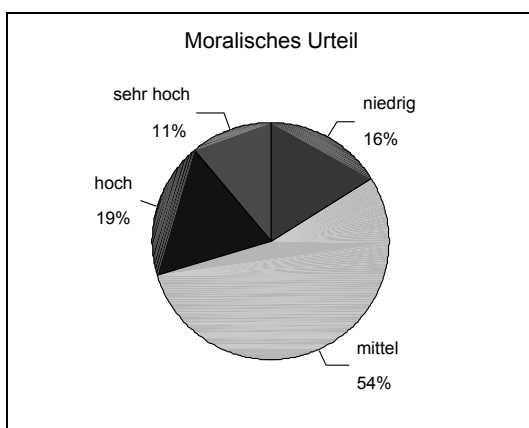


Abb. 21: Prozentuelle Verteilung der moralischen Urteilsfähigkeit

4.4. Prüfung der Hypothesen

Zunächst war von Interesse, ob sich die ermittelten Werte bei den beiden Gesamtstichproben unter den Faking-Instruktionen von jenen in der Honest-Bedingung unterscheiden. Zur Berechnung dieser Frage sowie zu allen weiteren Berechnungen wurden die Sten-Werte der Globalskalen des 16 PF-R herangezogen. Die Sten-Werte können einen Wert von 1 bis 10 annehmen, die Teilung erfolgte deshalb bei 5,5. Die Stichprobengröße betrug 89 Personen in der Faking-bad-Bedingung sowie 81 Personen in der Faking-good-Bedingung.

Allgemeines zu den nachfolgenden Berechnungen: Für die gesamte Auswertung der Untersuchung wurde ein Alpha-Niveau von 5 % zugrunde gelegt. Die Normalverteilung wurde grundsätzlich immer vorab mittels des Kolmogorov-Smirnov-Anpassungstests überprüft. Wenn die Daten normal verteilt sind werden t-Tests für abhängige Stichproben berechnet. Sind die Daten nicht normal verteilt, so werden zur Berechnung Wilcoxon-Tests herangezogen. Für die Gruppenvergleiche (Kap. 4.4.1. bis Kap. 4.4.8.) wurde für jede Gruppe einzeln (Geschlecht, Altersklassen, Bildungsklassen, ...) eine Teilung über die Stichprobe vorgenommen. Die Normalverteilung der Daten wurde innerhalb der Gruppen geprüft. Signifikante Ergebnisse sind in den Tabellen grau unterlegt.

Hypothese H1A₁ postuliert, dass sich Personen in der Faking-bad-Bedingung introvertierter darstellen. Um diese Hypothese zu überprüfen werden die ermittelten Werte auf der Skala Extraversion unter der honest-Bedingung ($\mu = 4.87$) mit jenen der Faking-bad-Bedingung verglichen ($\mu = 2.51$). Da die Daten nicht normal verteilt sind, wurde der Wilcoxon-Test (Sign. = 0,000 für $Z = -6,254$) zur Berechnung herangezogen. Somit wird die Nullhypothese abgelehnt und die Alternativhypothese angenommen.

Hypothese H1B₁ besagt, dass Personen ihre Testwerte auf der Skala Ängstlichkeit erhöhen. Die „ehrlichen“ Testwerte ($\mu = 5.72$) wurden mit den gefakten Testwerten ($\mu = 7.82$) verglichen. Ein Wilcoxon-Test (Sign. = 0,000 für $Z = -6,684$) wurde zur Berechnung herangezogen, da die Daten nicht normal verteilt sind. Die Alternativhypothese kann auch in diesem Fall angenommen werden.

Laut **Hypothese H1C₁** sollten sich die Probanden in der Faking-bad-Bedingung unbeherrschter darstellen. Es erfolgte ein Vergleich der Werte auf den Skalen zur Selbstkontrolle unter der ehrlichen Bedingung ($\mu = 4.76$) mit den verfälschten Werten ($\mu = 2.07$). Wieder sind die Daten nicht normal verteilt, weshalb abermals der Wilcoxon-Test (Sign. = 0,000 für $Z = -7,106$) Verwendung fand. Auch hier trifft die Alternativhypothese zu.

Die **Hypothese H1D₁** besagt, dass die Testwerte der Probanden auf der Skala Unabhängigkeit in der Faking-bad-Bedingung ($\mu = 6.33$) verglichen mit ihren tatsächlichen Testwerten ($\mu = 5.52$) erhöht sind. Die Daten sind nicht normal verteilt, also wurde wieder der Wilcoxon-Test (Sign. = 0,015 für $Z = -2,433$) zur Berechnung herangezogen. Die Ergebnisse sprechen dafür, dass auch hier die Alternativhypothese anzunehmen ist.

Laut **Hypothese H1E₁** werden die Probanden ihre Testwerte auf der Skala Unnachgiebigkeit verglichen mit ihrem True-Score ($\mu = 4.04$) in der Faking-bad-Bedingung ($\mu = 3.18$) herabsetzen. Die Daten sind nicht normal verteilt. Die Ergebnisse des berechneten Wilcoxon-Tests (Sign. = 0,001 für $Z = -3,423$) bestätigen die Alternativhypothese.

Zusammenfassend lässt sich sagen: Die True-Scores unterscheiden sich von den verfälschten Werten bei allen fünf Globalskalen signifikant. Die Mittelwerte zeigen signifikant in die laut Alternativhypothesen postulierte Richtung. Die Probanden haben sich in der **Faking-bad-Bedingung** gemäß der Alternativhypothesen *introvertierter, ängstlicher, unbeherrschter, unabhängiger* und *empfindlicher* dargestellt.

In der **Faking-good-Bedingung** sind die Daten von vier Skalen nicht normal verteilt. Für diese Skalen wurde der Wilcoxon-Test berechnet. Einzig die Skala Unnachgiebigkeit hatte Normalverteilung, somit wurde für diese Skala der t-Test für abhängige Stichproben berechnet.

Die Probanden stellen sich in der Faking-good-Bedingung extravertierter dar, so besagt es die **Hypothese H2A₁**. Es erfolgte ein Vergleich der Testwerte auf der Skala Extraversion. Die „ehrlichen“ Testwerte ($\mu = 4.77$) unterscheiden sich von den gefakten Testwerten ($\mu = 6.35$) laut Wilcoxon-Test (Sign. = 0,000 für $Z = -4,908$) signifikant. Die Nullhypothese wird demnach zugunsten der Alternativhypothese verworfen.

Hypothese H2B₁ postuliert, dass sich Personen in der Faking-good-Bedingung als weniger ängstlich darstellen. Zur Prüfung dieser Annahme wurde der True-Score auf der Skala Ängstlichkeit ($\mu = 5.64$) mit dem verfälschten Testwert ($\mu = 2.59$) verglichen. Die Ergebnisse des Wilcoxon-Tests (Sign. = 0,000 für $Z = -6,716$) sprechen für die Ablehnung der Nullhypothese bei gleichzeitiger Annahme der Alternativhypothese.

Laut **Hypothese H2C₁** werden Probanden ihre Werte auf der Skala Selbstkontrolle in der Faking-good-Bedingung erhöhen. Der Vergleich des True-Scores ($\mu = 5.28$) mit dem Score der Faking-Bedingung ($\mu = 8.65$) mittels Wilcoxon-Test (Sign. = 0,000 für $Z = -6,795$) bestätigt diese Annahme. Die Alternativhypothese kann als bestätigt angesehen werden.

Die Probanden werden in der Faking-good-Bedingung höhere Skalenwerte bei der Skala Unabhängigkeit aufweisen, so besagt es die **Hypothese H2D₁**. Der Vergleich der Skalenwerte unter der ehrlichen Bedingung ($\mu = 5.60$) mit dem Skalenwert der Faking-good-Bedingung ($\mu = 8.02$) mittels Wilcoxon-Test (Sign. = 0,000 für $Z = -5,805$) bestätigt dies. Somit kann die Alternativhypothese angenommen werden.

In der Faking-good-Bedingung werden die Versuchspersonen ihre Skalenwerte auf der Skala Unnachgiebigkeit erhöhen, so lautet **Hypothese H2E₁**. Der ehrliche Wert ($\mu = 4.25$) liegt unter dem gefakten Wert ($\mu = 5.52$). Die Berechnung mittels t-Test für abhängige Stichproben ($T = -4,003$; $df = 80$;

$p = .000$) lieferte ein signifikantes Ergebnis. In diesem Fall kann ebenfalls die Alternativhypothese angenommen werden.

Auch hier unterscheiden sich die verfälschten Werte von den tatsächlichen Werten bei allen fünf Globalskalen signifikant. Die Probanden stellten sich gemäß der Alternativhypothesen in der Faking-good-Bedingung *extravertierter, gelassener, selbstkontrollierter, unabhängiger* sowie *unnachgiebiger* dar.

4.4.1. Geschlecht

Hypothese H1A₁ postuliert, dass sich Personen in der Faking-bad-Bedingung introvertierter darstellen. Um diese Hypothese zu überprüfen werden die ermittelten Werte der männlichen Probanden auf der Skala Extraversion unter der Honest-Bedingung ($\mu = 5.00$) mit jenen der Faking-bad-Bedingung verglichen ($\mu = 2.98$). Da die Daten nicht normal verteilt sind, wurde der Wilcoxon-Test (Sign. = 0,001 für $Z = -3,426$) zur Berechnung herangezogen. Bei den weiblichen Teilnehmern werden zur Prüfung der Hypothese die ermittelten Werte auf der Skala Extraversion unter der Honest-Bedingung ($\mu = 4.75$) ebenfalls mit jenen der Faking-bad-Bedingung verglichen ($\mu = 2.10$). Die Daten sind nicht normal verteilt, deshalb wurde der Wilcoxon-Test (Sign. = 0,000 für $Z = -5,436$) berechnet. Somit wird bei beiden Geschlechtern die Nullhypothese abgelehnt und die Alternativhypothese angenommen.

Die **Hypothese H1B₁** besagt, dass sich Personen in der Faking-bad-Bedingung ängstlicher darstellen. Um dies zu prüfen, wurden die Werte der Männer in der Honest-Bedingung ($\mu = 6.12$) mit jenen der Faking-bad-Bedingung ($\mu = 7.73$) verglichen. Ebenso erfolgte ein Vergleich der ehrlichen Werte ($\mu = 5.38$) mit jenen aus der Faking-bad-Bedingung ($\mu = 7.90$) bei den Frauen. Die Daten sind nicht normal verteilt, es wurde also jeweils ein Wilcoxon-Test (Sign. = 0,000 für $Z = -3,824$) bei Männern sowie ein Wilcoxon-Test (Sign. = 0,000 für $Z = -5,507$) bei Frauen zur Berechnung herangezogen. Die Ergebnisse sprechen dafür, dass sowohl bei Männern als auch bei Frauen die Alternativhypothese angenommen werden kann.

Laut **Hypothese H1C₁** werden die Versuchspersonen dahingehend verfälschen, dass sie in der Faking-Bedingung unbeherrschter erscheinen. Die Daten entsprechen nicht der Normalverteilung, deshalb wurden Wilcoxon-Tests berechnet. Bei Männern sind die tatsächlichen Werte auf der Skala Selbstkontrolle ($\mu = 4.80$) höher als in der Faking-bad-Bedingung ($\mu = 2.41$). Auch bei Frauen ist der „ehrliche“ Testwert ($\mu = 4.73$) höher als der gefakte ($\mu = 1.77$). Sowohl die Ergebnisse des Wilcoxon-Test bei Männern (Sign. = 0,000 für $Z = -4,621$) als auch bei Frauen (Sign. = 0,000 für $Z = -5,299$) sprechen dafür, die Alternativhypothese anzunehmen.

Hypothese H1D₁ postuliert, dass in der Faking-Bedingung die Testwerte auf der Skala Unabhängigkeit erhöht werden. Beim Vergleich des „ehrlichen“ Testwertes der Männer ($\mu = 5.73$) mit dem verfälschten Testwert ($\mu = 5.02$) wurde jedoch Gegenteiliges beobachtet. Die männlichen Probanden verfälschten in eine andere Richtung, sie stellten sich anpassungsbereiter dar. Es wurde ein t-Test für abhängige

Stichproben ($T = 1,729$; $df = 40$; $p = .092$) berechnet. Die Nullhypothese muss beibehalten werden. Ein Vergleich der True-Scores ($\mu = 5.33$) mit den gefakten Scores ($\mu = 7.44$) bei den Frauen erbrachte das erwartete Ergebnis. Der t-Test für abhängige Stichproben ($T = -5,884$; $df = 47$; $p = .000$) war signifikant. Bei den weiblichen Teilnehmern wird die Alternativhypothese angenommen.

Die Probanden stellen sich beim Verfälschen der Skala Unnachgiebigkeit als empfänglicher dar, so besagt es die **Hypothese H1E₁**. Der verfälschte Wert ($\mu = 3.56$) ist bei den männlichen Probanden niedriger als der tatsächliche Wert ($\mu = 4.37$), die Daten sind nicht normal verteilt. Der berechnete Wilcoxon-Test (Sign. = 0,049 für $Z = -1,973$) ist gerade noch signifikant, deshalb kann die Alternativhypothese angenommen werden. Auch bei den Frauen ist der verfälschte Wert ($\mu = 2.85$) niedriger als der True-Score ($\mu = 3.77$), die Daten sind nicht normal verteilt. Es wurde ein Wilcoxon-Test (Sign. = 0,005 für $Z = -2,796$) berechnet. Auch hier wird die Alternativhypothese angenommen.

Alle Differenzen sind signifikant mit Ausnahme der Skala Unabhängigkeit in der **Faking-bad-Bedingung** bei Männern. Sieht man sich die Mittelwerte genauer an, so fällt auf, dass Männer hier in die andere Richtung verfälscht haben (sie haben sich etwas anpassungsbereiter dargestellt, aber eben auch nicht signifikant). Auch bei der Skala Unnachgiebigkeit in der Faking-bad-Bedingung ist auffallend, dass die Differenz zwischen den Testwerten bei Männern gerade noch signifikant ausfällt.

In der **Faking-good-Bedingung** sind die Daten von drei Skalen nicht normal verteilt. Für diese Skalen wurde der Wilcoxon-Test berechnet. Die Skalen Extraversion und Unnachgiebigkeit wiesen Normalverteilung auf, deshalb wurde für diese Skalen der t-Test für abhängige Stichproben zur Berechnung herangezogen.

Will man einen guten Eindruck erzeugen, so stellt man sich extravertiert da, so besagt es die **Hypothese H2A₁**. Der tatsächliche Score ist sowohl bei Männern ($\mu = 5.08$) als auch bei Frauen ($\mu = 4.50$) niedriger als der gefakte Wert bei Männern ($\mu = 6.59$) und Frauen ($\mu = 6.14$). Die berechneten t-Tests für abhängige Stichproben erbrachten sowohl bei den männlichen ($T = -3,765$; $df = 36$; $p = .001$) wie auch bei den weiblichen Probanden ($T = -4,054$; $df = 43$; $p = .000$) signifikante Ergebnisse. Somit kann die Alternativhypothese für beide Geschlechter angenommen werden.

Die Probanden werden laut **Hypothese H2B₁** bemüht sein in der Faking-good-Bedingung wenig ängstlich zu erscheinen. Männer sind tatsächlich ängstlicher ($\mu = 5.76$) als wenn sie in eine positive Richtung verfälschen ($\mu = 3.05$). Auch bei Frauen zeigten sich in der ehrlichen Bedingung ($\mu = 5.55$) höhere Werte für Ängstlichkeit als in der gefakten Bedingung ($\mu = 2.20$). Die Wilcoxon-Tests, welche zur Berechnung herangezogen wurden, sprechen sowohl bei Männern (Sign. = 0,000 für $Z = -4,342$) als auch bei Frauen (Sign. = 0,000 für $Z = -5,159$) für die Annahme der Alternativhypothese.

Verfälscht man in eine positive Richtung, so wird man gemäß **Hypothese H2C₁** bemüht sein selbstkontrollierter zu erscheinen. Männer stellten sich in der Faking-Situation ($\mu = 8.59$)

selbstkontrollierter dar verglichen mit ihrem „ehrlichen“ Testwert ($\mu = 5.22$). Bei Frauen trifft das auch zu. Sie versuchten beim faking-good ($\mu = 8.70$) den Eindruck zu erwecken, sie wären selbstkontrollierter, als dies tatsächlich ($\mu = 5.34$) der Fall ist. Die Ergebnisse der Wilcoxon-Tests der männlichen (Sign. = 0,000 für $Z = -4,775$) und weiblichen Versuchspersonen (Sign. = 0,000 für $Z = -4,840$) sprechen für die Annahme der Alternativhypothese.

In der Faking-good-Bedingung werden sich die Probanden unabhängiger darstellen, so besagt es die **Hypothese H2D₁**. Der Skalenwert der Honest-Bedingung ($\mu = 5.97$) ist bei Männern niedriger als der Skalenwert der Faking-Bedingung ($\mu = 7.41$). Frauen erreichen ebenfalls in der Honest-Bedingung ($\mu = 5.30$) einen niedrigeren Skalenwert im Vergleich zur Faking-Bedingung ($\mu = 8.55$). Die berechneten Wilcoxon-Tests fielen auch hier bei den männlichen (Sign. = 0,005 für $Z = -2,793$) und weiblichen Probanden (Sign. = 0,000 für $Z = -4,946$) signifikant aus. Die Nullhypothese kann somit zugunsten der Alternativhypothese verworfen werden.

Gemäß der **Hypothese H2E₁** stellen sich die Versuchspersonen in der Faking-good-Bedingung unnachgiebiger dar. Der Skalenwert der Honest-Bedingung ist bei Männern ($\mu = 4.49$) und bei Frauen ($\mu = 4.05$) niedriger als der Skalenwert der Faking-Bedingung bei Männern ($\mu = 5.68$) und bei Frauen ($\mu = 5.39$). Die Ergebnisse der t-Tests für abhängige Stichproben sprechen sowohl bei Männern ($T = -2,758$; $df = 36$; $p = .009$) sowie bei Frauen ($T = -2,895$; $df = 43$; $p = .006$) für die Annahme der Alternativhypothese.

In der **Faking-good-Bedingung** konnten somit für beide Geschlechter alle Alternativhypothesen bestätigt werden.

4.4.2. Alter

Hypothese H1A₁ postuliert, dass sich Personen in der Faking-bad-Bedingung introvertierter darstellen. Die Daten sind nicht normal verteilt, somit wurde der Wilcoxon-Test zur Berechnung herangezogen. Bei allen Altersklassen sind wie erwartet die Werte für Extraversion in der Faking-bad-Bedingung niedriger. Jedoch kann nur für die Altersklassen der 17 bis 19jährigen, der 30 bis 39jährigen und der 40 bis 56jährigen die Alternativhypothese angenommen werden. Für die Altersklassen der 20 bis 29jährigen fiel der Unterschied nicht signifikant aus, hier wird die Nullhypothese beibehalten. Die Ergebnisse sind in der Tabelle 38 veranschaulicht.

Altersklassen	μE		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
17-19 (n = 38)	5,11	2,21	-4,715	0,000	H1A ₁
20-29 (n = 22)	4,00	2,36	-1,879	0,060	H1A ₁
30-39 (n = 13)	5,31	2,54	-3,210	0,001	H1A ₁
40-56 (n = 16)	5,13	3,38	-2,627	0,009	H1A ₁

Tab. 38: Prüfung der H1A₁ innerhalb der Altersklassen

Hypothese H1B₁ besagt, dass Personen in der Faking-bad-Bedingung bemüht sind ängstlich zu erscheinen. Zur Prüfung der Hypothese wurden t-Tests für abhängige Stichproben berechnet. Bei allen Altersklassen fiel der Unterschied zwischen dem ehrlichen und dem verfälschten Skalenwert signifikant aus (siehe Tab. 39). Somit wird die Alternativhypothese angenommen.

Altersklassen	μAE		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
17-19 (n = 38)	6,32	7,74	-4,304	37	0,000	H1B ₁
20-29 (n = 22)	5,32	8,09	-5,444	21	0,000	H1B ₁
30-39 (n = 13)	5,38	7,69	-4,308	12	0,001	H1B ₁
40-56 (n = 16)	5,13	7,75	-5,766	15	0,000	H1B ₁

Tab. 39: Prüfung der H1B₁ innerhalb der Altersklassen

Laut **Hypothese H1C₁** werden die Versuchspersonen ihre Werte auf der Skala Selbstkontrolle in der Faking-bad-Bedingung herabsetzen. Dies wurde mittels Wilcoxon-Tests geprüft, da die Daten nicht normal verteilt sind. Die Ergebnisse (siehe Tab. 40) sprechen bei allen Altersklassen für die Annahme der Alternativhypothese.

Altersklassen	μSK		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
17-19 (n = 38)	4,55	2,47	-4,114	0,000	H1C ₁
20-29 (n = 22)	5,05	1,68	-3,782	0,000	H1C ₁
30-39 (n = 13)	4,46	1,54	-3,095	0,002	H1C ₁
40-56 (n = 16)	5,13	2,06	-3,268	0,001	H1C ₁

Tab. 40: Prüfung der H1C₁ innerhalb der Altersklassen

Hypothese H1D₁ postuliert, dass sich Personen in der Faking-bad-Bedingung unabhängiger darstellen als sie sind. Lediglich für die Altersklassen der 17 bis 19jährigen sowie der 20 bis 29jährigen trifft diese Annahme zu. Die berechneten t-Tests für abhängige Stichproben fielen signifikant aus. Für diese beiden Altersklassen kann die Alternativhypothese angenommen werden. Bei der Altersklasse der 30 bis 39jährigen und der 40 bis 56jährigen sind die Werte für Unabhängigkeit in der Faking-bad-Bedingung niedriger. Im Falle der Altersklasse der 30 bis 39jährigen fiel der berechnete t-Test für abhängige Stichproben sogar signifikant aus, das Ergebnis weist aber wie in der Altersklasse der 40 bis 56jährigen nicht in die erwartete Richtung, deshalb wird die Nullhypothese beibehalten. Eine detaillierte Darstellung der Ergebnisse liefert die Tabelle 41.

Altersklassen	μ UA		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
17-19 (n = 38)	5,29	6,79	-3,681	37	0,001	H1D ₁
20-29 (n = 22)	4,95	6,91	-3,514	21	0,002	H1D ₁
30-39 (n = 13)	7,38	5,31	3,038	12	0,010	H1D ₁
40-56 (n = 16)	5,31	5,25	0,088	15	0,931	H1D ₁

Tab. 41: Prüfung der H1D₁ innerhalb der Altersklassen

Die Probanden stellen sich beim Verfälschen der Skala Unnachgiebigkeit als empfänglicher dar, so besagt es die **Hypothese H1E₁**. Allen Altersklassen ist es gelungen in die erwartete Richtung zu verfälschen, bei der Altersklasse der 17 bis 19jährigen ist der verfälschte Wert jedoch nur gering kleiner. Der berechnete t-Test für abhängige Stichproben lieferte kein signifikantes Ergebnis, die Nullhypothese wird beibehalten. Bei den übrigen drei Altersklassen fielen die t-Tests für abhängige Stichproben signifikant aus, deshalb wird bei diesen die Nullhypothese zugunsten der Alternativhypothese verworfen (siehe Tab. 42).

Altersklassen	μ UN		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
17-19 (n = 38)	3,76	3,71	0,162	37	0,872	H1E ₁
20-29 (n = 22)	4,23	2,82	2,845	21	0,010	H1E ₁
30-39 (n = 13)	4,46	3,00	2,714	12	0,019	H1E ₁
40-56 (n = 16)	4,13	2,56	2,894	15	0,011	H1E ₁

Tab. 42: Prüfung der H1E₁ innerhalb der Altersklassen

Zusammenfassend kann man sagen, dass in der Faking-bad-Bedingung jede der Altersklassen ein einziges Mal kein signifikantes Ergebnis hatte.

Werden Probanden durch eine Fragebogeninstruktion aufgefordert sich gut darzustellen, werden sie versuchen extravertierter zu erscheinen, so besagt es die **Hypothese H2A₁**. In allen Altersklassen war der Wert auf der Skala Extraversion in der Faking-good-Bedingung höher als in der Honest-Bedingung. Die berechneten t-Tests fielen bei den 17 bis 19jährigen als auch bei den 30 bis 39jährigen Probanden nicht signifikant aus, hier wird also die Nullhypothese beibehalten. Die Ergebnisse der t-Tests für abhängige Stichproben sprechen bei den 20 bis 29jährigen und 40 bis 56jährigen Probanden für die Verwerfung der Nullhypothese zugunsten der Alternativhypothese (siehe Tab. 43).

Altersklassen	μ E		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
17-19 (n = 31)	5,16	6,00	-1,864	30	0,072	H2A ₁
20-29 (n = 21)	3,57	6,10	-4,322	20	0,000	H2A ₁
30-39 (n = 13)	5,31	7,00	-2,065	12	0,061	H2A ₁
40-56 (n = 16)	5,13	6,81	-3,645	15	0,002	H2A ₁

Tab. 43: Prüfung der H2A₁ innerhalb der Altersklassen

Laut **Hypothese H2B₁** sollte der tatsächliche Wert der Skala Ängstlichkeit höher sein als der in eine positive Richtung verfälschte Wert. Die Prüfung dieser Annahme erfolgte durch die Berechnung von

Wilcoxon-Tests geprüft, da die Daten nicht normal verteilt sind. Bei allen Altersklassen sprechen die Ergebnisse (siehe Tab. 44) dafür die Alternativhypothese anzunehmen.

Altersklassen	μ AE		Wilcoxon		Hypothese
	HO	FG	Z	Sign.	
17-19 (n = 31)	6,13	2,61	-3,678	0,000	H2B ₁
20-29 (n = 21)	5,48	2,29	-3,945	0,000	H2B ₁
30-39 (n = 13)	5,38	2,46	-3,083	0,002	H2B ₁
40-56 (n = 16)	5,13	3,06	-3,063	0,002	H2B ₁

Tab. 44: Prüfung der H2B₁ innerhalb der Altersklassen

In der Faking-good-Bedingung werden die Versuchspersonen darum bemüht sein sich als selbstkontrolliert darzustellen, so besagt es die **Hypothese H2C₁**. Dies wurde mittels Wilcoxon-Tests geprüft, da die Daten nicht normal verteilt sind. Hier sprechen die Ergebnisse (siehe Tab. 45) bei allen Altersklassen für die Annahme der Alternativhypothese.

Altersklassen	μ SK		Wilcoxon		Hypothese
	HO	FG	Z	Sign.	
17-19 (n = 31)	5,74	8,97	-3,802	0,000	H2C ₁
20-29 (n = 21)	5,24	8,81	-3,715	0,000	H2C ₁
30-39 (n = 13)	4,46	8,69	-3,064	0,002	H2C ₁
40-56 (n = 16)	5,13	7,81	-3,118	0,002	H2C ₁

Tab. 45: Prüfung der H2C₁ innerhalb der Altersklassen

Hypothese H2D₁ postuliert, dass die Probanden in der Faking-good-Bedingung ihre Werte auf der Skala Unabhängigkeit erhöhen. Bei allen Altersklassen konnte dies bestätigt werden. Die berechneten t-Tests für abhängige Stichproben fielen bei den 17 bis 19jährigen, 20 bis 29jährigen und den 40 bis 56jährigen Versuchspersonen signifikant aus, für diese drei Altersklassen spricht alles für die Annahme der Alternativhypothese. Einzig die Ergebnisse der Altersklasse der 30 bis 39jährigen waren nicht signifikant, für diese Altersklasse wird die Nullhypothese beibehalten. In Tabelle 46 sind die Ergebnisse dargestellt.

Altersklassen	μ UA		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
17-19 (n = 31)	5,39	7,81	-3,838	30	0,001	H2D ₁
20-29 (n = 21)	5,05	8,33	-6,191	20	0,000	H2D ₁
30-39 (n = 13)	7,38	8,54	-1,931	12	0,077	H2D ₁
40-56 (n = 16)	5,31	7,63	-3,706	15	0,002	H2D ₁

Tab. 46: Prüfung der H2D₁ innerhalb der Altersklassen

Laut **Hypothese H2E₁** werden die Versuchspersonen ihre Werte auf der Skala Unnachgiebigkeit in der Faking-good-Bedingung erhöhen. Lediglich für die Altersklassen der 17 bis 19jährigen, der 20 bis 29jährigen sowie der 30 bis 39jährigen trifft diese Annahme zu. Die berechneten t-Tests für abhängige Stichproben fielen jedoch nur für die Altersklassen der 17 bis 19jährigen und der 20 bis 29jährigen signifikant aus. Für diese beiden Altersklassen kann die Alternativhypothese angenommen werden. Bei der Altersklasse der 40 bis 56jährigen sind die Werte für Unnachgiebigkeit in der Faking-good-

Bedingung niedriger. Die Nullhypothese wird beibehalten. Im Falle der Altersklasse der 30 bis 39jährigen fiel der berechnete t-Test für abhängige Stichproben nicht signifikant aus, deshalb wird auch hier die Nullhypothese beibehalten. Die Ergebnisse sind in Tabelle 47 dargestellt.

Altersklassen	μ UN		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
17-19 (n = 31)	4,23	6,68	-4,955	30	0,000	H2E ₁
20-29 (n = 21)	4,24	5,57	-2,597	20	0,017	H2E ₁
30-39 (n = 13)	4,46	4,85	-0,433	12	0,673	H2E ₁
40-56 (n = 16)	4,13	3,75	0,594	15	0,562	H2E ₁

Tab. 47: Prüfung der H2E₁ innerhalb der Altersklassen

4.4.3. Bildung

Hypothese H1A₁ besagt, dass sich Personen in der Faking-bad-Bedingung introvertierter darstellen. Die Ergebnisse der berechneten Wilcoxon-Tests sprechen (siehe Tab. 48) bei allen Bildungsklassen für die Annahme der Alternativhypothese.

Bildungsklassen	μ E		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
Pflichtschule (n = 42)	5,31	2,12	-5,112	0,000	H1A ₁
Berufsausbildung (n = 30)	4,83	3,43	-2,594	0,009	H1A ₁
Reifeprüfung (n = 17)	3,82	1,82	-2,636	0,008	H1A ₁

Tab. 48: Prüfung der H1A₁ innerhalb der Bildungsklassen

Personen sind in der Faking-bad-Bedingung bemüht ängstlich zu erscheinen, so lautet **Hypothese H1B₁**. Zur Prüfung der Annahme wurden Wilcoxon-Tests berechnet. Hier sprechen die Ergebnisse (siehe Tab. 49) bei allen Bildungsklassen für die Annahme der Alternativhypothese.

Bildungsklassen	μ AE		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
Pflichtschule (n = 42)	6,21	7,71	-3,965	0,000	H1B ₁
Berufsausbildung (n = 30)	5,17	7,53	-4,033	0,000	H1B ₁
Reifeprüfung (n = 17)	5,47	8,59	-3,532	0,000	H1B ₁

Tab. 49: Prüfung der H1B₁ innerhalb der Bildungsklassen

Laut **Hypothese H1C₁** werden die Werte auf der Skala Selbstkontrolle in der Faking-bad-Bedingung niedriger sein, als sie tatsächlich sind. Dies wurde mittels Wilcoxon-Tests geprüft, die Ergebnisse sprechen (siehe Tab. 50) bei allen Bildungsklassen dafür die Alternativhypothese anzunehmen.

Bildungsklassen	μ SK		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
Pflichtschule (n = 42)	4,50	2,40	-4,385	0,000	H1C ₁
Berufsausbildung (n = 30)	4,90	1,87	-4,470	0,000	H1C ₁
Reifeprüfung (n = 17)	5,18	1,59	-3,439	0,001	H1C ₁

Tab. 50: Prüfung der H1C₁ innerhalb der Bildungsklassen

Laut **Hypothese H1D₁** werden die Probanden in der Faking-bad-Bedingung bemüht sein unabhängiger zu erscheinen. Dies trifft auch bei allen Bildungsklassen zu. Die berechneten t-Tests für abhängige Stichproben lieferten jedoch nur für die Bildungsklasse „Pflichtschule“ ein signifikantes Ergebnis. In diesem Falle wird die Alternativhypothese angenommen. Die Unterschiede zwischen den ehrlichen und verfälschten Skalenwerten waren bei den Bildungsklassen „Berufsausbildung“ und „Matura“ zu gering, die hierzu berechneten t-Tests für abhängige Stichproben waren nicht signifikant. Für diese beiden Bildungsklassen muss somit die Nullhypothese beibehalten werden (siehe Tab. 51).

Bildungsklassen	μ UA		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
Pflichtschule (n = 42)	5,36	6,60	-3,333	41	0,002	H1D ₁
Berufsausbildung (n = 30)	6,10	6,30	-0,320	29	0,751	H1D ₁
Reifeprüfung (n = 17)	4,88	5,71	-1,138	16	0,272	H1D ₁

Tab. 51: Prüfung der H1D₁ innerhalb der Bildungsklassen

Die Probanden stellen sich beim Verfälschen der Skala Unnachgiebigkeit als empfänglicher dar, so lautet die **Hypothese H1E₁**. Bei den vorliegenden Ergebnissen ist auffällig, dass hier scheinbar keine Verfälschung in der Bildungsklasse „Pflichtschule“ stattfand. Der Mittelwert ist in beiden Antwortbedingungen gleich. Somit wird für diese Bildungsklasse die Nullhypothese beibehalten. Den beiden „höheren“ Bildungsklassen gelang es jedoch recht gut in die erwartete Richtung zu verfälschen, die Ergebnisse der t-Tests für abhängige Stichproben sprechen für die Annahme der Alternativhypothese. In Tabelle 52 sind die Ergebnisse genauer dargestellt.

Bildungsklassen	μ UN		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
Pflichtschule (n = 42)	3,64	3,64	0,000	41	1,000	H1E ₁
Berufsausbildung (n = 30)	4,60	2,90	4,090	29	0,000	H1E ₁
Reifeprüfung (n = 17)	4,06	2,53	3,147	16	0,006	H1E ₁

Tab. 52: Prüfung der H1E₁ innerhalb der Bildungsklassen

In der Faking-good-Bedingung sind die Skalenwerte für Extraversion höher als in der Honest-Bedingung, so besagt es **Hypothese H2A₁**. Die Mittelwerte der Probanden bestätigen diese Annahme. Jedoch war der Unterschied in der Bildungsklasse „Pflichtschule“ nicht signifikant, wie der t-Test für abhängige Stichproben zeigte. Die Nullhypothese muss für die „Pflichtschüler“ beibehalten werden. Die t-Tests der

beiden anderen Bildungsklassen waren allerdings hoch signifikant, womit die Alternativhypothese für diese Bildungsklassen angenommen wird. Die Ergebnisse sind in Tabelle 53 dargestellt.

Bildungsklassen	μE		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
Pflichtschule (n = 33)	5,24	6,00	-1,776	32	0,085	H2A ₁
Berufsausbildung (n = 31)	4,77	6,48	-4,169	30	0,000	H2A ₁
Reifeprüfung (n = 17)	3,82	6,76	-4,411	16	0,000	H2A ₁

Tab. 53: Prüfung der H2A₁ innerhalb der Bildungsklassen

Hypothese H2B₁ lautet, dass Personen wenig ängstlich erscheinen wollen in der Faking-good-Bedingung. Die gefakten Werte auf der Skala Ängstlichkeit sollten somit niedriger sein als die ehrlichen Werte. Die berechneten Wilcoxon-Tests fielen hoch signifikant aus, somit wurde die Alternativhypothese für alle drei Bildungsklassen bestätigt. Eine genauere Darstellung der Ergebnisse findet sich in Tabelle 54.

Bildungsklassen	μAE		Wilcoxon		Hypothese
	HO	FG	Z	Sign.	
Pflichtschule (n = 33)	6,09	2,79	-3,745	0,000	H2B ₁
Berufsausbildung (n = 31)	5,26	2,68	-4,585	0,000	H2B ₁
Reifeprüfung (n = 17)	5,47	2,06	-3,536	0,000	H2B ₁

Tab. 54: Prüfung der H2B₁ innerhalb der Bildungsklassen

Laut **Hypothese H2C₁** werden die Probanden ihre Werte auf der Skala Selbstkontrolle in der Faking-good-Bedingung erhöhen. Ein Vergleich der Mittelwerte bestätigt diese Annahme. Die Ergebnisse der berechneten Wilcoxon-Tests waren hoch signifikant, die Alternativhypothese wird für alle drei Bildungsklassen angenommen. Die Ergebnisse sind in Tabelle 55 dargestellt.

Bildungsklassen	μSK		Wilcoxon		Hypothese
	HO	FG	Z	Sign.	
Pflichtschule (n = 33)	5,67	8,55	-3,773	0,000	H2C ₁
Berufsausbildung (n = 31)	4,94	8,52	-4,661	0,000	H2C ₁
Reifeprüfung (n = 17)	5,18	9,12	-3,316	0,001	H2C ₁

Tab. 55: Prüfung der H2C₁ innerhalb der Bildungsklassen

Hypothese H2D₁ postuliert, dass die Probanden in der Faking-good-Bedingung ihre Werte auf der Skala Unabhängigkeit erhöhen werden. Dies trifft auf alle drei Bildungsklassen zu. Die berechneten t-Tests fielen hoch signifikant aus (siehe Tab. 56), das spricht für die Annahme der Alternativhypothese.

Bildungsklassen	μ UA		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
Pflichtschule (n = 33)	5,55	7,94	-3,950	32	0,000	H2D ₁
Berufsausbildung (n = 31)	6,06	7,97	-4,416	30	0,000	H2D ₁
Reifeprüfung (n = 17)	4,88	8,29	-5,994	16	0,000	H2D ₁

Tab. 56: Prüfung der H2D₁ innerhalb der Bildungsklassen

Laut **Hypothese H2E₁** werden die Probanden in der Faking-good-Bedingung höhere Werte auf der Skala Unnachgiebigkeit erzielen als in der Honest-Bedingung. Dies kann ein Vergleich der Mittelwerte bestätigen (siehe Tab. 57). Zur Prüfung der Hypothese wurden t-Tests für abhängige Stichproben berechnet. Im Falle der Bildungsklassen „Pflichtschule“ und „Reifeprüfung“ sprechen die Ergebnisse für die Annahme der Alternativhypothese. Bei der Bildungsklasse „Berufsausbildung“ unterscheiden sich die Mittelwerte jedoch nur geringfügig voneinander, sodass hier die Nullhypothese beibehalten werden muss.

Bildungsklassen	μ UN		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
Pflichtschule (n = 33)	4,09	6,24	-4,269	32	0,000	H2E ₁
Berufsausbildung (n = 31)	4,52	4,68	-0,331	30	0,743	H2E ₁
Reifeprüfung (n = 17)	4,06	5,65	-2,618	16	0,019	H2E ₁

Tab. 57: Prüfung der H2E₁ innerhalb der Bildungsklassen

4.4.4. Extraversion versus Introversion

Die Mittelwerte gehen in der Faking-bad-Bedingung sowohl bei Intro- als auch bei Extravertierten in die laut Alternativhypothesen erwartete Richtung. Die berechneten Wilcoxon-Tests zeigten signifikante Differenzen für die Globalskalen Extraversion (**H1A₁**), Ängstlichkeit (**H1B₁**), Selbstkontrolle (**H1C₁**) und Unnachgiebigkeit (**H1E₁**). Die Probanden haben sich in der Faking-bad-Bedingung gemäß der Alternativhypothesen - verglichen mit den Werten aus der Honest-Bedingung - introvertierter, ängstlicher, unbeherrschter und empfänglicher dargestellt. Die Ergebnisse sprechen für die Annahme dieser vier Alternativhypothesen. Allein die Differenz der Mittelwerte für die Skala Unabhängigkeit ist laut berechnetem t-Test für abhängige Stichproben nicht signifikant. Hier muss die Alternativhypothese (**H1D₁**) verworfen werden, die Nullhypothese wird beibehalten (siehe Tab. 58 und Tab. 59).

Hypothesenprüfung für Introvertierte (n = 53)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
3,49	2,51	-3,160	0,002				H1A ₁
5,96	7,74	-4,481	0,000				H1B ₁
5,06	2,21	-5,425	0,000				H1C ₁
5,57	6,43			-2,002	52	0,051	H1D ₁
4,23	3,25	-2,764	0,006				H1E ₁

Tab. 58: Ergebnisse der Introvertierten in der Faking-bad-Stichprobe

Hypothesenprüfung für Extravertierte (n = 36)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
6,89	2,50	-5,249	0,000				H1A ₁
5,36	7,94	-4,947	0,000				H1B ₁
4,33	1,86	-4,651	0,000				H1C ₁
5,44	6,17			-1,731	35	0,092	H1D ₁
3,78	3,08	-2,020	0,043				H1E ₁

Tab. 59: Ergebnisse der Extravertierten in der Faking-bad-Stichprobe

In der Faking-good-Bedingung unterscheiden sich im Falle der Introvertierten die verfälschten Werte von den tatsächlichen Werten bei allen fünf Globalskalen signifikant. Gemäß der Alternativhypothesen stellten sich die Probanden in der Faking-good-Bedingung extravertierter (**H2A₁**), gelassener (**H2B₁**), selbstkontrollierter (**H2C₁**), unabhängiger (**H2D₁**) sowie unnachgiebiger (**H2E₁**) dar. In Tabelle 60 sind die Mittelwerte sowie die Ergebnisse der Wilcoxon-Tests und des t-Tests für abhängige Stichproben dargestellt.

Hypothesenprüfung für Introvertierte (n = 49)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FG	Z	Sign.	T	df	Sign.	
3,41	6,35	-5,867	0,000				H2A ₁
6,00	2,53	-5,515	0,000				H2B ₁
5,45	8,82	-5,102	0,000				H2C ₁
5,49	8,12	-4,542	0,000				H2D ₁
4,22	5,53			-3,173	48	0,003	H2E ₁

Tab. 60: Ergebnisse der Introvertierten in der Faking-good-Stichprobe

Bei den Extravertierten fällt vor allem auf, dass ihre Extraversions-Werte in der Faking-good-Bedingung entgegen der Erwartungen etwas niedriger sind als in der Honest-Bedingung (**H2A₁**). Hier muss die Alternativhypothese zugunsten der Nullhypothese verworfen werden. Ansonsten waren auch die extravertierten Probanden dazu in der Lage sich in der Faking-Bedingung signifikant gelassener (**H2B₁**), selbstkontrollierter (**H2C₁**), unabhängiger (**H2D₁**) sowie unnachgiebiger (**H2E₁**) darzustellen. Eine

Übersicht über die Mittelwerte, die Wilcoxon-Tests und den t-Test für abhängige Stichproben gibt Tabelle 61.

Hypothesenprüfung für Extravertierte (n = 32)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FG	Z	Sign.	T	df	Sign.	
6,84	6,34	-1,263	0,206				H2A ₁
5,09	2,69	-3,870	0,000				H2B ₁
5,03	8,41	-4,513	0,000				H2C ₁
5,78	7,88	-3,645	0,000				H2D ₁
4,28	5,50			-2,402	31	0,023	H2E ₁

Tab. 61: Ergebnisse der Extravertierten in der Faking-good-Stichprobe

4.4.5. Emotionale Stabilität vs. Emotionale Instabilität

In der Faking-bad-Bedingung ist es der Gruppe der emotional instabilen Probanden gelungen in die von den Alternativhypothesen erwarteten Richtungen zu verfälschen. Der Vergleich der Mittelwerte mittels Wilcoxon-Tests fiel signifikant aus für die Globalskalen Extraversion (**H1A₁**), Ängstlichkeit (**H1B₁**), Selbstkontrolle (**H1C₁**) und Unnachgiebigkeit (**H1E₁**). Gemäß der Alternativhypothesen hat sich diese Gruppe in der Faking-bad-Bedingung introvertierter, ängstlicher, unbeherrschter und empfänglicher darstellen können. Die Ergebnisse sprechen für die Annahme dieser vier Alternativhypothesen. Der t-Test für abhängige Stichproben für die Skala Unabhängigkeit (**H1D₁**) war ebenfalls signifikant. Auch hier spricht alles für eine Annahme der Alternativhypothese (siehe Tab. 62).

Hypothesenprüfung für Emotional Instabile (n = 40)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
4,38	2,50	-3,596	0,000				H1A ₁
6,88	7,65	-2,301	0,021				H1B ₁
4,55	2,17	-4,384	0,000				H1C ₁
5,33	6,68			-3,083	39	0,004	H1D ₁
4,18	3,33	-2,329	0,020				H1E ₁

Tab. 62: Ergebnisse der Emotional Instabilen in der Faking-bad-Stichprobe

Die Mittelwerte gehen in der Faking-bad-Bedingung bei der Gruppe der emotional stabilen Personen in die laut Alternativhypothesen erwartete Richtung. Signifikante Differenzen zeigten hierbei die berechneten Wilcoxon-Tests für die Globalskalen Extraversion (**H1A₁**), Ängstlichkeit (**H1B₁**), Selbstkontrolle (**H1C₁**) und Unnachgiebigkeit (**H1E₁**). Den Probanden gelang es somit haben sich introvertierter, ängstlicher, unbeherrschter und empfänglicher darzustellen. Diese vier Alternativhypothesen können angenommen werden. Der t-Test für abhängige Stichproben erbrachte für die Skala Unabhängigkeit (**H1D₁**) kein signifikantes Ergebnis. Hier muss die Alternativhypothese verworfen werden, die Nullhypothese wird beibehalten (siehe Tab. 63).

Hypothesenprüfung für Emotional Stabile (n = 49)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
5,27	2,51	-5,175	0,000				H1A ₁
4,78	7,96	-6,077	0,000				H1B ₁
4,94	1,98	-5,598	0,000				H1C ₁
5,67	6,04			-0,872	48	0,388	H1D ₁
3,94	3,06	-2,413	0,016				H1E ₁

Tab. 63: Ergebnisse der Emotional Stablen in der Faking-bad-Stichprobe

In der Faking-good-Bedingung ist es emotional stabilen und emotional instabilen Personen gleichermaßen gelungen in die erwartete Richtung zu verfälschen. Die Werte der Faking-Bedingung unterscheiden sich von den Werten der Honest-Bedingung bei den interessierenden fünf Globalskalen signifikant. Die Probanden stellten sich gemäß der Alternativhypothesen in der Faking-good-Bedingung extravertierter (H2A₁), gelassener (H2B₁), selbstkontrollierter (H2C₁), unabhängiger (H2D₁) sowie unnachgiebiger (H2E₁) dar. Somit können in der Faking-good-Bedingung alle Alternativhypothesen angenommen werden. Tabelle 64 und Tabelle 65 bieten eine Übersicht über die Mittelwerte und die Ergebnisse der Wilcoxon-Tests und t-Tests.

Hypothesenprüfung für Emotional Instabile (n = 36)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FG	Z	Sign.	T	df	Sign.	
4,17	6,47			-4,825	35	0,000	H2A ₁
6,97	2,44	-5,183	0,000				H2B ₁
5,22	8,89	-4,795	0,000				H2C ₁
5,14	8,08	-4,576	0,000				H2D ₁
4,25	5,97			-3,712	35	0,001	H2E ₁

Tab. 64: Ergebnisse der Emotional Instabilen in der Faking-good-Stichprobe

Hypothesenprüfung für Emotional Stabile (n = 45)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FG	Z	Sign.	T	df	Sign.	
5,24	6,24			-3,131	44	0,003	H2A ₁
4,58	2,71	-4,309	0,000				H2B ₁
5,33	8,47	-4,846	0,000				H2C ₁
5,98	7,98	-3,687	0,000				H2D ₁
4,24	5,16			-2,109	44	0,041	H2E ₁

Tab. 65: Ergebnisse der Emotional Stablen in der Faking-good-Stichprobe

4.4.6. Repression vs. Sensitization

In der Faking-bad-Bedingung stellten sich die Probanden im Vergleich zur Honest-Bedingung erwartungsgemäß introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher dar. Lediglich die berechneten Wilcoxon-Tests für die Skala Selbstkontrolle (H1C₁) fielen sowohl bei

Repressern und Sensitizern signifikant aus, die Alternativhypothese wird demnach angenommen. Die berechneten t-Tests für abhängige Stichproben zeigten bei den Skalen Extraversion (**H1A₁**) und Unnachgiebigkeit (**H1E₁**) keine signifikanten Unterschiede der Mittelwerte, somit werden hier die Nullhypothesen beibehalten. Nur die Represser haben auf der Skala Ängstlichkeit (**H1B₁**) ihren Wert in der Faking-Bedingung signifikant erhöhen können. Im Falle der Represser wird die Alternativhypothese angenommen, während sie bei der Gruppe der Sensitizer zugunsten der Nullhypothese verworfen werden muss. Bei der Skala Unabhängigkeit (**H1D₁**) waren es dann allerdings die Sensitizer, deren Faking-Werte signifikant höher waren als deren Honest-Werte. Für die Sensitizer spricht das Ergebnis für die Annahme der Alternativhypothese, während bei den Repressern die Nullhypothese beibehalten werden muss (siehe Tab. 66 und Tab. 67).

Hypothesenprüfung für Represser (n = 10)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
4,70	3,00			1,497	9	0,169	H1A ₁
5,10	8,00			-4,529	9	0,001	H1B ₁
4,40	1,50	-2,676	0,007				H1C ₁
5,50	6,00			-0,582	9	0,575	H1D ₁
3,60	3,00			0,627	9	0,546	H1E ₁

Tab. 66: Ergebnisse der Represser in der Faking-bad-Stichprobe

Hypothesenprüfung für Sensitizer (n = 14)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
4,50	2,79			1,949	13	0,073	H1A ₁
7,07	7,71			-1,351	13	0,200	H1B ₁
4,21	1,86	-2,966	0,003				H1C ₁
4,86	6,57			-2,350	13	0,035	H1D ₁
3,86	3,14			0,989	13	0,341	H1E ₁

Tab. 67: Ergebnisse der Sensitizer in der Faking-bad-Stichprobe

Die Probanden stellten sich gemäß der Alternativhypothesen in der Faking-good-Bedingung extravertierter (**H2A₁**), gelassener (**H2B₁**), selbstkontrollierter (**H2C₁**) und unabhängiger (**H2D₁**) dar. Die berechneten t-Tests zeigten, dass sich die Mittelwerte der Honest- und der Faking-good-Bedingung signifikant voneinander unterscheiden. Die vier Alternativhypothesen werden somit angenommen. Represser weisen zwar in der Faking-Bedingung höhere Testwerte auf der Skala Unnachgiebigkeit (**H2E₁**) auf, der t-Test war jedoch nicht signifikant. Sensitizer verfälschten ihre Werte hier in eine andere Richtung, d. h. sie stellten sich in der Faking-Bedingung empfänglicher dar (jedoch nur geringfügig). Sowohl bei Repressern als auch bei Sensitizern muss die Nullhypothese beibehalten werden (siehe Tab. 68 und Tab. 69).

Hypothesenprüfung für Represser (n = 7)					
μ		t-Test abhängig			Hypothesen
HO	FG	T	df	Sign.	
4,43	6,71	-4,824	6	0,003	H2A ₁
5,14	1,86	5,811	6	0,001	H2B ₁
4,86	9,43	-8,656	6	0,000	H2C ₁
4,86	8,14	-3,798	6	0,009	H2D ₁
3,86	5,29	-1,698	6	0,140	H2E ₁

Tab. 68: Ergebnisse der Represser in der Faking-good-Stichprobe

Hypothesenprüfung für Sensitizer (n = 10)					
μ		t-Test abhängig			Hypothesen
HO	FG	T	df	Sign.	
3,30	6,20	-3,059	9	0,014	H2A ₁
7,10	2,80	3,924	9	0,003	H2B ₁
4,80	8,80	-5,367	9	0,000	H2C ₁
4,70	8,00	-3,637	9	0,005	H2D ₁
4,20	4,10	0,103	9	0,920	H2E ₁

Tab. 69: Ergebnisse der Sensitizer in der Faking-good-Stichprobe

4.4.7. Moralische Urteilsfähigkeit

Die Probanden werden in der Faking-bad-Bedingung auf der Skala Extraversion niedrigere Werte aufweisen als in der Honest-Bedingung, da sie sich introvertierter darstellen, so lautet die **Hypothese H1A₁**. Sieht man sich die jeweiligen Skalen-Mittelwerte an, so kann man diese Tendenz ganz klar erkennen. Die berechneten Wilcoxon-Tests zeigten bei den Gruppen der Personen mit niedriger, mittlerer und hoher moralischer Urteilsfähigkeit signifikante Ergebnisse, für diese Gruppen spricht demnach alles für die Annahme der Alternativhypothese. Die Differenz der Mittelwerte war in der Gruppe der Probanden mit sehr hoher moralischer Urteilsfähigkeit nicht signifikant, hier wird die Nullhypothese beibehalten (die Ergebnisse sind in Tab. 70 dargestellt).

Moralisches Urteil	μE		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
niedrig (n = 18)	5,00	3,11	-2,562	0,010	H1A ₁
mittel (n = 46)	4,74	2,28	-4,799	0,000	H1A ₁
hoch (n = 16)	5,50	2,50	-3,208	0,002	H1A ₁
sehr hoch (n = 9)	4,11	2,44	-1,185	0,236	H1A ₁

Tab. 70: Prüfung der H1A₁ innerhalb der MU-Klassen

Hypothese H1B₁ besagt, dass Personen in der Faking-bad-Bedingung bemüht sind ängstlich zu erscheinen. T-Tests für abhängige Stichproben wurden zur Prüfung dieser Hypothese berechnet. Der Unterschied zwischen dem ehrlichen und dem verfälschten Skalenwert fiel bei allen Gruppen signifikant aus (siehe Tab. 71). Somit wird die Alternativhypothese angenommen.

Moralisches Urteil	μ AE		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
niedrig (n = 18)	5,72	7,11	-2,782	17	0,013	H1B ₁
mittel (n = 46)	5,59	8,02	-8,273	45	0,000	H1B ₁
hoch (n = 16)	6,19	7,81	-2,855	15	0,012	H1B ₁
sehr hoch (n = 9)	5,56	8,22	-3,336	8	0,010	H1B ₁

Tab. 71: Prüfung der H1B₁ innerhalb der MU-Klassen

Laut **Hypothese H1C₁** werden die Werte auf der Skala Selbstkontrolle in der Faking-bad-Bedingung niedriger sein als in der Honest-Bedingung. Dies wurde mittels Wilcoxon-Tests geprüft, die Ergebnisse (siehe Tab. 72) sprechen bei allen Klassen des moralischen Urteils dafür die Alternativhypothese anzunehmen.

Moralisches Urteil	μ SK		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
niedrig (n = 18)	4,78	2,50	-2,644	0,008	H1C ₁
mittel (n = 46)	4,89	2,00	-5,374	0,000	H1C ₁
hoch (n = 16)	4,38	2,06	-3,075	0,002	H1C ₁
sehr hoch (n = 9)	4,78	1,56	-2,561	0,010	H1C ₁

Tab. 72: Prüfung der H1C₁ innerhalb der MU-Klassen

Hypothese H1D₁ postuliert, dass sich Personen in der Faking-bad-Bedingung unabhängiger darstellen als sie sind. Diese Tendenz lässt sich bei einem Vergleich der Mittelwerte erkennen (siehe dazu Tab. 73). Einzig der t-Test für abhängige Stichproben für die Gruppe der Personen mit mittlerer moralischer Urteilsfähigkeit war signifikant, hier wird die Alternativhypothese angenommen. Für die anderen drei Gruppen gilt weiterhin die Nullhypothese.

Moralisches Urteil	μ UA		t-Test abhängig			Hypothese
	HO	FB	T	df	Sign.	
niedrig (n = 18)	5,94	7,06	-1,368	17	0,189	H1D ₁
mittel (n = 46)	5,39	6,26	-2,161	45	0,036	H1D ₁
hoch (n = 16)	5,75	5,81	-0,082	15	0,936	H1D ₁
sehr hoch (n = 9)	4,89	6,11	-1,504	8	0,171	H1D ₁

Tab. 73: Prüfung der H1D₁ innerhalb der MU-Klassen

Die Probanden stellen sich beim Verfälschen der Skala Unnachgiebigkeit in der Faking-bad-Bedingung als empfänglicher dar, so besagt es die **Hypothese H1E₁**. Hier zeigt sich dasselbe Bild wie schon bei der Skala Unabhängigkeit. Wieder war nur die Personengruppe mit mittlerer moralischer Urteilsfähigkeit dazu imstande signifikant zu verfälschen, weshalb auch nur für diese Gruppe die Alternativhypothese angenommen. In den Gruppen der Probanden mit niedriger, hoher sowie sehr hoher moralischer Urteilsfähigkeit muss die Nullhypothese beibehalten werden (siehe Tab. 74).

Moralisches Urteil	μ UN		Wilcoxon		Hypothese
	HO	FB	Z	Sign.	
niedrig (n = 18)	4,22	3,33	-1,237	0,216	H1E ₁
mittel (n = 46)	4,02	3,09	-2,798	0,005	H1E ₁
hoch (n = 16)	4,06	3,63	-0,783	0,433	H1E ₁
sehr hoch (n = 9)	3,78	2,56	-1,535	0,125	H1E ₁

Tab. 74: Prüfung der H1E₁ innerhalb der MU-Klassen

In der Faking-bad-Bedingung waren es die Probanden mit mittlerer moralischer Urteilsfähigkeit, welche am besten ihre Antworten verfälschten.

Will man einen guten Eindruck erzeugen, wird man versuchen sich extravertiert darzustellen, so besagt es die **Hypothese H2A₁**. Der tatsächliche Score ist bei allen vier Gruppierungen durchwegs niedriger als der verfälschte Wert (siehe Tab. 75). Die berechneten t-Tests für abhängige Stichproben erbrachten bei den Probanden mit niedriger und mittlerer moralischer Urteilsfähigkeit signifikante Ergebnisse. Hingegen für die Gruppen der Probanden mit hoher und sehr hoher moralischer Urteilsfähigkeit waren die Ergebnisse der t-Tests nicht signifikant. Somit kann die Nullhypothese zugunsten der Alternativhypothese verworfen werden für Personen mit niedriger und mittlerer moralischer Urteilsfähigkeit, die Nullhypothese wird beibehalten bei Personen, deren moralische Urteilsfähigkeit hoch bzw. sehr hoch ist.

Moralisches Urteil	μ E		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
niedrig (n = 13)	4,54	6,46	-2,775	12	0,017	H2A ₁
mittel (n = 44)	4,64	6,39	-4,973	43	0,000	H2A ₁
hoch (n = 15)	5,40	5,93	-0,673	14	0,512	H2A ₁
sehr hoch (n = 9)	4,67	6,67	-2,155	8	0,063	H2A ₁

Tab. 75: Prüfung der H2A₁ innerhalb der MU-Klassen

Die Probanden werden laut **Hypothese H2B₁** bemüht sein in der Faking-good-Bedingung wenig ängstlich zu erscheinen. Bei allen vier Gruppen zeigten sich in der ehrlichen Bedingung höhere Werte für Ängstlichkeit als in der gefakten Bedingung. Die Wilcoxon-Tests, welche zur Berechnung herangezogen wurden, sprechen für die Annahme der Alternativhypothese. Eine Übersicht hierzu liefert die Tabelle 76.

Moralisches Urteil	μ AE		Wilcoxon		Hypothese
	HO	FG	Z	Sign.	
niedrig (n = 13)	5,69	3,38	-2,919	0,004	H2B ₁
mittel (n = 44)	5,68	2,41	-5,353	0,000	H2B ₁
hoch (n = 15)	5,67	2,80	-2,108	0,035	H2B ₁
sehr hoch (n = 9)	5,33	2,00	-2,527	0,012	H2B ₁

Tab. 76: Prüfung der H2B₁ innerhalb der MU-Klassen

Verfälscht man in eine positive Richtung, so wird man gemäß **Hypothese H2C₁** bemüht sein selbstkontrollierter zu erscheinen. Alle vier Gruppen waren dazu imstande sich beim faking-good selbstkontrollierter darzustellen. Die Ergebnisse der Wilcoxon-Tests sprechen bei allen Gruppen für die Annahme der Alternativhypothese (siehe Tab. 77).

Moralisches Urteil	μ SK		Wilcoxon		Hypothese
	HO	FG	Z	Sign.	
niedrig (n = 13)	5,15	8,69	-2,910	0,004	H2C ₁
mittel (n = 44)	5,32	8,68	-5,299	0,000	H2C ₁
hoch (n = 15)	5,40	8,47	-2,308	0,021	H2C ₁
sehr hoch (n = 9)	5,11	8,78	-2,401	0,016	H2C ₁

Tab. 77: Prüfung der H2C₁ innerhalb der MU-Klassen

In der Faking-good-Bedingung werden sich die Probanden unabhängiger darstellen, so besagt es die **Hypothese H2D₁**. Der Skalenwert der Honest-Bedingung ist erwartungsgemäß bei jeder Gruppe niedriger als der Skalenwert der Faking-Bedingung (dargestellt in Tab. 78). Die berechneten t-Tests fielen für Personen mit mittlerer und sehr hoher moralischer Urteilsfähigkeit hoch signifikant aus. Somit kann für diese Probanden die Nullhypothese zugunsten der Alternativhypothese verworfen werden. Im Falle der Probanden, deren moralische Urteilsfähigkeit niedrig bzw. hoch ist, muss aufgrund nicht signifikanter t-Tests die Nullhypothese beibehalten werden.

Moralisches Urteil	μ UA		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
niedrig (n = 13)	6,46	7,62	-1,865	12	0,087	H2D ₁
mittel (n = 44)	5,27	7,98	-7,252	43	0,000	H2D ₁
hoch (n = 15)	6,53	8,07	-1,384	14	0,188	H2D ₁
sehr hoch (n = 9)	4,44	8,78	-7,506	8	0,000	H2D ₁

Tab. 78: Prüfung der H2D₁ innerhalb der MU-Klassen

Gemäß der **Hypothese H2E₁** stellen sich die Versuchspersonen in der Faking-good-Bedingung unnachgiebiger dar. Der Skalenwert der Honest-Bedingung ist wie erwartet bei allen vier Gruppen niedriger als der Skalenwert der Faking-Bedingung. Die Ergebnisse der t-Tests für abhängige Stichproben sprechen sowohl bei der Gruppe mit mittlerer moralischer Urteilsfähigkeit sowie bei der Gruppe mit sehr hoher moralischer Urteilsfähigkeit für die Annahme der Alternativhypothese. Für die Gruppen, deren

moralische Urteilsfähigkeit niedrig bzw. hoch ist, muss aufgrund der Ergebnisse der t-Tests die Nullhypothese beibehalten werden (eine Übersicht bietet Tab. 79).

Moralisches Urteil	μ UN		t-Test abhängig			Hypothese
	HO	FG	T	df	Sign.	
niedrig (n = 13)	4,77	5,38	-0,682	12	0,508	H2E ₁
mittel (n = 44)	4,48	5,70	-3,434	43	0,001	H2E ₁
hoch (n = 15)	3,47	4,80	-1,281	14	0,221	H2E ₁
sehr hoch (n = 9)	3,67	6,00	-3,055	8	0,016	H2E ₁

Tab. 79: Prüfung der H2E₁ innerhalb der MU-Klassen

Auch in der Faking-good-Bedingung waren es somit die Probanden mit mittlerer moralischer Urteilsfähigkeit, welche am besten ihre Antworten verfälschten.

4.4.8. Positive Selbstdarstellung

Personen mit wenig Tendenz zu positiver Selbstdarstellung und Personen mit mittlerer Tendenz zu positiver Selbstdarstellung haben ihre Werte in der Faking-bad-Bedingung dahingehend verfälscht, dass sie introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher erscheinen. Für die Globalskalen Extraversion (H1A₁), Ängstlichkeit (H1B₁) und Selbstkontrolle (H1C₁) lieferten die berechneten Wilcoxon-Tests hoch signifikante Ergebnisse, was für die Annahme der Alternativhypothese spricht. Die berechneten t-Tests für abhängige Stichproben waren bei der Skala Unabhängigkeit (H1D₁) bei beiden Gruppen nicht signifikant, somit muss die Nullhypothese beibehalten werden. Personen, die wenig dazu tendieren sich positiv darzustellen, waren fähiger im Verfälschen der Skala Unnachgiebigkeit (H1E₁), hier fiel der berechnete t-Test signifikant aus. Die Alternativhypothese wird für diese Gruppe angenommen, während bei Personen mit mittlerer Tendenz zu positiver Selbstdarstellung die Nullhypothese gültig ist. Die Tabellen 80 und 81 bieten hierzu eine genauere Übersicht.

Hypothesenprüfung für „geringe positive Selbstdarstellung“ (n = 52)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
4,48	2,67	-4,074	0,000				H1A ₁
5,71	7,83	-5,202	0,000				H1B ₁
4,79	1,85	-5,767	0,000				H1C ₁
5,62	6,33			-1,742	51	0,088	H1D ₁
3,98	2,90	-2,913	0,004				H1E ₁

Tab. 80: Ergebnisse der Probanden mit geringer IM-Tendenz (FB)

Hypothesenprüfung für „mittlere positive Selbstdarstellung“ (n = 37)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FB	Z	Sign.	T	df	Sign.	
5,41	2,27	-4,807	0,000				H1A ₁
5,73	7,81	-4,234	0,000				H1B ₁
4,73	2,38	-4,190	0,000				H1C ₁
5,38	6,32			-2,014	36	0,052	H1D ₁
4,14	3,57	-1,707	0,088				H1E ₁

Tab. 81: Ergebnisse der Probanden mit mittlerer IM-Tendenz (FB)

Sowohl Personen, die wenig zu positiver Selbstdarstellung tendieren, als auch Personen mit mittlerer Tendenz zu positiver Selbstdarstellung haben ihre Testwerte in die erwartete Richtung verfälschen können. Die Werte der Honest-Bedingung unterscheiden sich von jenen der Faking-good-Bedingung bei allen fünf Globalskalen signifikant. Die Probanden stellten sich gemäß der Alternativhypothesen in der Faking-good-Bedingung extravertierter (H2A₁), gelassener (H2B₁), selbstkontrollierter (H2C₁), unabhängiger (H2D₁) sowie unnachgiebiger (H2E₁) dar. Somit können in der Faking-good-Bedingung alle Alternativhypothesen angenommen werden. Eine genaue Darstellung der Ergebnisse findet sich in Tabelle 82 und Tabelle 83.

Hypothesenprüfung für „geringe positive Selbstdarstellung“ (n = 49)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FG	Z	Sign.	T	df	Sign.	
4,45	6,24	-4,190	0,000				H2A ₁
5,67	2,84	-5,643	0,000				H2B ₁
5,18	8,51	-5,507	0,000				H2C ₁
5,59	7,94			-5,741	48	0,000	H2D ₁
4,37	5,31			-2,241	48	0,030	H2E ₁

Tab. 82: Ergebnisse der Probanden mit geringer IM-Tendenz (FG)

Hypothesenprüfung für „mittlere positive Selbstdarstellung“ (n = 32)							
μ		Wilcoxon		t-Test abhängig			Hypothesen
HO	FG	Z	Sign.	T	df	Sign.	
5,25	6,50	-2,680	0,007				H2A ₁
5,59	2,22	-3,767	0,000				H2B ₁
5,44	8,88	-4,049	0,000				H2C ₁
5,63	8,16			-4,765	31	0,000	H2D ₁
4,06	5,84			-3,720	31	0,001	H2E ₁

Tab. 83: Ergebnisse der Probanden mit mittlerer IM-Tendenz (FG)

5. Diskussion

5.1. Untersuchungsergebnisse

In der vorliegenden Arbeit wurde die Fakeability aus differentialpsychologischer Sicht untersucht. Die Untersuchung stützt sich auf ein *Within-Subjects-Design*. Sie ist so angelegt, dass ein und dieselbe Versuchsperson denselben Fragebogen (16 PF-R) mehrmals beantworten muss, einmal wahrheitsgetreu sowie unter einer faking-good- und/oder faking-bad-Instruktion. Die Honest-Instruktion ist die gewöhnliche Standartinstruktion des Fragebogens.

An der vorliegenden Untersuchung nahmen 145 Personen teil. Der Großteil der Probanden bestand aus Schülern, deren Lehrer für die Durchführung je zwei Schulstunden zur Verfügung stellten. Der Rest der Stichprobe setzte sich vorwiegend aus dem Bekanntenkreis der Autorin zusammen. Die Testdurchführung erfolgte in Papier-Bleistift-Form und dauerte mindestens 90 Minuten. Jeder Proband der Bekanntenstichprobe musste den 16 PF-R dreimal ausfüllen, jeder Proband der Schülerstichprobe zumindest zweimal, unter jeweils verschiedenen schriftlichen Instruktionsbedingungen (honest, fake good, fake bad). Es gab zehn verschiedene Reihenfolgen der Vorgabe, um eventuelle Testeffekte auszubalancieren. Die Probanden wurden zufällig einer der Vorgabereihenfolgen zugeteilt.

Als Grundlage zur Erhebung der Persönlichkeit dienen die Messung des Extra- und Introversionsgrades mittels 16 PF-R, die Skalenwerte der Probanden auf der Primärskala „Emotionale Stabilität“ des 16 PF-R, die Messung von Repression versus Sensitization mittels RSCI, die Erfassung der moralischen Urteilsfähigkeit mittels MUT sowie die Messung der Tendenz zur positiven Selbstdarstellung durch die IMS.

Nach Ausschluss von unvollständigen und auffälligen Datensätzen liegen insgesamt 112 vollständige Datensätze für die Auswertung der Fragestellung vor. Es erfolgte eine Trennung der Gesamtstichprobe in die beiden Faking-Bedingungen. Die Berechnungen erfolgten jeweils für die Faking-bad-Stichprobe ($n = 89$) und die Faking-good-Stichprobe ($n = 81$) getrennt.

5.1.1. Differentialpsychologische Aspekte

Beim Verfälschen in eine positive Richtung waren *Männer und Frauen* gleichermaßen erfolgreich. In der *Faking-good-Stichprobe* unterscheiden sich die verfälschten Werte von den tatsächlichen Werten bei allen fünf Globalskalen signifikant. Den Probanden ist es somit geglückt sich beim Verfälschen der Antworten extravertierter, gelassener, selbstkontrollierter, unabhängiger sowie unnachgiebiger darzustellen.

Geschlechtsspezifische Unterschiede bezüglich der Fakeability gab es nur *in der Faking-bad-Stichprobe*. Sieht man sich die Mittelwerte genauer an, so fällt auf, dass Männer bei der Skala Unabhängigkeit in die andere Richtung verfälscht haben. Beim Vergleich des „ehrlichen“ Testwertes der Männer ($\mu_{UA\ HO} = 5.73$) mit dem verfälschten Testwert ($\mu_{UA\ FB} = 5.02$) ist zu erkennen, sie stellten sich anpassungsbereiter dar. Auch die Differenz zwischen den Testwerten der männlichen Probanden bei der Skala Unnachgiebigkeit in der Faking-bad-Bedingung fällt gerade noch signifikant aus (Wilcoxon-Test: Sign. = 0,049 für $Z = -1,973$).

Der Versuch sich selbst schlechter darzustellen dürfte den *Männern* dieser Stichprobe größere *Probleme* bereitet haben als den Frauen. Ein möglicher Grund für die Verfälschung des Testwertes auf der Skala Unabhängigkeit in die andere Richtung, könnte jedoch auch der sein, dass die männlichen Teilnehmer Anpassungsbereitschaft für eine Schwäche halten und sich deshalb beim faking-bad anpassungsbereiter dargestellt haben.

In der Faking-bad-Stichprobe hatte jede **Altersklasse** ein einziges Mal kein signifikantes Ergebnis. Bei den 20 bis 29jährigen betrifft das die Skala Extraversion. In der Gruppe der 30 bis 39jährigen wie auch in der Altersklasse der 40 bis 56jährigen erbrachte ein Vergleich der Mittelwerte für die Skala Unabhängigkeit nicht das erwartete Ergebnis. Beide Altersklassen verfälschten hier in die falsche Richtung, sie stellten sich anpassungsbereiter dar, im Falle der 30 bis 39jährigen sogar signifikant. Bei der Altersklasse der 17 bis 19jährigen wiederum ist der verfälschte Wert für Unnachgiebigkeit nur geringfügig kleiner als der ehrliche Wert.

Bezüglich des Zusammenhangs zwischen Fakeability und dem Alter kann man die Aussage treffen, dass *beim faking-bad* die Fähigkeit zur Verfälschung *ziemlich ausgewogen* ist.

In der Faking-good-Bedingung schien es, als hätte die Altersklasse der 30 bis 39jährigen die größeren Schwierigkeiten erfolgreich in die vorgegebene Richtung zu verfälschen. Für diese Altersklassen musste drei Mal die Nullhypothese beibehalten werden. Die Altersklasse hatte ihre Werte in der Faking-Situation zwar in die erwartete Richtung verfälscht, jedoch nicht signifikant. Betroffen sind die Skalen Extraversion ($\mu_E\ HO = 5.31$; $\mu_E\ FG = 7.00$), Unabhängigkeit ($\mu_{UA}\ HO = 7.38$; $\mu_{UA}\ FG = 8.54$) sowie Unnachgiebigkeit ($\mu_{UN}\ HO = 4.46$; $\mu_{UN}\ FG = 4.85$).

In der *Faking-good-Stichprobe* war somit die *Altersklasse der 30 bis 39jährigen weniger effektiv* im Verfälschen als die übrigen Altersklassen.

In der Faking-bad-Stichprobe fiel das Ergebnis für die Skala Unabhängigkeit nicht signifikant aus für die **Bildungsklassen** „Berufsausbildung“ ($\mu_{UA}\ HO = 6.10$; $\mu_{UA}\ FB = 6.30$) und „Reifeprüfung“ ($\mu_{UA}\ HO = 4.88$; $\mu_{UA}\ FB = 5.71$). Die Verfälschung erfolgte zwar in die erwartete Richtung, jedoch war der Unterschied zwischen den Mittelwerten zu gering. Bei der Skala Unnachgiebigkeit wiederum erreichte die Bildungsgruppe „Pflichtschule“ kein signifikantes Ergebnis, wobei hier auffällt, dass scheinbar keine

Verfälschung stattfand. Der ehrliche und der verfälschte Mittelwert nehmen denselben Wert an ($\mu = 3.64$).

Alle Bildungsklassen verfälschten in der Faking-good-Stichprobe in die erwartete Richtung. Bei der Bildungsklasse „Pflichtschule“ war der Unterschied zwischen dem ehrlichen und dem verfälschten Wert nicht signifikant betreffend die Skala Extraversion. Die Bildungsklasse „Berufsausbildung“ erzielte kein signifikantes Ergebnis bei der Verfälschung der Skala Unnachgiebigkeit.

Zusammenfassend sprechen die Ergebnisse dafür, dass es *keine bildungsspezifischen Unterschiede* hinsichtlich der Fakeability gibt.

Was die Faking-bad-Stichprobe betrifft, so gehen die Mittelwerte sowohl bei **Intro-** als auch bei **Extravertierten** in die erwartete Richtung. Es gelang beiden Gruppen dahingehend zu verfälschen, dass sie introvertierter, ängstlicher, unbeherrschter und empfänglicher wirkten. Nur für die Verfälschung der Skala Unabhängigkeit musste bei extravertierten ($\mu E HO = 5.44$; $\mu E FB = 6.17$) und introvertierten Personen ($\mu E HO = 5.57$; $\mu E FB = 6.43$) die Nullhypothese beibehalten werden.

Bei der Gruppe der Introvertierten konnten in der *Faking-good-Stichprobe* alle Alternativhypothesen angenommen werden, hier unterscheiden sich die verfälschten Werte von den Honest-Werten bei den fünf Globalskalen signifikant. Im Falle der Extravertierten trifft dies nur für vier Globalskalen zu. Einzig bei der Skala Extraversion gelang es ihnen nicht ihre Extraversions-Werte beim Verfälschen zu erhöhen ($\mu E HO = 6.84$; $\mu E FB = 6.34$).

Die Ergebnisse sprechen insgesamt dafür, dass sich **Extravertierte** nicht von **Introvertierten** hinsichtlich ihrer Fakeability unterscheiden.

Der Gruppe der **Emotional Instabilen** ist es in der *Faking-bad-Bedingung* gelungen signifikant in die von der Autorin erwarteten Richtungen zu verfälschen. Auch bei der Gruppe der **Emotional Stabilen** weisen die Mittelwerte in die vorab angenommenen Richtungen. Einzig der t-Test für abhängige Stichproben erbrachte für die Skala Unabhängigkeit ($\mu UA HO = 5.67$; $\mu UA FB = 6.04$) kein signifikantes Ergebnis.

In der *Faking-good-Stichprobe* ist es sowohl emotional stabilen wie auch emotional instabilen Personen gelungen in die erwartete Richtung zu verfälschen. Die Werte der Faking-Bedingung unterscheiden sich von den Werten der Honest-Bedingung bei beiden Gruppen für alle Globalskalen signifikant.

Die Ergebnisse lassen darauf schließen, dass sich **emotional stabile** und **emotional instabile** Personen bezüglich ihrer Fakeability nicht voneinander unterscheiden.

In der Faking-bad-Stichprobe stellten sich **Represser** und **Sensitizer** im Vergleich zur Honest-Bedingung erwartungsgemäß introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher dar. Ein

Vergleich der Mittelwerte fiel nur bei der Skala Selbstkontrolle für beide Gruppen signifikant aus. Bei den Skalen Extraversion und Unnachgiebigkeit konnten keine signifikanten Unterschiede der Mittelwerte festgestellt werden. Nur die Represser konnten auf der Skala Ängstlichkeit ihren Wert in der Faking-Bedingung signifikant erhöhen, bei der Skala Unabhängigkeit waren es die Sensitizer, deren Faking-Werte signifikant höher waren als deren Honest-Werte.

In der Faking-good-Bedingung ist es sowohl Repressern wie Sensitizern gelungen sich signifikant extravertierter, gelassener, selbstkontrollierter und unabhängiger darzustellen. Represser weisen zwar in der Faking-Bedingung höhere Testwerte auf der Skala Unnachgiebigkeit auf, jedoch nicht signifikant. Sensitizer stellten sich geringfügig empfänglicher dar, d. h. sie verfälschten ihre Werte hier in eine andere Richtung.

Represser scheinen somit nicht fähiger im Verfälschen von Persönlichkeitsfragebogen zu sein als **Sensitizer**. Die Ergebnisse sind hier jedoch mit Vorsicht zu interpretieren, da die Stichprobengrößen gering sind, mit einem Minimum von $n = 7$ und einem Maximum von $n = 14$.

Die Autorin ist davon ausgegangen, dass sich Personen beim faking-bad in signifikanter Weise ängstlich und unbeherrscht darstellen. Dies konnte für alle vier Gruppen, die nach der Höhe ihrer **moralischen Urteilsfähigkeit** eingestuft worden sind, bestätigt werden. Ebenso stellten sich alle vier Gruppen introvertiert dar, einzig im Falle der Gruppe mit sehr hoher moralischer Urteilsfähigkeit war das Ergebnis nicht signifikant. Obwohl alle vier Gruppen in die Richtung verfälschten, dass sie einen unabhängigen und empfänglichen Eindruck hinterlassen, gelang dies nur der Gruppe mit mittlerer moralischer Urteilsfähigkeit signifikant.

Die vier Gruppen zur moralischen Urteilsfähigkeit in der Faking-good-Stichprobe konnten sich alle signifikant gelassener und selbstkontrollierter darstellen. Ein Vergleich der Mittelwerte bei der Skala Extraversion brachte für die Probanden mit niedriger bzw. mittlerer moralischer Urteilsfähigkeit ein signifikantes Ergebnis, sie gaben sich in der Faking-Situation extravertierter. Die Stichprobe mit mittlerer sowie die Stichprobe mit sehr hoher moralischer Urteilsfähigkeit schaffte es sich signifikant unabhängiger und unnachgiebiger darzustellen.

Einzig Probanden mit *mittlerer moralischer Urteilsfähigkeit* gelang es in der Faking-good- und der Faking-bad-Bedingung alle Testwerte signifikant zu verfälschen. Für diese Personengruppe konnten die Annahmen aller Alternativhypothesen bestätigt werden. Bei Personen mit mittlerer moralischer Urteilsfähigkeit scheint es sich demnach um die besseren Faker zu handeln. Die Autorin möchte an dieser Stelle jedoch darauf hinweisen, dass die Stichproben dieser Gruppe weitaus größer waren ($n = 46$ in der Faking-bad-Stichprobe bzw. $n = 44$ in der Faking-good-Stichprobe) als die der niedrigen, hohen und sehr hohen moralischen Urteilsfähigkeit. Man sollte die Ergebnisse also mit Vorsicht interpretieren.

Für die Untersuchung der **Tendenz zur positiven Selbstdarstellung** im Zusammenhang mit der Fakeability konnten beide Gruppen ihre Werte in der Faking-bad-Stichprobe dahingehend verfälschen, dass sie introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher erscheinen. Ein

Vergleich der Mittelwerte war bei der Skala Unabhängigkeit ($\mu_{UA} HO = 5.62$; $\mu_{UA} FB = 6.33$) für die Personengruppe mit geringen IM-Tendenzen nicht signifikant. Sie waren allerdings fähiger im Verfälschen der Skala Unnachgiebigkeit, hier fiel der berechnete t-Test signifikant aus. Bei Personen mit mittlerer Tendenz zu positiver Selbstdarstellung musste die Nullhypothese für die Skalen Unabhängigkeit ($\mu_{UA} HO = 5.38$; $\mu_{UA} FB = 6.32$) und Unnachgiebigkeit ($\mu_{UN} HO = 4.14$; $\mu_{UN} FB = 3.57$) beibehalten werden.

Sowohl Personen, die wenig zu positiver Selbstdarstellung tendieren, als auch Personen mit mittlerer Tendenz zu positiver Selbstdarstellung haben ihre Testwerte in der Faking-good-Stichprobe in die erwartete Richtung verfälschen können. Die Werte der Honest-Bedingung unterscheiden sich von jenen der Faking-good-Bedingung bei allen fünf Globalskalen signifikant.

Die Ergebnisse sprechen nicht dafür, dass Personen mit mittlerer Tendenz zu positiver Selbstdarstellung fähiger im Verfälschen sind als Personen, die nur eine geringe Tendenz zur positiven Selbstdarstellung aufweisen.

5.1.2. Allgemeine Fakeability

Es stellte sich natürlich auch die Frage, ob sich die ermittelten Werte der Faking-Instruktionen von jenen in der Honest-Bedingung über alle Versuchspersonen hinweg unterscheiden.

Hier zeigte sich, dass sich die verfälschten Werte in der Faking-bad-Stichprobe signifikant von den True-Scores bei allen fünf Globalskalen unterscheiden. Wie es die Autorin erwartet hatte, ist es den Probanden beim Verfälschen gelungen sich introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher darzustellen.

In der Faking-good-Stichprobe zeigte sich dasselbe Bild. Auch hier unterscheiden sich bei allen fünf Globalskalen die verfälschten Werte von den tatsächlichen Werten signifikant. Den Probanden ist es somit geglückt sich beim Verfälschen der Antworten extravertierter, gelassener, selbstkontrollierter, unabhängiger sowie unnachgiebiger darzustellen.

Die Ergebnisse sprechen für die Annahme aller Alternativhypothesen bei den beiden Gesamtstichproben. Somit kann eine allgemeine Fakeability als bestätigt angesehen werden. Die Probanden waren sowohl beim faking-bad als auch beim faking-good sehr gut dazu in der Lage ihr Testprofil in die instruierte Richtung abzuändern.

Jede der Alternativhypothesen wurde 22-mal auf ihre Gültigkeit geprüft. Nachfolgend sind die Hypothesen nach der Anzahl der signifikanten Ergebnisse in absteigender Reihenfolge gereiht: In allen Fällen fiel die Prüfung der Hypothesen $H1C_1$, $H2B_1$ und $H2C_1$ signifikant aus. Die Prüfung der $H1B_1$ ergab nur einmal kein signifikantes Ergebnis. In 19 Fällen kann die $H2D_1$ aufgrund signifikanter Ergebnisse angenommen werden. Für die $H1A_1$ fiel die Prüfung insgesamt 18-mal signifikant aus. Die Prüfung der $H2A_1$ erbrachte 16 signifikante Ergebnisse, die Prüfung der $H2E_1$ 15 signifikante Ergebnisse

und die Prüfung der $H1E_1$ schließlich war 14-mal signifikant. Lediglich die $H1D_1$ war nur in acht Fällen signifikant.

Die größten Schwierigkeiten hatten die Probanden demzufolge in der Faking-bad-Bedingung bei der Skala Unabhängigkeit. Ein möglicher Grund dafür könnte sein, dass die Probanden Anpassungsbereitschaft für eine Schwäche halten und sich deshalb beim faking-bad anpassungsbereiter darstellten.

Zusammenfassend lässt sich noch sagen, dass es den Probanden beim Faken meist gelungen ist ihre Testwerte in die erwarteten Richtungen zu ändern, wenn auch nicht immer signifikant.

5.2. Abschließende Bemerkungen

Das Untersuchungsdesign wird im Folgenden genau unter die Lupe genommen. Kritikpunkte werden angeführt. Danach folgen Anregungen für zukünftige Untersuchungen.

5.2.1. Kritikpunkte

Ein größerer Pool an Probanden wäre bestimmt von Vorteil gewesen. Außerdem wäre eine homogenere Stichprobe empfehlenswert. Dadurch könnten Alters- und Bildungsunterschiede konstant gehalten werden und eine bessere Vergleichbarkeit wäre gegeben.

Die Untersuchungsdauer von mindestens 90 Minuten könnte sich bei einigen Versuchspersonen hinsichtlich der Motivation zur Mitarbeit negativ ausgewirkt haben. Die einmalige Bearbeitung des 16 PF-R nimmt schon 30 bis 45 Minuten in Anspruch.

Dennoch spricht alles für die Vorgabe des 16 PF-R, da dieser Fragebogen eine IM-Skala, eine Akquieszenzskala sowie eine Infrequenzskala beinhaltet. Das ermöglichte es auffällige Datensätze nicht in die Berechnungen mit einzubeziehen. Außerdem wurden mit den drei übrigen Fragebogen Verfahren gefunden, deren Bearbeitung nicht allzu lange in Anspruch nimmt.

Es ist möglich, dass die Probanden nicht genügend motiviert waren ihre Antworten in die vorgegebenen Richtungen zu verfälschen, da sie sich für die Arbeitsplätze im realen Leben nicht interessieren. Kluger, Reilly und Russell (1991) weisen darauf hin, dass in einer Felduntersuchung die Motivation zur Verfälschung ein sehr wichtiger Faktor ist. Laut Corr und Gray (1995) sind bei Durchschaubarkeit des Messprinzips nur dann verfälschte Antworten zu erwarten, wenn die Probanden auch dazu motiviert sind.

Die Repräsentativität dieser Stichprobe für die Gesamtbevölkerung erscheint insofern problematisch, da die Versuchspersonen der Studie hauptsächlich Schüler im Alter von 17 bis 19 Jahren waren.

5.2.2. Anregungen für zukünftige Untersuchungen

Viswesvaran und Ones (1999) sehen einen Weg für fruchtbare zukünftige Forschung beim Nachforschen, ob einzelne Unterschiede bezüglich der Fakeability im Zusammenhang stehen mit bestimmten Eigenschaften der Personen. Angenommen, dass Fakeability soziale Intelligenz oder eine Form der Anpassungsfähigkeit widerspiegelt, so könnten einzelne Unterschiede bezüglich der Fakeability zur Vorhersage einer erfolgreichen Jobleistung beitragen.

6. Zusammenfassung

Die Erfassung und Messung der Persönlichkeit sind schon seit jeher ein großes Interessens- und Forschungsgebiet der Psychologie. Die Differentielle Psychologie befasst sich mit der Art des Erlebens und Verhaltens des Individuums sowie den Ursachen und Bedingungen von Verhaltensunterschieden zwischen den Menschen (Amelang & Bartussek, 2001). Die Persönlichkeitspsychologie hat sich aus der Differentiellen Psychologie und aus der Charakterologie heraus entwickelt. Persönlichkeitspsychologie wird sowohl als Teil der Differentiellen Psychologie als auch als eigenständige Grunddisziplin der Psychologie verstanden. Ausgangspunkt der Persönlichkeitspsychologie ist der einzelne Mensch, das Individuum, die Person.

Die Erfassung von Persönlichkeitseigenschaften erfolgt zumeist mittels Persönlichkeitsfragebogen. Die Frage, ob Individuen ihre Antworten in einem Persönlichkeitsfragebogen verfälschen können, beschäftigt Psychologen seit vielen Jahren. Man hat festgestellt, dass Persönlichkeitsfragebogen sensitiv sind gegenüber einer absichtlichen Verfälschung der Antworten durch die befragten Personen. Die Verfälschung kann sowohl in eine positive Richtung (*fake good*) als auch in eine negative Richtung (*fake bad*) gehen (Amelang & Bartussek, 2001). Dieses Phänomen ist unter einer Vielzahl von Namen einschließlich *Response distortion*, *Impression Management*, *soziale Erwünschtheit* sowie *Selbstdarstellung* untersucht worden (Rosse, Stechner, Levin & Miller, 1998; Viswesvaran & Ones, 1999).

Eine Möglichkeit zur Erfassung der Fakeability beinhaltet das Vergleichen der Antworten von ein und derselben Person unter unterschiedlichen Instruktionsbedingungen. Hier wird auf die Bestimmung der maximalen Grenzen der Fakeability von verschiedenen Persönlichkeitsskalen gezielt (*Within-Subjects-Design*). Ein anderer Ansatz schließt das Vergleichen der, von verschiedenen Gruppen, erhaltenen Antworten mit ein. Man spricht dann von einem *Between-Subjects-Design*. Hierzu gehört auch die Erforschung der Fakeability in tatsächlichen Bewerbungssituationen. Die vorliegende Arbeit konzentriert sich nur auf die Quantitätsbestimmung der Effekte von Instruktionsbedingungen in einem Within-Subjects-Design (Viswesvaran & Ones, 1999).

Unterschiedliche Instruktionsbedingungen sind in Untersuchungen, welche das Thema Verfälschung zum Inhalt hatten, benutzt worden, um festzustellen, ob Personen ihre Antworten auf einen Persönlichkeitsfragebogen verfälschen können. Die drei allgemein verglichenen Fragebogen-Instruktionen sind: *a) respond honestly* (ehrlich antworten), *b) fake good* oder so zu antworten, dass man einen möglichst *vorteilhaften Eindruck* hinterlässt *c) fake bad* oder so zu antworten, dass man um einen *ungünstigen Eindruck* macht (Viswesvaran & Ones, 1999).

Die Fehlerfaktoren im Persönlichkeitsfragebogen (Response-Sets) werden von Amelang und Bartussek (2001) unterteilt in: Absichtliche Verstellung (Verfälschung), soziale Erwünschtheit, Akquieszenz und weitere Fehlerfaktoren (bspw. Bevorzugung extremer Antwortkategorien)

Das Testkuratorium (1986, S. 359) definiert das Gütekriterium Verfälschbarkeit als das „Ausmaß, in dem ein Test die individuelle Kontrolle über Art und Inhalt der verlangten bzw. gelieferten Informationen ermöglicht.“

Persönlichkeitsfragebogen sind wie eingangs schon erwähnt also „sensitiv gegenüber einer absichtlichen Verfälschung in *jeder* Richtung. Instruktionen zum *faking good* bzw. *faking bad* haben Resultate zur Folge, die sich nicht nur voneinander, sondern auch jeweils von den unter Normalinstruktion erhaltenen Mittelwerten unterscheiden“ (Amelang & Bartussek, 2001, S. 171). Das Ausmaß der Verfälschung hängt jedoch vom Anlass der Testdurchführung ab (Häcker & Stapf, 1998). Karner (2002) sieht Probleme bei der Verwendung von Persönlichkeitsfragebogen, wenn die dabei erhaltenen Scores wichtige Konsequenzen für die Beantwortenden haben.

Darüber hinaus stellt anscheinend auch die instruktionsgemäße Übernahme verschiedener Rollen für die Probanden keinerlei Probleme dar z. B. wenn sie dazu aufgefordert werden, „die Fragen so zu beantworten, wie dieses typische Angehörige einzelner Berufsgruppen ... vermutlich täten, und es zweckmäßig wäre, um eine in diesen Bereichen ausgeschriebene Stelle zu erhalten“ (Amelang & Bartussek, 2001, S. 171).

Die zentrale Fragestellung der Untersuchung lautet: Besteht ein Zusammenhang zwischen der Fähigkeit einer Person zur Verfälschung von Persönlichkeitsfragebogen und spezifischen Merkmalen dieser Person? Die Untersuchung soll klären inwieweit demographische Daten wie Alter, Geschlecht und Bildung sowie Persönlichkeitseigenschaften wie Extra- und Introversion, Emotionale Stabilität vs. Emotionale Instabilität, Repression vs. Sensitization, moralische Urteilsfähigkeit und die Tendenz zur positiven Selbstdarstellung mit der Fakeability in Zusammenhang stehen.

Die vorliegende Untersuchung stützt sich auf ein *Within-Subjects-Design*. Sie ist so angelegt, dass ein und dieselbe Versuchsperson denselben Fragebogen mehrmals beantworten muss, einmal wahrheitsgemäß sowie unter einer *faking-good* und/oder *faking-bad*-Instruktion. Die *Honest*-Instruktion ist die gewöhnliche Standartinstruktion des Fragebogens.

An der vorliegenden Untersuchung nahmen 145 Personen teil. Der Großteil der Probanden bestand aus Schülern, deren Lehrer für die Durchführung je 2 Schulstunden zur Verfügung stellten. Der Rest der Stichprobe setzte sich vorwiegend aus dem Bekanntenkreis der Autorin zusammen. Die Testdurchführung erfolgte in Papier-Bleistift-Form und dauerte mindestens 90 Minuten. Jeder Proband der Bekanntenstichprobe musste den 16 PF-R dreimal ausfüllen, jeder Proband der Schülerstichprobe zumindest zweimal, unter jeweils verschiedenen schriftlichen Instruktionsbedingungen (*honest*, *fake*

good, fake bad). Es gab zehn verschiedene Reihenfolgen der Vorgabe, um eventuelle Testeffekte auszubalancieren. Die Probanden wurden zufällig einer der Vorgabereihenfolgen zugeteilt.

Als Grundlage zur Erhebung der Persönlichkeit dienen die Messung des Extra- und Introversionsgrades mittels 16 PF-R, die Skalenwerte der Probanden auf der Primärskala „Emotionale Stabilität“ des 16 PF-R, die Messung von Repression versus Sensitization mittels RSCI, die Erfassung der moralischen Urteilsfähigkeit mittels MUT sowie die Messung der Tendenz zur positiven Selbstdarstellung durch die IMS.

Die Probanden hatten in der Faking-bad-Bedingung den 16 PF-R so ausfüllen, als ob sie vom AMS zu einer Bewerbung für eine Stelle als Küchengehilfin müssen und diesen Job nicht haben wollen. Es wird erwartet, dass sich die Probanden in der Faking-bad-Bedingung demnach introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher darstellen.

Die Versuchspersonen hatten in der Faking-good-Bedingung den 16 PF-R so auszufüllen, als ob sie sich für eine Stelle als Geschäftsführerin bewerben und dieser Job erstrebenswert für sie ist. Die Probanden sollten sich demnach in der Faking-good-Bedingung extravertierter, gelassener, selbstkontrollierter, unabhängiger sowie unnachgiebiger darstellen.

Nach Ausschluss von unvollständigen und auffälligen Datensätzen liegen insgesamt 112 vollständige Datensätze für die Auswertung der Fragestellung vor. Es erfolgte eine Trennung der Gesamtstichprobe in die beiden Faking-Bedingungen. Die Berechnungen erfolgten jeweils für die Faking-bad-Stichprobe ($n = 89$) und die Faking-good-Stichprobe ($n = 81$) getrennt.

Die *unabhängigen Variablen* sind demographische Daten wie das Alter (jung versus alt), das Geschlecht (männlich versus weiblich) sowie der Bildungsgrad (niedrige versus hohe Bildung) der Probanden. Die weiteren *unabhängigen Variablen* stellen die Persönlichkeitseigenschaften wie Extra- versus Introversion, Emotionale Stabilität versus Emotionale Instabilität, Repression versus Sensitization, niedrige versus hohe moralische Urteilsfähigkeit sowie niedrige versus hohe Tendenz zur positiven Selbstdarstellung dar. Als *abhängige Variablen* fungieren die Stenwerte auf den Globalskalen des 16 PF-R. Hier stellt der honest-Wert die Baseline dar und lässt somit einen Vergleich mit den gefakten Werten zu.

Beim Verfälschen in eine positive Richtung waren Männer und Frauen gleichermaßen erfolgreich. Der Versuch sich selbst schlechter darzustellen dürfte den Männern dieser Stichprobe allerdings größere Probleme bereitet haben als den Frauen.

Bezüglich des Zusammenhangs zwischen Fakeability und dem Alter kann die Aussage getroffen werden, dass beim faking-bad die Fähigkeit zur Verfälschung ziemlich ausgewogen ist. In der Faking-good-

Stichprobe war die Altersklasse der 30 bis 39jährigen weniger effektiv im Verfälschen als die übrigen Altersklassen (17 bis 19jährige, 20 bis 29jährige, 40 bis 56jährige).

Die Ergebnisse sprechen dafür, dass es keine bildungsspezifischen Unterschiede hinsichtlich der Fakeability gibt.

Insgesamt gesehen bescheinigen die Ergebnisse, dass sich Extravertierte nicht von Introvertierten hinsichtlich ihrer Fakeability unterscheiden.

Die Ergebnisse lassen darauf schließen, dass sich emotional stabile und emotional instabile Personen bezüglich ihrer Fakeability nicht voneinander unterscheiden.

Represser scheinen nicht fähiger im Verfälschen von Persönlichkeitsfragebogen zu sein als Sensitizer. Die Ergebnisse sind hier jedoch mit Vorsicht zu interpretieren, da die Stichprobengrößen gering sind, mit einem Minimum von $n = 7$ und einem Maximum von $n = 14$.

Einzig Probanden mit *mittlerer moralischer Urteilsfähigkeit* gelang es in der Faking-good- und der Faking-bad-Bedingung alle Testwerte signifikant zu verfälschen. Für diese Personengruppe konnten die Annahmen aller Alternativhypothesen bestätigt werden. Bei Personen mit mittlerer moralischer Urteilsfähigkeit scheint es sich demnach um die besseren Faker zu handeln. Die Autorin möchte an dieser Stelle jedoch darauf hinweisen, dass die Stichproben dieser Gruppe weitaus größer waren ($n = 46$ in der Faking-bad-Stichprobe bzw. $n = 44$ in der Faking-good-Stichprobe) als die der niedrigen, hohen und sehr hohen moralischen Urteilsfähigkeit. Man sollte die Ergebnisse also mit Vorsicht interpretieren.

Die Ergebnisse sprechen nicht dafür, dass Personen mit mittlerer Tendenz zu positiver Selbstdarstellung fähiger im Verfälschen sind als Personen, die nur eine geringe Tendenz zur positiven Selbstdarstellung aufweisen.

Es stellte sich natürlich auch die Frage, ob sich die ermittelten Werte der Faking-Instruktionen von jenen in der Honest-Bedingung über alle Versuchspersonen hinweg unterscheiden. Durch die Ergebnisse der vorliegenden Untersuchung konnte eine generelle Fakeability bestätigt werden. Die Probanden waren sowohl beim faking-bad als auch beim faking-good sehr gut dazu in der Lage ihr Testprofil in die instruierte Richtung abzuändern. Die Ergebnisse sprechen für die Annahme aller Alternativhypothesen bei den beiden Gesamtstichproben.

Es zeigte sich, dass sich die verfälschten Werte in der Faking-bad-Stichprobe signifikant von den True-Scores bei allen fünf Globalskalen unterscheiden. Wie es die Autorin erwartet hatte, ist es den Probanden beim Verfälschen gelungen sich introvertierter, ängstlicher, unbeherrschter, unabhängiger und empfänglicher darzustellen.

In der Faking-good-Stichprobe zeigte sich dasselbe Bild. Auch hier unterscheiden sich bei allen fünf Globalskalen die verfälschten Werte von den tatsächlichen Werten signifikant. Den Probanden ist es somit geglückt sich beim Verfälschen der Antworten extravertierter, gelassener, selbstkontrollierter, unabhängiger sowie unnachgiebiger darzustellen.

Die größten Schwierigkeiten hatten die Probanden in der Faking-bad-Bedingung bei der Skala Unabhängigkeit. Ein möglicher Grund dafür könnte sein, dass die Probanden Anpassungsbereitschaft für eine Schwäche halten und sich deshalb beim faking-bad – entgegen den Erwartungen – anpassungsbereiter darstellten.

Zusammenfassend lässt sich noch sagen, dass es den Probanden beim Faken meist gelungen ist ihre Testwerte in die erwarteten Richtungen zu ändern, wenn auch nicht immer signifikant.

Literaturverzeichnis

- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5. aktualisierte und erweiterte Auflage). Stuttgart: Kohlhammer.
- Amelang, M., Schäfer, A. & Yousfi, S. (2002). Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of fake good instructions. *Psychologische Beiträge*, 44, 24-41.
- Bijttebier, P., Vertommen, H. & Steene, G. V. (2001). Assessment of cognitive coping styles: A closer look at situation-response inventories. *Clinical Psychology Review*, 21 (1), 85-104.
- Corr, P. J. & Gray, J. A. (1995). Relationship between attributional style and Lie scores in an occupational sample motivated to fake good. *Personality and Individual Differences*, 18 (3), 433-435.)
- Dilchert, S., Ones, D. S., Viswesvaran, C. & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, 48 (3), 209-225.
- Elliot, S., Lawty-Jones, M. & Jackson, C. (1996). Effect of dissimulation on self-report and objective measures of personality. *Personality and Individual Differences*, 21 (3), 335-343.
- Eysenck, H. J. & Eysenck, S. B. G. (1970). *Personality Structure and Measurement* (2nd ed.). London: Routledge & Kegan.
- Eysenck, H. J. (1975). *Die Ungleichheit der Menschen*. München: Paul List Verlag KG.
- Eysenck, H. J. (1977). *Sexualität und Persönlichkeit*. Wien: Europaverlag.
- Eysenck, H.J. & Eysenck, M.W. (1987). *Persönlichkeit und Individualität. Ein naturwissenschaftliches Paradigma*. Weinheim: Psychologie Verlags Union.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400.
- Furnham, A. (1990). The fakeability of the 16 PF, Myers-Briggs and Firo-B personality measures. *Personality and Individual Differences*, 11, 711-716.
- Furnham, A., Petrides, K. V. & Bowdage, S. S. (2002). The effects of different types of social desirability on the identification of repressors. *Personality and Individual Differences*, 33, 119-130.

- Graham, K. E., McDaniel, M. A., Douglas, E. F. & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology*, 16 (4), 573-592.
- Griffith, R. L. & McDaniel, M. A. (in press). The nature of deception and applicant faking behavior. In R. L. Griffith (Hrsg.), *A closer examination of applicant faking behavior*. Greenwich, CT: Information Age Publishing.
- Habermas, J. (1996). *Moralbewußtsein und kommunikatives Handeln* (6. Aufl.). Frankfurt: Suhrkamp.
- Häcker, H., Schwenkmezger, P. & Utz, H. (1979). Über die Verfälschbarkeit von Persönlichkeitsfragebogen und Objektiven Persönlichkeitstests unter SD-Instruktion und in einer Auslesesituation. *Diagnostica*, 10 (1), 7-23.
- Häcker, H. & Stapf, K. H. (1998). *Dorsch Psychologisches Wörterbuch* (13. überarbeitete und erweiterte Aufl.). Bern: Verlag Hans Huber.
- Hall, C. S., Lindzey, G. & Campbell, J. B. (1997). *Theories of Personality* (fourth edition). New York: Wiley.
- Heidbrink, H. (1991). *Stufen der Moral: Zur Gültigkeit der kognitiven Entwicklungstheorie Lawrence Kohlbergs*. München: Quintessenz-Verlag.
- Herkner, W. (1991). *Lehrbuch Sozialpsychologie* (5., korr. Und stark erw. Aufl.). Bern: Verlag Hans Huber.
- Höfer, I., Wallbott, H. & Scherer, K. (1985). Messung multimodaler Streßindikatoren in Belastungssituationen: Person- und Situationsfaktoren. In H. W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen*. Weinheim: VCH.
- Hofmann, K. & Kubinger, K. D. (2001). Herkömmliche Persönlichkeitsfragebogen und Objektive Persönlichkeitstests im „Wettstreit“ um (Un-)Verfälschbarkeit. *Report Psychologie*, 26, 298-304.
- Huwe, S., Hennig, J. & Netter, P. (1996). Das Repression-Sensitization-Coping-Inventar (RSCI). *Diagnostica*, 42 (2), 157-174.
- Jackson, C. J. & Francis, L. J. (1997). Interpreting the correlation between neuroticism and lie scale scores. *Personality and Individual Differences*, 26, 59-63.
- Karner, T. (2002). The volunteer effect of answering personality questionnaires. *Psychologische Beiträge*, 44, 42-49.
- Khorramdel, L. & Kubinger, K. D. (2006). The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure. *Psychology Science*, 48 (3), 378-397.

- Kluger, A. N., Reilly, R. R. & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, 76 (6), 889-896.
- Kohlberg, L. (1974). *Zur kognitiven Entwicklung des Kindes*. Frankfurt am Main: Suhrkamp.
- Kohlberg, L. (1981). *Essays on moral development. Volume 1: The philosophy of moral development. Moral Stages and the Idea of Justice*. San Francisco: Harper & Row.
- Kohlberg, L. (1995). *Die Psychologie der Moralentwicklung*. Frankfurt am Main: Suhrkamp.
- Kohlberg, L., Levine, C. & Hewer, A. (1983). *Contributions to human development. Volume 10*. Basel: S. Karger.
- Kohrramdel, L. & Kubinger, K. D. (2006). The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure. *Psychology Science*, 48 (3), 378-397.
- Krohne, H. W. (1975). *Angst und Angstverarbeitung*. Stuttgart: Kohlhammer.
- Krohne, H. W. (1976). *Theorien zur Angst* (1. Aufl.). Stuttgart: Kohlhammer.
- Krohne, H. W. (1978). Individual Differences in Coping with Stress and Anxiety. In C. D. Spielberger & I. G. Sarason (Hrsg.), *Stress and Anxiety. Volume V*. Washington: Hemisphere.
- Krohne, H. W. (1985a). Das Konzept der Angstbewältigung. In H. W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen*. Weinheim: VCH.
- Krohne, H. W. (1985b). Entwicklungsbedingungen von Ängstlichkeit und Angstbewältigung: Ein Zweiprozeß-Modell elterlicher Erziehungswirkung. In H. W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen*. Weinheim: VCH.
- Krohne, H. W. (1993a). Attention and Avoidance. Two Central Strategies in Coping with Aversiveness. In H. W. Krohne (Hrsg.), *Attention and Avoidance. Strategies in Coping with Aversiveness*. Seattle: Hogrefe & Huber Publishers.
- Krohne, H. W. (1993b). Vigilance and Cognitive Avoidance as Concepts in Coping Research. In H. W. Krohne (Hrsg.), *Attention and Avoidance. Strategies in Coping with Aversiveness*. Seattle: Hogrefe & Huber Publishers.
- Krohne, H. W. (1996). *Angst und Angstbewältigung*. Stuttgart: Kohlhammer.
- Krohne, H. W. & Rogner, J. (1985). Mehrvariablen-Diagnostik in der Bewältigungsforschung. In H. W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen*. Weinheim: VCH.

- Krohne, H. W., Wigand, A. & Kiehl, E. (1985). Konstruktion eines multidimensionalen Instruments zur Erfassung von Angstbewältigungstendenzen. In H. W. Krohne (Hrsg.), *Angstbewältigung in Leistungssituationen*. Weinheim: VCH.
- Krohne, H. W., Schmukle, S. C., Burns, L. R., Egloff, B. & Spielberger, C. D. (2001). The measurement of coping in achievement situations: an international comparison. *Personality and Individual Differences*, 30, 1225-1243.
- Kubinger, K. D. (2002). On faking personality inventories. *Psychologische Beiträge*, 44, 10-16.
- Lind, G. (1978). Wie mißt man moralisches Urteil? Probleme und Möglichkeiten der Messung eines komplexen Konstrukts. In G. Portele (Hrsg.), *Sozialisation und Moral: neuere Ansätze zur moral. Entwicklung und Erziehung*. Weinheim: Beltz.
- Lind, G. (1989). Die Entwicklung moralischer Gefühle durch Vernunft und Dialog. In G. Lind & G. Pollitt-Gerlach, (Hrsg.), *Moral in „unmoralischer“ Zeit: Zu einer partnerschaftlichen Ethik in Erziehung und Gesellschaft*. Heidelberg: Roland Asanger.
- Lind, G. (1993). *Moral und Bildung: Zur Kritik von Kohlbergs Theorie der moralisch-kognitiven Entwicklung*. Heidelberg: Roland Asanger.
- Lind, G. (2000). *Inhalt und Struktur des moralischen Urteilens*. Unveröffentlichte Dissertation (Korrigierte Neuauflage), Universität Konstanz.
- Lind, G. (2002a). *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung* (Zweite, ergänzte Auflage). Berlin: Logos.
- Lind, G. (2002b). *The Meaning and Measurement of Moral Judgement Competence – A Dual Aspect Model*. Konstanz: Universität. (Article in Press)
- Lind, G. (2002c). *Introduction to the Moral Judgment Test (MTJ)*. Konstanz: Universität (Article in Press)
- Lind, G. & Wakenhut, R. (1985). Testing Moral Judgement Competence. In G. Lind, H. A. Hartmann & R. Wakenhut (Hrsg.), *Moral development and the social environment. Studies in the philosophy and psychology of moral judgement and education* (S. 79-105). Chicago: Precedent Publishing Inc.
- Mahar, D., Cologon, J. & Duke, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences*, 18 (5), 605-609.
- Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science*, 48 (3), 226-246.

- Martin, B. A., Bowen, C.-C. & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, 32, 247-256.
- Mersman, J. L. & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences*, 24 (2), 217-227.
- Mittenecker, E. (1982). Subjektive Tests zur Messung der Persönlichkeit. In K.-J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie, Psychologische Diagnostik: Persönlichkeitsdiagnostik* (S. 57-131). Göttingen: Hogrefe.
- Mummendey, H. D. (1990). *Psychologie der Selbstdarstellung*. Göttingen: Hogrefe.
- Mummendey, H. D. & Eifler, S. (1994). Ein Fragebogen zur Erfassung „positiver“ Selbstdarstellung (Impression-Management-Skala). In H. D. Mummendey (Hrsg.), *Bielefelder Arbeiten zur Sozialpsychologie, Nr. 170* (S. 1-10). Bielefeld: Universität.
- Mummendey, H. D. (1995). *Die Fragebogen-Methode: Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung* (2. korrigierte Aufl.). Göttingen: Hogrefe.
- Nguyen, N. T., Biderman, M. D. & McDaniel, M. A. (2005). Effects of response instructions on faking a Situational Judgment Test. *International Journal of Selection and Assessment*, 13 (4), 250-260.
- 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.
- Paulhus, D. L. (1984). Two-Component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Peeters, H. & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students success: The influence of faking. *Educational and Psychological Measurement*, 65 (1), 70-89.
- Pervin, L. A. (2000). *Persönlichkeitstheorien* (4. Aufl.). München: Reinhardt.
- Rollett, B. (1997). *Lernen und Lehren. Eine Einführung in die Pädagogische Psychologie und ihre entwicklungspsychologischen Grundlagen* (5. überarb. Auflage). Wien: WUV.
- Rosse, J. G., Stechner, M. D., Levin, R. A. & Miller, J. L. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.
- Schmidt, L. R. (1975). *Objektive Persönlichkeitsmessung in diagnostischer und klinischer Psychologie*. Weinheim: Beltz.
- Schmidt, L. R. & Schwenkmezger, P. (1990). Objektive Persönlichkeitstests: Perspektiven für die Diagnostik. In P. Bartussek & M. Amelang (Hrsg.), *Fortschritte der Differentiellen Psychologie und*

- Psychologischen Diagnostik. Festschrift zum 60. Geburtstag von Professor Dr. Kurt Pawlik* (S. 229-239). Göttingen: Hogrefe.
- 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.
- Schneewind, K. – A. & Graf, J. (1998). *Der 16-Persönlichkeits-Faktoren-Test Revidierte Fassung. 16 PF-R*. Bern: Huber.
- Schneider, J. F. & Hübner, R. (1980). Einfluß von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebogen. *Zeitschrift für experimentelle und angewandte Psychologie*, 27 (4), 565-579.
- Seiwald, B. B. (2002). Replicability and generalizability of Kubinger's results: Some more studies on faking personality inventories. *Psychologische Beiträge*, 44, 17-23.
- Snell, A. F., Sydel, E. J. & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9 (2), 219-242.
- Testkuratorium (der Föderation deutscher Psychologenverbände) (1986). Mitteilung. *Diagnostica*, 32, 358-360.
- Topping, G. D. & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, 23 (1), 117-124.
- Viswesvaran, C. & Ones, D. S. (1999). Meta-analyses of fakability estimates: implications for personality measurement. *Educational and Psychological Measurement*, 59 (2), 197-210.
- 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.
- Zimbardo, P. G. (1995). *Psychologie* (6. neu bearb. u. erw. Aufl.). Berlin: Springer.

Lebenslauf

Christa Brunflicker, geboren am 01. August 1978 in Neunkirchen.

Ausbildung:

1984 – 1988	Volksschule Neunkirchen
1988 – 1992	Hauptschule Neunkirchen
1992 – 1997	Höhere Lehranstalt für Mode und Bekleidungstechnik Wiener Neustadt
seit 1997	Studium der Psychologie an der Universität Wien

Fachpraktika:

Juni – Sept. 2002	HILL International, Wien 6-wöchiges Fachpraktikum für PsychologInnen
Juni – Sept. 2007	Kolpinghaus „Gemeinsam leben“, Wien Freiwilliges Fachpraktikum für PsychologInnen

Berufliche Tätigkeit:

März – Aug. 2004	Trenkwalder Personaldienste AG, Wien Datenarchivierung und Beschlagwortung
Sept. 2004	Österreichisches Siedlungswerk, Wien Postbearbeitung und Beschlagwortung
Sept. 2004 – Sept. 2005	Trenkwalder Personaldienste AG, Wien Datenarchivierung und Postbearbeitung
Juli – Sept. 2006	Modellspielwaren GmbH, Gloggnitz Mitarbeit in der Fertigung