

Diplomarbeit

Validierung des Persönlichkeitsverfahrens *Big Five Plus One (B5PO):*

Eine Validierung anhand von Extremgruppen und Überprüfung eines möglichen State-Einflusses

Lisbeth Weitensfelder

angestrebter akademischer Grad
Magistra der Naturwissenschaften (Mag.^a rer. nat.)

Wien, im Oktober 2008

Studienkennzahl lt. Studienbuchblatt:	A298
Studienrichtung lt. Studienbuchblatt:	Psychologie
Betreuer:	o.Univ.-Prof. Dr. Mag. Klaus D. Kubinger

Danksagung

Da eine Aufzählung aller Beteiligten aus dem privaten Kreis, die mit der Diplomarbeit mittelbar durch Unterstützung, geduldiges Zuhören, Korrekturlesen und nicht zuletzt Entbehrungen in Verbindung stehen, den Rahmen sprengen würde, möchte ich all jenen aus dem Freundes- und Familienkreis ein gemeinsames „Dankeschön“ aussprechen.

Namentlich erwähnen möchte ich jedoch diejenigen, die mit der Erstellung dieser Diplomarbeit unmittelbar zu tun hatten:

An erster Stelle gilt mein Dank Univ.-Prof. Dr. Mag. K.D. Kubinger für dessen hervorragende Betreuung.

Des Weiteren möchte ich mich bei all jenen bedanken, die diese Untersuchung ermöglicht haben: Dr. Gernot Schuhfried GmbH durch die Zurverfügungstellung der benötigten Hard- und Software, den beteiligten Schulen (und Schüler/innen) in Wien und Niederösterreich sowie allen engagierten AHS-Lehrer/innen und Direktor/innen ebendieser, und für die Genehmigung den Landes- und Stadtschulräten Wien und Niederösterreich.

Weiters möchte ich meinem Kollegen Andreas Schneiderbauer für die hervorragende Zusammenarbeit bei der gemeinsamen Datenerhebung danken und allen beteiligten Kollegen, Kolleginnen, Freunden und Freundinnen aus dem Forschungsseminar für deren Rat, Unterstützung und Hilfe.

Meiner Schwester Ute, die liebevoll und originell das für das Experiment benötigte Suchbild zeichnete, ebenfalls ein herzliches Dankeschön.

Abstract

Aus einer Kombination verschiedener Konzepte der Validierung wurde in der vorliegenden Arbeit eine Validierung des Persönlichkeitsverfahrens *Big Five Plus One (B5PO)* hinsichtlich zweier wesentlicher Punkte vorgenommen: Einerseits wurde eine Extremgruppenvalidierung vorgenommen, wobei die Extremgruppen durch Expertenratings von offenen Selbstbeschreibungen gebildet wurden, sowie eine mehr am Konstrukt der Traits orientierte Validierung durch ein Experiment, um zu überprüfen, ob die in dem Persönlichkeitsverfahren gemessenen Traits von States überlagert seien. Die Extremgruppenvalidierung gelang für fünf der sechs Skalen („Extraversion“, „Verträglichkeit“, „Gewissenhaftigkeit“, „Emotionale Kontrolle“ und „Empathie“). Nicht gelungen ist die Validierung in Bezug auf die Skala „Offenheit“. Auch zeigte sich eine State-Abhängigkeit der Skala „Gewissenhaftigkeit“ nach Bearbeiten eines Suchbild-Rätsels. Eine weitere State-Abhängigkeit der Skala „Empathie“ trat bei weiblichen Testpersonen zutage, nachdem sie in offenem Antwortformat ihre Persönlichkeit beschreiben sollten. Als Nebenergebnis zeigten sich signifikante Unterschiede in der Selbstdarstellung auf den Skalen „Empathie“ und „Emotionale Kontrolle“ des Persönlichkeitsinventars zwischen Schülerinnen und Schülern.

Out of a combination of different concepts of validation, a validation of the personality inventory *Big Five Plus One (B5PO)* was accomplished, concerning two major goals: On the one hand a validation via extreme groups was conducted, whereas the extreme groups were composed through expert-ratings from self-descriptions in an essay, on the other hand an experiment was set up to show whether the measured traits were influenced by states, which was rather a validation based on the construct of traits. The validation via extreme groups worked out well for five scales of the personality inventory (namely “extraversion”, “agreeableness”, “consciousness”, “emotional stability” and “empathy”) but failed for another (“openness to experience”). Furthermore, a state influence on the scale “consciousness”, caused by the work on a puzzle picture, could be shown. Another state influence on the scale “empathy” was found for female test persons only, caused by the self-description. As a by-product, significant differences in the inventory’s self-portrayal of males and females for the scales “empathy” and “emotional stability” could be shown.

Inhaltsverzeichnis

Danksagung	3
Abstract	5
I Einleitung und Fragestellung der vorliegenden Arbeit	11
II Theoretischer Teil	13
1. States und Traits	14
1.1 Traits	14
1.2 States	16
1.3 Anforderungen an Verfahren zur Erfassung von States vs. Traits	17
1.4 Zum Zusammenhang von States und Traits	18
1.5 Exkurs: Die Ergebnisse von Schwarz und Clore	19
2. Die Big Five der Persönlichkeit	22
2.1 Der lexikalische Ansatz	22
2.1.1 Geschichtliche Entwicklung	23
2.1.2 Von Cattell unabhängige Studien	24
2.2 Die Big Five im Überblick	25
2.2.1 Extraversion vs. Introversion	25
2.2.2 Verträglichkeit	26
2.2.3 Gewissenhaftigkeit	26
2.2.4 Neurotizismus vs. Emotionale Stabilität	26
2.2.5 Offenheit	27
2.3 Die Big Five als States?	28
2.4 Kritik am Big Five-Modell	28
3. Empathie	30
4. Das Persönlichkeitsverfahren <i>Big Five Plus One (B5PO)</i>	31
4.1 Theoretische Grundlagen	31
4.1.1 Die Persönlichkeitspsychologie von George A. Kelly	32
4.1.2 Antwortformate	33
4.1.3 Selbstbild/Selbstkonzept, Selbstdarstellung und soziale Erwünschtheit	33
4.2 Testaufbau und Verrechnung	35
4.2.1 Vorgabe	35

4.2.2 Verrechnung.....	36
4.3 Kritik	36
5. Validität	38
5.1 Inhaltliche Validität.....	38
5.2 Kriterienbezogene Validität.....	38
5.3 Konstruktvalidität.....	39
5.3.1 Der Ansatz von Campbell und Fiske	39
5.3.2 Nicht-korrelative Ansätze	40
6. Inhaltsanalyse	41
6.1 Die quantitative Inhaltsanalyse	41
6.2 Die qualitative Inhaltsanalyse	42
III Empirischer Teil.....	45
7. Fragestellung und Hypothesen.....	46
7.1 Validierung mit Hilfe der Extremgruppen.....	46
7.2 Erfassung eines möglichen State-Einflusses.....	49
8. Methode und Stichprobenberechnung	51
9. Versuchsplan	53
9.1 Materialien und Vorgabe	53
9.1.1 Das Verfahren <i>B5PO</i>	53
9.1.2 Aufsatz.....	54
9.1.2 Suchbild	55
9.1.3 Überblick	56
9.2 Einteilung der Versuchsgruppen.....	57
10. Stichprobe.....	58
10.1 Stichproben für die Extremgruppenvalidierung.....	58
10.2 Stichprobe für die Überprüfung eines möglichen State-Einflusses.....	60
11. Entwicklung eines Kategoriensystems für die Extremgruppenbildung	63
11.1 Qualitative Auswertungen	63
11.2 Globalbeurteilung und Rangreihung	69
11.3 Schlüsselwörterkodierung.....	71
11.4 Extremgruppenbildung	73
11.4.1 Skala „Extraversion“	75
11.4.2 Skala „Verträglichkeit“	75
11.4.3 Skala „Gewissenhaftigkeit“	76

11.4.4 Skala „Emotionale Kontrolle“	76
11.4.5 Skala „Offenheit“	77
11.4.6 Skala „Empathie“	77
12. Interpolieren fehlender <i>T</i> -Werte im <i>B5PO</i>	79
13. Ergebnisse	82
13.1 Extremgruppenvergleiche.....	82
13.1.1 Extraversion	82
13.1.2 Verträglichkeit.....	83
13.1.3 Gewissenhaftigkeit	84
13.1.4 Emotionale Kontrolle.....	85
13.1.5 Offenheit	85
13.1.6 Empathie	86
13.2 Überprüfung einer möglichen State-Abhängigkeit	87
13.2.1 Überprüfung eines möglichen State-Einflusses durch den Aufsatz.....	87
13.2.2 Überprüfung eines State-Einflusses durch die Suchbildvorgabe.....	94
14. Diskussion und Ausblick	101
15. Zusammenfassung	105
IV Literatur	107
V Tabellenverzeichnis	111
VI Abbildungsverzeichnis	113
VII Anhang.....	115
Anhang A: Materialien und offizielle Schreiben	116
Anhang B: Auswertungshilfen	124
Lebenslauf (Curriculum Vitae)	135

I Einleitung und Fragestellung der vorliegenden Arbeit

Ziel der vorliegenden Arbeit ist die Validierung des Persönlichkeitsinventars *Big Five Plus One (B5PO)*; Holocher-Ertl, Kubinger & Menghin, 2003).

Validität ist – neben Objektivität und Reliabilität (Messgenauigkeit) – eines der Hauptgütekriterien eines Tests. „Ein Test ist dann valide bzw. »gültig«, wenn er tatsächlich diejenige Eigenschaft misst, welche er zu messen beabsichtigt“ (Kubinger, 2003a, S. 199).

Dies ist leider durchwegs keine Selbstverständlichkeit (vgl. ebd.), und gerade aus diesem Grund ist es sowohl für die Grundlagenforschung als auch für die Praxis, in der die Verfahren angewandt werden, wichtig, zu überprüfen, ob die intendierten Eigenschaften tatsächlich erfasst werden.

Diese Eigenschaften, die der *B5PO* zu messen beabsichtigt, sind folgende sechs Rasch-Modell-konformen Dimensionen: die Big Five *Extraversion*, *Verträglichkeit*, *Gewissenhaftigkeit*, *Emotionale Kontrolle* und *Offenheit* sowie als weitere Dimension *Empathie*. All diese Dimensionen werden durch die Selbsteinschätzung der Testpersonen anhand von bipolaren Eigenschaftswörtern gewonnen.

Im Zuge der Validierung sollen nun nach einer Selbstbeschreibung der Testpersonen in offenem Antwortformat Extremgruppen zu den jeweiligen Dimensionen gebildet werden. Daraufhin soll betrachtet werden, wie „gut“ das Verfahren *B5PO* zwischen diesen Gruppen unterscheiden kann. In diesem Sinne handelt es sich also einerseits um eine „Extremgruppenvalidierung“, andererseits fließen aber auch Punkte des so genannten Multi-Trait-Multi-Method-Ansatzes (MTMM-Ansatz) zur Validierung ein, nämlich insofern, als versucht wird, mit verschiedenen Methoden dasselbe zu erfassen.

Die Eigenschaften, die im Verfahren *B5PO* gemessen werden sollen, sind so genannte *Traits* (vgl. Holocher-Ertl, Kubinger & Menghin, 2003, S. 5), also zeitlich relativ stabile Eigenschaften. Im Gegensatz dazu kann man die zeitlich weniger stabilen *States* als Zustände unterscheiden.

Wie im theoretischen Teil dieser Arbeit dargestellt wird, sind diese Dimensionen jedoch nicht unabhängig voneinander. Nun stellt sich natürlich die Frage, inwieweit durch das Persönlichkeitsverfahren *B5PO* tatsächlich Traits erfasst werden, und das Ergebnis nicht (oder

zumindest nicht in inhaltlich interessierendem Ausmaß) von Situationseigenschaften überlagert wird.

Eine starke Beeinflussung durch States wäre für die Praxis ein nicht unerhebliches Problem, da es bedeuten würde, dass das Testergebnis durch unmittelbar zuvor Erlebtes verzerrt würde, wodurch die Aussagekraft des Verfahrens eingeschränkt wird.

Daher soll auch auf diese Frage im Rahmen dieser Arbeit Antwort gegeben werden. Entsprechend lautet die Forschungsfrage, die hier beantwortet werden soll: „Erfasst das Persönlichkeitsinventar *B5PO* jene sechs Dimensionen, die es messen soll, indem es fähig ist zwischen den jeweiligen Extremgruppen zu unterscheiden, und ist das Ergebnis jenes Persönlichkeitsinventars zumindest nicht in relevantem Ausmaß von momentanen zeitlichen Zuständen beeinflusst?“

In den Kapiteln 1–6 werden hierfür die theoretischen Grundlagen, auf denen diese Arbeit basiert, wiedergegeben. Diese reichen von States und Traits über die theoretischen Grundlagen des zu validierenden Verfahrens bis hin zu den Theorien und Methoden, mit deren Hilfe schließlich validiert werden soll. Mit Kapitel 7 beginnt der empirische Teil, in dem die Untersuchung dargestellt und die Forschungsfrage beantwortet werden soll.

II Theoretischer Teil

1. States und Traits

Bereits Cattell unterschied 1950 zwischen *States* und *Traits*, wobei letztere zeitlich relativ stabil sind, während States zeitlich fluktuieren (vgl. Amelang et al., 2006, S. 61). Was genau mit diesen beiden Begriffen gemeint ist und wie sie zusammen hängen, sei im Folgenden kurz dargestellt.

1.1 Traits

Im Grunde genommen sind Traits das, was die meisten Menschen meinen, wenn sie von Persönlichkeitseigenschaften sprechen. Die in Wörterbüchern häufig für den Trait-Begriff verwendete Übersetzung „Charakterzug“ greift jedoch etwas kurz und wurde nunmehr „in der psychologischen Wissenschaft durch den Begriff der Eigenschaft abgelöst“ (Schmitt, 2003, S. 424): So werden unter Traits nicht nur die Persönlichkeitseigenschaften in einem engeren Sinn verstanden, sondern auch Leistungs- oder Gefühlseigenschaften, Interessen, Eigenschaften des Selbstkonzepts u.v.m. (vgl. Schmitt, 2003, S. 424).

Eine bekannte Definition von Traits stammt von Guilford (1974, zit. nach Amelang et al., 2006, S. 47): „Die Persönlichkeit eines Individuums ist seine einzigartige Struktur von Persönlichkeitszügen (Traits) ... Ein Trait ist jeder abstrahierbare und relativ konstante Persönlichkeitszug, hinsichtlich dessen eine Person von anderen Personen unterscheidbar ist.“ Dabei sind bei den Traits morphologische und physiologische Persönlichkeitszüge abgrenzbar, sowie „Bedürfnisse, Interessen, Einstellungen, Eignungen und die Temperamente“ (Amelang et al., 2006, S. 47). Die Summe aller Traits ergibt für die Mehrzahl der Eigenschaftstheoretiker die Persönlichkeit (vgl. ebd., S. 55).

Amelang et al. (2006) beschreiben Traits als „relativ breite und zeitlich stabile Dispositionen zu bestimmten Verhaltensweisen, die konsistent in verschiedenen Situationen auftreten“ (S. 54). In dieser Definition sind zwei der drei wesentlichen Merkmale erwähnt, die ein Merkmal erfüllen muss, um als Trait zu gelten (vgl. Schmitt, 2003, S. 425):

- Stabilität über einen längeren Zeitraum¹
- Verhaltenskonsistenz in unterschiedlichen Situationen

Das dritte Merkmal ist in der Definition von Amelang und Bartussek nicht enthalten, jedoch in der Definition von Guilford („von anderen Personen unterscheidbar“) und heißt *Distinktheit*:

¹ Anm.: Dieses Kriterium findet sich auch bei Guilford, vgl. „... jeder relativ konstante Persönlichkeitszug ...“

- „Personen müssen sich in einem nicht trivialen Ausmaß in dem Merkmal unterscheiden, sonst hätte es keinen Informationswert“ (Schmitt, 2003, S. 425).

Traits sind latent, das heißt, sie selbst sind nicht beobachtbar, sondern nur das Verhalten, dem sie zugrunde liegen – und aus diesem müssen die Traits dann erschlossen werden (vgl. Schmitt, 2003, S. 424). Oder, in den Worten von Amelang et al. (2006, S. 57): „Traits sind abstrakte Kategorien für (konkret beobachtbare) Verhaltensweisen“. Dabei kann man – je nachdem, welche Verhaltensweisen man bündelt – schmale oder weite Traits erhalten, wobei man die Breite als Kontinuum ansehen kann. So wäre das unterste (schmälste) Extrem des Kontinuums eine einzige Verhaltensweise, z.B. andere zu schlagen. Wenn man hierzu noch treten, würgen etc. hinzufügen würde, würde daraus schon ein breiterer Trait resultieren (körperliche Aggressivität). Durch weitere Aufnahme von Verhaltensweisen wie z.B. verbalen Gemeinheiten könnte man dieses Konzept zu einer sehr breiten Disposition (allgemeine Aggressivität) erweitern (vgl. Amelang et al., 2006, S. 57.). „Ob relativ schmale oder breite Traits von Vorteil sind, hängt von den jeweiligen Zielen ab, die damit verfolgt werden. Schmale Traits sind intern homogen und erlauben gewöhnlich genauere Vorhersagen spezifischer Verhaltensweisen“ (ebd.). Allerdings können sehr schmale Traits auch Verhaltensweisen beinhalten, die nur sehr selten auftreten, wogegen breite Traits eher „die Vorhersagbarkeit von sehr vielen Verhaltensweisen in verschiedenem Kontext“ versprechen (ebd.).

Zur Stabilität von Traits

Zur relativen Stabilität über einen längeren Zeitraum ist eine Studie von Schuerger, Zarrella und Hotz (1989) anzuführen, die den Abfall in der Stabilität im Laufe der Zeit untersuchten: Bereits in einer früheren Studie (Schuerger, Tate & Tavernelli, 1982) wurde ein Abfall in der Reliabilität beobachtet: „The most significant of our findings was that tests of general personality all show a similar pattern of decline in reliability (stability) over time, a pattern of exponential decay asymptotic at about .50“ (Schuerger, Zarrella & Hotz, 1989, S. 777). Auch in der Studie von 1989 zeigte sich ein Abfall in der Reliabilität, wobei der wichtigste Bestimmungsfaktor der Test-Retest-Reliabilität die Länge des Intervalls war; die Reliabilität fällt dabei im ersten Jahr stark ab und verflacht dann. Auch das Alter der Testpersonen hatte einen messbaren Einfluss, besonders bei Studien mit längeren Intervallen (vgl. ebd., S. 780). Damit übereinstimmend fassen auch Pervin, Cervone und John (2005, S. 343) zusammen:

Derzeit scheinen die Daten folgenden Schluss nahe zu legen: (1) Die Persönlichkeit ist über kürzere Zeiträume hinweg stabiler als über längere Zeiträume. (2) Die Persönlichkeit ist im Erwachsenenalter stabiler als in der Kindheit. (3) Auch wenn es Belege für eine allgemeine Stabilität der Wesenszüge gibt, bestehen während der Entwicklung individuelle Unterschiede in der Stabilität. (4) Auch wenn es Belege für eine allgemeine Stabilität der Wesenszüge gibt, sind die Grenzen der Umwelteinflüsse bei der Veränderung während der Kindheit und im Erwachsenenleben nach wie vor zu ermitteln.

1.2 States

Das wesentliche Merkmal, das einen State von einem Trait unterscheidet, ist jenes der Stabilität. Distinktheit und Konsistenz sind zwar gegeben, das Stabilitätskriterium jedoch nicht: Es handelt sich um einen Zustand (vgl. Schmitt, 2003, S. 426).

„Die States entsprechen in grober Annäherung dem umgangssprachlichen Stimmungsbegriff; es handelt sich hierbei um temporäre Zustände von Aktivierung, Entspannung, guter Stimmung und dgl.“ (Amelang et al., 2006, S. 61).

Was das Auftreten betrifft, sind States „nur aus vergleichsweise kurzzeitig auftretenden Verhaltensweisen erschließbar“ (Amelang, 2003, S. 97), während die Indikatoren für Traits „zahlreich und in vielen Situationen beobachtbar“ sind (ebd., S. 98). Aus diesen Gründen sind Instrumente zur Erfassung von Traits in Relation von größerer Bedeutung als jene zur Erfassung von States (vgl. ebd., S. 98), dennoch gibt es Einiges an Inventar zur Erfassung der „momentanen“ Eigenschaften. Dabei handelt es sich zumeist um Listen von Eigenschaftswörtern, die die Testpersonen „in alternativer Form (auf den momentanen Zustand zutreffend oder nicht) oder anhand abgestufter Skalen“ bearbeiten sollen (Amelang et al., 2006, S. 61).

Um States zu erfassen, wäre es notwendig, „dass solche Variablen einen gemeinsamen Faktor oder eine Skala konstituieren, die über Messzeitpunkte oder Situationen hinweg miteinander korrelieren“ (Amelang & Schmidt-Atzert, 2006, S. 287). Diese Faktoren würden dann entweder nur für die betreffende Person Gültigkeit besitzen, oder – im Fall, dass die Daten über Testpersonen aggregiert wurden – sie würden für eine fiktive Durchschnittsperson gelten. Reliabilität wäre demnach als Maß dafür aufzufassen, inwiefern intraindividuelle Unterschiede reproduzierbar sind (vgl. ebd.). Generell wären zur Erfassung von State-Tests daher Längsschnittstudien oder Untersuchungen in unterschiedlichen Situationen erforderlich; allerdings basieren die meisten State-Tests „aber auf Querschnittsanalysen und auf nur einmaliger Vorgabe der Items mit nachgeschalteten R-Analysen, also der Korrelation zwischen Variablen über Probanden. ... Damit werden die stabilen interindividuellen Unterschiede zu Lasten intraindividueller Veränderungen kontraproduktiv maximiert“ (ebd.).

Im deutschsprachigen Raum sind vor allem die Skalen von Janke und Debus (1978) zur Erfassung von States bekannt, wobei meist Dimensionen wie gehobene vs. gedrückte Stimmung, Extravertiertheit vs. Introvertiertheit, Aktivierung vs. Desaktivierung o.ä. zentral sind (vgl. Amelang et al., 2006, S. 61).

Allerdings gibt es auch Inventar, das versucht, nicht nur State-, sondern auch Trait-Eigenschaften (getrennt voneinander) zu erfassen – man denke hier an die Beispiele *STAI* (State-Trait-Angstinventar, Laux et al., 1981) oder *STAXI* (State-Trait-Ärgerausdrucks-Inventar, Schwenkmezger et al., 1992). Des Weiteren gibt es das – bislang unveröffentlichte – State-Trait-Persönlichkeitsinventar *STPI* von Spielberger, das gleich drei verschiedene State-Trait-Skalen messen soll: „Wie für das Angstkonstrukt können auch Ärger und Neugier gesondert in Form von Eigenschafts- und Zustandsskalen erfasst werden“ (Schwenkmezger et al., 1992, S. 8).

1.3 Anforderungen an Verfahren zur Erfassung von States vs. Traits

Laut Laux et al. (1981) sollen für State- bzw. Trait-Tests folgende Kriterien erfüllt werden:

- a) Trait-Tests sollen eine hohe Retest-Reliabilität aufweisen, da sie ein Merkmal messen, das zeitlich überdauernd ist. Das genaue Gegenteil gilt für State-Tests, „da sie sich auf die Erfassung von zeitlich instabilen fluktuierenden Merkmalen richten“ (S. 9).
- b) Trait-Tests sollen im Gegensatz zu State-Tests „nicht auf kurzfristige situative Änderungen reagieren, während State-Tests jenen Bedingungen gegenüber sensitiv sein sollen“ (ebd.).
- c) Korrelationen zwischen verschiedenen Trait-Tests, die den gleichen Bereich messen, müssen höhere Werte erzielen als Korrelationen zwischen State- und Trait-Tests des gleichen Bereichs (vgl. ebd.).
- d) Der letzte von Laux et al. angeführte Punkt bezieht sich vor allem auf den *STAI* selbst, nämlich die Frage, ob die Differenzierung in State-Angst und Trait-Angst ausschließlich durch die Instruktion bedingt sei. Empirische Untersuchungen zum Instruktionseinfluss widersprachen dem jedoch (vgl. ebd., S. 10). Allgemeiner könnte man diese Anforderung folgendermaßen formulieren: States bzw. Traits sollen kein instruktionsbedingtes Artefakt darstellen. Untersuchungen, die dies überprüfen, wären im Sinn der Validität von Vorteil, wenn nicht gar Voraussetzung, um einem gewissen wissenschaftlichen Anspruch zu genügen.

1.4 Zum Zusammenhang von States und Traits

Prinzipiell kann man States und Traits nicht vollkommen getrennt voneinander betrachten. So schreiben z.B. Amelang et al. (2006, S. 62):

In der Tat handelt es sich bei Traits häufig um die Summation von State-Einheiten, weshalb die Übergänge zwischen aktuellen und habituellen Zuständen fließend sind ... Dieses wird auch deutlich, wenn man von Traits als den *relativ* stabilen und überdauernden, von States hingegen als den *relativ* temporären Charakteristika spricht. Es scheint sich um ein Kontinuum zu handeln, auf dem die Unterscheidung noch problematischer wird, wenn man auch änderungssensitive Traits konzidiert ... Von da aus ist es nur noch ein kleiner Schritt zu der Auffassung, dass jedes psychologische Attribut sowohl Trait- als auch State-Komponenten aufweist ...

In einer solchen Auffassung würde der Zusammenhang zwischen States und Traits darin bestehen, dass Eigenschaften bzw. Traits aus mittleren Zuständen über die Zeit gesehen bestehen, also „gleichsam aus der Summation von State-Einheiten resultieren“ (Amelang & Schmidt-Atzert, 2006, S. 287). So wäre die Eigenschaft der Ängstlichkeit einer Person die durchschnittliche Ängstlichkeit ebendieser Person in vielen Situationen (vgl. Schmitt, 2003, S. 426).

Neben dieser Erklärung gibt es noch ein zweites Modell, das den Zusammenhang zwischen States und Traits erklärt, wobei es für beide Modelle empirische Belege gibt (vgl. ebd.). Nach diesem zweiten Modell wird das Verhaltenspotential von Traits in bestimmten, funktional äquivalenten, Situationen aktiviert: Die Eigenschaft Ängstlichkeit wird als latente Bereitschaft gesehen, „in einer bedrohlichen Situation Angst (als State) zu entwickeln“ (ebd.). Dabei kommt es zu einer synergetischen Interaktion zwischen Traits und States, wenn diese funktional äquivalent sind: So ist „bei ängstlichen Personen (trait) der Unterschied in der Angst (State) zwischen einer neutralen und einer bedrohlichen Situation größer als bei unängstlichen Personen“ (ebd.). Allerdings ist nicht in jeder Situation die State-Angst bei Personen mit einer ausgeprägteren Trait-Angst notwendigerweise höher: So berichten Laux et al. in ihrem Manual (1981, S. 8), dass die Zustandsangst bei Hochängstlichen nur in solchen Situationen höher sei, in denen der Selbstwert der betreffenden Personen beeinträchtigt oder in Frage gestellt werde (sog. *ich-involvierende* Situationen). In Situationen mit physischer Gefährdung (z.B. schmerzhaft Reize) ergebe sich bei Hochängstlichen „kein unterschiedlicher Anstieg in der Zustandsangst“ (ebd.). Dadurch kann man zwar an keiner globalen Angstneigung festhalten, „da selbstwertrelevante Stress-Situationen jedoch viel häufiger vorkommen als physische Gefährdungssituationen, bleibt Spielberger (in

Vorbereitung) bei der Aussage, daß Hochängstliche (erfasst mit dem STAI) dazu tendieren, insgesamt mehr Situationen als bedrohlich einzuschätzen“ (ebd.).

1.5 Exkurs: Die Ergebnisse von Schwarz und Clore

Üblicherweise nicht zur Diskussion von States und Traits selbst, sondern zu einer Reihe anderer Themen (z.B. Lebenszufriedenheit, affektive States, Missattributionen, ...) findet man die berühmten Ergebnisse von Schwarz und Clore.

Eine Reihe von Untersuchungen verschiedener Autoren zeigte, dass affektive Zustände eine Bandbreite an Beurteilungen beeinflussen können (vgl. Schwarz & Clore, 1988, S. 44). In Experimenten untersuchten die beiden Forscher daher den Einfluss der Stimmung auf die subjektive Lebenszufriedenheit mit dem ganzen bisherigen Leben. Dabei gibt es mehrere Theorien, die man als Begründung für den Stimmungseinfluss heranziehen könnte:

- a) Nach der Theorie der Stimmungskongruenz ist es leichter, sich positives Material in Erinnerung zu rufen, wenn man selbst in positiver Stimmung ist. Umgekehrt verhält es sich mit negativem Material (vgl. Schwarz & Clore, 1988, S. 45). „When an individual who is in a pronounced mood state retrieves relevant information from memory, the mood-congruent subset of the potentially relevant information is more accessible, and therefore comes to mind more easily“ (ebd.). Es kommt zu einer Selektion, da Personen kaum *alle* potentiell relevanten Informationen berücksichtigen, sondern den Suchprozess beenden, sobald genügend Information zur Verfügung steht, um ein Urteil zu bilden (vgl. ebd.). Allerdings birgt die Theorie der Stimmungskongruenz einiges an Problemen zur Erklärung von Bewertungen: So wird stimmungskongruenter Einfluss vor allem für „self-referenced material“ beobachtet. „Moreover, mood congruency may be limited to relatively unstructured material and tends to be difficult to find when material is presented in narrative form, such that positive and negative elements are interconnected ... or otherwise well organized“ (Schwarz & Clore, 1988, S. 46). Stimmungseffekte auf Bewertungen sind aber relativ robust und zeigen ein Muster, das mit Vorhersagen der Stimmungskongruenz nicht vereinbar ist (vgl. ebd.).
- b) Ausgehend von Tversky und Kahnemann (1973) wäre eine mögliche Erklärung des Einflusses von Stimmung auf die Theorie der Verfügbarkeitsheuristik zurückzuführen: Auch wenn die Forscher sich mehr mit der Verfügbarkeit von Wörtern,

Kombinationsmöglichkeiten oder Ereignissen beschäftigten (siehe hierzu Tversky & Kahnemann, 1973), lassen sich ihre Ergebnisse auch auf Stimmungen anwenden: Das Denken an ein positives (oder negatives) Ereignis erleichtert die Verfügbarkeit solcher Ereignisse im Gedächtnis und führt daher zu einer Überschätzung der Häufigkeit derartiger Begebenheiten.

Die Theorie der Verfügbarkeitsheuristik steht nicht im Widerspruch zur Theorie der Stimmungskongruenz. Das Denken an ein glückliches oder trauriges Ereignis „not only may change a person's mood but also may increase the availability ... of similarly events ... Either of these factors could influence the person's judgements of general well-being“ (Schwarz & Clore, 1983, S. 513).

- c) Schwarz und Clore schlagen eine gänzlich andere Theorie vor, weshalb Stimmung einen Einfluss auf Bewertungen haben kann: Zwar meinen sie nicht, dass das Phänomen der Stimmungskongruenz nicht existieren würde, sie halten es in diesem Zusammenhang jedoch für überbewertet. Stattdessen meinen sie, dass Personen ihre erhaltenen affektiven Reaktionen als relevante Information heranziehen würden, wenn sie Bewertungen vornehmen: „For example, when asked how »likeable« Peter is, individuals may interpret this question to refer to *their feelings* toward Peter“ (Schwarz & Clore, 1988, S. 46). Studien über die Bewertung des Lebens als Ganzes unterstützen die Theorie von Schwarz und Clore: Testpersonen, die an einem sonnigen Tag am Telefon interviewt wurden, berichteten, dass es ihnen global besser ginge als solche, die an einem regnerischen Tag interviewt wurden – aber nur, wenn ihre Aufmerksamkeit nicht auf das Wetter gerichtet wurde. Als Erklärung dient, dass Personen ihre momentanen Gefühle als zur Urteilszeit als Indikator für die Antwort verwenden, nicht jedoch dann, wenn der Informationswert dieser Gefühle hinterfragt wird. Da es sich um solch einen breiten Bereich wie Lebenszufriedenheit handelt, widersprechen diese Ergebnisse der Theorie der Stimmungskongruenz: “Attributing one's current feelings to the weather should simply limit the range of issues about which the feelings seem informative” (ebd., S. 49).

Ein ähnliches Ergebnis erzielten die Forscher, als sie bei Psychologiestudent/innen nach Beschreiben eines positiven bzw. negativen Erlebnisses die Lebenszufriedenheit erhoben und einigen von ihnen die Möglichkeit gaben, ihre (positive oder negative) Stimmung auf den Testraum zurückzuführen (vgl. Schwarz & Clore, 1983). Bei Personen, die ihre *schlechte* Stimmung auf den Testraum zurückführen konnten,

wurde die Bewertung der Lebenszufriedenheit nicht durch das Beschreiben eines schlechten Ereignisses beeinflusst. Anders verhielt es sich bei den gut gelaunten Testpersonen; diese berichteten trotz des Hinweises, dass der Testraum sie möglicherweise in eine gute Stimmung versetzt, eine hohe Lebenszufriedenheit (vgl. ebd.). Wäre die berichtete Lebenszufriedenheit nur auf leichtere Verfügbarkeit aufgrund von Stimmungskongruenz gegeben, sollte es keine Rolle spielen, auf welche Erklärung die Testpersonen ihre Stimmung zurückführen (vgl. ebd., S. 514).

Bekannt ist auch die Studie von Schwarz (1983), in der gute Stimmung durch das „zufällige“ Finden einer Münze auf einem Kopiergerät erzeugt wurde. „Subjects who found a dime subsequently reported higher satisfaction and happiness with their life as a whole than subjects who did not find a dime“ (Schwarz & Clore, 1988, S. 48).

Warum diese Studien nun an dieser Stelle angeführt sind, hat folgenden Grund: Wenn die Beurteilung der Lebenszufriedenheit (im weiteren Sinn als überdauernde Eigenschaft, also als Trait angesehen) durch eine kleine Veränderung des Zustands (State) derart beeinflusst werden kann, wäre es interessant zu wissen, inwieweit dies auch auf Persönlichkeitseigenschaften zutreffen würde.

Da States und Traits in einem Zusammenhang stehen (vgl. 1.4), stellt sich die Frage, inwieweit Persönlichkeitsverfahren (insbesondere Fragebögen), selbst wenn sie Traits erfassen sollen, nicht doch in einem relevanten Anteil von States beeinflusst werden. Darauf soll (zumindest was das Verfahren *B5PO* betrifft) im empirischen Teil eine Antwort gegeben werden.

2. Die Big Five der Persönlichkeit

Psychologisch-diagnostische Testverfahren zur Erfassung der Persönlichkeit im engeren Sinn beziehen sich „heutzutage ziemlich universell und das sogar international auf das »*Big Five*-Persönlichkeitsmodell«“ (Kubinger, 2006, S. 213).

Diese faktorenanalytisch immer wieder bestätigten fünf Faktoren der Persönlichkeit sind:

- Neurotizismus (bzw. Emotionale Stabilität, sofern man diese Skala positiv formuliert),
- Extraversion,
- Offenheit für Erfahrung,
- Verträglichkeit und
- Gewissenhaftigkeit (vgl. ebd.).

Diese fünf Faktoren sollen eine Taxonomie darstellen, aber „sind nicht dazu gedacht, die vielen spezifischen Begriffe für Traits zu ersetzen, die alle ihre eigenen Nuancen und Bedeutungsabstufungen besitzen“ (Zimbardo & Gerrig, 2004, S. 607).

Auch wenn das Big-Five-Modell der Persönlichkeit auf den so genannten „lexikalischen Ansatz“ zurückgeht, ist die Idee, dass Sprache im Leben wichtige Phänomene abbildet, nicht spezifisch für diesen Ansatz (vgl. De Raad, 2000, S. 16). So schrieb zum Beispiel der Kognitionswissenschaftler Miller 1991:

When an idea is important, people are likely to have a word for it. Mountain people will have a word for mountain; people who live on the plains and have never seen a mountain will not have such a word. The more important something is, moreover, the more words there are likely to be (Miller, 1991, zit. nach De Raad, 2000, S. 2).

Zwar kann Alltagssprache nicht einfach ungefragt ins Wissenschaftliche übersetzt werden, da sie z.B. zu unpräzise oder mehrdeutig sein mag, sie kann jedoch einen Anfangspunkt für eine wissenschaftliche Sprache bilden (vgl. De Raad, 2000, S. 2). Wie dieser Wissenschaftlichkeit Genüge geleistet wird, sei im Folgenden kurz dargestellt.

2.1 Der lexikalische Ansatz

Kurz und prägnant auf den Punkt gebracht wird die lexikalische Hypothese von Goldberg:

Those individual differences that are of most significance in daily transactions of persons with each other will eventually become encoded into their language. The more important is such a difference, the more people will notice it and wish to talk of it, with the result that eventually they will invent a word for it (Goldberg, 1981, zit. nach De Raad, 2000, S. 15).

Der erste Teil dieser Definition, nämlich jener der individuellen Unterschiede, ist jedoch jener Teil, der im Zuge dieser Hypothese am meisten ignoriert wurde (vgl. De Raad, 2000, S. 15).

Die lexikalische Hypothese „is a specification, tailored to accommodate ... (1) the definition of the domain of traits, (2) the reflection of traits in language, and (3), as a sort of codicil to (2), a claim on the relationship between actual traits and their lingual counterparts“ (De Raad, 2000, S. 15).

Wie bereits oben erwähnt, ist die Idee, dass individuelle Unterschiede in der Sprache repräsentiert sind, nicht spezifisch für den lexikalischen Ansatz. Spezifisch ist jedoch die Annahme, dass *alle* diese Unterschiede in der Sprache Ausdruck finden (vgl. De Raad, S. 16). Allport und Odbert, die für die Entwicklung² des lexikalischen Ansatzes eine große Rolle spielen (s.u.), fanden an dieser Auffassung jedoch selbst einiges auszusetzen. Einer ihrer Kritikpunkte sei hier wiedergegeben: „Many traits never receive a name. Littérateurs, psychologists and the makers of slang continually invent new designations for human nature, thus proving our current vocabulary to be insufficient“ (Allport & Odbert, 1936, zit. nach De Raad, 2000, S. 17).

2.1.1 Geschichtliche Entwicklung

Geschichtlich geht der lexikalische Ansatz bis auf Klages im Jahr 1926 zurück (vgl. Amelang et al., 2006, S. 276).

Für die Entwicklung des Big Five-Modells spielt aber die Arbeit von Allport und Odbert eine große Rolle, die 1936 „im Rahmen einer »psycholexikalischen Studie« alle persönlichkeitsrelevanten Begriffe aus »Webster’s New International Dictionary« (1925er Ausgabe) heraus“ suchten (Holocher-Ertl, 2003, S. 26).

Von den 17953 Begriffen, die Allport und Odbert fanden, konnten ca. 4500 der Kategorie „personal traits“ und 100 der Kategorie „passing activities and temporary states“ zugeordnet werden. Dieser große Datensatz wurde in späterer Folge von Cattell auf 35 Cluster reduziert, die je sechs bis zwölf Elemente enthielten. Diese Cluster dienten ihm schließlich als Variablen für eine Fremdbeurteilung und bildeten die Grundlage seiner 12 Faktoren. Viele nachfolgende Wissenschaftler nutzten die Cattell-Variablen als Grundlage, auch wenn die 35 Cluster dabei häufig modifiziert wurden (vgl. Amelang et al., 2006, S. 276).

Den eigentlichen Ausgangspunkt für die Big Five bildeten jedoch die Studien von Tupes und Christal (1958, 1961). Sie reanalysierten Korrelationsmatrizen von acht verschiedenen

² Trotz ihrer großen Rolle, die Allport und Odbert spielten, formulierten sie nie explizit so etwas wie eine lexikalische Hypothese, diese wurde erst von Cattell formuliert (vgl. De Raad, 2000, S. 16).

Stichproben (zwei davon stammten von Cattell, zwei von Fiske). Trotz der verschiedenen Stichproben fanden sie immer wieder fünf gemeinsame Faktoren; diese wurden später von Goldberg „Big Five“ genannt (vgl. Amelang et al., 2006, S. 276f.).

2.1.2 Von Cattell unabhängige Studien

Neben Studien, die auf Cattells Datensätzen beruhen, gibt es aber auch Studien, die Datensätze verwenden, die auf die Zeit vor Cattell zurückgehen (Conley, 1985; Field & Millsap, 1989; vgl. Amelang et al., 2006, S. 277); diese fanden jedoch keine genaue Entsprechung zu den Norman-Faktoren. Auch einige weitere Studien mit anderen Datensätzen können „ebenfalls keinen Anspruch auf Repräsentativität für die Gesamtpersönlichkeit erheben. Auch lässt die starke Überlappung der Variablenätze verschiedener Autoren keinen klaren Schluss über die Generalisierbarkeit der Ergebnisse über verschiedene Variablenauswahlen zu“ (Amelang et al., 2006, S. 277).

Neuere Taxonomien der Persönlichkeit versuchen daher, dem Anspruch auf Repräsentativität besser gerecht zu werden, wobei die Studien von Norman, Goldberg und Peabody von besonderer Bedeutung sind. Norman (1967) arbeitete mit dem „Webster’s Third New International Dictionary“, wobei er die Liste von Allport und Odbert nur um 172 Begriffe erweitern musste. Diese Liste wurde weiter reduziert und kategorisiert, und die resultierenden 1566 stabilen Persönlichkeitsmerkmale wurden in zehn Klassen (die Pole der Big Five) sortiert (vgl. Amelang et al., 2006, S. 277f.). „Die Klassifikation Normans kann allerdings ... vor allem auf Grund des Vorsortierens auf die zehn Pole des Fünf-Faktoren-Modells, als zu willkürlich kritisiert werden“ (ebd., S. 278).

Auch Peabody und Goldberg arbeiteten nicht mit den Cattell’schen Datensätzen und fanden unter verschiedenen Bedingungen immer wieder fünf Faktoren. Im Vergleich mit der Variablenauswahl Cattells kamen sie weiters zum dem Ergebnis, dass manche Bereiche („Conscientiousness“, „Impulse Control“, „Intellect“) „in den Cattell-Variablen gegenüber ihrem Vorkommen in der natürlichen Sprache unterrepräsentiert sind“ (ebd.); das Gegenteil gilt jedoch für „Neurotizismus“ und „Culture“.

Ostendorf und Angleitner (1994) machen darauf aufmerksam, dass bei all diesen Studien durch das Vorselektieren der Forscher der Verdacht besteht, dass implizite Modellvorstellungen das Ergebnis beeinflussen und plädieren daher für eine „Zufallsstichprobe aus der Population aller persönlichkeitsbeschreibenden Begriffe“ (Amelang et al., 2006, S. 279). Solche Zufallsstichproben wurden laut Amelang et al. beispielsweise von Tellegen und Waller (1987) vorgenommen.

Von besonderem Interesse sind jedoch Studien im nicht-englischsprachigen Raum, da Persönlichkeitstaxonomien „nicht nur Phänomene beschreiben sollten, die sich auf die Situation in einer Sprachgemeinschaft beschränken“ (Amelang et al., 2006, S. 279). So scheint sich im deutschen Sprachraum das Modell der Big Five bisher gut zu bewähren, und auch in aktuellen taxonomischen Studien sowohl aus dem europäischen als auch aus dem nicht-europäischen Sprachraum „ließ sich das Fünf-Faktoren-Modell im Allgemeinen gut replizieren, wenn auch in einigen Untersuchungen andere bzw. anders interpretierte Faktoren als die »Big Five« gefunden wurden“ (ebd.).

2.2 Die Big Five im Überblick

Aufgrund der Bedeutung der Big Five soll im Folgenden eine kurze Beschreibung dieser fünf Faktoren unter den hierzulande üblichen Begriffen gegeben werden.

2.2.1 Extraversion vs. Introversion

Das Konstruktpaar Extraversion-Introversion hat eine lange und viel diskutierte Geschichte; so meinten schon Guilford und Braly 1930 (zit. nach De Raad, 2000, S. 88): „No single pair of traits of personality has been quite so widely discussed and studied as that of extroversion and introversion“.

Auch wenn es schon aus der Zeit vor Jung Bezüge auf dieses Konstruktpaar gibt, so geht doch das Verständnis dieses Paares stark auf Jung zurück: „To Jung (1917) Extraversion is the outward turning of psychic energy toward the external world, while Introversion refers to the inward flow of psychic energy towards the depths of the psyche“ (De Raad, 2000, S. 88).

Auch wenn Jung die Pole „Extraversion“ und „Introversion“ in einem typologisierenden Sinn verwendete, gab es für ihn doch keine rein Extra- oder Introvertierten, sondern es handelte sich in seinem Verständnis mehr um ein Kontinuum. Sein Verständnis dieses Konstruktpaares war sehr vielschichtig und bezog sich auf die vier Funktionsbereiche „Thinking“, „Feeling“, „Sensing“ und „Intuiting“. Das volle Bild des Konstruktpaares hat so viele Charakteristika und scheint so breit, dass es beinahe alle fünf Faktoren der Big Five abzudecken scheint (vgl. De Raad, 2000, S. 88–89). Eine vielschichtige Natur wie im Jungianischen Verständnis erscheint auch in Guilford's Fünf-Faktoren-Modell, allerdings ohne die in Jungs Auffassung erscheinenden Widersprüche. Im weiteren geschichtlichen Verlauf des Konstruktpaares wurde der Kern des Verständnisses (die nach außen gerichtete psychische Energie) beibehalten, verschiedene Forscher betrachteten dabei aber verschiedene Facetten und verwendeten unterschiedliche Bezeichnungen: So sprach Fiske (1949) von „Confident Self Expression“,

Tupes und Christal (1961) und Goldberg (1990) verwendeten den Begriff „Surgency“ und Norman (1963) „Extroversion or Surgency“. „Different selections of trait terms in different taxonomies may lead to different interpretations, partially also on the basis of taste, but more important is that interpretations of constructs are governed by context of use“ (De Raad, 2000, S. 90).

2.2.2 Verträglichkeit

Interessanterweise ist Verträglichkeit jene Dimension der Big Five, die die kleinste Geschichte aufweist, und das, obwohl sie Konstrukte wie Liebe, Höflichkeit, Solidarität etc. umfasst. Es handelt sich dabei um jene Dimension, die vermutlich am meisten mit zwischenmenschlichen Beziehungen verbunden ist (vgl. De Raad, 2000, S. 91.). „According to Hogan (1983), Agreeableness enables individuals to cope with problems associated with communal living“ (ebd.).

Die Dimension der Verträglichkeit spielte im psycholexikalischen Ansatz eine zwiespältige Rolle: „The psycholexical dilemma has been whether so-called purely evaluative terms (nice, awful) should be part of a scientific trait-vocabulary...“ (ebd.). Daher wurden von einigen Taxonomern derartige Begriffe entfernt, von anderen wieder beibehalten. So oder so hat sich Verträglichkeit als eine der Dimensionen in den Big Five gezeigt, auch wenn – wie schon bei Extraversion – wieder verschiedenste Namen für diese Dimension gegeben wurden: So sprach z.B. Fiske (1949) von „Conformity“ und Hogan (1983) von „Likability“ (vgl. De Raad, 2000, S. 92).

2.2.3 Gewissenhaftigkeit

Das Konstrukt der Gewissenhaftigkeit beinhaltet den Antrieb, etwas zu erreichen und die dafür notwendigen Charakteristika, wie beispielsweise geplant zu handeln (vgl. De Raad, 2000, S. 92). „After Agreeableness, the relatively strongest growth of studies during the last decade of the twentieth century can be observed for Conscientiousness“ (ebd., S. 93). Verschiedenste Studien ließen darauf schließen, dass gerade dieser Faktor der Big Five mit Job Performance in Zusammenhang steht (vgl. ebd.).

2.2.4 Neurotizismus vs. Emotionale Stabilität

Laut psychologischem Wörterbuch „Dorsch“ (14. Aufl.) hängt Neurotizismus „mit Intensität und Kontrolle emotionaler Reaktionen und Abläufe zusammen“ (Häcker & Stapf, 2004, S. 645). Der Name dieser Dimension wurde ebenfalls einige Male geändert, erscheint

heutzutage aber beinahe durchgängig als Neurotizismus oder Emotionale Stabilität. Dabei wird in der psycholexikalischen Tradition vor allem von Emotionaler Stabilität gesprochen; diese Bezeichnung wird häufiger in Zusammenhängen verwendet, in denen Emotionale Stabilität als Ressource gesehen wird (vgl. De Raad, 2000, S. 95). „»Neuroticism« is given priority in the tradition of personality questionnaire construction (cf. McCrae & John, 1992) where neurotic behavior is considered a problem” (ebd.).

Das erste Verfahren zur Erhebung von Neurotizismus ist das „Personal Data Sheet” von Woodworth aus dem Jahr 1917, das während des ersten Weltkrieges entwickelt wurde, um die Fähigkeit von Soldaten „to cope with military stresses” zu erfassen (ebd., S. 94).

2.2.5 Offenheit

Im Englischen ist dieser Faktor mit „Intellect and Openness to Experience” bezeichnet – ein Faktor und eine Bezeichnung, bei der sich die Geister scheiden. De Raad (2000, S. 96) schreibt hierzu: „Feelings are usually running highest for the Fifth of the Big Five (De Raad & Van Heck, 1994). This not only refers to the naming of the factor but also to its origin and to its relevance as a personality trait factor”. So war beispielsweise Zuckermann (1994) dagegen, dass Intelligenz ein Bereich der Persönlichkeit sei (vgl. De Raad, 2000, S. 96). Die zweigeteilte Einstellung zu dieser Dimension findet sich auch in der niederländischen und deutschen Taxonomie dieses Begriffs: „In the Dutch taxonomy ... the fifth factor had a rebellious coloring, due to such adjectives as »rebellious«, »critical«, »plain«, and »mutinous«. ... Yet, Intellect did not become the primary theme of the Dutch fifth factor“ (ebd., S. 97).

Offenheit ist auch der schwächste Punkt, was den interkulturellen Vergleich betrifft: So berichten Pervin, Cervone und John von einer quantitativen Überprüfung De Raads aus dem Jahr 2005, in der viele europäische Studien verglichen wurden. „Dabei kam man zu dem Schluss, dass die mit den Großen Fünf vergleichbaren Faktoren in den meisten dieser Sprachen festgestellt wurden, wobei die Beweise für den Faktor Offenheit ... jedoch am wenigsten zwingend sind“ (Pervin, Cervone & John, 2005, S. 326). Auch in nichtwestlichen Kulturen, in denen nur wenige Studien durchgeführt wurden, war Offenheit die Replizierbarkeit betreffend am schwächsten (vgl. ebd.). Im Deutschen hingegen kommt der Faktor jenem aus dem amerikanischen Englisch näher, De Raad (2000) meint gar, dass dieser Faktor als Intellekt bezeichnet würde (vgl. S. 51). „This German fifth factor can be well understood as a more or less direct result of the inclusion of the categories »Abilities, talents, or their absence« in the German trait inventory” (ebd.).

2.3 Die Big Five als States?

Es liegen ausführliche Untersuchungen vor, wie gut die Big Five Unterschiede in Traits erklären; die Hypothese, dass sie auch zu intraindividuellen Unterschieden in States beitragen, wurde jedoch nicht getestet (vgl. Borkenau & Ostendorf, 1998, S. 204). Genau dieser Hypothese gingen Borkenau und Ostendorf (1998) daher nach; sie ließen 22 Psychologiestudierende allabendlich an 90 aufeinanderfolgenden Tagen Adjektivskalen mit insgesamt 30 Items zu den Big Five bearbeiten.

Gearbeitet wurde mit der P-Technik, also „factoring of intraindividual correlations between variables across persons“ (Borkenau & Ostendorf, 1998, S. 203), im Gegensatz zu der R-Technik, mit der interindividuelle Unterschiede betrachtet werden. Die P-Faktor Struktur sollte dabei mit der R-Faktor Struktur verglichen werden, weshalb auch die R-Faktor Struktur der 30 Adjektive gebildet wurde (vgl. ebd., S. 206). Es zeigte sich, dass die Big Five auch in der P-Faktor Analyse gefunden wurden: „We recovered the Big Five also in longitudinal variations in states if we averaged P-correlations, but not if individual participants' P-correlations, were factor-analyzed“ (ebd., S. 217). Dass die Big Five nur bei gemittelten P-Korrelationen gefunden wurden mag daran liegen, dass relativ wenige Daten vorhanden waren. Die Tatsache, dass die Big Five in der P-Faktor Analyse jedoch überhaupt gefunden wurden, führt zu Kritik am Modell der Big Five (ebd., S. 219):

... Our findings show that the five-factor model is not »wedded« to the trait approach. A more appropriate analogy is that the five-factor model is a child of the trait approach: It would probably not exist without the trait approach. ... And as far as the five-factor model is useful to describe variations in psychological states over time, it has to be sharply distinguished from the trait approach.

2.4 Kritik am Big Five-Modell

Einer der Kritikpunkte am Big Five-Modell ist sicherlich, dass die Big Five in so vielen verschiedenen Theorien und Sprach- und Kulturkreisen vorkommen, niemals jedoch einheitlich sind, wie De Raad treffend zur Sprache bringt (2000, S. 45):

The hermeneutics of this model of personality description has, on the one hand, led to a sea (OCEAN) of Big Five's and Big Five views, evoking in John (1990) such questions as “*which* Big Five?” or “*whose* Big Five?”. This bewildering situation is expressive of the Babel-like context for personality assessment (John, 1990) with so much diversity of concepts and so little theoretical guidance.

Als weiterer Kritikpunkt kann angeführt werden, dass die Big Five ebenso oft bestätigt wie auch widerlegt wurden: So listet Block (1995) eine Menge an methodischen, empirischen,

semantischen und theoretischen Kritikpunkten am Fünf-Faktoren-Modell auf. Er fasst zusammen (S. 187):

The algorithmic method of factor analysis may not provide dimensions that are incisive. The “discovery” of *the* five factors may be influenced by unrecognized constraints on the variable sets analyzed. Lexical analyses are based on questionable conceptual and methodological assumptions, and have achieved uncertain results. The questionnaire version of the FFA³ has not demonstrated the special merits and sufficiencies of the five factors settled upon.

Schon in älteren Studien finden sich Abweichungen von den Big Five der Persönlichkeit; so waren es selbst bei Costa und McCrae ursprünglich nur drei Faktoren (N-E-O), und auch in jüngeren Studien gibt es einige Abweichungen: So fand z.B. Becker (1999) als sechsten Faktor Hedonismus/Spontaneität vs. Arbeitsorientierung (vgl. Kubinger, Hofmann & Litzenberger, 2002). Becker entwickelte daraufhin in den folgenden Jahren ein Alternativmodell, ein four-plus-X Faktoren Modell mit den vier Faktoren Neurotizismus, Extraversion/Offenheit, Verträglichkeit und Gewissenhaftigkeit, wobei die globale Existenz dieser vier Dimensionen postuliert. Je nach Umfang der Persönlichkeitsinventare können auch mehr, eben four-plus-X Faktoren, zur Erklärung von Gesamtvarianz und Korrelationen von Bedeutung sein (vgl. Holocher-Ertl, 2003, S. 28). „Wenn allerdings mehr als vier Faktoren extrahiert werden, so steigt laut Becker die Wahrscheinlichkeit, dass mehrere dieser Faktoren hoch miteinander korrelieren und zwei oder mehr Super-Faktoren extrahiert werden können“ (ebd.).

Kritik an der Messung der Big Five mittels Fragebogenverfahren ist in Kapitel 4.3 zu finden.

³ Anm.: FFA steht für „5-factor approach“

3. Empathie

Da sich „Empathie“ im Persönlichkeitsverfahren *Big Five Plus One* als eigenständiger Faktor herausstellte, soll auch diese Dimension kurz dargestellt werden.

Das Psychologische Wörterbuch „Dorsch“ (14. Aufl.) liefert zu „Empathie“ folgende Erklärung: „Phänomenal ist E. die Erfahrung, unmittelbar der Gefühlslage eines anderen teilhaftig zu werden und sie dadurch zu verstehen. Trotz dieser Teilhabe bleibt das Gefühl aber anschaulich dem anderen zugehörig. Darin unterscheidet sich Empathie von Gefühlsansteckung“ (Häcker & Stapf, 2004, S. 243).

Das Konzept der Empathie ist vielschichtig an Definitionen und Erhebungsinstrumenten. „Begrifflich herrscht in Publikationen mit Überblickscharakter zumindest insoweit Einigkeit, als sowohl »kognitive Empathie« als auch »affektive Empathie« dem Konzept zugerechnet werden“ (Holz-Ebeling & Steinmetz, 1995, S. 11). Allerdings haben sich viele Autoren dem einen oder anderen Konstruktverständnis (kognitiv bzw. affektiv) verschrieben, wobei eine saubere Trennung der Varianten aber selten vorgenommen wird (vgl. ebd.).

Wallbott (2000) unterscheidet sogar drei Arten von Empathie, wobei er als dritte Art mehr oder weniger die im Psychologischen Wörterbuch „Dorsch“ ausgeschlossene Gefühlsansteckung hinzunimmt: So wird zwischen affektiver Empathie (dem Einfühlen in andere Personen), kognitiver Empathie (Perspektivenübernahme, eher ein rationaler Prozess) und „Inferenz durch Analogie“ (dem Rückschließen von sich auf andere) unterschieden (vgl. Wallbott, 2000, S. 370). „Wiggers hat gezeigt, dass wohl alle drei Prozesse (affektive Empathie, kognitive Empathie und wohl auch emotionale Ansteckung) als Teil empathischer Reaktionen verstanden werden müssen. Dabei favorisiert er jedoch den Aspekt der kognitiven Empathie“ (ebd.).

Zur Erfassung der Empathie gibt es zwei verschiedene Arten von Verfahren: Verhaltensstichproben oder Fragebogenverfahren (vgl. ebd., S. 371). Zur Brauchbarkeit der Fragebogenverfahren liegt eine Studie von Holz-Ebeling und Steinmetz (1995) vor, in der vier aus dem englischsprachigen Bereich stammende Fragebogenverfahren untersucht wurden. Zwei davon zeigten (fast) keine substantielle Validität, die anderen beiden nur in eingeschränktem Maß, wobei bei den letzteren beiden Fragebogen weniger als die Hälfte der Items tatsächlich geeignet erschien. „Bei allen vier Fragebogen erschienen die Iteminhalte nicht oder nur bedingt vereinbar mit dem jeweiligen der Konstruktion zugrundeliegenden theoretischen Begriffsverständnis“ (Holz-Ebeling & Steinmetz, 1995, S. 27).

4. Das Persönlichkeitsverfahren *Big Five Plus One (B5PO)*

Das von Holocher-Ertl, Kubinger und Menghin entwickelte Persönlichkeitsinventar *Big Five Plus One (B5PO)* ist ein computergestütztes, Rasch-Modell-konformes Verfahren, das neben den Big Five der Persönlichkeit auch Empathie erfasst (vgl. Holocher-Ertl, Kubinger & Menghin, 2003, S. 3). Die Testpersonen müssen sich dabei mit Hilfe eines Antwortbalkens zwischen bipolaren Eigenschaftswörtern selbst einschätzen.

Die im Folgenden dargestellte Einteilung in „Theoretische Grundlagen“ und „Testaufbau und Verrechnung“ ist insofern nicht exakt zu trennen, als im Kapitel „4.2 Testaufbau und Verrechnung“ natürlich theoretische Überlegungen dahinterstehen. Für eine übersichtlichere Darstellung wurden jedoch alle theoretischen Hintergründe in einem Kapitel zusammengefasst, sodass unter „4.2 Testaufbau und Verrechnung“ nur noch eine reine Beschreibung des Verfahrens vorzufinden ist.

4.1 Theoretische Grundlagen

Eine der theoretischen Grundlagen dieses Verfahrens ist die Theorie der Persönlichkeit als Struktur von Eigenschaften bzw. Traits (s. Kap. 1.1); entsprechend wurde in diesem Verfahren auch die Erhebung von Traits angestrebt (vgl. Holocher-Ertl, Kubinger & Menghin, 2003, S. 5).

Auf die Grundlage des Big Five-Persönlichkeitsmodells (s. Kap. 2) wurde bereits ausführlich eingegangen, daher wird es hier nicht nochmals dargestellt.

Weiters basiert das Verfahren auf der zugrunde liegenden Annahme, dass Unterschiede zwischen verschiedenen Personen auch immer in Eigenschaftswörtern ausgedrückt werden.

So schreibt Holocher-Ertl (2003, S. 7f.):

Unterschiede zwischen Personen werden zumeist mittels *Eigenschaftsbezeichnungen* ausgedrückt, seien diese in erster Linie umgangssprachlich (z.B. die Bezeichnungen ehrlich, lieb, freundlich) oder eher wissenschaftlich gefärbt (z.B. extravertiert, introvertiert, dominant). Die meist verbreitete und wohl auch ökonomische Art der Persönlichkeitsbeschreibung ist die *direkte Zuschreibung von Eigenschaftswörtern zu Personen* ...

Demzufolge wurden zur Selbstbeschreibung im *B5PO* später Eigenschaftswörter herangezogen.

Die Itemformulierung wurde dabei bewusst in Form nur eines einzigen Adjektivs gewählt (und nicht in Form einer Frage, Situation o.ä. formuliert), „um bei der Darbietung ein Minimum an möglicherweise störenden Bedeutungen mitzutransportieren“ (Holoher-Ertl, Kubinger & Menghin, 2003, S. 6).

4.1.1 Die Persönlichkeitspsychologie von George A. Kelly

Nach der Persönlichkeitstheorie von George A. Kelly ist „das Leben ... eine Repräsentation oder Konstruktion der Realität, und das befähigt den Menschen, sich selbst zu formen“ (Holoher-Ertl, 2003, S. 29). Das Begreifen der Wirklichkeit stellt nach Kelly ein Erforschen, einen Konstruktionsprozess, dar, für den sogenannte „Konstrukte“ herangezogen werden. Diese Konstrukte sind Hypothesen über das Bestehen von bestimmten Klassen von Dingen/Elementen, die sich untereinander ähnlich sind (Ähnlichkeitspol) und von anderen Dingen unterscheiden (Gegensatzpol). Um ein Konstrukt zu verstehen, müssen sowohl Ähnlichkeits- als auch Gegensatzpol bekannt sein (vgl. Holoher-Ertl, S. 29).

Dabei kann man die Konstrukte nicht nur auf andere Dinge, sondern auch auf sich selbst anwenden, wobei „das Konstruktsystem ... insofern die Persönlichkeit des Individuums [ist], als das Interpretationssystem der Grundstein des ganzen psychischen Lebens ist: das Konstruktsystem gestaltet die Weise, in der man denkt, fühlt oder sich verhält“ (Holoher-Ertl, 2003, S. 30).

Nach dieser Theorie hat Kelly den *Role-Construct-Repertory-Test* (kurz: *Rep-Test*) entwickelt - ein Verfahren, bei dem „die befragte Person vollkommen frei ist, zu entscheiden und darüber zu berichten, wie sie die Welt erfasst“ (ebd., S. 34). Dazu werden der Testperson Gruppen von Elementen vorgelegt (in der Regel drei Elemente pro Gruppe) und die Testperson wird darüber befragt, in welcher Weise sich zwei dieser Elemente ähnlich, dabei aber vom dritten verschieden sind (vgl. ebd.).

Diese Repertory-Grid-Technik wurde auch für die Itemgewinnung des Verfahrens *B5PO* verwendet, die herangezogenen Elemente waren Personen, wobei angenommen wurde, „dass die Konstrukte, die von der Testperson verbalisiert werden, auch diejenigen sind, die sie verwendet, um ihre Welt zu analysieren“ (ebd., S. 38). In Bezug auf die Elemente wurde beachtet, dass beide Geschlechter vertreten sind, des Weiteren, dass positive und negative Rollenbezüge, persönliche Nähe und sowohl instrumentelle als auch emotionale Beziehungen vorkommen (vgl. ebd.). Auch ein Element „Ich, wie ich bin“ war notwendig, um Eigenschaftswörter zu generieren, die die eigene Person beschreiben. Die mittels dieses Verfahrens generierten Eigenschaftswörter sind zur Beschreibung von Persönlichkeiten

geeignet und können sowohl Gemeinsamkeiten von Personen beschreiben als auch zwischen ihnen differenzieren (vgl. ebd.).

4.1.2 Antwortformate

Generell kann man zwischen freiem und gebundenem Antwortformat unterscheiden, wobei beim freien Antwortformat die Testperson vollkommen frei ist, ihre Antwort „in einem allgemein verständlichen Zeichensystem selbst“ zu formulieren (Holoher-Ertl, 2003, S. 11).

Üblicherweise werden bei Persönlichkeitsverfahren gebundene Antwortformate verwendet, wobei man zwischen dichotomen, mehrkategorialen und kontinuierlichen Antwortformaten unterscheiden kann. Beim dichotomen Antwortformat (z.B. „ja“ vs. „nein“) wird die Testperson gezwungen, zwischen nur zwei Antwortmöglichkeiten zu wählen.

Der Nachteil, der sich hierbei ergeben kann (siehe z.B. eine Studie von Karner, 1993), ist, dass Personen, die sich zwischen zwei Alternativen entscheiden müssen und keine differenzierten Abstufungen treffen können, obwohl sie das wollen, auf diese »Freiheitsbeschränkung« willkürlich und unter Umständen ihren echten Eigenschaften zuwiderlaufend antworten (Holoher-Ertl, Kubinger & Menghin, 2003, S. 5f.).

4.1.3 Selbstbild/Selbstkonzept, Selbstdarstellung und soziale Erwünschtheit

Abgesehen vom Punkt der sozialen Erwünschtheit wurden die folgenden Punkte nicht als Theorie bei der Testentwicklung angeführt; sie seien hier jedoch kurz erwähnt, da es sich bei dem Verfahren ja um einen Fragebogen handelt, der auf einer Selbsteinschätzung beruht und die Validierung in der hier vorliegenden Arbeit einerseits anhand einer Selbsteinschätzung vorgenommen wird und andererseits anhand einer Stimmungsänderung, die möglicherweise kurzfristig einen Einfluss auf das Selbstbild hat.

Selbstbild/Selbstkonzept

Herkner (2001, S. 361) definiert: „Das Selbstbild (oder *Selbstkonzept*) ist das im Langzeitgedächtnis gespeicherte Wissen eines Menschen über sich selbst.“ Dieses komplexe Wissenssystem enthält sowohl universelle Anteile (in dem viele Menschen einer Kultur übereinstimmen) als auch individuell charakteristisches Wissen (vgl. Asendorpf, 2007, S. 263). Letzteres beinhaltet z.B. Wissen über den eigenen Namen, Persönlichkeitseigenschaften o.ä. und macht den größeren Anteil des Selbstkonzepts aus (vgl. ebd.).

Zur Frage, ob das Selbstbild stabil oder veränderlich ist, gibt es unterschiedliche Argumente. So sprechen für ein änderungsresistentes Selbstbild beispielsweise die Ergebnisse zur

selbstgefälligen Attribution: „Personen mit positivem Selbstbild führen eigene Erfolge in höherem Maß auf interne Ursachen (vor allem Fähigkeiten) zurück als Erfolge des Gegners. Bei eigenen Mißerfolgen spielt der (externe) Zufall eine größere Rolle als bei Mißerfolgen des Gegners“ (Herkner, 2001, S. 362). Durch solche Attributionen wird der Selbstwert stabilisiert (vgl. ebd.); im Zuge der hier vorliegenden Arbeit soll jedoch versucht werden, einer externen Zuschreibung von Misserfolgen entgegenzuwirken (vgl. Kap. 9).

Neben Ergebnissen zur Stabilität des Selbstbildes gibt es auch solche, die für ein variables Selbstbild sprechen, beispielsweise der starke Einfluss von Stimmungen auf Selbstwahrnehmung und Selbstbewertungen (vgl. Herkner, 2001, S. 362). Der scheinbare Widerspruch zwischen Stabilität und Variabilität löst sich auf, wenn man das Arbeitsselbstbild mit einbezieht, also den gerade aktivierten Teil des Selbstbildes: Zwar kann man davon ausgehen, dass das gesamte Selbstbild relativ stabil ist, aber nur die jeweils aktivierten Teile des Selbstbildes die Informationsverarbeitung und das Verhalten steuern (vgl. ebd., S. 362f.).

Selbstdarstellung

In Bezug auf Selbstdarstellung kann man zwei Tendenzen unterscheiden: Die Tendenz zur Selbstwerterhöhung und jene der Selbstkonsistenz. Der selbstwertdienliche Aspekt der Selbstdarstellung wäre dabei, sich so darzustellen, dass einem die erhaltenen Rückmeldungen anderer Personen schmeicheln. Nach dem Konsistenzaspekt hingegen stellen sich Personen so dar, wie sie tatsächlich zu sein glauben, um soziale Rückmeldungen zu erhalten, die das Selbstkonzept bestätigen (vgl. Asendorpf, 2007, S. 272). Allerdings muss es dabei nicht nur um eine Darstellung gegenüber anderen Personen gehen, denn in einem etwas weiter gefassten Sinn ist man selbst auch Publikum der Selbstdarstellung (vgl. Herkner, 2001, S. 375).

Soziale Erwünschtheit

Da Persönlichkeitseigenschaften meist wertbehaftet sind, besteht bei Eigenschaftsbeurteilungen immer die Gefahr, dass solche Urteile in sozial erwünschte Richtung hin verfälscht sind (vgl. Asendorpf, 2007, S. 142). Die Tendenz, bei Persönlichkeitsfragebögen sozial erwünscht zu antworten, ist vor allem dann zu erwarten, wenn die Ergebnisse mit Konsequenzen verbunden sind: „So muss in Fällen von Selektion und Konkurrenzauslese von einer absichtlichen Verfälschung in sozial erwünschte Richtung ausgegangen werden ...“ (Holoher-Ertl, 2003, S. 15).

Allerdings muss es nicht heißen, dass in anderen Fällen keine soziale Erwünschtheit auftritt, denn: „Sackheim und Gur (1978) zeigten, dass diese Eigenschaft⁴ bei genauerer Analyse aus zwei relativ unabhängigen Faktoren besteht: Tendenz zur Selbsttäuschung ... und Tendenz zur Fremdtäuschung ...“ (Asendorpf, 2007, S. 142).

4.2 Testaufbau und Verrechnung

Der Test besteht aus insgesamt 66 bipolaren Eigenschaftspaaren (z.B. weltoffen – engstirnig), die den Testpersonen auf einer Analogskala vorgegeben werden. Diese wird in späterer Folge dichotom verrechnet.

4.2.1 Vorgabe

Nach einer Instruktion und einem Beispielitem wird den Testpersonen jedes der 66 Eigenschaftspaare auf einer eigenen Bildschirmseite vorgegeben. Die Selbsteinschätzung der Testpersonen erfolgt dabei dadurch, dass sie eine Markierung auf dem Antwortbalken zwischen den beiden Worten verschieben und so ihre Zustimmung zu einem der beiden Worte ausdrücken. Zur Beantwortung der Items haben die Personen beliebig viel Zeit; eine Korrektur der Antwort zum jeweils vorherigen Item ist möglich, indem ein „Zurück“-Button betätigt wird (vgl. Holocher-Ertl, Kubinger & Menghin, 2003, S. 23).

Die positiven und negativen Eigenschaften sind sowohl links als auch rechts auf der Skala platziert, um durch diese unterschiedliche Polung einer möglichen Ja-Sage-Tendenz und absichtlichen Verfälschung entgegenzuwirken (vgl. Holocher-Ertl, 2003, S. 61).

Auch Extremantworten und das ständige Geben von mittleren Antworten wurden zu verhindern versucht: Wenn eine Testperson den ursprünglich genau mittig angesiedelten Balken gar nicht bewegt, erscheint die Rückmeldung: „Wenn diese Eigenschaft wirklich überhaupt nicht auf Sie zutrifft, drücken Sie bitte nochmals auf »Weiter«“ (Holocher-Ertl, Kubinger & Menghin, 2003, S. 23). Gibt die Testperson hingegen Extremantworten, erscheint im Item nach der fünften Extremeinschätzung eine Rückmeldung, in der die Testperson um eine ehrliche Selbsteinschätzung und Vermeiden von unrealistischen Extremantworten gebeten wird. Dieses Feedback erscheint jedoch nur, wenn sich die Testperson nicht bereits nach dem 30. Item befindet – damit wurde versucht zu verhindern, dass die Testperson diese Meldung nicht erst beim vorletzten Item bekommt, wenn sie darauf nicht mehr adäquat reagieren kann und ein solches Feedback nur zu Verunsicherung führen würde (vgl. ebd.).

⁴ Anm.: gemeint ist die Tendenz zur sozialen Erwünschtheit

4.2.2 Verrechnung

Um den Test nach dem dichotomen logistischen Modell nach Rasch auswerten zu können, war eine Dichotomisierung der Items erforderlich, wobei diese für jede Skala einzeln erfolgte. Bei manchen Skalen fand sich eine linksschiefe Verteilung, die mit einer Tendenz, sich sozial erwünscht darzustellen, erklärt werden kann. Entsprechend erfolgte die Dichotomisierung nicht notwendigerweise 50:50, sondern bspw. 30:60 oder 20:80 (vgl. Holocher-Ertl, 2003, S. 84). Bei der Skala „Emotionale Kontrolle“ war eine rechtsschiefe Verteilung zu beobachten, die ebenfalls mit der Tendenz, sich sozial erwünscht darzustellen, erklärt werden kann (vgl. Holocher-Ertl, 2003, S. 111).

Die Dichotomisierung erfolgte

- für die Dimensionen „Gefühlsbetontheit“ (Bezeichnung im *B5PO*: „Empathie“), „Offenheit“ und „Extraversion“ im Verhältnis 20:80,
- für die Dimension „Sorgfältigkeit“ (Bezeichnung im *B5PO*: „Gewissenhaftigkeit“) im Verhältnis 30:70,
- für die Dimension „Verträglichkeit“ im Verhältnis 50:50 und
- für die Dimension „Emotionale Kontrolle“ im Verhältnis 60:40 (vgl. Holocher-Ertl, 2003, S. 85ff.).

4.3 Kritik

Ein grundsätzlicher Kritikpunkt des Persönlichkeitsverfahrens *B5PO* liegt darin, dass es sich um einen Fragebogen handelt, weshalb Testwerte auch vor diesem Hintergrund interpretiert werden sollen: „Persönlichkeitsfragebögen erfassen nicht irgendwelche »wirklichen« Persönlichkeitseigenschaften, sondern die von einer Person bevorzugte Wahrnehmung und Darstellung ihrer selbst“ (Amelang & Schmidt-Atzert, 2006, S. 250). Das Problem der Verfälschbarkeit liegt dabei in der Augenscheinvalidität, also der Durchschaubarkeit, was erfasst werden soll. Kubinger schreibt hierzu (2003b, S. 429): „Während eine hohe Augenscheinvalidität bei Intelligenz- und Leistungstests in der Hauptsache nicht schadet und Projektive Verfahren kaum Augenscheinvalidität besitzen, sind Persönlichkeitsfragebögen kritisch: Weil zumeist durchschaubar, sind sie auch verfälschbar!“

Als zusätzlicher, etwas weiterführender Kritikpunkt kann angeführt werden, dass die Erfassung der Persönlichkeit in Fragebogenverfahren alleine nicht als ausreichend angesehen werden kann, wie Kubinger, Hofmann und Litzenberger (2002) in einer Studie zeigten, in der

sie den Persönlichkeitsfragebogen *NEO-FFI* anhand objektiver Persönlichkeitstests und eines semiprojektiven Verfahrens validierten: „Am Beispiel des Big Five-Persönlichkeitsmodells wird offensichtlich, dass die mit Persönlichkeitsfragebogen erfassten Eigenschaften grundsätzlich nicht ausreichen, um die Variabilitäten zwischen verschiedenen Persönlichkeiten zu erklären. Persönlichkeitsdiagnostik verlangt also nach einem nicht-einseitigen erfassungsmethodischen Zugang, sondern nach einem multi-methodischen“ (S. 65). Ein nicht intendiertes, aber doch sehr wichtiges Nebenergebnis der Studie zeigte weiters, dass die Skalen des *NEO-FFI* untereinander korrelierten, „besonders Neurotizismus scheint sich von Extraversion und Gewissenhaftigkeit nicht gut abgrenzen zu lassen“ (vgl. ebd., S. 64). Damit ist auch die Grundannahme, dass es sich bei den Big Five (und somit den Skalen im Verfahren *Big Five Plus One*) um voneinander unabhängige Dimensionen handelt, in Frage gestellt.

5. Validität

Validität gibt die Genauigkeit an, mit dem ein Test tatsächlich misst, was er messen soll. Ein Test, der etwas anderes misst als vermutet, ist unbrauchbar – wenn er es auch noch so genau misst (*Reliabilität*). Leider ist eben dies in der Psychologie durchaus keine Selbstverständlichkeit; so schreibt hierzu Kubinger (2003a, S. 199):

Anders als etwa in der Physik ist es bei psychologischen Messungen nicht selbstverständlich, dass das formale Ergebnis mit dem inhaltlichen übereinstimmt; Messungen in *cm* drücken immer eine Distanz zwischen zwei Punkten aus, Messungen in Schulnoten aber nicht notwendigerweise die erbrachten Leistungen in einem Prüfungsfach, sondern oft (auch) Fleiß u.ä.; Messungen in einem Test zur Raumvorstellung können (auch) vom verbalen Instruktionsverständnis abhängen!

Man kann zwischen drei verschiedenen Arten der Validität unterscheiden: inhaltliche Validität, kriterienbezogene Validität und Konstruktvalidität.

5.1 Inhaltliche Validität

Inhaltliche Validität ist dann gegeben, wenn der Test das zu erfassende Merkmal repräsentiert, also selbst das optimale Kriterium darstellt; dies wäre z.B. bei einer Schreibprobe für die Erfassung der Schnelligkeit und Genauigkeit beim Maschineschreiben der Fall (vgl. Lienert & Raatz, 1994, S. 10). Für diesen Validitätsaspekt lässt sich im Allgemeinen keine Maßzahl ermitteln (vgl. ebd., S. 11). Im Fall der hier vorliegenden Arbeit wäre der offenen Selbstbeschreibung der Testpersonen in einem Aufsatz ein gewisses Maß an inhaltlicher Validität zuzusprechen.

In der Praxis äußert sich inhaltliche Validität vor allem durch sogenannte „Experten-Ratings“: Dabei prüfen Experten „kraft ihrer Sachkompetenz, ob das konkrete Testmaterial (jedes einzelne Item) eine vorgegebene operationale Definition der Messintention erfüllt“ (Kubinger, 2003a, S. 199).

5.2 Kriterienbezogene Validität

Unter kriterienbezogener Validität, oder auch Kriteriumsvalidität, versteht man die Übereinstimmung des zu validierenden Verfahrens mit einem sogenannten Außenkriterium (vgl. Lienert & Raatz, 1994; Kubinger, 2003a; Kubinger, 2006). Dabei hat die

Kriteriumsvalidität den Vorteil, dass sie der einzige Validierungsansatz ist, der eine statistische Kennzahl liefert (vgl. Kubinger, 2006, S. 61).

Das Außenkriterium, mit dem der Test übereinstimmen soll, kann entweder ein anderer Test sein (in diesem Fall spricht man von „Übereinstimmungsvalidität“ oder „konkurrenter Validität“) oder ein Ereignis in der Zukunft sein („prognostische Validität“ bzw. „Vorhersagevalidität“; vgl. Kubinger, 2003a, S. 200).

Das Kriterium sollte „einen ausreichend validen Repräsentanten für das Persönlichkeitsmerkmal“ darstellen (Lienert & Raatz, 1994, S. 11). Dies kann jedoch problematisch sein; zieht man als Außenkriterium z.B. einen bereits vorhandenen Test heran, der den üblichen Gütekriterien nicht entspricht, ist er natürlich auch als Außenkriterium nicht geeignet. Nimmt man aber einen – im Sinne der Gütekriterien – „guten“ Test, stellt sich die Frage, wozu man überhaupt einen neuen Test validiert, denn dann besteht für diesen ja kein Bedarf (vgl. Kubinger, 2003a, S.200).

5.3 Konstruktvalidität

Wie für die inhaltliche Validität lässt sich auch für die Konstruktvalidität keine Maßzahl ermitteln (vgl. Lienert & Raatz, 1994, S. 11). Diese Form der Validierung zielt direkt auf die psychologische Analyse des Tests ab; dadurch ist diese Form weniger pragmatisch orientiert, sondern die Bedeutung liegt in einer „theoretischen Klärung dessen, was der betreffende Test misst“ (ebd., S. 226). Unter Konstrukten versteht man dabei „allgemein anerkannte, aber eben nicht direkt beobachtbare »Phänomene« wie z.B. Intelligenz, Angst oder Stress“ (Kubinger, 2006, S. 53).

Es gibt mehrere Möglichkeiten und Ansätze für die Konstruktvalidierung eines Tests: den klassischen Ansatz über die Faktorenanalyse, Campells und Fiskes Ansatz über die MTMM-Matrix und nicht-korrelative Ansätze. Der klassische Ansatz über die Faktorenanalyse spielt für die hier vorliegende Arbeit keine Rolle, daher sollen nur die anderen beiden kurz erwähnt werden.

5.3.1 Der Ansatz von Campbell und Fiske

Die Überlegung von Campbell und Fiske ist, mehrere Konstrukte unter mehreren Bedingungen zu erfassen, und zwar so, „dass nicht nur konstruktnahe und konstruktferne Tests zusätzlich zum eigentlichen interessierenden Test untersucht werden, sondern alle Tests auch unter verschiedenen Methoden/Bedingungen“ (Kubinger, 2006, S. 56). Durch die daraus bestimmbaren Korrelationen kann man eine Matrix (Multi-trait-multi-method-Matrix, kurz

MTMM-Matrix) bilden, in der man zwischen hetero- und monomethodischen und hetero- und monotrait Korrelationen unterscheiden kann (vgl. ebd.).

Für eine Validierung diagnostischer Verfahren mit Hilfe dieses Ansatzes wären folgende drei Punkte vorgesehen (vgl. hierzu Kubinger, Hofmann & Litzenberger, 2002, S. 63):

1. Korrelationen von Verfahren, die mit unterschiedlichen Methoden dieselbe Eigenschaft erfassen, sollen hoch sein (hetero-methodische-mono-trait-Korrelationen).
2. Die hetero-methodischen-mono-trait-Korrelationen müssen höher sein als Korrelationen von Verfahren, die mit derselben Methode unterschiedliche Merkmale erfassen (mono-methodische-hetero-trait-Korrelationen).
3. Die unter (1) erwähnten hetero-methodischen-mono-trait-Korrelationen müssen weiters höher sein als Korrelationen zwischen Verfahren, die mit unterschiedlichen Methoden unterschiedliche Merkmale erfassen (hetero-methodische-hetero-trait-Korrelationen).

„Der Vorteil von Multitrait-Multimethod-Analysen ist vor allem, dass Methodeneinflüsse auf die Eigenschaftsmessungen systematisch untersucht werden können“ (Asendorpf, 2007, S. 139). Dabei ist der Fall kritisch, dass Heterotrait-Monomethod-Korrelationen höher sind als Monotrait-Heteromethod-Korrelationen; die Daten sind dann mehr durch die Methode als durch den Inhalt bestimmt (vgl. ebd.).

5.3.2 Nicht-korrelative Ansätze

Zu den nicht-korrelativen Ansätzen zählen vor allem zwei Möglichkeiten, zu überprüfen, inwiefern das mit dem Test intendierte Konstrukt tatsächlich gemessen wird: Eine Möglichkeit ist die Methode des „lauten Denkens“, in der durch Selbst- oder Fremdbeobachtung geprüft werden soll, ob das Konstrukt erfasst wird. Die zweite Möglichkeit besteht in der Feststellung von Unterschieden im Test bei Personengruppen, die das Konstrukt in unterschiedlichem Ausmaß repräsentieren sollen – beispielsweise durch Extremgruppenvalidierung (vgl. Kubinger, 2003a, S. 199). „Wenn ein Test valide sein soll, dann muss er mit seiner Diagnose zumindest zwischen zwei Gruppen differenzieren“ (Kubinger, 2006, S. 64).

6. Inhaltsanalyse

Eine Inhaltsanalyse ist „...eine empirische Methode zur systematischen, intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen“ (Früh, 1991, zit. nach Lissmann, 2001, S. 16).

Nach Mayring (2003, S. 12f.) sind die Spezifika der Inhaltsanalyse als sozialwissenschaftliche Methode

- ein Analysieren von Kommunikation
- bei fixierter (festgehaltener/protokollierter) Kommunikation,
- wobei die Inhaltsanalyse systematisch vorgeht,
- regelgeleitet und
- theoriegeleitet,
- wodurch Rückschlüsse auf bestimmte Aspekte der Kommunikation gezogen werden sollen.

Häufig wird zwischen qualitativer und quantitativer Inhaltsanalyse unterschieden, wobei diese beiden Arten nicht als Gegensatz gesehen werden müssen: So beinhaltet eine Inhaltsanalyse sowohl qualitative als auch quantitative Aspekte, wobei das Aufspüren von relevanten Textstellen als qualitativ, das anschließende Zählen dieser Textstellen dann als quantitativ gilt (vgl. Lissmann, S. 16). Auch Mayring (2005) schreibt zur qualitativen Inhaltsanalyse, dass sie ein gutes Beispiel für Verknüpfung zwischen quantitativen und qualitativen Auswerteverfahren sei: „Denn die Schritte der Kategorienbildung und der Zuordnung von Kategorien zum Text sind eindeutig qualitative Schritte (vgl. Mayring 2002), in aller Regel werden dann aber Kategorienhäufigkeiten erhoben und quantitativ analysiert“ (S. 10).

6.1 Die quantitative Inhaltsanalyse

In der quantitativen Inhaltsanalyse wird versucht, das „Wortmaterial hinsichtlich bestimmter Aspekte (stilistische, grammatische, inhaltliche, pragmatische Merkmale) zu quantifizieren“ (Bortz & Döring, 2005, S. 147). Einzelne Textteile sollen dabei ausgewählten Kategorien zugeordnet werden, und die Eigenschaft eines Textes ergibt sich dadurch, wie viele Teile in die jeweiligen Kategorien fallen (vgl. ebd., S. 148).

Vorläufer der quantitativen Inhaltsanalyse (in Form von Häufigkeitsanalysen im Alten Testament) lassen sich bis ins 7. Jahrhundert zurückverfolgen (vgl. Lissmann, 2001, S. 14). Eine sozialwissenschaftliche Methode wurde die quantitative Inhaltsanalyse in den 20er und 30er Jahren des 20. Jahrhunderts; im Zuge der Kritik an der Einseitigkeit dieser Methode in

den 70er Jahren kamen qualitative Auswertungsmethoden hinzu oder wurden mit quantitativen Methoden kombiniert (vgl. Bortz & Döring, 2005, S. 149).

Kritiker/innen der quantitativen Inhaltsanalyse meinen, dass die oberflächliche „Wortzählerei“ den komplexen Bedeutungsgehalt von Texten nicht erfassen könne. Tatsächlich kann die Orientierung an nur einzelnen Begriffen zu Fehlschlüssen führen; so wurde z.B. in Deutschland – im Bemühen pornographische Inhalte aus den Datennetzen zu entfernen – aufgrund des Schlüsselbegriffes „Sex“ „auch die computervermittelte Auseinandersetzung über sexuellen Missbrauch blockiert“ (Bortz & Döring, 2005, S. 149).

Allerdings kann diese „Wortzählerei“ auch bei der Auswertung zu Problemen führen; als Beispiel sei die Häufigkeitsanalyse oder auch Frequenzanalyse angeführt, die einfachste Grundtechnik⁵ inhaltsanalytischen Arbeitens, bei der im Material bestimmte Elemente ausgezählt werden und hinsichtlich ihrer Häufigkeit mit dem Auftreten anderer Elemente verglichen werden (vgl. Mayring, 2003, S. 19). Für solche Häufigkeitsanalysen ist ein umfassendes Kategoriensystem (sog. Wörterbuch) notwendig, wobei dieses – bei computergestützter Auswertung – natürlich die verschiedenen grammatikalischen Abwandlungen erkennen können muss. Probleme sind aber die Mehrdeutigkeit mancher Begriffe oder deren inhaltliche Färbung durch den Kontext, Dialektfärbungen, substitutive Wörter (z.B. „davon“ – von was?) und „die Extensionsbestimmung durch den Kontext (bei »keine Angst«, »wenig Angst« und »viel Angst« wird jeweils einmal ‚Angst‘ gezählt)“ (Mayring, 2003, S. 14).

6.2 Die qualitative Inhaltsanalyse

In den letzten Jahrzehnten wurden in den Human- und Sozialwissenschaften immer komplexere quantitative Auswertungsmodelle mit rigiden Anforderungen an das Datenmaterial vorgeschlagen (vgl. Mayring, 2003, S. 9). „Andererseits forderte gerade dies zur Kritik heraus, zu Forderungen nach offenen Erhebungsmethoden, bei denen die Befragten wirklich zur Sprache kommen“ (ebd.).

Bei qualitativen Auswertungsverfahren wird versucht, „die manifesten und latenten Inhalte des Materials in ihrem sozialen Kontext und Bedeutungsfeld zu interpretieren“ (Bortz & Döring, 2005, S. 329). Dabei spielen Interpretationen und Deutungen eine große Rolle, wobei versucht wird, sich in die Lage der Agierenden hineinzusetzen. Interpretiert wird dabei

⁵ Als weitere Grundtechniken inhaltsanalytischen Arbeitens führt Mayring *Valenz- und Intensitätsanalysen* (Skalierung von Textteilen) und *Kontingenzanalysen* an (vgl. S. 14f.).

verbales bzw. nicht-numerisches Material, wobei die Interpretationen konsensfähig sein müssen, das heißt, von mehreren Forscher/innen, Expert/innen, Laien und/oder den Betroffenen selbst als zutreffend angesehen werden müssen. Damit streben qualitative Inhaltsanalysen eine intersubjektiv nachvollziehbare und möglichst erschöpfende Interpretation an (vgl. Bortz & Döring, 2005, S. 329ff.).

Bei der Interpretation steht das Kategoriensystem im Mittelpunkt, wobei dieses entweder induktiv aus dem Material gewonnen werden kann oder deduktiv bzw. theoriegeleitet erstellt wird. In der Praxis sind Mischformen häufig, bei denen zunächst a priori ein grobes Kategoriensystem erstellt wird, das dann nach Durchsicht des Materials ergänzt und verfeinert wird (vgl. ebd., S. 330).

Für die Kategorien als Analyseraster gilt dabei, dass sie abstrakter sein müssen als das Material, das sie ordnen sollen (vgl. Reinhoffer, 2005, S. 125). „Mindestens folgende Kriterien sind an Kategorien anzulegen: Sie sollen präzise benannt und definiert, mit Ankerbeispielen versehen ... und logisch gegeneinander abgegrenzt sein ...“ (ebd.).

III Empirischer Teil

7. Fragestellung und Hypothesen

Wie bereits in der Einleitung erwähnt, lautet die Fragestellung der vorliegenden Arbeit: „Erfasst das Persönlichkeitsinventar *B5PO* jene sechs Dimensionen, die es messen soll, indem es fähig ist zwischen jeweiligen Extremgruppen zu unterscheiden, und ist das Ergebnis jenes Persönlichkeitsinventars zumindest nicht in relevantem Ausmaß von momentanen zeitlichen Zuständen beeinflusst?“

Der erste Teil der Fragestellung bezieht sich darauf, ob der *B5PO* in der Lage ist, zwischen verschiedenen Extremgruppen zu unterscheiden, daher soll zunächst auf den diesbezüglichen Teil und seine Hypothesen etwas genauer eingegangen werden.

7.1 Validierung mit Hilfe der Extremgruppen

Die Bildung der Extremgruppen erfolgt mittels Aufsatz, in dem die Testpersonen gebeten werden, ihre Persönlichkeit möglichst genau und umfassend zu beschreiben. Dieser Aufsatz wird in späterer Folge mit Hilfe eines eigens dafür erstellten Kategoriensystems in Hinblick auf die Big Five der Persönlichkeit sowie in Hinblick auf Empathie ausgewertet. Daraufhin werden je Persönlichkeitsdimension Extremgruppen gebildet, wobei circa⁶ die mittleren 50% entfallen. Für die Skala „Extraversion“ werden so z.B. nur die introvertiertesten 25% und die extravertiertesten 25% berücksichtigt.

Im Anschluss werden diese Extremgruppen anhand ihrer Werte im Persönlichkeitsverfahren *B5PO* verglichen, wobei überprüft wird, ob sich diese Gruppen im Test um mindestens eine Standardabweichung (10 *T*-Werte) voneinander unterscheiden. Ein kleinerer Unterschied wird als irrelevant betrachtet, da es sich ja um die beiden Extrempole handelt.

Die zugrunde liegenden Hypothesen lauten daher:

1. Hypothese:

H₀: Die Testpersonen, die sich in einem Aufsatz als extravertiert beschreiben, stellen sich im Persönlichkeitsverfahren Big Five Plus One nicht oder in geringerem Ausmaß als 10 T-Werte extravertierter dar als jene Testpersonen, die sich im Aufsatz als introvertiert beschreiben.

⁶ Der Wert kann nur näherungsweise angegeben werden, da die endgültige Bildung der Extremgruppen erst nach Auswertung der Aufsätze möglich war. Sicherheitshalber wurde die Stichprobe auch geringfügig vergrößert, da bereits während der Datenerhebung sichtbar war, dass sich viele Testpersonen nicht auf allen Skalen beschrieben; dabei lag die Priorität stärker auf einer optimalen Zahlengröße der Extremgruppen als auf einer exakten Prozentsatzgröße. Zur endgültigen Extremgruppeneinteilung siehe Kap. 11.4.

H₁: Die Testpersonen, die sich in einem Aufsatz als extravertiert beschreiben, unterscheiden sich hinsichtlich ihrer Testwerte auf der Skala „Extraversion“ des Persönlichkeitsinventars Big Five Plus One um mindestens 10 T-Werte in Richtung mehr extravertiert von den Testpersonen, die sich im Aufsatz als introvertiert beschreiben.

2. Hypothese:

H₀: Die Testpersonen, die sich in einem Aufsatz als verträglich beschreiben, stellen sich im Persönlichkeitsverfahren Big Five Plus One nicht oder in geringerem Ausmaß als 10 T-Werte verträglicher dar als jene Testpersonen, die sich im Aufsatz als wenig verträglich beschreiben.

H₁: Die Testpersonen, die sich in einem Aufsatz als verträglich beschreiben, unterscheiden sich hinsichtlich ihrer Testwerte auf der Skala „Verträglichkeit“ des Persönlichkeitsinventars Big Five Plus One um mindestens 10 T-Werte in Richtung mehr verträglich von den Testpersonen, die sich im Aufsatz als wenig verträglich beschreiben.

3. Hypothese:

H₀: Die Testpersonen, die sich in einem Aufsatz als gewissenhaft beschreiben, stellen sich im Persönlichkeitsverfahren Big Five Plus One nicht oder in geringerem Ausmaß als 10 T-Werte gewissenhafter dar als jene Testpersonen, die sich im Aufsatz als wenig gewissenhaft beschreiben.

H₁: Die Testpersonen, die sich in einem Aufsatz als gewissenhaft beschreiben, unterscheiden sich hinsichtlich ihrer Testwerte auf der Skala „Gewissenhaftigkeit“ des Persönlichkeitsinventars Big Five Plus One um mindestens 10 T-Werte in Richtung mehr gewissenhaft von den Testpersonen, die sich im Aufsatz als wenig gewissenhaft beschreiben.

4. Hypothese:

H₀: Die Testpersonen, die sich in einem Aufsatz als emotional kontrolliert beschreiben, stellen sich im Persönlichkeitsverfahren Big Five Plus One nicht oder in geringerem Ausmaß als 10 T-Werte emotional kontrollierter dar als jene Testpersonen, die sich im Aufsatz als wenig emotional kontrolliert beschreiben.

H₁: Die Testpersonen, die sich in einem Aufsatz als emotional kontrolliert beschreiben, unterscheiden sich hinsichtlich ihrer Testwerte auf der Skala „Emotionale Kontrolle“ des Persönlichkeitsinventars Big Five Plus One um mindestens 10 T-Werte in Richtung mehr emotional kontrolliert von den Testpersonen, die sich im Aufsatz als wenig emotional kontrolliert beschreiben.

5. Hypothese:

H₀: Die Testpersonen, die sich in einem Aufsatz als offen für Erfahrung beschreiben, stellen sich im Persönlichkeitsverfahren Big Five Plus One nicht oder in geringerem Ausmaß als 10 T-Werte offener für Erfahrung dar als jene Testpersonen, die sich im Aufsatz als wenig offen für Erfahrung beschreiben.

H₁: Die Testpersonen, die sich in einem Aufsatz als offen für Erfahrung beschreiben, unterscheiden sich hinsichtlich ihrer Testwerte auf der Skala „Offenheit“ des Persönlichkeitsinventars Big Five Plus One um mindestens 10 T-Werte in Richtung mehr offen für Erfahrung von den Testpersonen, die sich im Aufsatz als wenig offen für Erfahrung beschreiben.

6. Hypothese:

H₀: Die Testpersonen, die sich in einem Aufsatz als empathisch beschreiben, stellen sich im Persönlichkeitsverfahren Big Five Plus One nicht oder in geringerem Ausmaß als 10 T-Werte empathischer dar als jene Testpersonen, die sich im Aufsatz als wenig empathisch beschreiben.

H₁: Die Testpersonen, die sich in einem Aufsatz als empathisch beschreiben, unterscheiden sich hinsichtlich ihrer Testwerte auf der Skala „Empathie“ des Persönlichkeitsinventars Big Five Plus One um mindestens 10 T-Werte in Richtung mehr empathisch von den Testpersonen, die sich im Aufsatz als wenig empathisch beschreiben.

Allerdings stellt sich die Frage, ob der Aufsatz selbst einen Einfluss auf die Testergebnisse im Persönlichkeitsinventar B5PO hatte. Dies könnte dadurch möglich sein, dass sich die Testpersonen im Sinn einer instrumentellen Reaktivität durch das Verfassen des Aufsatzes verstärkt gedanklich mit der eigenen Persönlichkeit auseinandersetzen und sich daraufhin im anschließenden Fragebogen anders beurteilten.

Daher lautet die **7. Hypothese**:

H₀: Testpersonen, die zusätzlich zur Bearbeitung des Verfahrens B5PO vorher einen Aufsatz über ihre Persönlichkeit verfassten, unterscheiden sich hinsichtlich ihrer Testwerte nicht von den Testpersonen, die zuvor keinen Aufsatz über ihre Persönlichkeit verfassten.

H₁: Testpersonen, die zusätzlich zur Bearbeitung des Verfahrens B5PO vorher einen Aufsatz über ihre Persönlichkeit verfassten, unterscheiden sich hinsichtlich ihrer Testwerte von den Testpersonen, die zuvor keinen Aufsatz über ihre Persönlichkeit verfassten.

Im Sinne des zweiten Teils der Fragestellung (bezogen auf die State-Abhängigkeit) würde die Annahme der siebenten Alternativhypothese für eine State-Abhängigkeit des Testergebnisses sprechen, da die Testpersonen dann durch das Schreiben des Aufsatzes in einen anderen Zustand versetzt wurden, wodurch sich eine andere Selbstdarstellung im Verfahren erklären könnte. Daher wird diese Hypothese später auch bei der Darstellung der State-Abhängigkeit beantwortet.

7.2 Erfassung eines möglichen State-Einflusses

Um eine State-Komponente herauszufiltern, ist es wichtig, einen Teil der Testpersonen systematisch in eine andere Stimmung zu versetzen als den anderen Teil. Im Anschluss daran soll das Verfahren *B5PO* vorgegeben werden, um zu überprüfen, ob sich diese Gruppen hinsichtlich ihrer Testwerte unterscheiden; dies würde für einen Einfluss von States auf das Testergebnis sprechen. Dabei wird davon ausgegangen, dass sich die randomisierten Versuchsgruppen vor einer Stimmungsänderung im Durchschnitt nicht voneinander unterscheiden, ein Unterschied in den Testergebnissen zwischen den Gruppen also nur durch die Stimmungsänderung im Sinne einer Änderung des Zustandes (State) bedingt sein kann.

Um eine Stimmungsänderung bei den Testpersonen der Versuchsgruppen zu erreichen soll eine Versuchsgruppe durch das Gefühl des Versagens frustriert werden: Bei einer Gruppentestung in Schulklassen werden Suchbilder ausgeteilt, die auf den ersten Blick völlig gleich aussehen, sich jedoch darin voneinander unterscheiden, dass eine Hälfte der Suchbilder lösbar, die andere Hälfte unlösbar ist. Die Frustration, die bei dem Teil der Testpersonen auftritt, die das unlösbare Rätsel bekommen (VG 2), soll noch dadurch verstärkt werden, dass sie mitbekommen sollen, dass andere (welche die lösbare Variante bekamen; VG 1) die Aufgabe bewältigen konnten.

Zusätzlich wird eine Kontrollgruppe benötigt, um sicherzustellen, dass nicht das Bearbeiten bzw. die Vorgabe des Rätsels selbst Einfluss auf die Testergebnisse hatte. In solch einem Fall wäre natürlich ein Unterschied zwischen den beiden Rätselgruppen zu relativieren, die State-Abhängigkeit des *B5PO* wäre dann aber trotzdem gegeben: Im Sinne einer reinen Erfassung der sechs Trait-Dimensionen sollten sich die Testpersonen – unabhängig davon, was sie vorher getan haben – im Mittel gleich auf den sechs Dimensionen darstellen. Daher würde nicht nur die Frustration, sondern auch eine vorherige Bearbeitung des Rätsels, sofern sie einen Einfluss auf die Darstellung im *B5PO* hat, gegen die ausschließliche Erfassung von Traits sprechen.

Die zugrunde liegenden Hypothesen für die Erfassung eines möglichen State-Einflusses lauten daher:

8. Hypothese:

H₀: Ein vorangehendes frustrierendes Ereignis hat keinen Einfluss auf die Darstellung der Testpersonen im Persönlichkeitsverfahren Big Five Plus One.

H₁: Bei vorangehender Frustration stellen sich die Testpersonen im Persönlichkeitsverfahren B5PO anders dar als jene Testpersonen, die zuvor nicht frustriert wurden.

Sowie für den Fall, dass die Bearbeitung des Rätsels selbst einen Einfluss auf die Selbstdarstellung der Testpersonen ausübt:

9. Hypothese:

H₀: Die Bearbeitung eines Rätsels vor Bearbeitung des Persönlichkeitsverfahrens B5PO hat keinen Einfluss auf die Selbstdarstellung der Testpersonen in diesem Verfahren.

H₁: Nach Bearbeitung eines Rätsels stellen sich die Testpersonen im Persönlichkeitsverfahren B5PO anders dar als jene Testpersonen, die kein Rätsel bearbeitet haben.

8. Methode und Stichprobenberechnung

Je nach Hypothese werden unterschiedliche Methoden für die Auswertung herangezogen.

Für die Validierung mithilfe der Extremgruppen (Hypothesen 1–6) wird je Dimension ein *t*-Test bei einseitiger Fragestellung gerechnet, sofern die Varianzen homogen sind, als abhängige Variablen dienen jeweils die *T*-Werte der interessierenden Skala des *B5PO*, unabhängige Variable ist die Zuteilung zur jeweiligen Extremgruppe. Sollte keine Gleichheit der Varianzen vorliegen, wird stattdessen der Welch-Test gerechnet (vgl. Kubinger & Rasch, 2006, S. 183). Die einseitige Fragestellung begründet sich aus der logisch gegebenen Gerichtetheit der Hypothesen: Testpersonen, die sich beispielsweise in einem offenen Antwortformat als extravertiert beschreiben, sollen sich im *B5PO* extravertierter (und nicht etwa introvertierter) darstellen als jene Testpersonen, die sich im Aufsatz introvertiert darstellen.

Für die Beantwortung der Hypothese 7 (ob das Schreiben des Aufsatzes selbst einen Einfluss auf die Ergebnisse im *B5PO* hatte) soll bei Homogenität der Varianzen ebenfalls ein *t*-Test herangezogen werden, diesmal jedoch zweiseitig, da es hier keine theoretisch ableitbare Annahme darüber gibt, in welche Richtung sich die Testpersonen anders darstellen würden. Als abhängige Variablen hierfür dienen die *T*-Werte der sechs Skalen des *B5PO*, unabhängige Variable ist die Versuchsbedingung (Aufsatz vs. kein Aufsatz).

Um einen möglichen State-Einfluss durch die Rätselvorgabe (Hypothese 9) zu überprüfen, wird (Homogenität der Varianzen vorausgesetzt) auf eine einfache Varianzanalyse als Methode zurückgegriffen, die *T*-Wert-Ausprägung der sechs Dimensionen des *B5PO* sind wieder die abhängigen Variablen. Für die Erfassung dieses State-Einflusses gibt es drei Gruppen: Versuchsgruppe (VG) 1, VG 2 und Kontrollgruppe (KG), die Gruppenzugehörigkeit ist die unabhängige Variable. Eine weitere Möglichkeit der Berechnung besteht auch darin, die beiden Versuchsgruppen zusammenzufassen und gemeinsam mit der Kontrollgruppe mit Hilfe eines *t*-Tests zu vergleichen.

Hypothese 8, der Einfluss durch die Frustrationsbedingung, lässt sich wieder mit einem *t*-Test bei zwei Gruppen (durch das Rätsel frustrierte sowie durch das Rätsel nicht frustrierte Schüler/innen) oder durch eine Varianzanalyse bei drei Gruppen (durch das Rätsel Frustrierte, durch das Rätsel nicht frustrierte Versuchsgruppen sowie eine Kontrollgruppe) überprüfen. Abhängige Variablen sind wieder die *T*-Werte des *B5PO*.

Stichprobenberechnung

Die Stichprobengröße wurde im Vorhinein unter Berücksichtigung der relevanten Effekte mit dem Statistikprogramm *CADEMO Light* (Version 3.27) berechnet.

Für die Validierung des Aufsatzes wird ein Unterschied von 10 T -Werten als relevanter Effekt betrachtet, da Extrempole miteinander verglichen werden. Es ergibt sich damit ein Stichprobenumfang von $n = 14$ pro Gruppe, also $n_{\Sigma} = 28$. Da die mittleren 50% jedoch ausgeschieden werden sollen und somit für die Berechnung nicht in Frage kommen, ergibt sich ein Gesamtumfang von $n_{\Sigma} = 56$.

Für den Einfluss von States im Sinne einer Frustration auf das Testergebnis wird eine Abweichung von mehr als zwei Drittel der Standardabweichung von T -Werten als relevant erachtet (7 T -Werte). Es errechnete sich für die Varianzanalyse eine Stichprobengröße von $n = 41$ pro Gruppe, gesamt also $n_{\Sigma} = 123$.

Die Stichprobengrößen für den Einfluss des Aufsatzes oder des Rätsels selbst würden weit unter den zu erhebenden Stichprobengrößen liegen ($n = 34$ pro Gruppe bei einer Effektgröße von $T = 7$), daher wird im umgekehrten Weg bei Vorliegen von Signifikanz die Effektgröße später dargestellt und deren Relevanz diskutiert. Der Einfluss von Aufsatz und Rätsel wird mit t -Tests berechnet, wobei die Versuchsbedingungen (Versuchsgruppen: frustrierte und nicht frustrierte sowie die Kontrollgruppe) die unabhängigen Variablen darstellen, abhängige Variablen sind die T -Werte der sechs Dimensionen des *B5PO*.

9. Versuchsplan

Zur Gewinnung der Testpersonen wurden verschiedene AHS in Wien und Niederösterreich angeschrieben. Das Informationsschreiben zur Untersuchung für den Stadt- und Landesschulrat findet sich in Anhang A. In Anhang A findet sich weiters die Einverständniserklärung für Eltern der teilnehmenden Schüler/innen; hier war es allerdings nicht immer möglich, alle Einverständniserklärungen zurückzuerhalten, teilweise nahmen zusätzliche Schulklassen auch spontan an der Testung teil. Da die Schüler/innen sich jedoch bereits in begrenzt rechtsfähigem Alter befanden, wurde ihnen die Kompetenz zugestanden, dies durchaus auch alleine entscheiden zu können.

Aufgrund der vorliegenden Normen für das Verfahren *B5PO* für Personen ab 16 Jahren wurden entsprechend nur Schüler/innen der 6. AHS und darüber getestet. Die Entscheidung, nur an allgemeinbildenden, nicht jedoch an berufsbildenden Schulen (HAK, HTL, ...) zu testen, wurde deshalb getroffen, weil Schüler/innen einer speziellen berufsbildenden Schule möglicherweise in verstärktem Maß gleiche Interessen und Persönlichkeitseigenschaften aufweisen könnten als Schüler/innen in einer allgemeinbildenden Schule, womit die Aussagekraft der Ergebnisse eingeschränkt wäre.

9.1 Materialien und Vorgabe

Je nachdem, welche Hypothesen überprüft werden sollten, wurden unterschiedliche Materialien vorgegeben. So gab es zwei grundsätzlich verschiedene Vorgangsweisen: Jene für die Hypothesen 1–7 und jene für die Hypothesen 8 und 9. Das Verfahren *B5PO* wurde in jedem Fall vorgegeben, für die Überprüfung der Hypothesen 1–7 bekam die Versuchsgruppe zusätzlich einen Aufsatz zu bearbeiten, für Überprüfung der Hypothesen 8 und 9 bekamen die Versuchsgruppen anstelle des Aufsatzes ein Suchbild vorgelegt.

9.1.1 Das Verfahren *B5PO*

Jeder Testperson, egal ob Versuchs- oder Kontrollgruppe, wurde dieses Verfahren vorgegeben. Den Versuchsgruppen bearbeiteten es an zweiter Stelle (nach Bearbeiten ihres Suchbildes oder Aufsatzes). Den Kontrollgruppen wurde gleich zu Beginn der Testung nur das Verfahren *B5PO* vorgegeben und danach noch ein anderes Verfahren für eine andere

Untersuchung⁷. Für die Validierung des Persönlichkeitsinventars sind bei den Kontrollgruppen jedoch nur die Ergebnisse bis zur Beendigung des *B5PO* von Relevanz. Das Verfahren wurde bereits in Kap. 4 des Theorieteils ausführlich beschrieben, so dass hier nicht erneut darauf eingegangen werden soll. Zur Veranschaulichung sei jedoch eines der 66 Items dargestellt (Abb. 1).

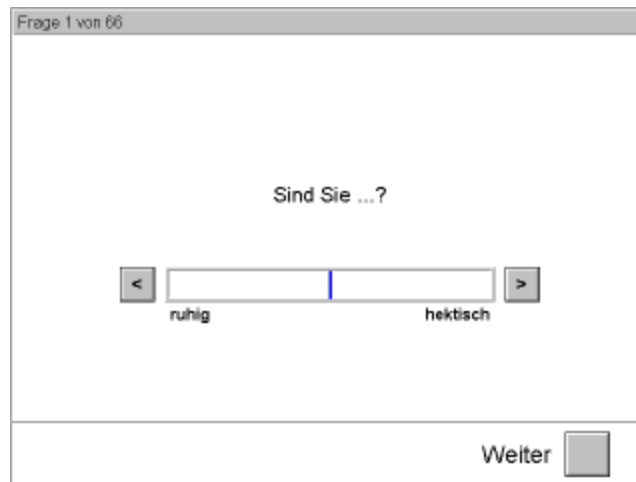


Abb. 1: Item 1 des Persönlichkeitsverfahrens B5PO

9.1.2 Aufsatz

In einem offenen Antwortformat sollten die Testpersonen sich selbst beschreiben. Die Instruktion dafür lautete:

Im Folgenden soll Ihre Persönlichkeit erfasst werden.

Bitte schreiben Sie einen kurzen Aufsatz von 1-2 Seiten mit dem Titel „Wer bin ich – wie bin ich?“.

In diesem Aufsatz sollen Sie versuchen, Ihre Persönlichkeit unter Verwendung von Eigenschaftswörtern möglichst genau und umfassend zu beschreiben (auch negative Eigenschaften).

Denken Sie daran, dass Sie Ihre Persönlichkeitseigenschaften beschreiben sollen, nicht Ihre Tätigkeit, Ihr Aussehen oder Ihre momentane Stimmung.

Die Auswertung erfolgt anonym, für die Zuordnung und den Vergleich mit anderen Verfahren ist es jedoch erforderlich, dass Sie den Ihnen zugeteilten Code angeben.

Vielen Dank für die Mitarbeit!

Ein Beispielbogen für den Aufsatz findet sich in Anhang A.

Der Text der Instruktion wurde absichtlich kurz und sehr offen und ohne Beispiele formuliert, um die Testpersonen nicht in eine Richtung zu drängen. Einzige Ausnahme war der Hinweis

⁷ Die Schüler/innen der Kontrollgruppe wurden im Anschluss an die Vorgabe des *B5PO* noch für eine andere Untersuchung benötigt (vgl. Schneiderbauer, 2008, in Vorbereitung), die allerdings nicht im Zusammenhang mit der hier vorliegenden Fragestellung steht, weshalb darauf nicht näher eingegangen werden soll.

auf das Erwähnen negativer Eigenschaften, um einem zu sehr sozial erwünschten Bild vorzubeugen.

Auch wenn die Testpersonen mündlich einheitlich geduzt wurden (entsprechend ihrem doch recht jungen Alter und dem Zeitgeist), wurde die Instruktion in formaler Form mit „Sie“ gewählt, um ein gewisses Maß an Wertschätzung bei höherem Anonymitätsgrad zu geben.

Die Vorgabedauer war unbeschränkt, um die Personen nicht unter Druck zu setzen und ihnen genügend Zeit zu geben, über sich zu reflektieren. Nach Ablauf einer gewissen Zeit (zumeist rund eine halbe Stunde), wenn ein großer Teil der Testpersonen den Aufsatz bereits fertig hatte, wurde bei Notwendigkeit darauf hingewiesen, dass der Aufsatz in nächster Zeit beendet werden sollte – aber wieder ohne Vorgabe einer Deadline und mit dem Kommentar, sich nicht zu stressen.

9.1.2 Suchbild

Für die Überprüfung der Hypothesen 8 und 9 wurde vor Bearbeitung des *B5PO* für die Versuchsgruppen anstelle des Aufsatzes jeweils ein Suchbild vorgegeben. Es gab dabei zwei verschiedene Versionen für das Suchbild, die abwechselnd ausgeteilt wurden, den Schüler/innen sollten sie jedoch wie ein und dieselbe Version erscheinen. Das Motiv des Bildrätsels war das gleiche, Unterschiede bestanden jedoch in der Anzahl eingebauter Fehler (gefunden werden sollten insgesamt acht Fehler): Eine Version beinhaltete zehn Unterschiede, die andere Version beinhaltete nur sechs. Durch die Version mit zehn Fehlern sollte denjenigen Testpersonen, die diese Version erhielten, die Möglichkeit eingeräumt werden, das Rätsel tatsächlich bewältigen zu können. Die Version mit nur sechs eingebauten Fehlern war hingegen unlösbar, wodurch die Testpersonen frustriert werden sollten. (Die beiden Versionen des Suchbildes inkl. Instruktion finden sich in Anhang A.)

Die Instruktion für das Suchbild lautete für beide Versionen einheitlich:

Auf der folgenden Seite finden Sie zwei Zeichnungen, die auf den ersten Blick gleich aussehen. Im unteren Bild verstecken sich jedoch einige Unterschiede.

Ihre Aufgabe ist es nun, die acht Unterschiede in der unteren Zeichnung zu finden und zu markieren. Sie haben dafür insgesamt acht Minuten Zeit.

Die Auswertung erfolgt anonym, für die Zuordnung und den Vergleich mit anderen Verfahren ist es jedoch erforderlich, dass Sie den Ihnen zugeteilten Code angeben.

Bitte tragen Sie Ihren Code gleich in das rechte obere Eck dieser Seite ein und blättern Sie erst um, wenn das Zeichen dazu gegeben wird.

Sobald Sie die acht Unterschiede gefunden und markiert haben, geben Sie bitte Ihren Bogen ab.

Vielen Dank für die Mitarbeit!

Der Hinweis auf Abgabe in der Instruktion wurde deshalb gewählt, damit die Versuchsgruppe, die das unlösbare Rätsel hatte, bemerkt, dass andere durchaus in der Lage sind, das Rätsel zu lösen, wodurch sie das Versagen auf die eigene Leistung zurückführen sollten. Durch dieses Gefühl des persönlichen Versagens sollte die Frustration in dieser Gruppe verstärkt werden.

9.1.3 Überblick

Im Folgenden sei in Form einer Tabelle (Tab. 1) überblicksartig dargestellt, wie sich der Ablauf der Untersuchung (inklusive Testvorgabe) für die einzelnen Versuchsgruppen gestaltete.

	Hypothesen 1-7		Hypothesen 8 und 9		
	VG	KG	VG1	VG2	KG
	Aufsatz	<i>B5PO</i>	Suchbild mit 10 Fehlern	Suchbild mit 6 Fehlern	<i>B5PO</i>
	<i>B5PO</i>	Tests für eine andere Untersuchung	<i>B5PO</i>	<i>B5PO</i>	Tests für eine andere Untersuchung
→ Tests im Ablauf	Fakultativ: Ausprobieren anderer Tests		Fakultativ: Ausprobieren anderer Tests	Fakultativ: Ausprobieren anderer Tests	

Tab. 1: Vorgabe der Materialien (Überblick)

Da die Datenerhebung gemeinsam mit einer anderen Untersuchung stattfand und die Testpersonen, die einen Aufsatz schreiben oder ein Rätsel bearbeiten sollten, auch gerne die Tests der parallel stattfindenden Untersuchung ausprobieren wollten, wurde allen Testpersonen die Möglichkeit gegeben, nach Beendigung des *B5PO* ebenfalls die Tests der anderen Untersuchung durchzuführen, wobei sie sich aussuchen durften, welche davon sie am meisten interessierten. Darauf wurden sie auch vor der Untersuchung hingewiesen, um einer möglichen Frustration und Non-Compliance („Die anderen dürfen so lustige Tests machen und ich muss einen Aufsatz schreiben, das interessiert mich gar nicht“) entgegenzuwirken.

9.2 Einteilung der Versuchsgruppen

Aufgrund der Tatsache, dass Computerverfahren verwendet wurden, fanden die Testungen jeweils in den Informatikräumen der Schulen statt. Die Klassen wurden dabei beim Eintritt in die PC-Räume in die Versuchsgruppen und Kontrollgruppen aufgeteilt: Jede Testperson bekam beim Eintreten eine Karte mit einem Code, wobei die Codes in abwechselnder, regelmäßiger Reihenfolge sortiert waren. Die Codes setzten sich aus Zahlen (Code-Nummer für Schule und laufende Testpersonen-Nummer) zusammen sowie aus einem Buchstaben, der die Zugehörigkeit zur jeweiligen Versuchs- bzw. Kontrollgruppe festlegte (für die hier vorliegende bzw. die parallel stattfindende Untersuchung). Dadurch sollten „Grüppchenbildungen“ in den jeweiligen Versuchsgruppen vermieden werden, vor allem aufgrund der Annahme, dass Freundesgruppen den Raum relativ gleichzeitig als Gruppe betreten würden und dadurch sonst eine Ex-post-facto-Situation auftreten würde.

Die Code-Karten dienten gleichzeitig der Wahrung der Anonymität, denn in weiterer Folge musste anstelle des Namens stets nur diese Nummer angegeben werden. Auch die Rückmeldung (jede Testperson hatte im Nachhinein die Möglichkeit, individuell ihr Testresultat zu erfahren) erfolgte mit Hilfe der Code-Kärtchen.

10. Stichprobe

Die Stichprobe wurde an insgesamt sieben AHS in Wien und Niederösterreich (Wien Umgebung, insgesamt drei Schulen) zwischen Mai und Dezember 2006 erhoben, größtenteils in den Informatikräumen der Schulen selbst und teilweise in den Computerdiagnostik-Räumen der Test- und Beratungsstelle der Fakultät für Psychologie.

Für die Fragestellung der vorliegenden Arbeit wurden insgesamt 370 Schüler/innen zwischen 15 und 19 Jahren getestet, wobei aufgrund von Problemen bei der Datenspeicherung in den Schulen 19 Datensätze verloren gingen, es bleiben also noch 351 Schüler/innen⁸.

Aus diesen 351 wurden die Stichproben für die Extremgruppenvalidierung und jene für die Überprüfung eines möglichen State-Einflusses gebildet. In der Regel überschneiden sich die Kontrollgruppen für die beiden Teilfragestellungen nicht, da aus pragmatischen Gründen Suchbild- und Aufsatztestungen getrennt durchgeführt wurden. Nur ein einziges Mal war eine gleichzeitige Testung möglich, weshalb sechs Testpersonen der Kontrollgruppen für die Extremgruppenvalidierung gleichzeitig Testpersonen der Kontrollgruppen für die Rätseltestungen sind.

10.1 Stichproben für die Extremgruppenvalidierung

Für die Extremgruppenvalidierung beträgt die gesamte Stichprobengröße 175, davon 62 in der Versuchsgruppe (jene, die einen Aufsatz bearbeiteten) und 113 in der gesamten Kontrollgruppe (nur das Verfahren *B5PO* wurde vorgegeben). Diese gesamte Kontrollgruppe setzt sich aus zwei Teilkontrollgruppen (KG 1: $n_1 = 60$, davon 6 Überschneidungen mit KG 2; KG 2: $n_2 = 59$, davon 6 Überschneidungen mit KG 1) zusammen, die jedoch alle dieselben Bedingungen hatten, randomisiert zugeteilt wurden und gleichzeitig mit der Versuchsgruppe getestet wurden. Der Grund, dass es zwei Kontrollgruppen gibt, liegt darin, dass im Zuge der gemeinsamen Datenerhebung mit einer anderen Untersuchung die Kontrollgruppen zumeist doppelt so groß ausfielen wie die Versuchsgruppen, weshalb sie gleich während der Testung in zwei Teilkontrollgruppen gesplittet wurden. Die Kontrollgruppe wird einzig dafür benötigt, um zu überprüfen, ob die Vorgabe bzw. Bearbeitung des Aufsatzes selbst einen Einfluss auf die Darstellung im *B5PO* hatte, hierfür kann also die gesamte Teilstichprobe herangezogen werden.

⁸ Im Folgenden wird die Stichprobengröße immer nur auf diejenigen Schüler/innen bezogen, von denen alle für die Fragestellung relevanten Daten vorhanden waren, die anderen wurden von der Berechnung ausgeschlossen.

Zur Verteilung der Geschlechter in Versuchs- und Kontrollgruppe ist zu sagen, dass sich bei der Versuchsgruppe und bei den beiden geteilten Kontrollgruppen keine signifikant unterschiedliche Geschlechterverteilung findet. Ein signifikanter Unterschied findet sich jedoch, wenn die beiden Kontrollgruppen zu einer gemeinsamen Kontrollgruppe zusammengefasst werden; in diesem Fall dominiert die Anzahl weiblicher Schüler/innen signifikant. Auch in der Gesamtstichprobe dominieren die weiblichen Testpersonen signifikant. Im Folgenden seien die Geschlechterverteilung (Tab. 2) und die SPSS-Outputs der signifikanten Chi-Quadrat-Tests (bei $\alpha = 5\%$) auf Gleichverteilung dargestellt (Tab. 3 und 4).

	<i>n</i> männlich	<i>n</i> weiblich	Signifikanz ($\alpha = 5\%$)
Versuchsgruppe	27 (44%)	35 (56%)	nein
Kontrollgruppe 1	26 (43%)	34 (57%)	nein
Kontrollgruppe 2	23 (39%)	36 (61%)	nein
gesamte KG	45 (40%)	68 (60%)	ja
gesamte Stichprobe	72 (41%)	103 (59%)	ja

Tab. 2: Geschlechterverteilung der Extremgruppenvalidierung (Prozentwerte gerundet)

Test Statistics

	Geschlecht
Chi-Square ^a	5,491
df	1
Asymp. Sig.	,019

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 87,5.

Tab. 3: Chi-Quadrat-Test auf Gleichverteilung der Geschlechter über die gesamte Stichprobe

Test Statistics

	Geschlecht
Chi-Square ^a	4,681
df	1
Asymp. Sig.	,030

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 56,5.

Tab. 4: Chi-Quadrat-Test auf Gleichverteilung der Geschlechter über die gesamte KG

Für die Bildung der Extremgruppen werden nur die Schüler/innen der Versuchsgruppe benötigt: Hier werden aus der Versuchsgruppe anhand des Aufsatzes sechsmal (auf allen

sechs Dimensionen) jeweils zwei Extremgruppen gebildet, circa die mittleren 50% werden dabei jedesmal ausgeschlossen. Da die Extremgruppenbildung erst bei der Auswertung der Aufsätze erfolgte, wird auf deren exakte Größe erst an der entsprechenden Stelle eingegangen (vgl. Kap. 11.4), weshalb der Ausschlusswert mit 50% nur rund angegeben werden kann; in Abbildung 2 soll jedoch ein Überblick über die Stichproben zur Extremgruppenvalidierung und ihr – nicht maßstabgetreues – Verhältnis zur Gesamtstichprobe dargestellt werden.

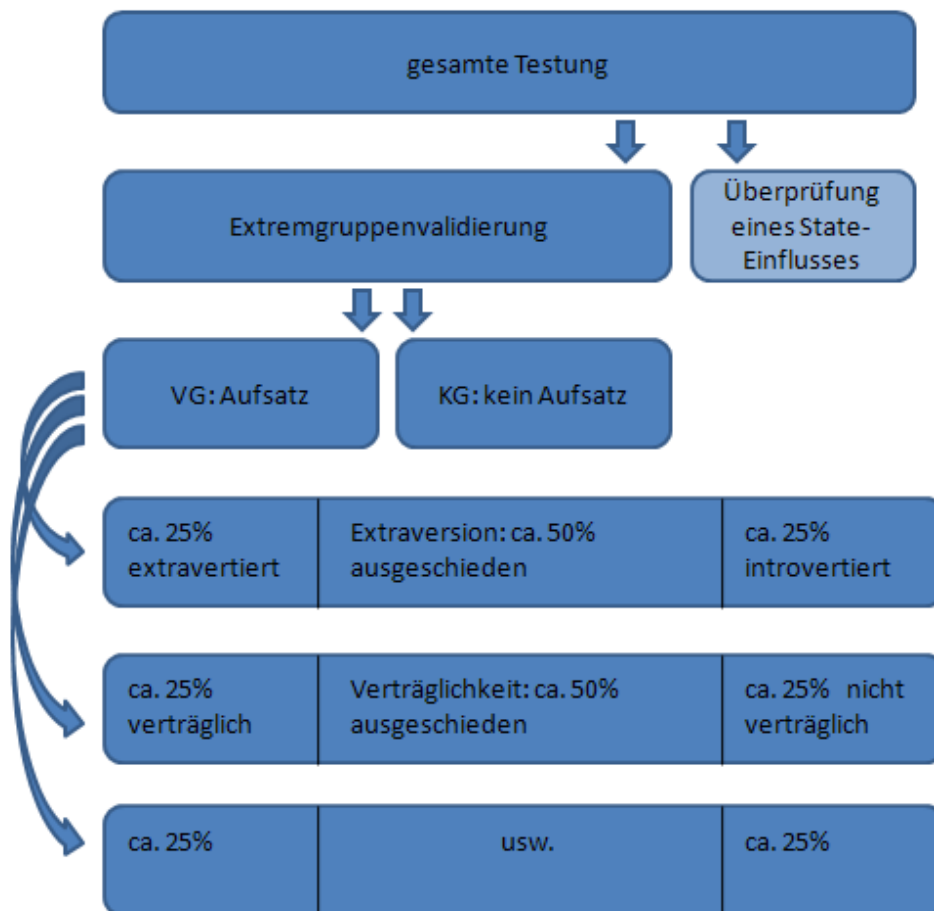


Abb. 2: Überblick über die Stichproben zur Extremgruppenvalidierung

10.2 Stichprobe für die Überprüfung eines möglichen State-Einflusses

Für die Überprüfung, ob der Aufsatz selbst einen State-Einfluss hat, wird mit der Stichprobe für die Extremgruppenvalidierung gerechnet; für die experimentelle Untersuchung mittels Suchbildvorgabe ist die Stichprobe jedoch eine andere. Hierfür wurden $n = 182$ Schüler/innen getestet, wobei es aufgrund der gemeinsam mit einer anderen Untersuchung⁹ durchgeführten Datenerhebung erneut zwei Kontrollgruppen gab, die sich teilweise überschneiden und die

⁹ vgl. Schneiderbauer, 2008, in Vorbereitung

nach dem Zufallsprinzip aus den selben Gruppen zugeteilt und beide gleichzeitig mit den Suchbild-Versuchsgruppen getestet wurden.

Insgesamt befanden sich 54 Schüler/innen in Versuchsgruppe 1 (Suchbild mit 10 Fehlern), hiervon wurden aber nur diejenigen benötigt, die das Rätsel lösen konnten, dies waren $n = 36$ Schüler/innen.

53 Schüler/innen befanden sich in Versuchsgruppe 2 (Suchbild mit 6 Fehlern); würde man diejenigen aus Versuchsgruppe 1 hinzuziehen, die das Rätsel trotz Möglichkeit nicht geschafft hatten, ergäben sich 71 Schüler/innen für diese Versuchsgruppe. Davon wurde jedoch abgesehen, da durch ein reines Hinzuziehen von Testpersonen, die die Aufgaben nicht bewältigen konnten, eine recht starke Leistungskomponente einfließen würde. Die Größe dieser Versuchsgruppe¹⁰ bleibt also bei $n = 53$.

Einzig für Hypothese 9, nämlich die Frage, ob die Rätselvorgabe selbst einen Einfluss auf die Selbstdarstellung im Persönlichkeitsverfahren hat, wird die gesamte Versuchsgruppe benötigt (also alle $n = 54 + 53 = 107$ Schüler/innen), ansonsten werden die beiden Teilversuchsgruppen herangezogen.

53 Schüler/innen befanden sich in Kontrollgruppe 1 und 56 in Kontrollgruppe 2, wobei sich die beiden Gruppen zum Teil überschneiden. Dies kam dadurch zustande, dass durch die gemeinsame Datenerhebung mit einer anderen Untersuchung manchmal ein gleichmäßiges Verhältnis von Versuchs- und Kontrollgruppe möglich war (wobei dies der Teil der Kontrollgruppen ist, der sich überschneidet), manchmal das Verhältnis jedoch bei 1:2 lag – in diesem Fall wurden dann die doppelt so großen Kontrollgruppen in zwei Teilkontrollgruppen aufgeteilt. Insgesamt befanden sich 75 Testpersonen in der gesamten Kontrollgruppen-Bedingung. Die gesamte Kontrollgruppe wird nur für Hypothese 9 benötigt, ansonsten kann mit den geteilten Kontrollgruppen gerechnet werden.

Zum genaueren Überblick sei auch hier die Stichprobe für die systematische Überprüfung einer State-Abhängigkeit schematisch dargestellt (Abb. 3).

¹⁰ Hierunter befand sich auch eine Austauschschülerin aus Frankreich, deren Deutsch nicht besonders gut war; sie wurde jedoch in der Stichprobe belassen, da sie in Wien zur Schule ging, und besonders in Wiener Schulen des Öfteren Schüler/innen mit weniger exzellenten Deutschkenntnissen getestet wurden. Sicherheitshalber wurden später sämtliche Tests nochmals durchgeführt, wobei sie ausgeschlossen wurde, an den Endresultaten änderte sich dadurch jedoch nichts.

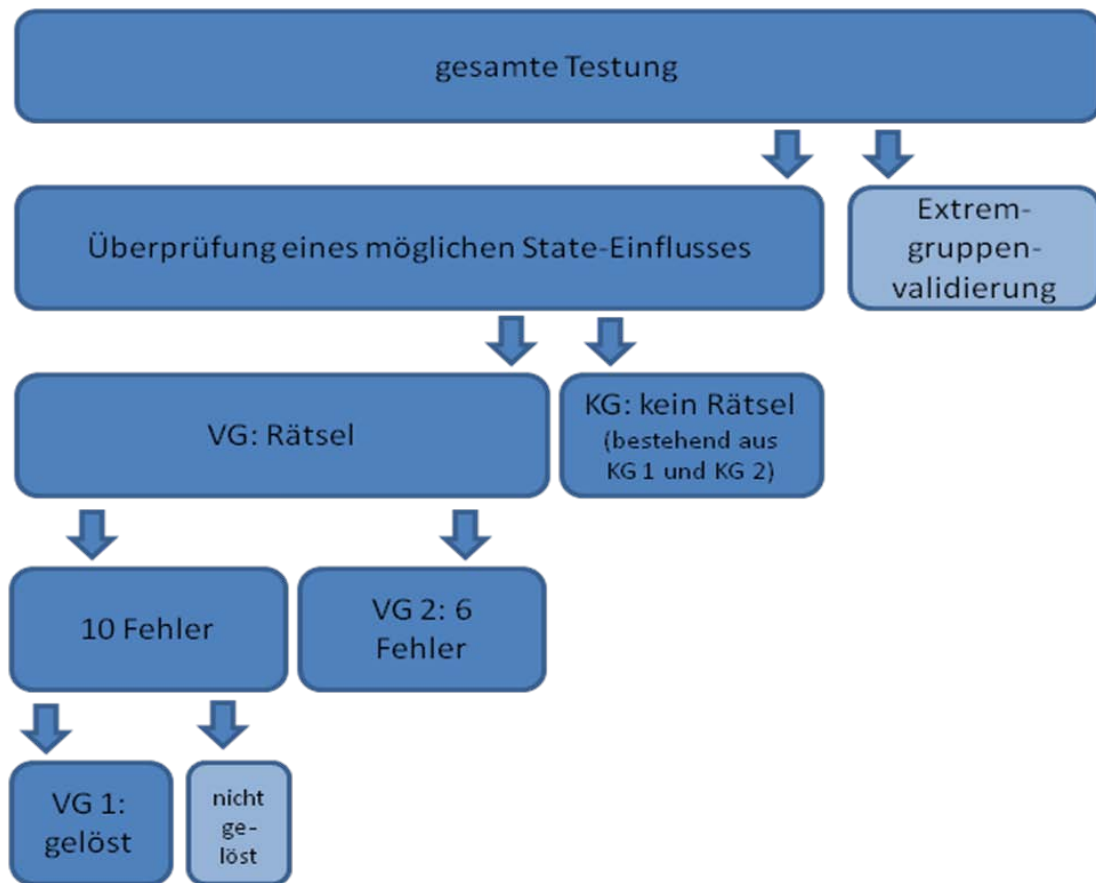


Abb. 3: Überblick über die Stichproben zur systematischen Überprüfung eines möglichen State-Einflusses

Auch die Stichprobe für die systematische Überprüfung der State-Abhängigkeit wird genauer hinsichtlich der Geschlechterverteilung betrachtet (s. Tab. 5). Bei keiner der Versuchs- oder Kontrollgruppen zeigt sich ein signifikanter Unterschied in der Verteilung der Geschlechter.

	<i>n</i> männlich	<i>n</i> weiblich
Versuchsgruppe 1	15 (42%)	21 (58%)
Versuchsgruppe 2	26 (49%)	27 (51%)
Kontrollgruppe 1	28 (53%)	25 (47%)
Kontrollgruppe 2	29 (52%)	27 (48%)
gesamte Versuchsgruppe	50 (47%)	57 (53%)
gesamte Kontrollgruppe	41 (55%)	34 (45%)

Tab. 5: Geschlechterverteilung der Stichprobe zum State-Einfluss (Prozentwerte gerundet)

11. Entwicklung eines Kategoriensystems für die Extremgruppenbildung

Für die Auswertung mittels Aufsatz wurde ein Kategoriensystem erstellt, das in Folge nach Durchsicht der ersten Aufsätze noch weiter modifiziert wurde, ebenso wie die Kodierrichtlinien.

11.1 Qualitative Auswertungen

Für die ursprüngliche Version des Kategoriensystems war vorgesehen, alle im *B5PO* verwendeten Worte oder Begriffe gleicher Bedeutung sowie Umschreibungen der Eigenschaften des *B5PO* auf einem Bogen zu einer der sechs Skalen zuzuordnen, wobei es je Skala fünf Abstufungen gab: Auf dem Pol positiv ladende, relativiert positiv ladende (dies betraf z.B. Äußerungen wie „eher“, „etwas“ und dergl.), neutrale, relativiert negativ ladende und negativ (auf dem Gegensatzpol) ladende Eigenschaften.

Für jede Skala wurde festgelegt, welcher Pol für die Verrechnung als „positiv“ angesehen wurde (z.B. Extraversion) oder als „negativ“ (z.B. Introversion), wobei dies keine Wertung darstellte, sondern einfach die Möglichkeit, Gegensätze darzustellen. Verneinungen einer negativen Ladung (z.B. Verneinung einer Introversion) wurden als positive Ladungen gewertet und umgekehrt.

Die Kodier- bzw. Auswertungsrichtlinien für diesen qualitativen Auswertungsschritt sowie die Grundversion des zugehörigen Auswertungsbogens¹¹ finden sich in Anhang B.

Nach Durchsicht der Aufsätze und Kodierung aller als relevant erscheinenden Begriffe wurden Punkte vergeben: Jede Ladung auf einer Skala zählte einen Punkt, wenn die zu kodierende Einheit ohne Relativierung genannt oder sogar verstärkt (z.B. „sehr“) wurde. Auch für relativierte Nennungen wurde ein Punkt vergeben, der jedoch in späterer Folge (bei Berechnung des Wertes) halbiert wurde. Für den Fall, dass neutrale Nennungen im Sinne von „weder-noch“ bei gegensätzlichen Eigenschaften vorkamen, gab es eine neutrale Kategorie mit null Punkten. Im Anschluss an diese Punktevergabe wurde ein Wert berechnet, der sich folgendermaßen zusammensetzte:

¹¹ In einem späteren Schritt wurden die nicht kodierbaren Einheiten noch etwas genauer aufgeschlüsselt.

$$\text{Wert} = AN_p + AN_{pr} \cdot 0,5 - AN_n - AN_{nr} \cdot 0,5$$

AN_p		Anzahl positiver Nennungen der Skala
AN_{pr}		Anzahl relativierter positiver Nennungen der Skala
AN_n		Anzahl negativer Nennungen der Skala
AN_{nr}		Anzahl relativierter negativer Nennungen der Skala

Dieser Wert wurde daraufhin noch in Bezug zur Anzahl gesamter Nennungen gesetzt, um Unterschiede zwischen Personen bewerten zu können, die sich in verschiedenen Dimensionen unterschiedlich ausführlich beschrieben. Der so erhaltene gewichtete Wert berechnet sich als

$$\text{Wert}_{\text{gew}} = \frac{\text{Wert}}{AN_{\text{ges}}}$$

AN_{ges}		Anzahl gesamter Nennungen der Skala
-------------------	-------	-------------------------------------

Weiters wurden noch „sozial erwünschte“ Werte berechnet, durch die berücksichtigt werden sollte, dass die positiv ladenden Pole (Extraversion, Verträglichkeit,...) sozial erwünschter erscheinen als die negativ ladenden; deshalb wurden für den sozial erwünschten Wert die negativen Pole eineinhalbmal so stark gewichtet wie die positiven, wobei dieses Maß willkürlich gewählt wurde. Natürlich wäre es auch möglich gewesen, die Gewichtung der Skalen an die Dichotomisierung der *B5PO*-Skalen anzugleichen; dies wäre jedoch eine zu starke Anlehnung an das zu validierende Verfahren gewesen.

Im ersten Probedurchlauf wurden fünf Aufsätze von zwei Beurteilerinnen kategorisiert; sämtliche hier und in Folge herangezogenen Beurteilenden waren Experten und Expertinnen, nämlich Diplomand/innen im Fach „Psychologie“. Als Zuordnungs- und Interpretationshilfe wurden die im *B5PO* verwendeten Wörter verwendet (66 Gegensatzpaare, also insgesamt 132 Eigenschaftswörter). Die Vorgangsweise bestand darin, jede Aussage (dies konnten Sätze sein oder teilweise auch nur Satzteile, wenn ein Satz aus mehreren Aussagen bestand) zu einer der sechs Dimensionen des *B5PO* zuzuweisen oder, falls dies nicht möglich war, als nicht kategorisierbar oder als zusätzliche, im *B5PO* nicht vorkommende Eigenschaft zu kennzeichnen.

Die Übereinstimmungen in diesem ersten Durchlauf waren nicht besonders gut, die sechs Dimensionen des *B5PO* sind in der folgenden Tabelle 6 dargestellt (Übereinstimmungen sind kursiv hervorgehoben):

Skala	Testperson	Beurteilerin 1			Beurteilerin 2		
		AN_{ges}	Wert	$Wert_{gew}$	AN_{ges}	Wert	$Wert_{gew}$
Extraversion	06002	8	4	0,5	5	2,5	0,5
	06005	5	1	0,2	6	2,5	0,42
	06014	0	0	0	1	1	1
	06017	4	2	0,5	5	-1	-0,2
	06020	4	4	1	2	2	1
Verträglichkeit	06002	2	-2	-1	2	-2	-1
	06005	2	0,5	0,25	6	2	0,33
	06014	5	-3	-0,6	5	-1,5	-0,3
	06017	4	-4	-1	3	-3	-1
	06020	1	-0,5	-0,5	2	-1,5	-0,75
Gewissenhaftigkeit	06002	4	4	1	3	1,5	0,5
	06005	7	-3	-0,43	10	-2,5	-0,25
	06014	5	5	1	7	7	1
	06017	0	0	0	5	-1	-0,2
	06020	1	1	1	1	1	1
Emotionale Kontrolle	06002	1	1	1	4	0	0
	06005	5	-1	-0,2	3	-1	-0,33
	06014	2	-2	-1	1	1	1
	06017	7	-3	-0,43	2	-2	-1
	06020	6	4,5	0,75	9	7,5	0,83
Offenheit	06002	3	3	1	1	1	1
	06005	1	1	1	3	3	1
	06014	4	4	1	3	3	1
	06017	2	2	1	2	2	1
	06020	1	1	1	2	2	1
Empathie	06002	2	2	1	4	4	1
	06005	2	2	1	4	4	1
	06014	4	4	1	5	4	0,8
	06017	2	2	1	3	3	1
	06020	0	0	0	3	3	1

Tab. 6: Erster Probedurchlauf der qualitativen Beurteilungsübereinstimmung

Betrachtet man diese Tabelle genauer, sieht man, dass viele Einschätzungen zwar tendenziell in die gleiche Richtung gehen, von 30 Möglichkeiten (sechs Dimensionen bei jeweils fünf Testpersonen) aber nur drei vollkommen in den Werten übereinstimmen (in der Tabelle kursiv hervorgehoben). Liest man hier die zugehörigen Aufsätze, sieht man, dass nur bei einem¹² dieser drei Aufsätze auch dieselben Aussagen auf die gleiche Art kodiert wurden; bei den anderen beiden ist die Übereinstimmung zwar im Gesamtwert gleich, jedoch beziehen sich die einzelnen Punktvorgaben teilweise auf unterschiedliche Aussagen.

¹² Es handelt sich um die Skala „Offenheit“ von Testperson Nr. 06017

Auch sieht man, dass viele Beurteilungen recht stark voneinander abweichen, wobei die Festlegung eines absoluten Schwellenwertes für einen absoluten „zu schlechten“ Wert keinen Sinn hat, da man diesen Wert auch im Kontext der anderen Werte betrachten muss: Würde man beispielsweise einen Unterschied im gewichteten Wert von $\geq 0,5$ als „stark“ beurteilen, würden sechs Einschätzungen diesen Wert erreichen, dies würde aber auch den Unterschied von Testperson 06014 der Skala „Extraversion“ beinhalten, der nur durch eine einzige Aussage bedingt ist. Hingegen beträgt der Unterschied im gewichteten Wert der Skala „Gewissenhaftigkeit“ bei Testperson Nr. 06017 nur 0,2; dieser Unterschied ist jedoch inhaltlich weit bedeutsamer, da hier gleich fünf Werte anders kodiert wurden.

Besonders dramatisch sind die Unterschiede bei „Extraversion“ von Testperson 06017 oder „Emotionaler Kontrolle“ von Testperson 06014; bei letzterer erreicht die Übereinstimmung (oder besser Nicht-Übereinstimmung) des gewichteten Werts den Maximalunterschied von 2.

Aufgrund der unterschiedlichen Beurteilungen im ersten Durchlauf wurde eine Synonymliste erstellt, in der alle Synonyme zu den im *B5PO* verwendeten Worten aus dem Duden Synonymwörterbuch (4. Auflage, 2006) herausgesucht wurden. Diese Liste sollte Zuordnungen und Interpretationen eindeutiger machen; insgesamt (unter Mitzählen der 132 Eigenschaftswörter des *B5PO* und aller doppelt und mehrfach vorkommenden Synonyme) belief sich die Liste auf 3238 Wörter. Nachdem jene ausgeschlossen wurden, die definitiv nicht persönlichkeitsrelevant waren, blieben noch 2570 Begriffe übrig (erneut einschließlich mehrfach vorkommender und im *B5PO* vorkommender Worte). Da viele der in der Liste mehrfach vorkommenden Synonyme auf mehreren Skalen luden, wurde in der Liste für zusätzliche Klarheit hinzugefügt, zu welchem Wort das genannte Wort synonym sein sollte und auch, welches der Gegenpol dazu sein sollte. Eine Beispielseite der Synonymliste findet sich in Anhang B.

Nun wurde unter Zuhilfenahme dieser Liste ein neuer Durchlauf gestartet, in dem weitere drei Aufsätze beurteilt wurden. (Um nicht noch Erinnerungsreste der ersten Beurteilung im Kopf zu haben, wurden andere Aufsätze herangezogen.) Auch die Beurteilungen in diesem zweiten Durchlauf stimmten nicht zufriedenstellend überein, hier wieder eine Übersicht (Tabelle 7):

Skala	Testperson	Beurteilerin 1			Beurteilerin 2		
		AN_{ges}	Wert	$Wert_{gew}$	AN_{ges}	Wert	$Wert_{gew}$
Extraversion	05084	3	3	1	5	5	1
	06008	2	2	1	3	3	1
	06011	4	-1	-0,25	9	-5	-0,56

Verträglichkeit	05084	7	-3	-0,43	5	-5	-1
	06008	2	-2	-1	1	-1	-1
	06011	2	0	0	1	-1	-1
Gewissenhaftigkeit	05084	0	0	0	0	0	0
	06008	1	-1	-1	3	-2,5	-0,83
	06011	1	-1	-1	1	-1	-1
Emot. Kontrolle	05084	1	-1	-1	0	0	0
	06008	3	-3	-1	1	-1	-1
	06011	3	3	1	0	0	0
Offenheit	05084	2	2	1	2	2	1
	06008	4	4	1	6	4	0,66
	06011	0	0	0	1	-1	-1
Empathie	05084	1	1	1	1	1	1
	06008	3	1	0,33	1	1	1
	06011	2	1,5	0,75	3	3	1

Tab. 7: Zweiter Probedurchlauf der Beurteilungsübereinstimmung

In diesem Durchlauf sieht man nun, dass vier von insgesamt 18 Beurteilungen genau übereinstimmen (kursiv hervorgehoben), wobei die Übereinstimmungen sich hier auch inhaltlich bei Kontrolle der Aufsätze zeigten, nur bei einer Übereinstimmung („Offenheit“ bei Testperson 05084) wurde jeweils eine der zugrundeliegenden Aussagen anders kodiert, die Übereinstimmung besteht hier also nur im Gesamtwert.

Dennoch sind die Übereinstimmungen insgesamt noch zu gering und auch einige der Urteile noch zu verschieden (z.B. bei Testperson 06011 die Skalen „Verträglichkeit“ und „Emotionale Kontrolle“).

In einem dritten Durchlauf (über insgesamt 5 Aufsätze) wurde daraufhin beschlossen, keine Interpretationen mehr vorzunehmen, sondern nur noch eindeutig zuordenbare Begriffe zu kodieren, mögliche Interpretationen sollten dabei gekennzeichnet und zugewiesen, jedoch nicht mitgerechnet werden. Als interpretationseindeutig sollten dabei jene Worte gelten, die entweder im *B5PO* verwendet wurden oder die Synonyme waren, die nur auf einer einzigen Skala luden (das heißt, die in der Synonymliste nur für eine einzige Skala ein Synonym darstellten). In der Auswertung der Bögen zeigte sich, dass bei zwei von drei Beurteilerinnen Interpretationen durchaus mitgerechnet wurden, durch die Kennzeichnung in den Aufsatzbögen konnten diese jedoch später herausgerechnet werden (wodurch probenhalber auf eine quantitative Analyse heruntergebrochen wurde). Aufgrund des Mitrechnens von Interpretationen macht eine vergleichende Darstellung aller drei Beurteilerinnen keinen Sinn, statt dessen sei hier ein Vergleich der Interpretationen und der gewichteten Werte nach Herausrechnen dieser Interpretationen dargestellt (Tab. 8):

Skala	TP-Nr.	Interpretation						nur eindeutig zuordenbare Kennzeichnungen gewertet			
		Beurt. 2			Beurt. 3			Beurt. 2		Beurt. 3	
		AN	Wert	gew.	AN	Wert	gew.	AN	gew.	AN	gew.
Extraversion	01016	5	4	0,8	6	2	0,33	2	0	3*	0,3*
	01019	7	3	0,43	4	0	0	2	1	2	1
	02002	4	2	0,5	4	4	1	0	-	2	1
	02008	2	1	0,5	6	-1	0,17	1	-1	1	-1
	02011	5	-1	-0,2	11	-0,5	-0,14	3	0,33	4	0
Verträglichkeit	01016	2	-2	-1	2	-2	-1	2	-1	2	-1
	01019	5	-3	-0,6	7*	-3*	-0,43*	4	-0,5	5*	-0,6*
	02002	7	-7	-1	6	-6	-1	4	-1	5	-1
	02008	3	-1	-0,33	2	-1,5	-0,75	1	-1	2	-1
	02011	1	1	1	4	2	0,5	0	-	1	-1
Gewissenhaftigkeit	01016	3	1	0,33	6	4	0,67	3	0,33	3	0,33
	01019	5	3	0,6	6	4	0,67	3	1	4	1
	02002	5	3	0,6	11	7	0,64	2	0	5	0,2
	02008	4	0	0	8	5,5	0,69	3	0,33	4	0,5
	02011	2	2	1	2	2	1	0	-	0	-
Emotionale Kontrolle	01016	0	-	-	0	-	-	0	-	0	-
	01019	4	-2	-0,5	2	-2	-1	1	-1	1	-1
	02002	1	-1	-1	2	-2	-1	1	-1	1	-1
	02008	1	-1	-1	1	-1	-1	1	-1	1	-1
	02011	6	-4	-0,67	6	-2	-0,33	4	-0,5	4	-1
Offenheit	01016	0	-	-	0	-	-	0	-	0	-
	01019	2	2	1	2	2	1	2	1	2	1
	02002	4	2	0,5	4	2	0,5	2	0	3	0,33
	02008	2	0	0	3	1	0,33	1	1	1	1
	02011	1	1	1	3	3	1	1	1	2	1
Empathie	01016	2	2	1	5	5	1	2	1	2	1
	01019	2	0	0	4*	-2*	-0,5*	2	0	2*	-1*
	02002	0	-	-	4	-2	-0,5	0	-	2	-1
	02008	2	2	1	4	4	1	0	-	0	-
	02011	1	1	1	3	3	1	0	-	0	-

Tab. 8: Übereinstimmung mit und ohne Interpretationen (dritter Durchlauf)

Mit * gekennzeichnete Werte sind jene, deren Kennzeichnung nicht eindeutig war oder deren Auszählung nicht nachvollziehbar war (TP 01019) – hier wurde korrigierend eingegriffen; „gew.“ steht für gewichteter Wert.

Da sich bei Betrachtung der bisherigen Versuche zeigt, dass die qualitativen Übereinstimmungen gänzlich nicht gut genug für eine Validierung sind (mangels Objektivität), wurde eine andere Variante der Kategorisierung versucht.

Im Vergleich mit und ohne Interpretationen¹³ der Tabelle 8 kann man sehen, dass die mangelnde Übereinstimmung bei der Beurteilung unter anderem durch Interpretationen begründet ist, was die Idee aufwirft, nur noch nach bestimmten Schlüsselwörtern zu suchen und diese zu kategorisieren (im Sinne einer quantitativen Inhaltsanalyse, vgl. Kap. 6.1). Hier stellt sich jedoch das Problem, dass gerade bei einer derartigen Form der Kategorisierung viele Inhalte und viel des Bedeutungskontextes verloren gehen. Daher sollte zunächst eine Globalbeurteilung der Aufsätze versucht werden.

11.2 Globalbeurteilung und Rangreihung

Auf einer 10stufigen Skala sollte (in einer globalen Einschätzung) von Expertinnen angegeben werden, wie ausgeprägt die jeweilige Dimension nach Darstellung der aufsatzschreibenden Testperson ist. Zusätzlich zu den Abstufungen von 1 bis 10 gab es eine Kategorie „0“, die nur dann vergeben werden durfte, wenn aufgrund des Aufsatzes keinerlei Hinweis über die gefragte Dimension vorliegt. Anleitung und Auswertungsbogen für die Globalbeurteilung finden sich in Anhang B.

Zusätzlich zu dieser Globalbeurteilung sollten die Testpersonen für jede Dimension hinsichtlich ihrer Ausprägung auf der Dimension ranggereiht werden. Dies wurde zunächst bei 10 Testpersonen vorgenommen, hier wieder eine Übersicht über die Übereinstimmung (kursiv dargestellt) von zwei Beurteilerinnen, zunächst die Globalbeurteilung (Tab. 9):

TP	Beurteilerin	EX	VER	GEW	EM. KO	OH	EMP
01001	1	6	2	2	0	<i>0</i>	9
	2	8	2	5	8	<i>0</i>	8
01004	1	8	3	<i>0</i>	3	<i>10</i>	<i>0</i>
	2	9	2	<i>0</i>	2	<i>10</i>	<i>0</i>
01007	1	5	6	4	9	6	10
	2	3	10	3	0	2	8
01010	1	10	5	5	0	10	5
	2	6	7	5	3	9	0
01013	1	8	7	7	6	8	6
	2	5	9	9	3	7	9

¹³ Da es allerdings auch bei den als interpretationseindeutig gekennzeichneten Worten Unterschiede gibt, kann man davon ausgehen, dass durchaus auch hier Interpretationen im Spiel waren und die Verrechnung somit nicht objektiv genug erfolgte.

01016	1	9	1	10	2	5	7
	2	10	3	10	0	5	4
01019	1	4	1	10	1	8	9
	2	2	1	10	2	8	8
01021	1	2	2	4	0	3	7
	2	4	5	7	6	3	10
01024	1	6	1	8	10	9	8
	2	7	3	8	7	8	7
01027	1	7	4	10	0	4	8
	2	3	4	8	4	3	6

Tab. 9: Übereinstimmungen in der Globalbeurteilung von zwei Beurteilerinnen

Betrachtet man Tabelle 9, kann man sehen, dass einige Urteile zwar tendenziell übereinstimmen, andere Urteile weichen jedoch sehr stark voneinander ab, insbesondere durch die Kategorie „0“, die eine andere Qualität als die Kategorien 1–10 darstellt. Dadurch ist ein Unterschied von 0 zu 1 ebenso wie von 0 zu beispielsweise 5 maximal.

Über alle Testpersonen berechnet liegt die Übereinstimmung ohne Berücksichtigung der Zufallswahrscheinlichkeit bei 0,233 (14 von 60 möglichen Fällen). Berechnet man einzelne Testpersonen, liegen die Übereinstimmungen bei maximal 0,5 (bei Testperson 01019 und 01004, hier stimmen drei der sechs Skalen exakt überein), dies wieder ohne Berücksichtigung der Zufallswahrscheinlichkeit. Die Wahrscheinlichkeit der vollkommenen Übereinstimmung für eine einzige Testperson geht jedoch gegen 0, da jede Testperson auf sechs in der Beurteilung unabhängigen Skalen eingeschätzt wird, deren Zufallsübereinstimmung jeweils ein Elftel wäre – sie ist also vernachlässigbar klein.

Ähnlich verhält es sich bei Übereinstimmung (max. 0,5 bei „Gewissenhaftigkeit“ und „Offenheit“) und Zufallsübereinstimmung für die einzelnen Skalen. So oder so sind jedoch die Übereinstimmungsmaße zu gering, um daran zu validieren.

In der Rangreihung waren die Ergebnisse ähnlich, die Übereinstimmungen der Rangplätze über alle Testpersonen gerechnet (wieder ohne Berücksichtigung der Zufallsübereinstimmung) beträgt hier sogar nur 0,133 (8 von 60 Möglichkeiten). Auch die Übereinstimmung für einzelne Testpersonen ist geringer (max. 0,33, also max. 2 der 6 möglichen Rangplätze gleich vergeben). Alle Rangurteile sind in Tabelle 10 dargestellt.

TP	Beurteilerin	EX	VER	GEW	EM. KO	OH	EMP
01001	1	7	7	9	0	0	2
	2	3	8	8	1	0	3
01004	1	4	5	0	4	2	0
	2	2	9	0	7	1	0
01007	1	8	2	7	2	6	1
	2	9	1	9	0	9	5
01010	1	1	3	6	0	1	0
	2	5	3	7	6	2	0
01013	1	3	1	5	3	4	8
	2	6	2	3	5	5	2
01016	1	2	8	3	5	7	7
	2	1	7	2	0	6	8
01019	1	9	10	1	6	5	3
	2	10	10	1	8	4	4
01021	1	10	6	8	0	9	6
	2	7	4	6	3	8	1
01024	1	6	9	4	1	3	5
	2	4	6	5	2	3	6
01027	1	5	4	2	0	8	4
	2	8	5	4	4	7	7

Tab. 10: Übereinstimmungen in der Rangreihung

Aufgrund dieser Ergebnisse wurde doch von einer Global- oder Rangreihung abgesehen und ein Versuch über Schlüsselwörterkodierung gestartet.

11.3 Schlüsselwörterkodierung

Um ein größeres Maß an Objektivität zu erreichen, wurde nun einzig anhand von Schlüsselwörtern kodiert, wofür die in Kap. 11.1 erwähnte Synonymliste herangezogen wurde. Es sollten nur solche Wörter kodiert werden, die eindeutig zuordenbar waren, also entweder im *B5PO* verwendete Eigenschaftswörter waren oder ein nur auf einer einzigen Skala ladendes Synonym hierzu, was bereits die Grundidee im dritten Versuch darstellte.

Zusätzlich durften, um mehr Kodiereinheiten zu erhalten, beispielsweise auch Hauptwörter kodiert werden, sofern sie sich eindeutig und wörtlich in ein Adjektiv der Liste transformieren ließen (z.B. „Freundlichkeit“ als „freundlich“), und auch Verneinungen durch „nicht“ oder die Vorsilbe „un-“ durften jeweils als der Gegenpol kodiert werden. Die genaue Instruktion sowie der Auswertebogen für die Schlüsselwortkodierung finden sich in Anhang B.

Dadurch wurde die Inhaltsanalyse zwar größtenteils quantitativ, da aber der Kontext mitberücksichtigt werden sollte (Berücksichtigung der Polung; nur Berücksichtigung von Schlüsselwörtern, die sich auf die eigene Persönlichkeit in der Gegenwart beziehen; Beschreibungen der Persönlichkeit in der Vergangenheit wurden ausgeschlossen) waren immer noch geringe qualitative Anteile vorhanden.

Wie auch bei der qualitativen Analyse im ersten Durchgang (Kap. 11.1) sollten ein Wert und ein gewichteter Wert berechnet werden, um die Anzahl genannter Schlüsselwörter zu berücksichtigen. Relativierte Aussagen wurden nicht mehr mitgerechnet, da dies erneut zu interpretativ wäre; es wurde nur noch nach Positiv- und Negativpol kodiert.

Die Kodierung wurde zunächst an fünf Personen versucht. Da die Schlüsselwörterkodierung nach strengen Regeln erfolgte und dadurch von den Globalurteilen im vorigen Durchgang unabhängig sein sollte, wurden die ersten fünf der im vorigen Durchgang kodierten Testpersonen gewählt. Dies bot den Vorteil, die Urteile der Schlüsselwörterkodierung später mit den Globalurteilen vergleichen zu können. In Tabelle 11 sind die Urteile (Übereinstimmungen wieder kursiv hervorgehoben) dargestellt; „x“ bedeutet, dass sich kein Schlüsselwort im Aufsatz fand, der Wert null hingegen, dass sich die Schlüsselwörter gegenseitig aufhoben.

TP	Beur- teilerin	EX Wert / Wert _{gew}	VER Wert / Wert _{gew}	GEW Wert / Wert _{gew}	EM. KO Wert / Wert _{gew}	OH Wert / Wert _{gew}	EMP Wert / Wert _{gew}
01001	1	x/x	-1/-0,33	-2/-1	x/x	x/x	6/1
	2	x/x	-1/-1	-2/-1	x/x	x/x	x/x
01004	1	2/1	-1/-1	x/x	x/x	2/1	x/x
	2	2/1	1/0,33	x/x	x/x	2/1	x/x
01007	1	-1/-1	x/x	0/0	0/0	x/x	4/1
	2	-1/-1	x/x	-1/-1	x/x	x/x	1/1

01010	1	1/1	-1/-1	-2/-1	-2/-1	1/1	x/x
	2	x/x	1/1	-1/-1	-1/-1	1/1	x/x
01013	1	3/1	1/1	-1/-1	-1/-1	x/x	1/1
	2	2/1	1/1	-1/-1	x/x	x/x	x/x

Tab. 11: Schlüsselwortkodierung über fünf Aufsätze

Wie man an den Werten in Tabelle 11 erkennen kann, stimmen manche Urteile in den gewichteten Werten überein, in den Werten selbst jedoch nicht. Das bedeutet, dass unterschiedlich viele Kodiereinheiten für die Urteile herangezogen wurden und das Urteil dadurch erneut nicht objektiv ist.

Vergleicht man weiters die Schlüsselwörterkodierung mit den Globalurteilen in Tabelle 9, sieht man, dass zwar tendenziell Übereinstimmungen bestehen, in einigen Fällen durch die Reduktion der Information auf Schlüsselworte das Urteil jedoch gänzlich anders ausfällt: So wurde beispielsweise Testperson 01001 auf der zehnstufigen Skala im Globalurteil als eher (Beurteilerin 1) bis sehr (Beurteilerin 2) extravertiert eingestuft. Nach der Schlüsselwörterkodierung ist hier jedoch gar kein Urteil möglich, da kein relevantes Schlüsselwort im Aufsatz erwähnt wurde.

Da das Globalurteil aufgrund all der kontextbezogenen Informationen vermutlich eher der „tatsächlichen“ Persönlichkeit der Testpersonen entspricht, erfolgte die Extremgruppenbildung nun auf gänzlich andere, jedoch wieder globale Weise.

11.4 Extremgruppenbildung

Aufgrund der gescheiterten analytischen Versuche erfolgte die Extremgruppenbildung in letzten Schritten durch einen weit pragmatischeren Zugang. Da für die Validierung an sich nur Extremgruppen benötigt werden, detailliertere Einschätzungen der Testpersonen jedoch nicht unbedingt notwendig sind, wurde von folgender Überlegung ausgegangen: Eine Dimension kann bei einer Testperson jeweils als extrem ausgeprägt beurteilt werden, wenn mehrere Beurteilende die Testperson auf ebendieser Dimension so einschätzen würden.

Es sollten nun alle Testpersonen nur noch hinsichtlich dessen global beurteilt werden, ob die jeweiligen Dimensionen bei ihnen extrem ausgeprägt sind oder eben nicht. Dabei handelt es sich streng genommen nicht mehr um eine quantitative oder qualitative Inhaltsanalyse, da nun der Aufsatz nicht mehr in Segmente zerlegt und analysiert wurde, sondern es wurde nur noch global qualitativ eingeschätzt, ob die jeweilige Testperson auf den interessierenden

Dimensionen durchschnittliche, extreme oder gar keine Ausprägungen zeigt. Das Kategoriensystem wurde je Skala auf nur vier Punkte herunter gebrochen: unterdurchschnittliche Ausprägung auf einer Skala, durchschnittliche Ausprägung, überdurchschnittlich hohe Ausprägung oder gar keine (kein mögliches Urteil). Als einzige Beurteilungshilfe sollte die Liste der im *B5PO* vorkommenden Eigenschaftsworte aus dem Manual dienen, um eine Vorstellung zu vermitteln, was mit den jeweiligen Skalen gemeint ist. Nach der Kategorienzuteilung sollte jede Testperson dann zu einer Extremgruppe verrechnet werden, wenn drei von drei Beurteilenden sie zu einem der Extrempole zuordnen. (Insgesamt gab es zehn Beurteiler/innen, die unterschiedlich viele Aufsätze bearbeiteten. Jeder Aufsatz wurde jedoch nur von drei Beurteiler/innen bewertet.) Wenn nur zwei der Beurteilenden die Testperson beispielsweise als überdurchschnittlich extravertiert einschätzen, ein dritter Beurteiler sie als durchschnittlich, wurde von der Überlegung ausgegangen, dass ihre Ausprägung auf der Skala vermutlich nicht stark genug ist, um als Extremgruppe zu gelten.

Die Anleitung und der Auswertungsbogen für diesen Schritt finden sich in Anhang B.

Nach erfolgter Durchführung mussten die Extremgruppen jedoch teilweise erweitert (z.B. bei Introversion), teilweise eingeschränkt werden (z.B. bei ausgeprägter Empathie), für die Erweiterung wurden dabei jeweils die nächststärkeren Ausprägungen hinzugenommen, bis die Extremgruppe annähernd der Größe entsprach, die zu Beginn mit dem Programm *CADEMO Light* berechnet worden war. Für das Verkleinern der Extremgruppen (dies war besonders gravierend in den Skalen „Empathie“ und „Offenheit“) wurden die zu großen Gruppen erneut von insgesamt jeweils zwei Beurteilerinnen durchgesehen, wobei jeweils eine gewisse Anzahl der sich weniger extrem darstellenden Schüler/innen aussortiert werden sollte. Die Übereinstimmungen, die beide Beurteilerinnen ausgeschieden hätten, wurden aus der zu großen Extremgruppe wieder herausgenommen.

All jene Testpersonen, die zumindest einmal als nicht beurteilbar bewertet wurden, wurden für die jeweilige Skala für erweiternde oder verkleinernde Schritte herausgenommen, da davon ausgegangen wurde, dass das Merkmal im Aufsatz nicht offensichtlich genug ist (sonst wäre es als Hinweis gewertet worden) und das Vorgehen daher zu interpretativ wäre. Begründete Ausnahmen hiervon finden sich für die Skalen „Offenheit“ (Kap. 11.4.5) und „Empathie“ (Kap 11.4.6).

Im Folgenden seien die letzten Schritte zur Bildung der endgültigen Extremgruppen und deren endgültig resultierende Größe je Skala detaillierter dargestellt.

11.4.1 Skala „Extraversion“

In der Extremgruppenbildung stellten sich $n = 14$ Schüler/innen nach Einschätzung von jeweils drei Beurteiler/innen extrem extravertiert dar. Diese Extremgruppe musste also weder verkleinert noch vergrößert werden. Allerdings stellten sich nur drei Schüler/innen übereinstimmend als extrem introvertiert dar, hier musste die Gruppe also noch vergrößert werden. Vier weitere Schüler/innen wurden von zumindest zwei Beurteiler/innen als sehr introvertiert eingeschätzt, eine dieser Testpersonen wurde jedoch ausgeschlossen, da ein Urteil über sie in sehr extravertierte Richtung vorlag. Um die Gruppe noch weiter zu vergrößern wurden all jene herangezogen, die von keinem einzigen Beurteilenden jemals als extrem extravertiert dargestellt wurden; es kamen damit noch weitere drei Testperson hinzu, die zweimal als durchschnittlich, einmal als introvertiert eingeschätzt wurden, sowie fünf jeweils von allen als durchschnittlich eingestufte Testpersonen. Insgesamt beträgt damit die Größe dieser Extremgruppe also $n = 3 + 3 + 3 + 5 = 14$.

11.4.2 Skala „Verträglichkeit“

Als Extremgruppe in die sehr verträgliche Richtung wurden nur sieben Testpersonen übereinstimmend zugeteilt. Eine weitere achte wurde von zwei Beurteilenden als sehr verträglich, einer dritten als durchschnittlich gesehen. Um die Gruppe weiter zu vergrößern sollten all jene hinzugezogen werden, die zumindest einmal als verträglich, jedoch nie als unverträglich oder nicht beurteilbar dargestellt wurden. Dies wären weitere neun Schüler/innen, wodurch die Extremgruppe hier zu groß ausfallen würde, benötigt wurden im Optimalfall nur sechs. Deshalb sollten in einem zusätzlichen Schritt zwei Beurteilerinnen einschätzen, welche drei davon noch am wenigsten verträglich seien und wen sie als vierte Person gereiht hätten (dies wurde gemacht, um eventuelle Unstimmigkeiten zwischen den Beurteilerinnen ausgleichen zu können; im Fall einer Unstimmigkeit bei den ersten drei Testpersonen sollte diese vierte Testperson aus der Extremgruppe herausfallen, falls die andere Beurteilerin sie herausgenommen hätte). Bei zwei Testpersonen waren sich die beiden Beurteilerinnen einig (bei einer der drei Erstgereihten sowie bei einer, die einmal bei den Erstgereihten und einmal viertgereiht gewertet wurde), die Extremgruppe beträgt somit $n = 7 + 1 + (9 - 2) = 15$ Schüler/innen.

Auf der anderen Extrempol-Seite wurden acht Schüler/innen übereinstimmend als extrem unverträglich eingestuft. Zwei weitere Testpersonen wurden zweimal als sehr unverträglich, einmal als durchschnittlich eingestuft. Die Schüler/innen, die zumindest einmal als sehr

unverträglich, nie jedoch als verträglich oder nicht beurteilbar eingestuft wurden, sind insgesamt sieben, also drei mehr als die optimale Stichprobengröße. Daher wurden hier analog zum anderen Extrempol zwei Beurteilerinnen beauftragt, drei Testpersonen herauszusuchen, die sie noch am verträglichsten einschätzen würden sowie eine vierte Person, die sie als nächste gereiht hätten. Bei einer Testperson waren sich die beiden Beurteilerinnen einig, zwei weitere wurden jeweils einmal ausgeschlossen und einmal als nächstgereiht angesehen. Insgesamt beläuft sich die Größe dieser Extremgruppe damit auf $n = 8 + 2 + (7 - 3) = 14$ Schüler/innen.

11.4.3 Skala „Gewissenhaftigkeit“

Auf dieser Skala wurden 21 Schüler/innen von drei Beurteiler/innen als extrem gewissenhaft eingestuft. Für eine Verkleinerung sollten sieben Testpersonen aussortiert werden, die noch am wenigsten gewissenhaft gesehen wurden, sowie (um Unstimmigkeiten auszugleichen) weitere drei, die als nächste in Richtung weniger Gewissenhaftigkeit gereiht würden.

Bei 5 Testpersonen waren sich die Beurteilerinnen einig, weitere 3 wurden einmal als weniger extrem als der Rest ausgeschlossen, einmal als nächstgereiht angegeben und wurden daher auch aus der Extremgruppe herausgenommen. Insgesamt beträgt die Gruppengröße somit $n = 21 - 5 - 3 = 13$.

Auf der anderen Extrempolseite wurden übereinstimmend nur drei Schüler/innen als sehr wenig gewissenhaft gesehen, durch Erweiterung um all jene, die zumindest einmal als gar nicht gewissenhaft, nie jedoch als gewissenhaft oder nicht bewertbar eingeschätzt wurden, kamen weitere 12 dazu. Insgesamt beträgt die Größe dieser Extremgruppe also $n = 15$.

11.4.4 Skala „Emotionale Kontrolle“

Auf der Skala Emotionale Kontrolle wurden nur zwei Testpersonen übereinstimmend als sehr emotional kontrolliert eingeschätzt. Unter Hinzunahme all jener, die nie in die neurotizistische Richtung eingeschätzt wurden, kommen weitere neun hinzu. Alle anderen Testpersonen wurden zumindest einmal als wenig emotional kontrolliert oder nicht beurteilbar bewertet. Um näher an die optimale Gruppengröße von 14 heranzukommen, wurden daher hier noch jene beiden Testpersonen hinzugenommen, die zwar einmal als neurotizistisch, aber zweimal als sehr emotional kontrolliert eingeschätzt wurden, insgesamt beträgt diese Extremgruppe also $n = 13$ Schüler/innen.

Auf dem anderen Extrempol wurden $n = 16$ Schüler/innen übereinstimmend als extrem wenig emotional kontrolliert eingeschätzt, deshalb sollten hier von zwei Beurteilerinnen zwei

Testpersonen aussortiert werden, die noch eher emotional kontrolliert eingeschätzt wurden sowie eine dritte, die als drittgeriht angesehen werden würde. Bei einer Testpersonen waren sich die Beurteilerinnen einig, die Stichprobengröße beträgt damit $n = 16 - 1 = 15$ Schüler/innen.

11.4.5 Skala „Offenheit“

Auf dieser Skala wurden gleich 25 Schüler/innen übereinstimmend als sehr offen für Erfahrungen eingeschätzt, die Extremgruppe musste also gravierend verkleinert werden. Analog zu den anderen Skalen sollten 11 Testpersonen aussortiert werden, die noch am wenigsten offen erschienen, sowie weitere vier, die die nächstgerihten gewesen wären. Bei 7 Testpersonen waren sich die Beurteilerinnen einig, weitere 5 wären zumindest einmal aussortiert worden und waren einmal als nächstgeriht angegeben. Diese Extremgruppe beträgt damit $n = 25 - 7 - 5 = 13$ Testpersonen.

Auf der anderen Extrempolseite gab es nur zwei Testpersonen, die als sehr wenig offen für Erfahrungen eingeschätzt wurden, eine weitere wurde zumindest von zwei Beurteilerinnen so eingeschätzt, von einer dritten als durchschnittlich, und noch eine vierte von allen als durchschnittlich. Alle anderen Testpersonen wurden zumindest von einer Testperson als überdurchschnittlich offen bewertet. Aufgrund der Tatsache, dass eine so stark linksschiefe Verteilung vorliegt, wurden jedoch auch noch jene Testpersonen hinzugezogen, bei denen das Urteil „sehr offen“ zumindest nur ein einziges Mal vorlag. Dies waren weitere sechs Schüler/innen. Da die Extremgruppe nun immer noch zu klein war, und alle verbleibenden Testpersonen, die auf dieser Skala übereinstimmend als beurteilbar angesehen wurden, zumindest zwei Einschätzungen in Richtung „extrem offen“ hatten, wurden ausnahmsweise noch drei Testpersonen zur Gruppe genommen, die einmal als nicht beurteilbar eingestuft wurden. Allerdings wurden zwei dieser Testpersonen auch jeweils einmal als durchschnittlich und einmal als extrem wenig offen beurteilt, die dritte zumindest einmal als durchschnittlich und nur einmal als sehr offen. Durch das jeweils einmal nicht vergebene Urteil mag die Einschätzung zwar stärker interpretativ sein, ist als Gegenpol zur extrem offenen Gruppe jedoch noch besser geeignet, als Testpersonen heranzuziehen, die zwei von drei Malen als sehr offen beurteilt wurden. Die Größe dieser Extremgruppe beträgt damit also $n = 13$.

11.4.6 Skala „Empathie“

Noch schiefer als bei der Skala „Offenheit“ stellt sich die Verteilung der Skala „Empathie“ dar: Hier wurden gleich 28 Testpersonen übereinstimmend als extrem empathisch

eingeschätzt, auch hier war also eine drastische Reduktion dieser Extremgruppe notwendig. Es sollten die 14 noch am wenigsten extrem empathischen Testpersonen herausgesucht werden sowie die vier, die nächstgereiht weniger empathisch eingeschätzt werden.

Bei 12 Testpersonen stimmten die beiden Beurteilerinnen überein, weitere 2 Testpersonen wurden zumindest von einer Beurteilerin ausgeschlossen, von einer weiteren als nächstgereiht angesehen. Es verbleiben somit $n = 28 - 12 - 2 = 14$ Testpersonen in dieser Extremgruppe.

Auch im Gegenpol erweisen sich die Einschätzungen verglichen mit der Skala „Offenheit“ als noch schiefer: So wurde das Urteil „extrem wenig empathisch“ bei den jeweils drei Urteilen auf 62 Aufsätzen, also 186 Möglichkeiten, überhaupt nur zweimal vergeben. Entsprechend werden als Extremgruppe für die wenig empathische Ausprägung zunächst jene Testpersonen herangezogen, die das Urteil „sehr empathisch“ nie oder nur maximal einmal hatten; dies waren nur acht Schüler/innen. Um die Gruppe noch etwas zu vergrößern, wurden analog dem Vorgehen bei der Skala „Offenheit“ auch jene Testpersonen herangezogen, die zwar ein Urteil in nicht bewertbarer Richtung hatten, jedoch auch maximal eines in Richtung sehr empathisch. Dies war bei weiteren vier Schüler/innen der Fall, die Größe dieser Extremgruppe beträgt somit endgültig $n = 12$.

12. Interpolieren fehlender *T*-Werte im *B5PO*

Da im *B5PO* nicht jedem Rohwert ein *T*-Wert zugeteilt ist, mussten die fehlenden *T*-Werte interpoliert werden.

Hierfür wurden im Statistikprogramm *SPSS* (Version 14.0, Evaluation Version) für jede Skala lineare, kubische, quadratische und exponentielle Kurven erstellt, die die vorhandenen *T*-Werte abbilden sollten. Die kubische Kurve wurde dabei als die am ehesten passende Form empfunden, die anschließende Berechnung der fehlenden *T*-Werte erfolgte mithilfe einer grafischen Lösung. Im Folgenden sind die kubischen Kurven und grafischen Berechnungen der fehlenden *T*-Werte dargestellt. Gerechnet wurde in späterer Folge mit auf ganze *T*-Werte gerundeten Zahlen.

Die folgenden Abbildungen (Abb. 4-7) stellen verkleinert die erstellten Kurven für die jeweiligen Dimensionen dar; am Schnittpunkt der Hilfslinien liegt die grafische Lösung des *T*-Werts, der im Folgenden auf ganze *T*-Werte gerundet wurde.

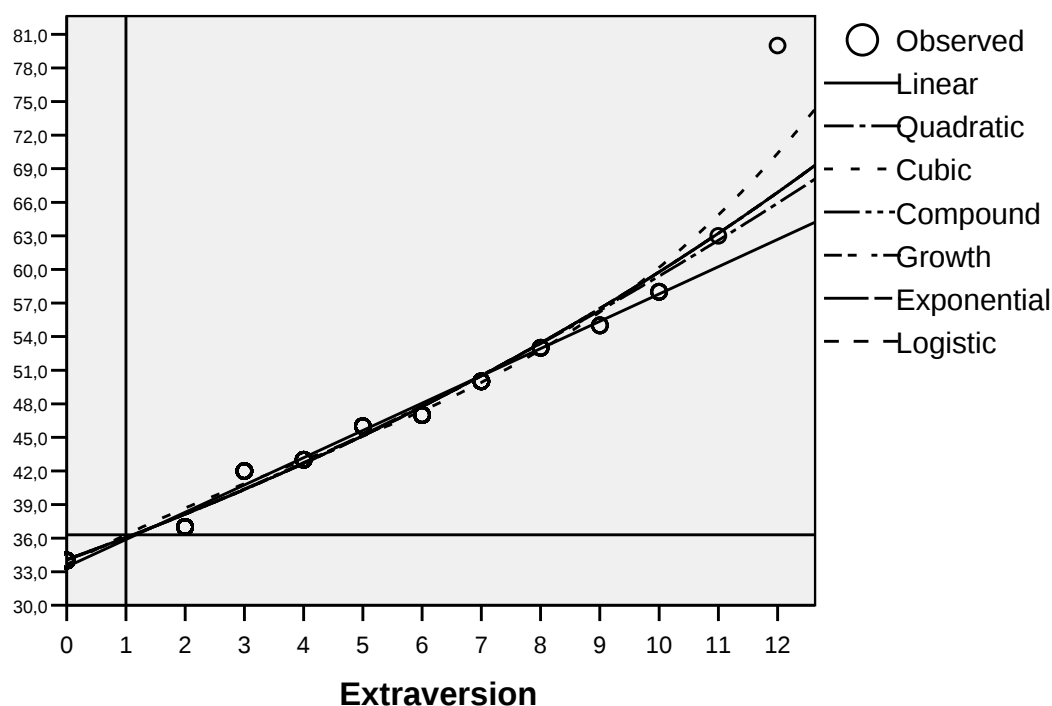


Abb. 4: Interpolieren der Skala „Extraversion“

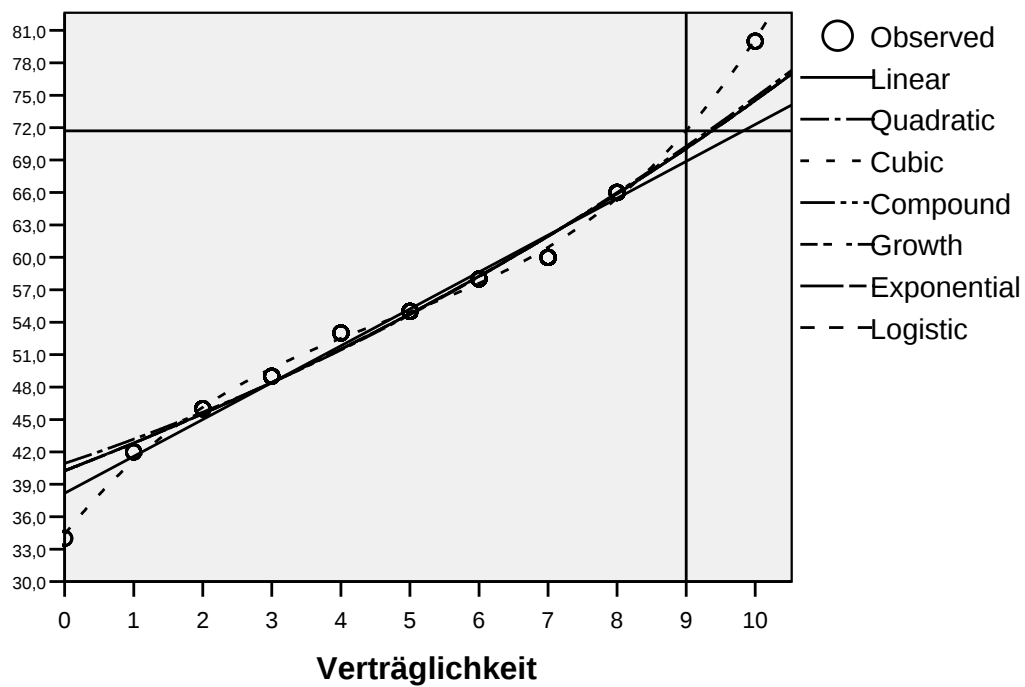


Abb. 5: Interpolieren der Skala „Verträglichkeit“

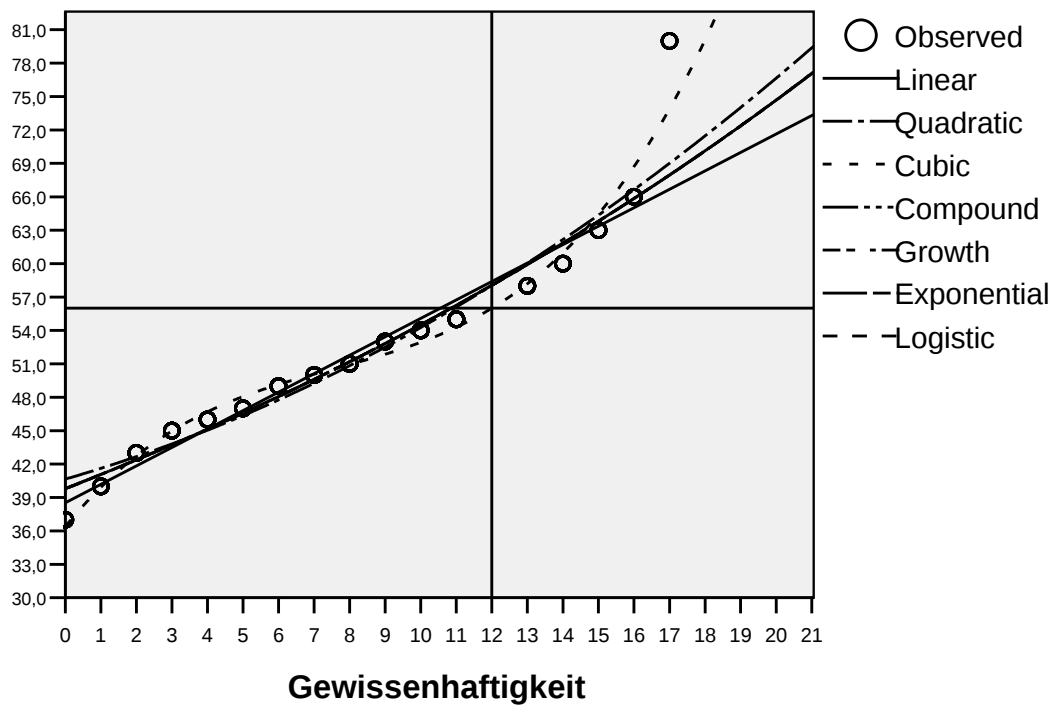


Abb. 6: Interpolieren der Skala „Gewissenhaftigkeit“

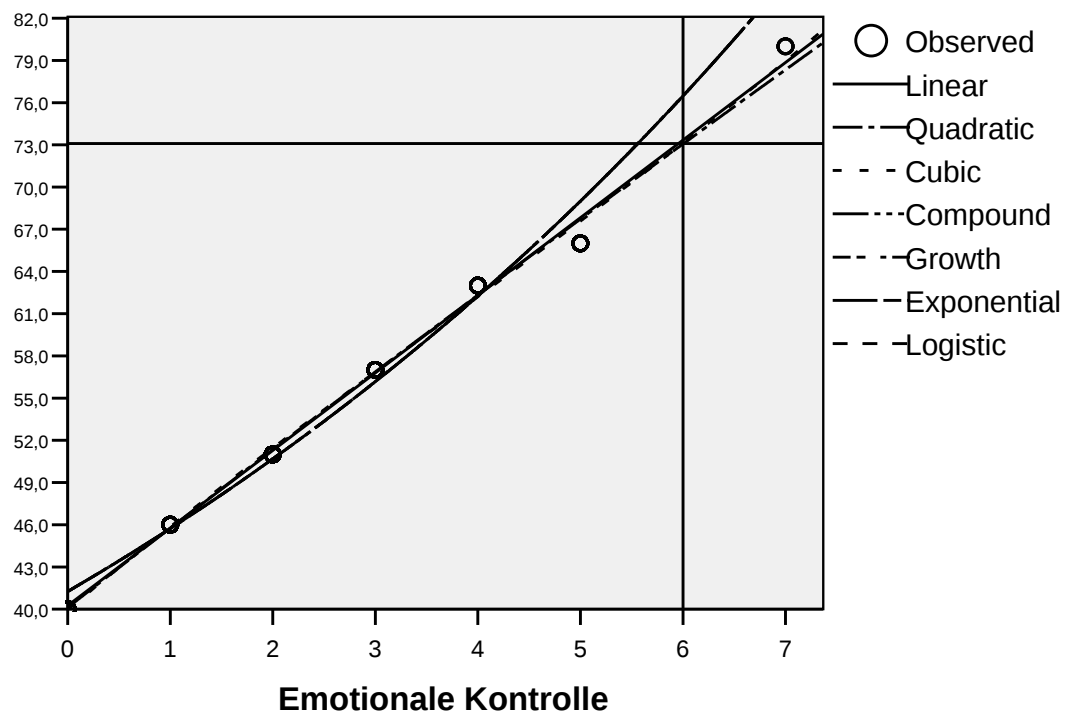


Abb. 7: Interpolieren der Skala „Emotionale Kontrolle“

Die sich anhand der grafischen Lösung ergebenden *T*-Werte wurden auf ganze Zahlen gerundet und betragen daher

- für den Rohwert 1 der Skala „Extraversion“ *T*-Wert 36,
- für den Rohwert 9 der Skala „Verträglichkeit“ *T*-Wert 72,
- für den Rohwert 12 der Skala „Gewissenhaftigkeit“ *T*-Wert 56,
- und für den Rohwert 6 der Skala „Emotionale Kontrolle“ *T*-Wert 73.

13. Ergebnisse

Zunächst werden die Ergebnisse der Extremgruppenvalidierung dargestellt, erst in Folge soll auf Ergebnisse zu einem möglichen State-Einfluss eingegangen werden. Die Berechnung erfolgte für alle Ergebnisse mit dem Statistikprogramm *SPSS* (Version 11.0).

13.1 Extremgruppenvergleiche

Anhand der Aufsätze wurden für jede Skala Extremgruppen gebildet, die dann hinsichtlich ihrer *T*-Werte im Persönlichkeitsverfahren *B5PO* verglichen werden. Um den Überblick zu wahren, sei jede Skala einzeln dargestellt.

13.1.1 Extraversion

Für die Skala „Extraversion“ beträgt der Umfang beider anhand des Aufsatzes ermittelten Extremgruppen jeweils $n = 14$ (vgl. Kap. 11.4.1). Rechnet man mit diesen Gruppen einen *t*-Test für unabhängige Stichproben mit dem *T*-Wert für Extraversion als abhängige Variable, ergibt sich ein signifikantes Ergebnis. Die nach dem offenen Antwortformat gebildeten Extremgruppen unterscheiden sich also hinsichtlich ihrer Werte auf der Skala „Extraversion“ signifikant voneinander. Der im Vorhinein festgelegte Unterschied von mindestens 10 *T*-Werten wurde zwar nicht ganz erreicht, liegt jedoch im 95%-Konfidenzintervall: Der Mittelwertunterschied beträgt knapp unter 9 *T*-Werten (der durchschnittliche Mittelwert der Introvertierten liegt bei 37,875, jener der Extravertierten bei 46,786), das Konfidenzintervall für den Mittelwertunterschied reicht von 5,297 *T*-Werten bis 12,56 *T*-Werten. Zur Veranschaulichung sei hier der *SPSS*-Output in Tabelle 12 dargestellt.

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower Upper
T-Wert Extraversion	Equal variances assumed	,153	,699	-5,054	26	,000	-8,929	1,7667	-12,5601 -5,2970
	Equal variances not assumed			-5,054	24,878	,000	-8,929	1,7667	-12,5681 -5,2891

Tab. 12: *t*-Test des Extremgruppenvergleichs für die Skala „Extraversion“

Zur Skala „Extraversion“ des *B5PO* ist zu sagen, dass eine stark rechtsschiefe Verteilung ins Auge fällt: Ein Mittelwert von rund $T = 47$ für die Extremgruppe der Extravertierten liegt sogar unter dem Erwartungswert von $T = 50$ für eine Gesamtstichprobe, dabei handelt es sich hier um eine Extremgruppe. Im Vergleich zur gesamten hier vorliegenden Stichprobe ist

allerdings auch ein *T*-Wert von 47 recht viel, da mehr als die Hälfte der Kontroll- und Versuchsgruppen des Aufsatzes *T*-Werte zwischen 34 (den im Verfahren niedrigstem erzielbaren Wert) und 42 aufweist, *T*-Werte von mehr als 50 weisen nicht einmal sieben Prozent auf (vgl. Tab. 13).

T-Wert Extraversion					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	34,0	27	15,4	15,4	15,4
	36,0	26	14,9	14,9	30,3
	37,0	29	16,6	16,6	46,9
	42,0	14	8,0	8,0	54,9
	43,0	18	10,3	10,3	65,1
	46,0	17	9,7	9,7	74,9
	47,0	17	9,7	9,7	84,6
	50,0	15	8,6	8,6	93,1
	53,0	4	2,3	2,3	95,4
	55,0	3	1,7	1,7	97,1
	58,0	3	1,7	1,7	98,9
	63,0	1	,6	,6	99,4
	80,0	1	,6	,6	100,0
	Total	175	100,0	100,0	

Tab. 13: Verteilung der *T*-Werte der Skala „Extraversion“

Aufgrund dieser schiefen Verteilung scheint der Unterschied um so stärker zu sein; so oder so war das Ergebnis jedoch signifikant und die Alternativhypothese kann angenommen werden: Die anhand des Aufsatzes gebildeten Extremgruppen zur Skala „Extraversion“ unterscheiden sich hinsichtlich ihrer Testwerte im *B5PO* auf ebendieser Skala deutlich voneinander.

13.1.2 Verträglichkeit

Auf der Skala „Verträglichkeit“ wurden $n = 15$ Schüler/innen anhand des Aufsatzes in die sehr verträgliche Richtung zugeteilt, $n = 14$ Schüler/innen wurden in die sehr unverträgliche Richtung zugeteilt.

Auch hier zeigte sich eine eher linksschiefe Verteilung: Der Mittelwert der als sehr verträglich eingestuften Extremgruppe liegt bei 64 *T*-Werten, jener der unverträglichen Extremgruppe mit 51,643 *T*-Werten jedoch ebenfalls über dem Erwartungsmittelwert. Trotz dieser linksschiefen Verteilung liegt der Unterschied deutlich über einer Standardabweichung und ist signifikant. Im Folgenden (Tab. 14) sei der *t*-Test zwischen den beiden Extremgruppen dargestellt.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Verträglichkeit	Equal variances assumed	1,127	,298	-5,240	27	,000	-12,357	2,3581	-17,1956	-7,5187
	Equal variances not assumed			-5,309	24,779	,000	-12,357	2,3274	-17,1527	-7,5615

Tab. 14: t-Test des Extremgruppenvergleichs für die Skala „Verträglichkeit“

Aufgrund dieses Ergebnisses kann für die zweite Hypothese (vgl. Kap. 7.1) die Nullhypothese verworfen und die Alternativhypothese angenommen werden: Die Extremgruppen, die anhand des Aufsatzes gebildet wurden, unterscheiden sich signifikant und in relevantem Ausmaß voneinander.

13.1.3 Gewissenhaftigkeit

Auch in der Skala „Gewissenhaftigkeit“ fanden sich signifikante Unterschiede zwischen den beiden mittels Aufsatz gebildeten Extremgruppen ($n = 13$ für die Extremgruppe der wenig gewissenhaften sowie $n = 16$ für die Extremgruppe der sehr gewissenhaften Testpersonen): Der Mittelwert der Gruppe, die als sehr gewissenhaft eingeschätzt wurde, liegt bei 59 T -Werten, jener der Gruppe, die als nicht gewissenhaft eingestuft wurde, bei 47,467 T -Werten. In Tab. 15 sei der t -Test zwischen den beiden Extremgruppen dargestellt.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Gewissenhaftigkeit	Equal variances assumed	1,118	,300	5,145	26	,000	11,533	2,2415	6,9259	16,1407
	Equal variances not assumed			4,956	18,509	,000	11,533	2,3274	6,6533	16,4134

Tab. 15: t-Test des Extremgruppenvergleichs für die Skala „Gewissenhaftigkeit“

Auch für diese Dimension kann der Extremgruppenvergleich also als gelungen betrachtet werden. Der Mittelwertunterschied liegt über dem festgelegten Mindestunterschied, die dritte Nullhypothese wird somit verworfen und die Alternativhypothese – nämlich, dass sich die Testpersonen, die sich im Aufsatz als gewissenhaft darstellen, auch im Persönlichkeitsverfahren als gewissenhafter darstellen – wird angenommen.

13.1.4 Emotionale Kontrolle

Auch in der Skala „Emotionale Kontrolle“ unterscheiden sich die mittels Aufsatz gebildeten Extremgruppen ($n = 13$ für emotional kontrollierte bzw. $n = 15$ für wenig emotional kontrollierte Schüler/innen) signifikant voneinander (siehe Tab. 16). Auch hier wurde der im Vorhinein festgelegte Mindestunterschied von 10 T -Werten nicht ganz erreicht (der Unterschied lag bei rund 8,8 T -Werten), liegt jedoch im 95%-Konfidenzintervall (vgl. Tab. 16). Der Mittelwert der Schüler/innen, die für die emotional kontrollierte Extremgruppe herangezogen wurden, liegt bei 59,462 T -Werten, jener der Schüler/innen, die sich im Aufsatz wenig emotional kontrolliert darstellten, bei 50,667. Erneut liegt also eine recht schiefe Verteilung vor.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Emotionale Kontrolle	Equal variances assumed	2,834	,104	2,608	26	,015	8,795	3,3718	1,8641	15,7256
	Equal variances not assumed			2,533	20,229	,020	8,795	3,4727	1,5562	16,0335

Tab. 16: t-Test des Extremgruppenvergleichs für die Skala „Emotionale Kontrolle“

Analog zur Skala „Extraversion“ liegt auch hier ein signifikanter Unterschied vor, wobei der mindestens geforderte Unterschied von 10 T -Werten (vgl. Kap. 7.1) im Konfidenzintervall liegt – die Nullhypothese muss somit verworfen werden.

13.1.5 Offenheit

In dieser Dimension lässt sich kein signifikanter Unterschied zwischen den mittels Aufsatz gebildeten Extremgruppen ($n = 13$ je Extremgruppe) feststellen. Der Mittelwertunterschied zwischen diesen beiden Extremgruppen war auch recht gering, nur etwas über drei T -Werte (vgl. Tab. 17).

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Offenheit	Equal variances assumed	2,517	,126	1,229	24	,231	3,231	2,6284	-2,1939	8,6554
	Equal variances not assumed			1,229	19,803	,233	3,231	2,6284	-2,2554	8,7169

Tab. 17: t-Test des Extremgruppenvergleichs für die Skala „Offenheit“

Auffällig ist bei dieser Skala ein (nicht signifikanter) Unterschied in der Standardabweichung, der etwas im Widerspruch zu den Problemen der Extremgruppenbildung steht: Bei Testpersonen der Skala, die laut Aufsatz als eindeutig offen eingeschätzt wurden, ist die Standardabweichung weit größer (8,098) als bei jenen, die als wenig offen eingeschätzt wurden (4,923). Im Gegensatz dazu wurde die Extremgruppe, die sich im Aufsatz als wenig offen darstellte, weit heterogener eingeschätzt (vgl. Kap. 11.4.5). Dies mag durch eine im Aufsatz sozial erwünschte Darstellung zu erklären sein, die bei der qualitativen Auswertung im Gegensatz zum Persönlichkeitsverfahren nicht durch eine verschobene Dichotomisierung ausgeglichen werden konnte.

13.1.6 Empathie

Für diese Skala beträgt die endgültige Größe der Extremgruppen $n = 14$ für sehr empathisch eingeschätzte und $n = 12$ für wenig empathisch eingeschätzte Schüler/innen. Es zeigt sich erneut ein signifikanter Unterschied zwischen den beiden mittels Aufsatz gebildeten Extremgruppen. Wieder ist der Unterschied geringer als der im Vorhinein festgelegten Mindestunterschied von 10 T -Werten, liegt aber bei vorher berechneter Stichprobengröße im 95%-Konfidenzintervall, weshalb auch hier die Alternativhypothese angenommen wird (vgl. Kap. 7.1): Die mittels Aufsatz gebildeten Extremgruppen zur Skala „Empathie“ unterschieden sich auf ebendieser Skala im *B5PO* signifikant voneinander. Bei der Extremgruppe, die sich im Aufsatz als sehr empathisch darstellte, liegt der Mittelwert bei 64,357 T -Werten, bei jener Extremgruppe, die sich als wenig empathisch darstellte, bei 55,417 T -Werten. Der SPSS-Output (inkl. Konfidenzintervall für den Mittelwertunterschied) ist in Tabelle 18 dargestellt.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Empathie	Equal variances assumed	,020	,889	2,304	24	,030	8,940	3,8796	,9333	16,9476
	Equal variances not assumed			2,290	22,728	,032	8,940	3,9034	,8602	17,0207

Tab. 18: t-Test des Extremgruppenvergleichs für die Skala „Empathie“

In Anbetracht der Tatsache, dass hier eine doch stark linksschiefe Verteilung vorliegt, scheint der Mittelwertunterschied um so stärker zu wiegen: Trotz verschobener Dichotomisierung bei der Testerstellung liegt der Mittelwert derjenigen Personen, die in der wenig empathischen Extremgruppe liegen, über dem Erwartungsmittelwert von $T = 50$. Dies mag teilweise darin

begründet sein, dass der niedrigste zu erreichende *T*-Wert dieser Skala 40 beträgt (vgl. Tabelle 19 zur Verteilung der *T*-Werte).

T-Wert Empathie					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	40,0	16	9,1	9,1	9,1
	45,0	17	9,7	9,7	18,9
	49,0	16	9,1	9,1	28,0
	51,0	16	9,1	9,1	37,1
	54,0	19	10,9	10,9	48,0
	55,0	25	14,3	14,3	62,3
	58,0	19	10,9	10,9	73,1
	60,0	16	9,1	9,1	82,3
	66,0	17	9,7	9,7	92,0
	80,0	14	8,0	8,0	100,0
	Total	175	100,0	100,0	

Tab. 19: Verteilung der *T*-Werte der Skala „Empathie“

13.2 Überprüfung einer möglichen State-Abhängigkeit

Zunächst sei Hypothese 7 (vgl. Kap. 7.1) beantwortet, da auch diese sich gewissermaßen auf eine State-Abhängigkeit bezieht. Erst danach soll auf die systematische Überprüfung eines State-Einflusses durch das Rätsel eingegangen werden.

13.2.1 Überprüfung eines möglichen State-Einflusses durch den Aufsatz

Für die Überprüfung, ob der Aufsatz selbst einen Einfluss auf das Testergebnis im Persönlichkeitsinventar *B5PO* hatte, standen aufgrund der Datenerhebung zwei Kontrollgruppen zur Verfügung, wobei sechs Testpersonen in beiden Kontrollgruppen vorkamen (vgl. Kap. 10). Diese Kontrollgruppen wurden randomisiert aus derselben Grundgesamtheit zugeteilt und hatten beide dieselben Bedingungen (beide starteten gleich zu Beginn mit dem Persönlichkeitsverfahren), während die Versuchsgruppe davor einen Aufsatz bearbeitete.

Es zeigt sich ein signifikanter Unterschied zwischen Versuchsgruppe und einer der beiden Kontrollgruppen hinsichtlich der *T*-Werte auf der Skala „Empathie“, jedoch nicht zwischen Versuchsgruppe und der anderen Kontrollgruppe – hier ist zwar deskriptiv ein Unterschied bemerkbar, dieser ist jedoch nicht groß genug, um signifikant zu werden. Mit beiden Kontrollgruppen gemeinsam verglichen ist der Mittelwertunterschied jedoch erneut signifikant. Der Unterschied geht jedes Mal in die Richtung, dass Testpersonen, die den

Aufsatz bearbeiteten, sich im Anschluss im Persönlichkeitsverfahren als höher empathisch einschätzen.

Im Folgenden seien in den Tabellen 20, 21 und 22 die SPSS-Outputs zwischen der Versuchsgruppe und der jeweiligen Kontrollgruppe dargestellt.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Extraversion	Equal variances assumed	,138	,711	,339	120	,735	,395	1,1639	-1,9099	2,6992
	Equal variances not assumed			,339	118,930	,735	,395	1,1651	-1,9125	2,7017
T-Wert Verträglichkeit	Equal variances assumed	,005	,942	1,311	120	,193	1,803	1,3755	-,9207	4,5261
	Equal variances not assumed			1,311	119,985	,192	1,803	1,3750	-,9197	4,5251
T-Wert Gewissenhaftigkeit	Equal variances assumed	1,064	,304	-,246	120	,806	-,396	1,6111	-3,5855	2,7941
	Equal variances not assumed			-,245	114,987	,807	-,396	1,6158	-3,5962	2,8048
T-Wert Emotionale Kontrolle	Equal variances assumed	,899	,345	1,017	120	,311	1,815	1,7845	-1,7188	5,3478
	Equal variances not assumed			1,016	118,955	,312	1,815	1,7863	-1,7226	5,3517
T-Wert Offenheit	Equal variances assumed	,002	,961	1,578	120	,117	2,160	1,3687	-,5496	4,8701
	Equal variances not assumed			1,580	119,938	,117	2,160	1,3674	-,5471	4,8676
T-Value Empathie	Equal variances assumed	,940	,334	2,334	120	,021	4,364	1,8696	,6624	8,0656
	Equal variances not assumed			2,338	119,427	,021	4,364	1,8664	,6684	8,0595

Tab. 20: t-Test des Vergleichs zwischen Versuchsgruppe und Kontrollgruppe 1

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Extraversion	Equal variances assumed	2,405	,124	,121	119	,904	,161	1,3301	-2,4724	2,7950
	Equal variances not assumed			,120	107,455	,904	,161	1,3394	-2,4938	2,8164
T-Wert Verträglichkeit	Equal variances assumed	2,040	,156	-1,034	119	,303	-1,572	1,5209	-4,5837	1,4393
	Equal variances not assumed			-1,030	113,972	,305	-1,572	1,5270	-4,5972	1,4528
T-Wert Gewissenhaftigkeit	Equal variances assumed	,017	,898	,178	119	,859	,252	1,4183	-2,5560	3,0607
	Equal variances not assumed			,178	118,911	,859	,252	1,4156	-2,5506	3,0553
T-Wert Emotionale Kontrolle	Equal variances assumed	,985	,323	-,091	119	,927	-,164	1,7984	-3,7252	3,3967
	Equal variances not assumed			-,091	117,374	,927	-,164	1,8014	-3,7318	3,4032
T-Wert Offenheit	Equal variances assumed	,001	,974	1,232	119	,220	1,727	1,4022	-1,0490	4,5039
	Equal variances not assumed			1,232	118,846	,220	1,727	1,4017	-1,0481	4,5030
T-Value Empathie	Equal variances assumed	1,196	,276	,986	119	,326	1,801	1,8257	-1,8141	5,4161
	Equal variances not assumed			,991	117,330	,324	1,801	1,8180	-1,7994	5,4014

Tab. 21: t-Test des Vergleichs zwischen Versuchsgruppe und Kontrollgruppe 2

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Extraversion	Equal variances assumed	1,273	,261	,507	173	,612	,568	1,1200	-1,6422	2,7790
	Equal variances not assumed			,536	146,436	,593	,568	1,0607	-1,5280	2,6647
T-Wert Verträglichkeit	Equal variances assumed	1,010	,316	-,035	173	,972	-,045	1,3084	-2,6277	2,5372
	Equal variances not assumed			-,036	138,090	,972	-,045	1,2662	-2,5489	2,4584
T-Wert Gewissenhaftigkeit	Equal variances assumed	,131	,718	-,014	173	,989	-,018	1,3012	-2,5868	2,5500
	Equal variances not assumed			-,014	128,684	,989	-,018	1,2908	-2,5723	2,5355
T-Wert Emotionale Kontrolle	Equal variances assumed	1,456	,229	,497	173	,620	,786	1,5817	-2,3361	3,9076
	Equal variances not assumed			,507	133,341	,613	,786	1,5497	-2,2794	3,8509
T-Wert Offenheit	Equal variances assumed	,008	,930	1,558	173	,121	1,879	1,2063	-,5016	4,2603
	Equal variances not assumed			1,546	122,908	,125	1,879	1,2156	-,5268	4,2856
T-Wert Empathie	Equal variances assumed	1,477	,226	2,067	173	,040	3,244	1,5698	,1460	6,3427
	Equal variances not assumed			1,984	111,566	,050	3,244	1,6354	,0038	6,4849

Tab. 22: t-Test des Vergleichs zwischen Versuchsgruppe und der gesamten Kontrollgruppe

Aufgrund dieses Ergebnisses empfiehlt es sich, die beiden Kontrollgruppen (jeweils ohne der sechs Testpersonen, die sich in beiden Gruppen überschneiden) nochmals detaillierter zu betrachten und miteinander zu vergleichen. Zur Veranschaulichung sei in der folgenden grafischen Darstellung mittels Boxplots der Mittelwertunterschied der Skala „Empathie“ dargestellt (berücksichtigt wurden nur die beiden unterschiedlichen Kontrollgruppen, Abb. 8), Ausreißer sind bei beiden annähernd gleich verteilt.

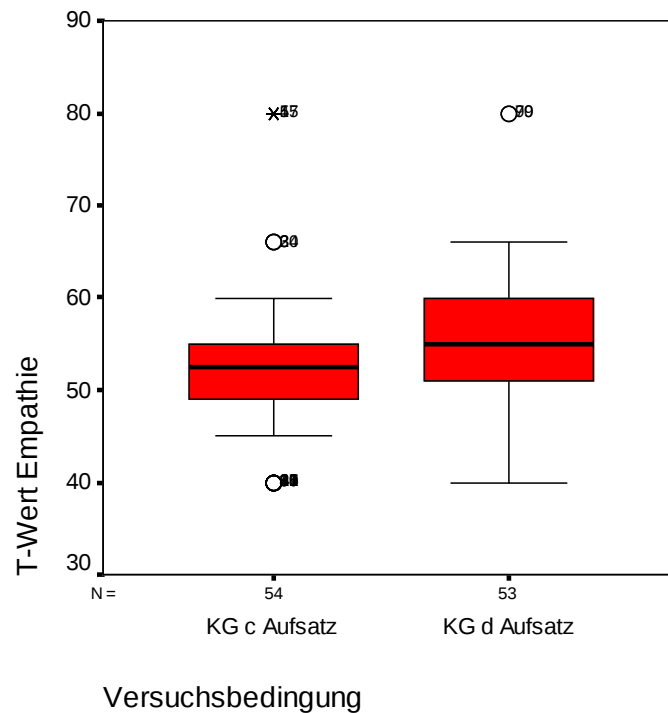


Abb. 8: Boxplots der beiden unterschiedlichen Kontrollgruppen für die Aufsatzbedingung

Da die Kontrollgruppen gleichmäßig aus der gleichen Grundgesamtheit zugewiesen wurden und dieselben Bedingungen hatten, stellt sich als erste Vermutung, dass sich die beiden Gruppen a priori durch Zufall unterscheiden. Hierzu wurde ein *t*-Test über alle sechs Dimensionen gerechnet. Es zeigt sich, dass in der interessierenden Dimension „Empathie“ kein signifikanter Unterschied vorliegt, jedoch in der Skala „Verträglichkeit“ (vgl. Tab. 23). Durch den in der Tabelle ersichtlichen nicht-signifikanten Unterschied der Skala „Empathie“ ist jedoch ein wesentlicher zufallsbedingter Unterschied zwischen den beiden Kontrollgruppen auf ebendieser Dimension ausgeschlossen. Der Unterschied zwischen den Vergleichen mit den verschiedenen Kontrollgruppen muss sich also anders begründen. Daher soll im Folgenden versucht werden, Gründe für den – teilweise signifikanten, bei einer Kontrollgruppe jedoch nicht signifikanten – Unterschied auf der Skala „Empathie“ zu finden.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Extraversion	Equal variances assumed	1,386	,242	-,169	105	,866	-,248	1,4652	-3,1533	2,6572
	Equal variances not assumed			-,169	98,187				-3,1624	2,6663
T-Wert Verträglichkeit	Equal variances assumed	1,815	,181	-2,265	105	,026	-3,760	1,6605	-7,0528	-,4678
	Equal variances not assumed			-2,261	101,371				-7,0595	-,4611
T-Wert Gewissenhaftigkeit	Equal variances assumed	1,910	,170	,468	105	,641	,723	1,5438	-2,3382	3,7840
	Equal variances not assumed			,470	95,643				-2,3321	3,7779
T-Wert Emotionale Kontrolle	Equal variances assumed	,002	,963	-1,107	105	,271	-2,203	1,9898	-6,1480	1,7427
	Equal variances not assumed			-1,107	104,936				-6,1482	1,7429
T-Wert Offenheit	Equal variances assumed	,005	,944	-,325	105	,746	-,484	1,4878	-3,4339	2,4660
	Equal variances not assumed			-,325	104,592				-3,4352	2,4674
T-Wert Empathie	Equal variances assumed	,015	,901	-1,595	105	,114	-2,845	1,7836	-6,3811	,6920
	Equal variances not assumed			-1,596	104,736				-6,3789	,6898

Tab. 23: Vergleich der Kontrollgruppen für die Extremgruppenbildung

Nachtragshypothese: Geschlechtsbedingte Unterschiede auf der Skala „Empathie“

Aufgrund der randomisierten Zuteilung zu verschiedenen Kontrollgruppen stellt sich die Frage, warum im Vergleich der Versuchsgruppe mit einer Kontrollgruppe ein signifikantes Ergebnis erzielt werden konnte, nicht jedoch im Vergleich mit der anderen Kontrollgruppe, wenn sowohl die Versuchsbedingungen gleich waren, als auch ein zufällig bestehender Unterschied zwischen den Kontrollgruppen selbst ausgeschlossen werden kann (vgl. Tab. 23).

Bei der Analyse der Geschlechterhäufigkeiten (vgl. Kap. 10.1) zeigt sich, dass zwar in den beiden Teilkontrollgruppen kein signifikanter Unterschied zwischen den Geschlechtern vorliegt, die höhere Anzahl von Frauen in der gesamten Kontrollgruppe jedoch signifikant ist. Möglicherweise lässt sich hierin ein Grund für den Unterschied finden, als Nachtragshypothese soll also untersucht werden, ob männliche und weibliche Schüler/innen sich hinsichtlich dieser Skala vielleicht unterschiedlich State-beeinflusst zeigen.

Zunächst wird untersucht, ob weibliche und männliche Schüler/innen sich auf dieser Skala unterschiedlich darstellen. Über die gesamte Aufsatz-Stichprobe gerechnet zeigt sich, dass weibliche Testpersonen sich signifikant empathischer darstellen als männliche Testpersonen, das Ergebnis ist in Tab. 24 dargestellt. Auch ein (an dieser Stelle nicht dargestellter) signifikanter Mittelwertunterschied der Skala „Emotionale Kontrolle“ erweist sich als

auffällig (der Mittelwert der Schüler liegt bei 58,444 *T*-Werten, jener der Schülerinnen bei 52,689 *T*-Werten), dies zeigt sich später nochmals (vgl. Exkurs in Kap. 13.2.2).

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
T-Wert Empathie	Equal variances assumed	2,793	,096	-3,025	173	,003	-4,554	1,5052	-7,5245 -1,5828
	Equal variances not assumed			-3,137	168,953	,002	-4,554	1,4517	-7,4194 -1,6879

Tab. 24: Vergleich von männlichen und weiblichen Testpersonen auf der Skala „Empathie“

Aufgrund des signifikanten Unterschiedes zwischen Schülerinnen und Schülern werden nun noch einmal jeweils für die einzelnen Geschlechter *t*-Tests gerechnet. Berücksichtigt man nur die weiblichen Testpersonen, stellt sich die Versuchsgruppe sowohl signifikant empathischer dar als die gesamte Kontrollgruppe (der Mittelwertunterschied beträgt 6,38 *T*-Werte), als auch als Kontrollgruppe 1 (signifikanter Mittelwertunterschied von rund 7,81 *T*-Werten) und Kontrollgruppe 2 (rund 5,02 *T*-Werte). Kein signifikanter Unterschied findet sich hingegen zwischen den beiden weiblichen Kontrollgruppen. Bei Schülerinnen liegt also ein signifikanter State-Einfluss durch die Aufsatzvorgabe vor.

Zur Demonstration sei der *t*-Test des noch geringsten signifikanten Unterschiedes dargestellt (vgl. Tab. 25).

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
T-Wert Empathie	Equal variances assumed	,525	,471	-2,061	69	,043	-5,016	2,4338	-9,8713 -,1605
	Equal variances not assumed			-2,058	67,994	,043	-5,016	2,4371	-9,8790 -,1527

Tab. 25: Vergleich weiblicher VG und weiblicher KG 2 auf der Skala „Empathie“

Der signifikante Unterschied zwischen Versuchs- und Kontrollgruppen bei ausschließlich weiblichen Testpersonen kann sich entweder durch die Anzahl weiblicher Testpersonen oder durch einen Einfluss des Aufsatzes begründen. Gegen einen reinen Einfluss der Anzahl weiblicher Testpersonen spricht jedoch, dass der Prozentsatz weiblicher Testpersonen in allen Gruppen und Teilgruppen annähernd gleich ist (zwischen 56 und 61%, vgl. Kap. 10.1) und im gesamtgeschlechtlichen Vergleich gerade der Unterschied zu Kontrollgruppe 2 mit dem

höchsten Frauenanteil nicht signifikant ist; zusätzlich stellt gerade diese Kontrollgruppe im rein weiblichen Vergleich auch den geringsten Effekt dar. Natürlich könnte der gesamtgeschlechtlich etwas zu geringe Unterschied zur Kontrollgruppe 2 auch durch Zufall zustande gekommen sein, hierauf soll unten bei der Darstellung der Testergebnisse männlicher Versuchsteilnehmer nochmals eingegangen werden. Jedenfalls müsste aufgrund der signifikanten Unterschiede im rein weiblichen Vergleich für die siebente Hypothese die Nullhypothese verworfen werden und die Alternativhypothese angenommen werden. Da die Fragestellung jedoch nicht geschlechtsspezifisch formuliert war, muss die Nullhypothese beibehalten werden.

Vergleicht man nun die männlichen Testpersonen, findet sich kein Unterschied, weder zwischen Versuchs- und Kontrollgruppen noch zwischen den Kontrollgruppen. Die Mittelwertunterschiede der Skala „Empathie“ bei männlichen Testpersonen lagen bei max. 2,335 *T*-Werten (bei Kontrollgruppe 2, hier stellten sich die Aufsatzschreibenden tendenziell weniger empathisch dar als jene ohne Aufsatz). Im Vergleich zwischen Versuchsgruppe und Kontrollgruppe 1 findet sich faktisch gar kein Unterschied (nur 0,93 *T*-Werte). Für männliche Testpersonen muss ein Einfluss des Aufsatzes also definitiv verneint werden.

Allerdings findet sich hier die Antwort auf die implizite Frage, warum im gesamtgeschlechtlichen Vergleich gerade der Unterschied zu Kontrollgruppe 2 nicht signifikant wird: Es ist genau diese Kontrollgruppe, in der der Effekt bei weiblichen Testpersonen (zufällig) etwas schwächer ist als bei der anderen Kontrollgruppe, bei männlichen Testpersonen findet sich jedoch (zufällig) eine etwas stärkere Tendenz als bei Kontrollgruppe 1 – und der Effekt bei weiblichen Testpersonen geht in die andere Richtung als die Tendenz bei männlichen Testpersonen.

In einem Fazit kann man also sagen, dass es durchaus einen State-Einfluss des Aufsatzes gibt – allerdings nur bei Schülerinnen! Dass dies im gesamtgeschlechtlichen Vergleich teilweise signifikant wird, liegt am Überhang weiblicher Versuchsteilnehmerinnen. Und dass dieser Effekt im Vergleich mit einer der beiden Kontrollgruppen nicht signifikant ist, liegt daran, dass der zufällig etwas geringere Effekt bei weiblichen Schülerinnen durch eine (zufällige) leichte, gegensätzlich gerichtete Tendenz bei männlichen Testpersonen noch etwas verschmälert wird.

13.2.2 Überprüfung eines State-Einflusses durch die Suchbildvorgabe

Auch für die systematische Überprüfung einer State-Abhängigkeit mittels vorheriger Frustration stehen aufgrund des Untersuchungsplans zwei Kontrollgruppen zur Verfügung. Zunächst sei Hypothese 8 beantwortet, nämlich, ob sich Testpersonen, die vorher frustriert wurden, sich anschließend im Persönlichkeitsverfahren *B5PO* anders darstellen. Hierzu standen drei Gruppen zur Verfügung: Testpersonen, die im Suchbild-Rätsel Erfolg hatten (jene, die die Möglichkeit dazu hatten, aber das Rätsel dennoch nicht lösen konnten, wurden ausgeschlossen), jene, die durch das Rätsel frustriert wurden sowie eine gesamte Kontrollgruppe, die noch gesplittet werden konnte (vgl. Kap. 10.2). Gerechnet werden soll mit einer einfachen Varianzanalyse. Da der Test auf Homogenität der Varianzen (Tab. 26) für die Skala „Empathie“ signifikant ist, wird diese für die weitere Beurteilung später extra betrachtet.

Test of Homogeneity of Variances				
	Levene Statistic	df1	df2	Sig.
T-Wert Extraversion	,648	2	161	,524
T-Wert Verträglichkeit	1,912	2	161	,151
T-Wert Gewissenhaftigkeit	,213	2	161	,808
T-Wert Emotionale Kontrolle	2,430	2	161	,091
T-Wert Offenheit	,118	2	161	,889
T-Value Empathie	6,190	2	161	,003

Tab. 26: Test auf Homogenität der Varianzen zwischen den Rätsel-Bedingungen

In der nachfolgenden einfachen Varianzanalyse (gerechnet wurde mit der gesamten Kontrollgruppe) zeigt sich, dass ein signifikanter Unterschied in der Skala „Gewissenhaftigkeit“ vorliegt¹⁴ (Tab. 27).

¹⁴ Auch für die Skala „Empathie“ findet sich ein signifikanter Unterschied, hier sind jedoch die Varianzen nicht homogen, daher wird an dieser Stelle von einer weiteren Interpretation abgesehen.

		ANOVA				
		Sum of Squares	df	Mean Square	F	Sig.
T-Wert Extraversion	Between Groups	84,419	2	42,209	1,188	,308
	Within Groups	5722,459	161	35,543		
	Total	5806,878	163			
T-Wert Verträglichkeit	Between Groups	65,568	2	32,784	,546	,580
	Within Groups	9668,695	161	60,054		
	Total	9734,262	163			
T-Wert Gewissenhaftigkeit	Between Groups	470,227	2	235,114	4,155	,017
	Within Groups	9109,303	161	56,580		
	Total	9579,530	163			
T-Wert Emotionale Kontrolle	Between Groups	111,273	2	55,636	,673	,512
	Within Groups	13313,282	161	82,691		
	Total	13424,555	163			
T-Wert Offenheit	Between Groups	13,530	2	6,765	,112	,894
	Within Groups	9715,714	161	60,346		
	Total	9729,244	163			
T-Value Empathie	Between Groups	561,073	2	280,536	3,268	,041
	Within Groups	13822,708	161	85,855		
	Total	14383,780	163			

Tab. 27: Einfache Varianzanalyse über die Rätsel-Bedingungen

Würde man statt der gesamten Kontrollgruppe an dieser Stelle die Teilkontrollgruppen heranziehen, läge nur bei einer Teilkontrollgruppe ein signifikanter Unterschied der Skala „Gewissenhaftigkeit“ vor, bei der zweiten ist der Unterschied (zufällig) etwas zu gering. Daher kann an dieser Stelle ein existierender Unterschied noch nicht postuliert werden.

In der Analyse des (nur eventuell bestehenden) Unterschiedes der Skala „Gewissenhaftigkeit“ zeigt sich, dass es nicht – wie angenommen – die Versuchsgruppe der frustrierten Schüler/innen ist, die sich anders darstellt als die Kontrollgruppe, sondern jene der im Rätsel Erfolgreichen¹⁵: In einem Boxplot (Abb. 9) seien die drei Gruppen anschaulich dargestellt, wobei die Testpersonen, die Erfolg haben konnten, ihn jedoch nicht erzielten, aus der Versuchsgruppe der Erfolgreichen ausgeschlossen sind.

¹⁵ Dies wird auch dadurch bestätigt, dass bei einer Varianzanalyse ohne Ausschluss der potentiell – aber nicht tatsächlich – erfolgreichen Testpersonen der Unterschied in der Skala „Gewissenhaftigkeit“ nicht signifikant wird.

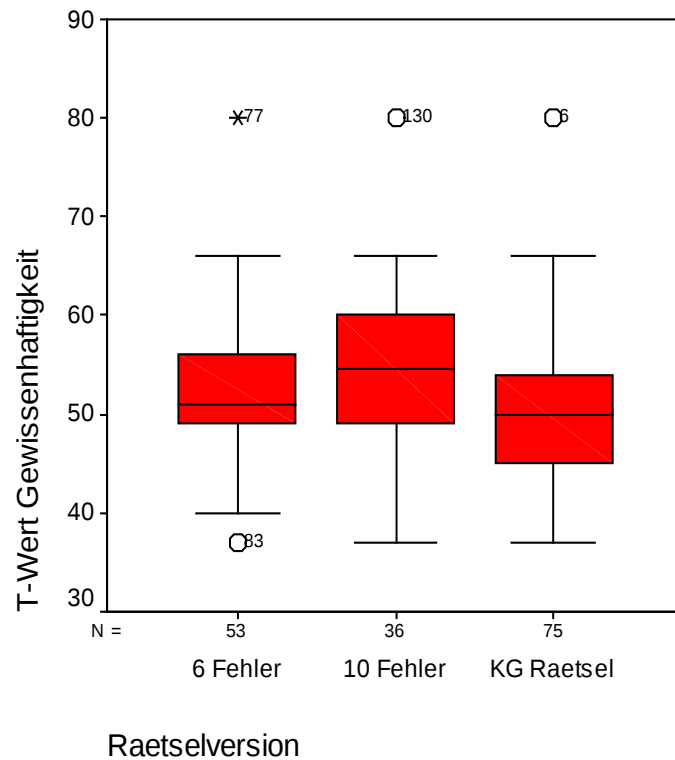


Abb. 9: Boxplots der verschiedenen Rätselbedingungen

Im Vergleich mit der gesamten Kontrollgruppe stellt sich die erfolgreiche Versuchsgruppe signifikant gewissenhafter dar (der Mittelwertunterschied beträgt 4,238 *T*-Werte, vgl. Tab. 28). Auch in Analysen mit den gesplitteten Kontrollgruppen stellt sich die erfolgreiche Versuchsgruppe immer gewissenhafter dar als die Kontrollgruppen, die Mittelwertunterschiede bei Splitten der Kontrollgruppen liegen bei rund 3,95 (KG 2) bis rund 5,02 (KG 1) *T*-Werten.

Die frustrierte Kontrollgruppe stellt sich hingegen nur im Vergleich mit KG 1 signifikant gewissenhafter dar (der Mittelwertunterschied beträgt rund 2,94 *T*-Werte), im Vergleich mit KG 2 liegt kein signifikantes Ergebnis vor, weshalb auch ein Vergleich mit der gesamten Kontrollgruppe nicht signifikant ausfällt.

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
T-Wert Gewissenhaftigkeit	Equal variances assumed	,434	,512	2,817	109	,006	4,283	1,5205	1,2698 7,2968
	Equal variances not assumed			2,699	62,130	,009	4,283	1,5873	1,1105 7,4562

Tab. 28: t-Test der Skala „Gewissenhaftigkeit“ zwischen erfolgreicher VG und gesamter KG

Zwischen den beiden Versuchsgruppen selbst (jenen, die das Rätsel erfolgreich bearbeiteten und jenen, die keine Chance auf Lösung hatten) findet sich auf dieser Skala kein signifikanter Unterschied. Zwar liegt ein kleiner Mittelwertunterschied vor, dieser ist jedoch zu gering (die Erfolgreichen erzielten einen Mittelwert von 54,417, die Frustrierten einen Mittelwert von 52,340). Der *t*-Test hierzu sei in Tab. 29 dargestellt.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
T-Wert Extraversion	Equal variances assumed	,014	,905	1,508	87	,135	1,984	1,3159	-,6313	4,5998
	Equal variances not assumed			1,490	72,181	,141	1,984	1,3315	-,6699	4,6385
T-Wert Verträglichkeit	Equal variances assumed	2,482	,119	,227	87	,821	,373	1,6427	-2,8919	3,6382
	Equal variances not assumed			,217	63,600	,829	,373	1,7160	-3,0554	3,8017
T-Wert Gewissenhaftigkeit	Equal variances assumed	,168	,683	1,234	87	,221	2,077	1,6836	-1,2693	5,4234
	Equal variances not assumed			1,217	71,714	,228	2,077	1,7065	-1,3249	5,4790
T-Wert Emotionale Kontrolle	Equal variances assumed	,425	,516	-1,157	87	,251	-2,089	1,8063	-5,6793	1,5011
	Equal variances not assumed			-1,198	83,322	,234	-2,089	1,7443	-5,5583	1,3801
T-Wert Offenheit	Equal variances assumed	,231	,632	,471	87	,639	,789	1,6746	-2,5396	4,1172
	Equal variances not assumed			,460	68,915	,647	,789	1,7149	-2,6325	4,2101
T-Value Empathie	Equal variances assumed	6,298	,014	2,153	87	,034	4,831	2,2432	,3721	9,2893
	Equal variances not assumed			2,005	56,642	,050	4,831	2,4091	,0060	9,6555

Tab. 29: *t*-Test über die 6 Skalen des B5PO zwischen frustrierter und erfolgreicher VG

Erneut zeigt sich allerdings (wie schon in der Varianzanalyse) ein signifikanter Unterschied auf der Skala „Empathie“, nämlich in der Richtung, dass Erfolgreiche sich etwas empathischer darstellen als solche, die keinen Erfolg haben konnten; da die Varianzen nicht homogen sind, wird (im Sinne des Welch-Tests) die Signifikanz für den Fall inhomogener Varianzen betrachtet. Diese liegt mit 0,050 genau grenzwertig und ist knapp signifikant. Wäre die achte Nullhypothese also für den Vergleich zwischen erfolgreichen und nicht erfolgreichen Testpersonen formuliert worden, müsste die Nullhypothese verworfen werden. Da der Effekt der Frustration aber generell erfragt wurde (vgl. Kap. 7.2), muss ein Vergleich zur Kontrollgruppe vorgenommen werden. Dabei zeigt sich, dass die Frustration keinen Einfluss auf die Skala „Empathie“ hat: Der Effekt ist nur im Vergleich mit Erfolgreichen signifikant, im Vergleich zur Gesamtkontrollgruppe fällt er hingegen nicht signifikant aus. Daher wird die achte Nullhypothese beibehalten; die Frustration hat keinen signifikanten Effekt auf die Selbstdarstellung im B5PO.

Exkurs: Berechnung eines Geschlechtereffekts auf die Skala „Empathie“

Aufgrund des tendenziellen Effekts des Erfolges auf die Skala „Empathie“ soll nun noch untersucht werden, ob sich vielleicht für die Geschlechter getrennt ein signifikanter Erfolgseffekt auf der Skala „Empathie“ zeigt. Zwar waren in der Stichprobe für die Rätseltestung die Unterschiede der Geschlechterverteilung nicht signifikant (vgl. Kap. 10.2), dennoch ist ein Einfluss durch das Geschlecht nicht auszuschließen, da in Versuchsgruppe 1 (den erfolgreichen Testpersonen) in Relation etwas mehr weibliche Testpersonen vorkamen als in Versuchsgruppe 2. Daher wurde zunächst für die gesamte Stichprobe der Suchbildtestung ein *t*-Test durchgeführt, wobei sich zeigte, dass auf den Skalen „Empathie“ und „Emotionale Kontrolle“ (wie auch schon bei der Stichprobe für die Aufsatztestung) signifikante Unterschiede zwischen männlichen und weiblichen Testpersonen bestanden (vgl. Tab. 30).

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Extraversion	Equal variances assumed	,437	,509	1,564	180	,120	1,396	,8923	-,3651	3,1563
	Equal variances not assumed			1,564	179,973					
T-Wert Verträglichkeit	Equal variances assumed	2,115	,148	-,037	180	,971	-,044	1,1948	-2,4017	2,3138
	Equal variances not assumed			-,037	178,145					
T-Wert Gewissenhaftigkeit	Equal variances assumed	,500	,481	-,674	180	,501	-,758	1,1244	-2,9769	1,4604
	Equal variances not assumed			-,674	177,823					
T-Wert Emotionale Kontrolle	Equal variances assumed	,968	,327	3,811	180	,000	5,022	1,3178	2,4217	7,6223
	Equal variances not assumed			3,811	178,059					
T-Wert Offenheit	Equal variances assumed	3,126	,079	-,605	180	,546	-,703	1,1627	-2,9977	1,5911
	Equal variances not assumed			-,605	172,924					
T-Value Empathie	Equal variances assumed	,034	,854	-2,063	180	,041	-2,868	1,3904	-5,6118	-,1245
	Equal variances not assumed			-2,063	179,890					

Tab. 30: Vergleich von männlichen und weiblichen Testpersonen der gesamten Stichprobe für die Suchbild-Testung

Aufgrund der signifikanten Unterschiede wird der Vergleich der Skala „Empathie“ von erfolgreichen Testpersonen und der Gesamtkontrollgruppe nun für Schüler und Schülerinnen getrennt berechnet. Es zeigt sich bei beiden Geschlechtern eine Tendenz, sich nach einem Erfolg empathischer darzustellen,

der Effekt ist jedoch nicht signifikant (im Vergleich zur gesamten Kontrollgruppe beträgt der Mittelwertunterschied rund 3,19 bis 4,37 *T*-Werte); es kann also von keinem geschlechtsspezifischen Effekt gesprochen werden.

Hypothese 9, dass das Rätsel selbst einen Einfluss auf die Darstellung im Persönlichkeitsverfahren hat, lässt sich bereits mit dem oben dargestellten Ergebnis der Varianzanalyse (Tab. 27) beantworten. Dennoch sei zur Absicherung nochmals ein *t*-Test gerechnet, in dem die beiden Rätselvorgaben zusammengefasst werden, diesmal mit den Bedingungen „Rätselvorgabe“ bzw. „keine Rätselvorgabe“ als abhängigen Variablen. Der SPSS-Output ist in Tabelle 31 dargestellt.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
T-Wert Extraversion	Equal variances assumed	1,065	,304	-,016	162	,987	-,015	,9384	-1,8681	1,8382
	Equal variances not assumed			-,016	159,826	,987	-,015	,9340	-1,8595	1,8295
T-Wert Verträglichkeit	Equal variances assumed	1,142	,287	-1,024	162	,307	-1,240	1,2111	-3,6316	1,1516
	Equal variances not assumed			-1,020	154,696	,309	-1,240	1,2158	-3,6418	1,1618
T-Wert Gewissenhaftigkeit	Equal variances assumed	,406	,525	2,579	162	,011	3,046	1,1813	,7136	5,3793
	Equal variances not assumed			2,597	160,750	,010	3,046	1,1729	,7303	5,3626
T-Wert Emotionale Kontrolle	Equal variances assumed	4,111	,044	,463	162	,644	,660	1,4259	-2,1562	3,4755
	Equal variances not assumed			,456	145,815	,649	,660	1,4461	-2,1984	3,5177
T-Wert Offenheit	Equal variances assumed	,018	,894	-,056	162	,955	-,069	1,2147	-2,4673	2,3301
	Equal variances not assumed			-,056	156,863	,955	-,069	1,2156	-2,4697	2,3325
T-Value Empathie	Equal variances assumed	5,627	,019	,829	162	,408	1,222	1,4739	-1,6883	4,1326
	Equal variances not assumed			,851	158,952	,396	1,222	1,4359	-1,6137	4,0580

Tab. 31: *t*-Test zwischen den Gruppen mit und ohne Rätselvorgabe

Aufgrund dieses Ergebnisses lässt sich sagen, dass Testpersonen, die zuvor ein Rätsel zu bearbeiten hatten, sich auf der Skala „Gewissenhaftigkeit“ signifikant anders darstellen als jene, die kein Rätsel zu bearbeiten hatten. Dies zeigt sich auch im zuvor gerechneten *t*-Test (vgl. Tab. 28) für den Einfluss von Erfolg, wobei an dieser Stelle nochmals festgehalten werden muss, dass es die in einer gewissenhaften Arbeit Erfolgreichen sind, die für den signifikanten Unterschied nach der Rätselvorgabe verantwortlich sind. Tendenziell stellen sich jedoch auch Testpersonen, die bei einer gewissenhaft zu bearbeitenden Aufgabe keinen Erfolg erzielen konnten, gewissenhafter dar.

Der Mittelwertunterschied zwischen Personen, die gewissenhaft arbeiteten (ohne Berücksichtigung von Erfolg und Misserfolg) und einer Kontrollgruppe liegt bei knapp über 3 *T*-Werten, ist also relativ gering. Da in der Hypothese jedoch kein Mindestunterschied festgelegt wurde, kann in diesem Fall die Nullhypothese dennoch verworfen und die Alternativhypothese angenommen werden.

14. Diskussion und Ausblick

Der Versuch einer Extremgruppenvalidierung des Persönlichkeitsverfahren *B5PO* sowie die Überprüfung eines möglichen State-Einflusses auf ebendieses Verfahren zeigen sowohl Vor- als auch Nachteile des Fragebogenverfahrens auf. Zunächst soll auf die Vorteile kurz eingegangen werden:

Bei der Extremgruppenbildung wurde zunächst der enorme Vorteil von genormten Verfahren gegenüber einem offenen Antwortformat offensichtlich. Das offene Antwortformat erwies sich in der Auswertung als widersprüchlich, das Gütekriterium der Objektivität war keineswegs gegeben, wie die vielfach auseinanderklaffenden Einschätzungen von verschiedenen Beurteilenden zeigen. Auch zeigte sich bei offenen Einschätzungen eine extrem linksschiefe Verteilung in einigen Dimensionen, besonders bei Empathie und Offenheit, da sich die Testpersonen, sofern sie sich auf diesen Dimensionen beschrieben, beinahe ausschließlich in sozial erwünschter Sicht darstellten. Diese linksschiefen Verteilungen waren bei der Dichotomisierung der Skalen des *B5PO* berücksichtigt worden, was einen weiteren Vorteil von genormten Verfahren aufzeigt. Die mangelnde Objektivität im offenen Antwortformat spiegelt jedoch auch einen Kritikpunkt dieser Diplomarbeit wieder, da deutlich sichtbar wurde, dass die Beurteilungen stark interpretationsabhängig sind. Ebenso kritisierbar im offenen Antwortformat ist, dass durch das freie Format auf manchen Dimensionen oftmals keine Einschätzungen möglich waren, auch wenn vielleicht bei einem expliziten Nachfragen durch einen Fragebogen die betreffende Testperson einer Extremgruppe zugeordnet worden wäre.

Das Gelingen der Extremgruppenvalidierung kann man als gut bezeichnen: In fünf der sechs Skalen liegen bedeutende Unterschiede zwischen den Extremgruppen vor.

Signifikante Unterschiede, die den zuvor festgelegten Mindestunterschied von 10 *T*-Werten überschreiten, finden sich für die Skalen „Verträglichkeit“ und „Gewissenhaftigkeit“.

Auch die Mittelwertunterschiede für die Skalen „Extraversion“, „Emotionale Kontrolle“ und „Empathie“ sind signifikant; hier erreichen die Effekte zwar nicht ganz den im vorhinein festgelegten Mindestunterschied von 10 *T*-Werten, die festgelegten Effektgrößen liegen jedoch jeweils im 95%-Konfidenzintervall. Besonders bei den Skalen „Extraversion“ und „Empathie“ fallen dabei extrem schiefe Verteilungen auf: So war in der vorliegenden Stichprobe ein Erreichen von *T*-Werten über 50 der Skala „Extraversion“ extrem selten, und das, obwohl der kleinste zu erreichende *T*-Wert 34 beträgt. Vor diesem Hintergrund ist der

signifikante Mittelwertunterschied von 8,9 *T*-Werten zwischen den als extrem extravertiert und extrem introvertiert eingeschätzten Testpersonen beachtlich. Die rechtsschiefe Verteilung lässt vermuten, dass hier bei der Dichotomisierung dieser Skala des *B5PO* die Korrektur der Linksschiefe zu stark vorgenommen wurde (vgl. Kap. 4.2.2).

Auch für die Skala „Empathie“ findet sich erneut eine schiefe Verteilung, wobei hier – ähnlich wie bei der Skala „Extraversion“ – faktisch keine große Bandbreite an Werten möglich ist, da der geringste zu erreichende *T*-Wert 40 beträgt.

Kein signifikanter Unterschied findet sich hingegen zwischen den Extremgruppen der Skala „Offenheit“, für diese Skala kann die Validierung keinesfalls als gelungen bezeichnet werden. Zusammenfassend kann hierzu gesagt werden, dass – abgesehen von der Skala „Offenheit“ – die Validierung gut gelang, jedoch ist eine Neunormierung oder zumindest ein Verschieben des Cut-Points bei der Dichotomisierung anzuraten, besonders für Skalen „Extraversion“ und „Empathie“.

Bezüglich der State-Abhängigkeit zeigt sich zunächst ein kleiner Effekt des Aufsatzes in der Skala „Empathie“, der jedoch nur bei einer von zwei Kontrollgruppen sowie im Vergleich mit der Gesamtkontrollgruppe signifikant ist. In einer näheren Analyse unter Berücksichtigung der Geschlechter zeigt sich, dass Schülerinnen sich signifikant empathischer darstellen als Schüler und zwischen Versuchs- und Kontrollgruppen bei weiblichen Schülerinnen immer ein signifikanter Unterschied vorliegt (in die Richtung gehend, dass Schülerinnen sich nach der Selbstbeschreibung im Aufsatz als empathischer einstufen). Eine reine Rückführung des gesamtgeschlechtlich teilweise signifikanten Unterschiedes auf die Geschlechterhäufigkeit ist auszuschließen, da der weibliche Überhang in den Versuchs- und Kontrollgruppen annähernd gleich war. Des Weiteren ist es gerade der Unterschied zu Kontrollgruppe 2, jener mit dem höchsten weiblichen Anteil, der im gesamtgeschlechtlichen Vergleich nicht signifikant war und der auch den geringsten Effekt im ausschließlich weiblichen Vergleich aufweist. Wahrscheinlicher erscheint daher, dass der gesamtgeschlechtlich zu geringe Effekt in dieser Kontrollgruppe dadurch begründet ist, dass die männlichen Testpersonen der Versuchsgruppe im Vergleich mit Kontrollgruppe 2 eine (nicht signifikante) Tendenz zeigen, sich genau gegensätzlich gepolt State-beeinflusst zu zeigen.

Die gegensätzlich gepolte Tendenz des Aufsatzes bei männlichen Testpersonen ist zu gering um signifikant zu werden, und bei Kontrollgruppe 1 (faktisch) gar nicht gegeben, und da die Fragestellung nicht geschlechtsspezifisch formuliert war, kann der State-Einfluss durch den Aufsatz gesamtgeschlechtlich nicht postuliert werden. Der im gesamtgeschlechtlichen

Vergleich teilweise gegebene signifikante Unterschied begründet sich aber insofern durch das Geschlecht, dass sich Frauen nach der Aufsatzbearbeitung signifikant empathischer darstellen. Dies eröffnet weiteren Forschungsbedarf, besonders in Bezug auf die Frage der (höheren) State-Abhängigkeit von Traits bei weiblichen Testpersonen und die eventuell unterschiedliche Richtung des State-Einflusses bei Männern und Frauen.

Zum Einfluss von Geschlecht sei an dieser Stelle auch erwähnt, dass ein signifikanter Unterschied zwischen männlichen und weiblichen Testpersonen nicht nur auf der Skala „Empathie“, sondern auch „Emotionale Kontrolle“ besteht, was für die Praxis auf jeden Fall beachtenswert erscheint.

Bei der systematischen Überprüfung eines State-Einflusses durch das Suchbild konnte gezeigt werden, dass die vorherige Bearbeitung eines Suchbildrätsels zu einem signifikanten Unterschied in der Darstellung auf der Skala „Gewissenhaftigkeit“ führt. Vergleicht man Personen, die das Suchbild erfolgreich lösten, mit Personen, die keinen Erfolg haben konnten und mit einer Kontrollgruppe, sieht man, dass diejenigen Personen, die Erfolg hatten, sich im Anschluss im Persönlichkeitsverfahren als gewissenhafter darstellen. Dies konnte bei zwei verschiedenen Kontrollgruppen gezeigt werden. Auch bei nicht erfolgreichen Testpersonen nach Rätselvorgabe zeigt sich eine leichte (nur bei einer Teilstichprobe signifikante) Tendenz in die Richtung, sich erfolgreicher darzustellen. Kritisch anzumerken ist dabei, dass die Testpersonen, die trotz möglicher Lösung keinen Erfolg erzielten, aus der Gruppe der erfolgreichen Testpersonen ausgeschlossen wurden: Dies mag gewissermaßen auch eine Leistungskomponente mit einfließen lassen, auch wenn versucht wurde, diese möglichst gering zu halten (indem die betreffenden Testpersonen auch nicht zur Gruppe der Frustrierten gerechnet wurden).

Betrachtet man die erfolgreichen und nicht erfolgreichen Versuchsgruppen gemeinsam und vergleicht sie mit einer Kontrollgruppe, liegt dadurch ein signifikanter Unterschied vor – beide tendieren also dazu, sich nach Bearbeitung einer gewissenhaft durchzuführenden Aufgabe gewissenhafter darzustellen. Durch die verstärkte (und signifikante) gewissenhaftere Darstellung bei Testpersonen, die dabei erfolgreich waren, wird der Unterschied im Gesamtvergleich signifikant. Ebenso fand sich eine Tendenz auf der Skala „Empathie“, sich nach einem Erfolg empathischer darzustellen, diese wurde jedoch nicht signifikant.

Die Hypothese, dass sich Testpersonen nach einem frustrierenden Ergebnis anders darstellen, musste jedoch verworfen werden. Es war das Erfolgserlebnis, nicht die Frustration, die für den signifikanten Effekt auf der Skala „Gewissenhaftigkeit“ verantwortlich war. Dies mag

zunächst überraschen, geht jedoch mit dem im Theorieteil (Kap. 4.1.3) erwähnten Phänomen einher, Erfolge intern zu attribuieren und Misserfolge extern zuzuschreiben. Zwar wurde versucht, eine externe Zuschreibung durch gleichzeitige Testung mit anderen Schüler/innen, die Erfolg hatten, zu verhindern, indem Schüler/innen der Frustrationsgruppe sehen sollten, dass andere die Aufgabe bewältigen konnten, dies scheint jedoch nicht ausreichend gewesen zu sein.

Dennoch liegt durch dieses Ergebnis ein State-Einfluss durch das Rätsel zumindest für die Skala „Gewissenhaftigkeit“ vor. Auch wenn dieser relativ gering ist, ist er doch insofern bedenklich, da in der Praxis durchaus schwerer wiegende Erfolge als das Bearbeiten eines Rätsels auftreten können und durch dieses Ergebnis – zumindest was das Verfahren *B5PO* betrifft – durchaus eine geringe State-Abhängigkeit von Persönlichkeitseigenschaften aufgezeigt werden konnte. Dies lässt darauf schließen, dass das (gewissenhafte) Bearbeiten einer zuvor gestellten Aufgabe möglicherweise einen geringen, zumindest kurzfristigen, Effekt auf dieser Skala in der Selbstdarstellung haben könnte.

Durch den oben erwähnten, bei weiblichen Testpersonen auftretenden Unterschied der Skala „Empathie“ nach Bearbeitung eines Aufsatzes ist zwar eine gesamtgeschlechtliche Aussage nicht zulässig, kombiniert mit dem Ergebnis des Rätsel-Einflusses lässt dies jedoch die Annahme zu, dass möglicherweise – je nachdem, was zuvor erlebt wurde und welche Dimension der Persönlichkeit das Erlebte am ehesten betrifft – die betreffende Dimension verzerrt sein könnte. Weiterer Forschungsbedarf, auch in Bezug auf andere Fragebogenverfahren, sei daher aufgezeigt.

Das Ergebnis der Diplomarbeit einschränkend zu erwähnen bleibt, dass ausschließlich Testpersonen einer relativ schmalen Altersklasse (AHS-Schüler/innen der 10. Schulstufe und darüber) getestet wurden, für einen altersübergreifenden Bereich besteht damit ebenfalls noch Forschungsbedarf.

15. Zusammenfassung

Die vorliegende Arbeit diente der Validierung des Persönlichkeitsverfahrens *Big Five Plus One* (Holocher-Ertl, Kubinger & Menghin, 2003). Diese wurde auf verschiedene Arten vorgenommen; zunächst wurde in einer Kombination verschiedenster Validierungsmethoden eine Extremgruppenvalidierung vorgenommen, wobei hier neben nicht-korrelativen Ansätzen auch Grundideen der MTMM-Matrix von Campbell und Fiske einfließen (nämlich die heteromethodische Erfassung von Traits), indem die Extremgruppen mit einer gänzlich anderen Methode in offenem Antwortformat gebildet wurden. Die zugrundeliegenden Hypothesen dieser Arbeit waren, dass die mittels offener Selbstbeschreibung gebildeten und je Dimension entgegengesetzten Extremgruppen sich hinsichtlich ihrer Testwerte im Persönlichkeitsinventar in der jeweils interessierenden Dimension um mindestens 10 *T*-Werte unterscheiden. Für die Extremgruppenbildung wurde zunächst versucht, die offenen Selbstbeschreibungen in Anlehnung an die qualitative Inhaltsanalyse und in ebenso starker Anlehnung an die Theorie des lexikalischen Ansatzes der Big Five zu bilden (vgl. z.B. De Raad, 2000). Dazu wurde eine eigene Eigenschaftswörterliste mit Synonymen des *Big Five Plus One* erstellt. Da sämtliche qualitativen Auswertungsversuche an mangelnder Objektivität scheiterten, wurden die Extremgruppen später ausschließlich mittels Experten-Ratings gebildet. Insgesamt standen für die Beantwortung dieser Fragestellung 175 AHS-Schüler/innen zur Verfügung, davon waren 62 in der aufsatzschreibenden Versuchsbedingung. Mit Hilfe der Kontrollgruppe wurde ein möglicher Effekt des Aufsatzes untersucht, der sich bei weiblichen Testpersonen als signifikant erwies.

Die Extremgruppenvalidierung kann für fünf der sechs Skalen als gelungen betrachtet werden; nicht gelungen ist sie für die Skala „Offenheit“.

In einem zweiten, mehr am Konstrukt der Traits orientierten Teil der Validierung wurde der Frage nachgegangen, inwiefern durch das Persönlichkeitsverfahren tatsächlich Traits gemessen werden und diese nicht durch States überlagert sind. Da States und Traits durchaus in einem Zusammenhang stehen (vgl. z.B. Amelang et al., 2006) und auch Borkenau und Ostendorf (1998) durchaus zeigen konnten, dass die Big Five auch Zustände über die Zeit beschreiben, stellt sich die Frage, inwieweit auch im interessierenden Persönlichkeitsverfahren eine Beeinflussung durch States sichtbar wird. Eine der weiteren grundlegenden Theorien war hierfür auch die Studie von Schwarz und Clore (1988), in der sie den Einfluss des Findens einer Münze auf einem Kopiergerät auf die Lebenszufriedenheit zeigten. Ausgehend von der Überlegung, dass offensichtlich schon kleine Ursachen einen

Einfluss auf Traits haben können, wurde in der vorliegenden Arbeit mithilfe eines Suchbild-Rätsels versucht, Testpersonen zu einem Erfolgserlebnis zu verhelfen oder sie zu frustrieren, um zu sehen, ob sie sich im Anschluss im Persönlichkeitsverfahren anders darstellen. Hierfür standen insgesamt 182 AHS-Schüler/innen zur Verfügung, davon 54 in der potentiell erfolgreichen Versuchsgruppe (davon wurden 18 ausgeschlossen, die nicht erfolgreich waren), 53 in der frustrierten Versuchsgruppe und 75 in der gesamten Kontrollgruppe. Die Kontrollgruppe diente der Überprüfung, ob das Rätsel selbst einen Einfluss hatte und konnte in zwei sich teilweise überschneidende Teilkontrollgruppen gesplittet werden.

Es zeigte sich ein signifikanter Einfluss des Rätsels auf die Skala „Gewissenhaftigkeit“, indem sich Testpersonen nach der Rätselbearbeitung als gewissenhafter darstellten als die gesamte Kontrollgruppe, wodurch ein State-Einfluss aufgezeigt werden konnte. Der Unterschied war durch die im Rätsel erfolgreichen Testpersonen bedingt; frustrierte Testpersonen stellten sich zwar tendenziell ebenfalls gewissenhafter dar, der Unterschied war jedoch nicht signifikant. Die Hypothese, dass Frustration einen Einfluss auf die Selbstdarstellung habe, musste also verworfen werden. Es waren im Gegenteil die Erfolgreichen, welche für den signifikanten Unterschied verantwortlich waren. Die Begründung mag in dem Phänomen liegen, Erfolge intern zu attribuieren und Misserfolge extern zuzuschreiben (vgl. Herkner, 2001).

IV Literatur

- Amelang, M. (2003). Differentielle Psychologie. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 96–102). Weinheim: Beltz PVU.
- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6., vollständig überarbeitete Aufl.). Stuttgart: Kohlhammer.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4., vollständig überarbeitete und erweiterte Auflage). Heidelberg: Springer.
- Asendorpf, J.B. (2007). *Psychologie der Persönlichkeit* (4., überarbeitete und aktualisierte Auflage). Heidelberg: Springer.
- Block, J. (1995). A Contrarian View of the Five-Factor Approach to Personality Description. *Psychological Bulletin*, 117, 185–215.
- Borkenau, P. & Ostendorf, F. (1998). The Big Five as States: How Useful Is the Five-Factor Model to Describe Intraindividual Variations over Time? *Journal of Research in Personality*, 32, 202–221.
- Bortz, J. & Döring, N. (2005). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3., überarbeitete Auflage, Nachdruck). Berlin: Springer.
- De Raad, B. (2000). *The Big Five Personality Factors. The Psycholexical Approach to Personality*. Seattle: Hogrefe & Huber Publishers.
- Häcker, H.O. & Stapf, K.-H. (2004). *Dorsch. Psychologisches Wörterbuch* (14., überarbeitete und erweiterte Auflage). Bern: Hans Huber.
- Herkner, W. (2001). *Lehrbuch Sozialpsychologie* (2., unveränderte Auflage). Bern: Hans Huber.

- Holocher-Ertl, S. (2003). *Das Big Five + One Persönlichkeitsinventar. Eine neue Adjektivliste zur Erfassung der Big Five Persönlichkeitsdimensionen sowie der Dimension Gefühlsbetontheit*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Holocher-Ertl, S., Kubinger, K.D. & Menghin, S. (2003). *Big Five Plus One Persönlichkeitsinventar (B5PO)*. Test: Software und Manual. Mödling: Dr. G. Schuhfried GmbH.
- Holz-Ebeling, F. & Steinmetz, M. (1995). Wie brauchbar sind die vorliegenden Fragebogen zur Messung von Empathie? Kritische Analysen unter Berücksichtigung der Iteminhalte. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 16, 11–32.
- Kubinger, K.D. (2003a). Gütekriterien. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 195–204). Weinheim: Beltz PVU.
- Kubinger, K.D. (2003b). (Un-)Verfälschbarkeit. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 429–432). Weinheim: Beltz PVU.
- Kubinger, K.D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K.D., Hofmann, K. & Litzenberger, M. (2002). Die Bedeutung einer multimethodischen Persönlichkeitsdiagnostik zur Wahrung der „Dimensionalität“ des Menschen. *Psychologie in Österreich*, 22, 60–66.
- Kubinger, K.D. & Rasch, D. (2006). *Statistik für das Psychologiestudium*. München: Elsevier.
- Laux, L., Glanzmann, P., Schaffner, P. & Spielberger, C.D. (1981). *State-Trait-Angstinventar (STAI)*. Test: Manual. Weinheim: Beltz.
- Lienert, G.A. & Ratz, U. (1994). *Testaufbau und Testanalyse* (5. völlig neubearbeitete und erweiterte Auflage). Weinheim: Beltz.

- Lissmann, U. (2001). *Inhaltsanalyse von Texten – ein Lehrbuch zur computerunterstützten und konventionellen Inhaltsanalyse* (2. aktualisierte und erweiterte Auflage). (Forschung, Statistik und Methoden, Bd. 2). Landau: Verlag Empirische Pädagogik.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (8. Auflage). Weinheim: Beltz.
- Mayring, P. (2005). Neuere Entwicklungen in der qualitativen Forschung und der Qualitativen Inhaltsanalyse. In P. Mayring & M. Gläser-Zikuda (Hrsg.). *Die Praxis der Qualitativen Inhaltsanalyse* (S. 7–19). Weinheim: Beltz.
- Pervin, L.A., Cervone, D. & John, O.P. (2005). *Persönlichkeitstheorien* (5., vollständig überarbeitete und erweiterte Auflage). München: Ernst Reinhardt.
- Reinhoffer, B. (2005). Lehrkräfte geben Auskunft über ihren Unterricht. In P. Mayring & M. Gläser-Zikuda (Hrsg.). *Die Praxis der Qualitativen Inhaltsanalyse* (S. 123–141). Weinheim: Beltz.
- Schmitt, M. (2003). *Trait*. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 424–426). Weinheim: Beltz PVU.
- Schuerger, J., Zarrella, K. & Hotz, A. (1989). Factors That Influence the Temporal Stability of Personality by Questionnaire. *Journal of Personality and Social Psychology*, 56, 777–783.
- Schwarz, N. & Clore, G.L. (1983). Mood, Misattribution, and Judgements of Well-Being: Informative and Directive Functions of Affective States. *Journal of Personality and Social Psychology*, 45, 513–523.
- Schwarz, N. & Clore, G.L. (1988). *How Do I Feel About It? The Informative Function of Affective States*. In K. Fiedler & J. Forgas (Hrsg.), *Affect, Cognition and Social Behavior. New Evidence and Integrative Attempts* (S. 44–62). Toronto: Hogrefe.

- Schwenkmezger, P., Hodapp, V. & Spielberger, C.D. (1992). *Das State-Trait-Ärgerausdrucksinventar STAXI*. Test: Manual. Bern: Hans Huber.
- Tversky, A. & Kahnemann, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5, 207–232.
- Wallbott, H.G. (2000). *Empathie*. In J.H. Otto, H.A. Euler & H. Mandl (Hrsg.), *Emotionspsychologie. Ein Handbuch* (S. 370–380). Weinheim: Beltz PVU.
- Zimbardo, P.G. & Gerrig, R.J. (2004). *Psychologie* (16., aktualisierte Auflage). München: Pearson Studium.

V Tabellenverzeichnis

Tab. 1: Vorgabe der Materialien (Überblick).....	56
Tab. 2: Geschlechterverteilung der Extremgruppenvalidierung (Prozentwerte gerundet).....	59
Tab. 3: Chi-Quadrat-Test auf Gleichverteilung der Geschlechter über die gesamte Stichprobe	59
Tab. 4: Chi-Quadrat-Test auf Gleichverteilung der Geschlechter über die gesamte KG.....	59
Tab. 5: Geschlechterverteilung der Stichprobe zum State-Einfluss (Prozentwerte gerundet).	62
Tab. 6: Erster Probedurchlauf der qualitativen Beurteilungsübereinstimmung.....	65
Tab. 7: Zweiter Probedurchlauf der Beurteilungsübereinstimmung.....	67
Tab. 8: Übereinstimmung mit und ohne Interpretationen (dritter Durchlauf)	68
Tab. 9: Übereinstimmungen in der Globalbeurteilung von zwei Beurteilerinnen.....	70
Tab. 10: Übereinstimmungen in der Rangreihung.....	71
Tab. 11: Schlüsselwortkodierung über fünf Aufsätze	73
Tab. 12: <i>t</i> -Test des Extremgruppenvergleichs für die Skala „Extraversion“	82
Tab. 13: Verteilung der <i>T</i> -Werte der Skala „Extraversion“	83
Tab. 14: <i>t</i> -Test des Extremgruppenvergleichs für die Skala „Verträglichkeit“	84
Tab. 15: <i>t</i> -Test des Extremgruppenvergleichs für die Skala „Gewissenhaftigkeit“	84
Tab. 16: <i>t</i> -Test des Extremgruppenvergleichs für die Skala „Emotionale Kontrolle“	85
Tab. 17: <i>t</i> -Test des Extremgruppenvergleichs für die Skala „Offenheit“	85
Tab. 18: <i>t</i> -Test des Extremgruppenvergleichs für die Skala „Empathie“	86
Tab. 19: Verteilung der <i>T</i> -Werte der Skala „Empathie“	87
Tab. 20: <i>t</i> -Test des Vergleichs zwischen Versuchsgruppe und Kontrollgruppe 1	88
Tab. 21: <i>t</i> -Test des Vergleichs zwischen Versuchsgruppe und Kontrollgruppe 2	88
Tab. 22: <i>t</i> -Test des Vergleichs zwischen Versuchsgruppe und der gesamten Kontrollgruppe	89
Tab. 23: Vergleich der Kontrollgruppen für die Extremgruppenbildung.....	91
Tab. 24: Vergleich von männlichen und weiblichen Testpersonen auf der Skala „Empathie“	92
Tab. 25: Vergleich weiblicher VG und weiblicher KG 2 auf der Skala „Empathie“	92

Tab. 26: Test auf Homogenität der Varianzen zwischen den Rätsel-Bedingungen	94
Tab. 27: Einfache Varianzanalyse über die Rätsel-Bedingungen.....	95
Tab. 28: <i>t</i> -Test der Skala „Gewissenhaftigkeit“ zwischen erfolgreicher VG und gesamter KG	96
Tab. 29: <i>t</i> -Test über die 6 Skalen des <i>B5PO</i> zwischen frustrierter und erfolgreicher VG	97
Tab. 30: Vergleich von männlichen und weiblichen Testpersonen der gesamten Stichprobe für die Suchbild-Testung	98
Tab. 31: <i>t</i> -Test zwischen den Gruppen mit und ohne Rätselvorgabe	99

VI Abbildungsverzeichnis

Abb. 1: Item 1 des Persönlichkeitsverfahrens <i>B5PO</i>	54
Abb. 2: Überblick über die Stichproben zur Extremgruppenvalidierung.....	60
Abb. 3: Überblick über die Stichproben zur systematischen Überprüfung eines möglichen State-Einflusses	62
Abb. 4: Interpolieren der Skala „Extraversion“	79
Abb. 5: Interpolieren der Skala „Verträglichkeit“	80
Abb. 6: Interpolieren der Skala „Gewissenhaftigkeit“	80
Abb. 7: Interpolieren der Skala „Emotionale Kontrolle“	81
Abb. 8: Boxplots der beiden unterschiedlichen Kontrollgruppen für die Aufsatzbedingung..	90
Abb. 9: Boxplots der verschiedenen Rätselbedingungen	96

VII Anhang

Anhang A: Materialien und offizielle Schreiben

- Anleitung für das Suchbild
- Suchbild mit 6 Fehlern
- Suchbild mit 10 Fehlern
- Aufsatzbogen mit Anleitung
- Kurzbeschreibung für den Landes- bzw. Stadtschulrat
- Einverständniserklärung

Code: _____

Anleitung

Auf der folgenden Seite finden Sie zwei Zeichnungen, die auf den ersten Blick gleich aussehen. Im unteren Bild verstecken sich jedoch einige Unterschiede.

Ihre Aufgabe ist es nun, die acht Unterschiede in der unteren Zeichnung zu finden und zu markieren. Sie haben dafür insgesamt acht Minuten Zeit.

Die Auswertung erfolgt anonym, für die Zuordnung und den Vergleich mit anderen Verfahren ist es jedoch erforderlich, dass Sie den Ihnen zugeteilten Code angeben.

Bitte tragen Sie Ihren Code gleich in das rechte obere Eck dieser Seite ein und blättern Sie erst um, wenn das Zeichen dazu gegeben wird.

Sobald Sie die acht Unterschiede gefunden und markiert haben, geben Sie bitte Ihren Bogen ab.

Vielen Dank für die Mitarbeit!





Anleitung

Im Folgenden soll Ihre Persönlichkeit erfasst werden.

Bitte schreiben Sie einen kurzen Aufsatz von 1-2 Seiten mit dem Titel „Wer bin ich – wie bin ich?“.

In diesem Aufsatz sollen Sie versuchen, Ihre Persönlichkeit unter Verwendung von Eigenschaftswörtern möglichst genau und umfassend zu beschreiben (auch negative Eigenschaften).

Denken Sie daran, dass Sie Ihre Persönlichkeitseigenschaften beschreiben sollen, nicht Ihre Tätigkeit, Ihr Aussehen oder Ihre momentane Stimmung.

Die Auswertung erfolgt anonym, für die Zuordnung und den Vergleich mit anderen Verfahren ist es jedoch erforderlich, dass Sie den Ihnen zugeteilten Code angeben.

Vielen Dank für die Mitarbeit!

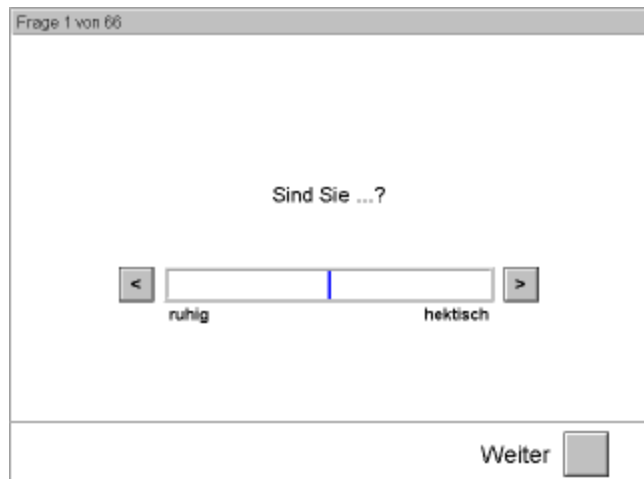
Code: _____

Wer bin ich – wie bin ich?

Kurzbeschreibung zur Validierung des B5PO (Weitensfelder Lisbeth)

Ziel dieses Teils der Untersuchung ist die Validierung des Persönlichkeitsinventars „Big Five Plus One“ (kurz B5PO); das heißt, die Überprüfung, ob dieses Verfahren tatsächlich misst, was es zu messen beansprucht.

In diesem B5PO sollen sich die SchülerInnen an insgesamt 60¹⁶ Eigenschaftspaaren (z.B. ruhig – hektisch) mithilfe eines Balkens selber einschätzen (Abbildung siehe rechts). Daraus resultierend erhält man die fünf Globalskalen der Persönlichkeit (Extraversion, Verträglichkeit, Gewissenhaftigkeit, Emotionale Kontrolle und Offenheit) sowie die zusätzliche Dimension Empathie.



Die Idee ist nun einerseits, den B5PO daran zu validieren, indem man einige SchülerInnen vor Vorgabe dieses Verfahrens einen kurzen Aufsatz schreiben lässt, in dem sie gebeten werden, ihre Persönlichkeit möglichst genau und umfassend zu beschreiben. Dieser Aufsatz soll im Anschluss an die Datenerhebung mit den Ergebnissen des B5PO verglichen werden. (Selbstverständlich erfolgt die Auswertung vertraulich und über Codes anonymisiert. Eine Vergabe von Codes ist jedoch erforderlich, da die Aufsätze den Ergebnissen im B5PO sonst nicht zugeordnet werden können.)

Weiters stellt sich aber auch die Frage, wie sehr die Selbstbeschreibung in diesem Persönlichkeitsverfahren von momentanen Stimmungslagen beeinflusst sein könnte. Um dies zu kontrollieren, würden einigen SchülerInnen vor Ausfüllen des Tests Rätsel vorgegeben werden. Die Hälfte dieser SchülerInnen würde leichte Rätsel bekommen, die andere Hälfte jedoch sehr schwere bzw. sogar unlösbare Rätsel. Dadurch soll bei diesen beiden Gruppen eine unterschiedliche Stimmung erzeugt werden, um im Anschluss zu prüfen, ob sich dies auf die Ergebnisse im B5PO auswirkt.

Selbstverständlich würden die SchülerInnen im Anschluss an die Testung über die Tatsache aufgeklärt, dass einige der Rätsel nicht lösbar waren, und auch der Grund für dieses Vorgehen würde ihnen erklärt werden.

Der Vorteil für SchülerInnen, bei einer derartigen Testung mitzumachen, liegt einerseits sicher darin, selber bereits einmal einen Persönlichkeits-„Test“ gemacht zu haben.

Andererseits liegt ein weiterer Vorteil aber auch darin, dass sie Psychologie einmal „erleben“ können und einen kleinen Einblick darüber gewinnen, wie über Experimente in den Sozial- und Humanwissenschaften Erkenntnisse gewonnen werden können.

¹⁶ [Anm.: Die tatsächliche Anzahl der Items betrug 66, da in der ursprünglich an Landes- und Stadtschulrat versandten Information jedoch ein Tippfehler vorlag, wurde er auch hier unkorrigiert übernommen.]

Liebe Eltern, liebe Mütter, liebe Väter!

Unsere Namen sind Andreas Schneiderbauer und Lisbeth Weitensfelder und wir studieren Psychologie an der Universität Wien.

Im Rahmen unserer Diplomarbeit führen wir eine Untersuchung durch, deren Ziele folgende sind:

- Messung des Einflusses von Persönlichkeit auf die Intelligenztestleistung,
- Überprüfung, ob das Persönlichkeitsverfahren „B5PO“ tatsächlich misst, was es zu messen beansprucht (=Validierung).

Bei den eingesetzten psychologischen Verfahren handelt es sich zum einen um ein Instrument zur Erfassung verschiedener Intelligenzleistungen und zum anderen um einen Fragebogen zur Erhebung bestimmter Persönlichkeitseigenschaften. Zusätzlich werden einige Schüler gebeten, einen kurzen Text zu verfassen oder eine Rätselaufgabe zu lösen.

Die eingesetzten Verfahren können im Rahmen einer Gruppenvorgabe erfolgen, wobei die benötigte Zeit zur Bearbeitung bei ca. 1 bis 2 Schulstunden liegen wird.

Alle Angaben sowie die Ergebnisse aus den psychologischen Verfahren werden natürlich vertraulich behandelt und die weitere Behandlung der Daten erfolgt vollständig anonymisiert. Die erhobenen Ergebnisse werden auch nicht weitergegeben und dienen ausschließlich dieser Untersuchung. Im Anschluss an die Untersuchung besteht für die SchülerInnen die Möglichkeit, eine kurze Rückmeldung zu ihren Ergebnissen zu erhalten.

Bei Einverständnis zur Teilnahme Ihres Kindes an der Untersuchung bitten wir Sie, diese Mitteilung zu unterschreiben und Ihrem Kind in die Schule mitzugeben.

Vielen herzlichen Dank für Ihre Unterstützung!

Mit freundlichen Grüßen,

Andreas Schneiderbauer

Lisbeth Weitensfelder

Ich bin mit der Teilnahme meines Kindes an der Untersuchung einverstanden.

Name meines Kindes: _____

Geburtsdatum: _____ Klasse: _____

Datum und Unterschrift des Erziehungsberechtigten

Anhang B: Auswertungshilfen

- Auswertungs- bzw. Kodierrichtlinien für die qualitative Auswertung
- Auswertungsbogen für die qualitative Auswertung (Grundversion)
- Beispielseite der Synonymliste
- Auswertungsbogen für die Globalbeurteilung inkl. Anleitung für Globalbeurteilung und Rangreihung
- Rangreihungsbogen
- Richtlinien für die Schlüsselwörtercodierung
- Auswertebogen für die Schlüsselwörtercodierung
- Auswertungsbogen inkl. Anleitung für die Extremgruppenzuordnung

Auswertungsrichtlinien für den Aufsatz

Als Nennungen für die jeweilige Skala zählen:

- Alle Eigenschaftswörter, die im B5PO genannt werden (jeweils als +/-)
- Alle Synonyme für diese Eigenschaftswörter
- Alle Umschreibungen der im B5PO genannten Eigenschaftswörter (= „Interpretation“)
- Für die Ladung auf (+) (z.B. Extraversion) zählen auch alle Verneinungen von Introversion (-); Umgekehrtes gilt für die Ladungen auf (-)

Vergabe von Punkten:

- 1 Punkt wird vergeben für alle Nennungen eines Eigenschaftswortes sowie für Verstärkungen dieses Eigenschaftswortes (z.B. „sehr“)
- 0,5 Punkte werden vergeben für alle Relativierungen (z.B. „etwas“, „eher“, „ein wenig“, etc.)
- 0 Punkte werden vergeben für neutrale Nennungen auf der Skala; z.B. „weder ... noch ...“, sofern die genannten Begriffe wirklich ein Gegensatzpaar bilden!! (Wenn geschrieben steht „ich bin weder ängstlich noch launisch“ sind dies zwei „+“-Ladungen für Emotionale Kontrolle!!)
- Umgang mit Widersprüchlichkeiten: Ein Beispiel: Wenn sich jemand beschreibt als „Ich bin kontaktfreudig, gehe aber selten auf Menschen zu“, wird dies einmal als (+), einmal als (-) auf der Extraversionsskala verrechnet. *Grund:* Beim B5PO hätte diese Person vermutlich auch „kontaktfreudig“ angegeben, daher zählt dies als (+). Der zweite Satzteil (Interpretation) muss jedoch inhaltlich eindeutig als (-) zählen. Insgesamt hebt sich diese Aussage damit auf, wird aufgrund der sozialen Erwünschtheit in der späteren Berechnung jedoch tendenziell als (-) verrechnet (siehe Auswertungsbogen). Diese Aussage nur vollkommen als (-) zu verrechnen, wäre jedoch auch unzuverlässig, da sich die Person trotzdem gewissermaßen als kontaktfreudig sieht (vermutlich als „trete gerne in Kontakt, wenn andere auf mich zukommen, gehe aber selbst nicht auf die Leute zu“).

Vorgangsweise:

- Es gibt eine Tabelle, in der sämtliche Eigenschaftswörter (inkl. „Polung“) der jeweiligen Skalen angegeben sind. In der Tabelle ist auch ersichtlich, ob es sich um ein „Originalwort“ aus dem B5PO oder um ein Synonym handelt.

Zwecks Nachvollziehbarkeit und Kontrollmöglichkeit soll folgendermaßen vorgegangen werden:

- Alle Eigenschaftswörter (oder deren Unterstreichungen) farbig unterstreichen und gleichzeitig in den Auswertungsbogen eintragen.
- Im B5PO genannte Adjektive, verwendete Synonyme und Interpretationen (von Umschreibungen) sind jeweils in einer anderen Farbe zu unterstreichen. Ebenso irrelevante Begriffe.

Farbvorschlag:

- o Gelb: Im B5PO vorkommende Adjektive
 - o Orange: Synonyme
 - o Rot: Interpretationen
 - o Grün: nicht vorkommende Skalen
- Mit Bleistift soll mit einem Kürzel bei der Unterstreichung gekennzeichnet werden, zu welcher Skala dies verrechnet wird (v.a. für spätere Kontrolle von Interpretationen wichtig). Der Gegenpol ist jeweils mit einem kleinen Minus zu versehen.
 - o EX: Extraversion (EX⁻: Introversion)
 - o V: Verträglichkeit
 - o G: Gewissenhaftigkeit
 - o EK: Emotionale Kontrolle
 - o O: Offenheit
 - o EMP: Empathie

Testpersonen-Code: _____

Skala	Anzahl Nennungen (+)	Anzahl rel. Nennungen (+)	Anz. neutrale Nennungen	Anzahl Nennungen (-)	Anzahl rel. Nennungen (-)	Anzahl ges. Nennungen der Skala	Wert*	Gew. Wert**	Wert SE SE	Gew. Wert SE SE*
Extraversion(+) vs. Introversion(-)										
Verträglichkeit(+) vs. Unverträglichkeit (-)										
Gewissenhaftigkeit(+)										
Em. Kontrolle(+) ...										
Offenheit(+) ...										
Empathie(+) ...										
Nennung zusätzlicher Eigenschaften										

*) Berechnung des Wertes: $\text{Anzahl Nennungen}(+) + \text{Anzahl rel. Nennungen}(+) * 0,5 - \text{Anzahl Nennungen}(-) + \text{Anzahl rel. Nennungen}(-) * 0,5$

**) Berechnung des gewichteten Werts = $\text{Wert} / \text{Anzahl gesamt Nennungen der Skala}$

SE) Berechnung des sozial erwünschten Werts SE: $\text{Anzahl Nennungen}(+) + \text{Anzahl rel. Nennungen}(+) * 0,5 - [\text{Anzahl Nennungen}(-) + \text{Anzahl rel. Nennungen}(-) * 0,5] * 1,5$

SE*) Berechnung des gewichteten Werts SE = $\text{Wert SE} / \text{Anzahl gesamt Nennungen der Skala}$

	A	B	C	D	E	F	G
	Eigenschaftswort	lädt auf Skala	(+) / (-)	Test/Synonym	Synonym für:	Gegenpol zu	and. Bed.
1							
274	begeistert	Empathie	+	Synonym	leidenschaftlich	emotionslos	y
275	behäbig	Offenheit	-	Synonym	langweilig	unternehmenslustig	
277	beharrlich	Extraversion	-	Synonym	verbissen	locker	
278	beharrlich	Verträglichkeit	-	Synonym	stur	nachgiebig	
279	beharrlich	Gewissenhaftigkeit	+	Synonym	zielstrebig	planlos	
280	beharrlich	Gewissenhaftigkeit	+	Synonym	konsequent	inkonsequent	
281	beharrlich	Gewissenhaftigkeit	+	Synonym	fleißig	faul	
282	beharrsam	Verträglichkeit	-	Synonym	stur	nachgiebig	
283	beharrsam	Gewissenhaftigkeit	+	Synonym	zielstrebig	planlos	
284	beharrsam	Gewissenhaftigkeit	+	Synonym	konsequent	inkonsequent	
285	beherrschend	Extraversion	+	Synonym	dominant	unterwürfig	
286	beherrscht	Extraversion	+	Synonym	ausgeglichen	launisch	
287	beherrscht	Gewissenhaftigkeit	+	Synonym	diszipliniert	undiszipliniert	
288	beherrscht	Emotionale Kontro	+	Synonym	ruhig	hektisch	
289	beherrscht	Emotionale Kontro	+	Synonym	gelassen	aufbrausend	
290	beherrscht	Emotionale Kontro	+	B5PO		unbeherrscht	
291	beherzt	Extraversion	+	Synonym	mutig	ängstlich	
292	behutsam	Gewissenhaftigkeit	+	Synonym	sorgfältig	schlampig	
293	behutsam	Gewissenhaftigkeit	+	Synonym	vorsichtig	unvorsichtig	y
294	behutsam	Empathie	+	Synonym	gefühlvoll	gefühllos	y
296	beiläufig	Gewissenhaftigkeit	-	Synonym	nachlässig	pflichtbewusst	
297	beizeiten	Gewissenhaftigkeit	+	Synonym	pünktlich	unpünktlich	
299	bemüht	Extraversion	+	Synonym	aktiv	passiv	
300	bemüht	Gewissenhaftigkeit	+	Synonym	fleißig	faul	
	Gesamt alphabetisch	B5PO alphabetisch	Synonyme alphabetisch	nach Eingabe			

Auszug aus der Synonymliste (Excel-Datei)

[Anm.: Spalte D zeigt, ob das Wort im Verfahren selbst vorkommt oder ein Synonym zu einem B5PO-Wort ist; in Spalte G ist gekennzeichnet, ob das Synonym eine etwas andere Unterbedeutung aufweist.]

Aufsatzauswertungen

Testpersonen-Code: _____
Beurteiler/in: _____

Schritt 1: Globalbeurteilung

Zunächst soll die Selbstbeschreibung der Testpersonen jeweils einmal überblicksartig durchgelesen werden. Im Anschluss soll der Aufsatz in Hinblick auf jede Skala extra durchgearbeitet werden: Auf einer Skala von 1-10 ist anzukreuzen, wie ausgeprägt der jeweilige Trait bei der vorliegenden Testperson ist (bei einer Maximalausprägung von 10). Einzig, wenn sich keinerlei Hinweis auf die jeweilige Skala findet, darf „0“ angekreuzt werden. Mittlere Ausprägungen (4-7) dürfen maximal in der Hälfte der vorliegenden Fälle vergeben werden. Rund ein Viertel soll sich auf die Ausprägungen 1-3 verteilen, ein weiteres Viertel auf die Ausprägungen 8-10.

Skala ↓	Ausprägung →	1	2	3	4	5	6	7	8	9	10	0
Extraversion(10) vs. Introversion (1)		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Verträglichkeit (10) vs. Unverträglichkeit (1)		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gewissenhaftigkeit (10)		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotionale Kontrolle (10)		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Offenheit (10)		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Empathie (10)		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Schritt 2: Rangreihung der vorliegenden Selbstbeschreibungen

Nach Globalbeurteilung aller vorliegenden Selbstbeschreibungen ist eine Rangreihung aller vorliegenden Testpersonen vorzunehmen. Hierfür gibt es einen eigenen Bogen. Die auf der jeweiligen Trait-Skala am stärksten ladende Testperson bekommt Rangplatz 1 auf dieser Skala, die nachfolgende Person Rangplatz 2 usw.

Rangplatz „0“ darf einzig dann vergeben werden, wenn sich keinerlei Hinweis für die Skala bei der Testperson finden sollte.

Rangreihungs-Bogen

Beurteiler/in: _____

Rangreihung der vorliegenden Selbstbeschreibungen:

Testpersonen-Code↓	Skala →	Extraversion	Verträglichkeit	Gewissenhaftigkeit	Emotionale Kontrolle	Offenheit	Empathie
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						
	Rangplatz:						

Richtlinien zur Auswertung mittels Schlüsselwörtern (Version2)¹⁷:

Grundvoraussetzungen für die Codierung:

- Codiert werden dürfen ausschließlich aktuelle Beschreibungen der eigenen Person oder auch bestimmter Ansichten, wenn man dadurch aus dem Zusammenhang einen Hinweis auf die Persönlichkeit des Schreibers bekommt (z.B. „Ich habe keine Angst vor der Zukunft, ich glaube, dass es später am Wichtigsten ist, einfach nur glücklich zu sein“ – hier darf – neben „nicht ängstlich – auch „glücklich“ ausnahmsweise codiert werden, da es zwar eine allgemeine Aussage ist, im Zusammenhang betrachtet jedoch auch eine Beschreibung der eigenen Persönlichkeit)..
- Beschreibungen der Persönlichkeit in der Vergangenheit, Beschreibungen anderer Personen, Dinge,... dürfen nicht codiert werden.

Codiert werden dürfen EINZIG Wörter, die in der Excel-Liste vorkommen und eindeutig zuordenbar sind. Wörter, die codiert werden dürfen, sind ausschließlich

- Eigenschaftswörter,
- Hauptwörter, wenn sie sich eindeutig in ein Eigenschaftswort umwandeln lassen (z.B. „Was mich auszeichnet, ist Freundlichkeit“ → „freundlich“.),
- Verben, wenn sie sich eindeutig und wörtlich in Eigenschaftswörter umwandeln lassen (z.B. „Ich gebe ungern nach“ → unnachgiebig)

Eindeutige Zuordnung ist der Fall, wenn:

- das Eigenschaftswort ein im B5PO verwendetes Wort ist (Hinweis in Excel-Spalte „Test/Synonym“); auch dann, wenn das Synonym möglicherweise auf mehreren Skalen lädt (z.B. „hilfsbereit“)
- wenn es sich um ein Synonym handelt, das nur ZU EINER EINZIGEN SKALA ein Synonym ist (z.B. „arrogant“).
- Synonyme, die auf mehreren Skalen laden, dürfen nicht codiert werden (z.B. „abgeklärt“).

Codierung als +/-:

- Verneinungen des Positiv-Pols werden als Negativ-Pol codiert (allerdings nur, wenn es eine Verneinung exakt des Positiv-Pols ist, z.B. durch „nicht“ oder „un-“; so darf „nicht interessiert“ als Negativ zu „interessiert“ gewertet werden, keinesfalls jedoch „nicht interessant“!)
- Entsprechend werden Verneinungen des Negativ-Pols als Positiv-Pol codiert.

Neutrale Nennungen dürfen nur in zwei Fällen vergeben werden:

- wenn es sich um ein eindeutiges „weder-noch“ von Gegensatzpolen handelt (z.B. weder freundlich noch unfreundlich)
- wenn es sich um einen eindeutigen Widerspruch handelt, in dem das vorher genannte zu codierende Wort gleich im Anschluss (also nicht irgendwann im Aufsatz) inhaltlich aufgehoben wird (z.B. „Ich bin einerseits kontaktfreudig, gehe aber selten oder nie auf Leute zu.“).

Doppelte Nennungen werden doppelt codiert, außer wenn sie SOFORT im Anschluss wiederholt oder näher erläutert wird (z.B. „ich kann stur sein, wirklich stur“ – wird nur einmal gewertet).

¹⁷ Anm.: Die zweite Version der Auswertungsrichtlinien wurde noch um einige Punkte erweitert, da während des Beurteilungsvorganges teilweise noch Unklarheiten auftraten. Die erste, kürzere Version findet sich am Auswertungsbogen selbst.

Auswertungsbogen-Schlüsselwörtercodierung

Beurteiler/in: _____ Testpersonen-Code: _____

Anmerkung zur Auswertung mittels Schlüsselwörtern: Codiert werden dürfen EINZIG Wörter, die in der Excel-Liste vorkommen und eindeutig zuordenbar sind. Wörter, die codiert werden dürfen, sind ausschließlich

- Eigenschaftswörter,
- Hauptwörter, wenn sie sich eindeutig in ein Eigenschaftswort umwandeln lassen (z.B. „Was mich auszeichnet, ist Freundlichkeit“ → „freundlich“),
- Verben, wenn sie sich eindeutig und wörtlich in Eigenschaftswörter umwandeln lassen (z.B. „Ich gebe ungern nach“ → „unnachgiebig“)

Eindeutige Zuordnung ist der Fall, wenn:

- das Eigenschaftswort ein im B5PO verwendetes Wort ist (Hinweis in Excel-Spalte „Test-Synonym“)
- wenn es sich um ein Synonym handelt, das nur ZU EINER EINZIGEN SKALA ein Synonym ist

Synonyme, die auf mehreren Skalen laden, dürfen nicht codiert werden (z.B. „abgeklärt“).

Neutrale Nennungen dürfen nur in zwei Fällen vergeben werden:

- wenn es sich um ein eindeutiges „weder-noch“ von Gegensatzpolen handelt (z.B. weder freundlich noch unfreundlich)
- wenn es sich um einen eindeutigen Widerspruch handelt, in dem das vorher genannte zu codierende Wort gleich im Anschluss (also nicht irgendwann im Aufsatz) inhaltlich aufgehoben wird (z.B. „Ich bin einerseits kontaktfreudig, gehe aber selten oder nie auf Leute zu.“)

Skala	Anzahl Nennungen (+)	Anz. neutrale	Anzahl Nennungen (-)	Nennungen/Skala gesamt	Wert*	Gew. Wert**
Extraversion(+) vs. Introversion(-)						
Verträglichkeit(+) vs. Unverträglichkeit (-)						
Gewissenhaftigkeit(+)...						
Em. Kontrolle(+) ...						
Offenheit(+) ...						
Empathie(+) ...						

*) Berechnung des Wertes: Anzahl Nennungen(+) - Anzahl Nennungen(-)

**) Berechnung des gewichteten Werts = Wert / Anzahl gesamt Nennungen der Skala

Beurteiler/in: _____

Anleitung: Jeder Aufsatz ist je Skala einzuschätzen. Durchschnitt sind die mittleren 50%, im Zweifelsfall ist auf Plus- oder Minuspol zuzuteilen. Nur, wenn sich keinerlei Hinweis auf die jeweilige Skala im Aufsatz findet, darf „keine Aussage möglich“ vergeben werden.

Plus-Pol: extravertiert, verträglich, gewissenhaft, emotional kontrolliert, offen, empathisch

Minus-Pol: introvertiert, unverträglich, nicht gewissenhaft, emotional instabil, nicht offen, nicht empathisch

TP NR. _____				
	+	Ø	-	keine Aussage möglich
Extraversion	☐	☐	☐	☐
Verträglichkeit	☐	☐	☐	☐
Gewissenhaftigkeit	☐	☐	☐	☐
Em. Kontrolle	☐	☐	☐	☐
Offenheit	☐	☐	☐	☐
Empathie	☐	☐	☐	☐

TP NR. _____				
	+	Ø	-	keine Aussage möglich
Extraversion	☐	☐	☐	☐
Verträglichkeit	☐	☐	☐	☐
Gewissenhaftigkeit	☐	☐	☐	☐
Em. Kontrolle	☐	☐	☐	☐
Offenheit	☐	☐	☐	☐
Empathie	☐	☐	☐	☐

TP NR. _____				
	+	Ø	-	keine Aussage möglich
Extraversion	☐	☐	☐	☐
Verträglichkeit	☐	☐	☐	☐
Gewissenhaftigkeit	☐	☐	☐	☐
Em. Kontrolle	☐	☐	☐	☐
Offenheit	☐	☐	☐	☐
Empathie	☐	☐	☐	☐

Lebenslauf (Curriculum Vitae)

Angaben zur Person

Name	Lisbeth Weitensfelder
Geburtsdatum	11.05.1982
Geburtsort	Klagenfurt
Staatsbürgerschaft	Österreich

Ausbildung

seit 2000/10	Diplomstudium Psychologie an der Universität Wien
seit 2000/10	Diplomstudium Betriebswirtschaft an der Wirtschaftsuniversität Wien
2008/02-2008/06	Erasmussemester an der Universidade Nova de Lisboa (Faculdade de Economia), Portugal
2002	Prüfung zur ÖVQ-Qualitätsbeauftragten im Gesundheitswesen (ÖVQ Training + Certification GmbH), Verlängerung des Zertifikats 2005
2000	Matura am Ingeborg-Bachmann-Gymnasium in Klagenfurt

Berufliche Tätigkeit

seit 2007/03	Studienassistentin an der Universität Wien, Fakultät für Psychologie, Arbeitsbereich Psychologische Diagnostik (Unterbrechung während des Erasmussemesters)
seit 2006/10	Werkvertragsnehmerin an der Universität Wien, Test- und Beratungsstelle des Arbeitsbereichs Psychologische Diagnostik
seit 2002	Qualitätsbeauftragte in der Ordination Dr. W. Weitensfelder, 9020 Klagenfurt (geringfügig beschäftigt)
2006/07-2006/09	Auslandsaufenthalt und Volunteertätigkeit bei der "Victoria Immigrant and Refugee Centre Society" in Victoria, BC (Kanada)
2006/01-2006/04	Pflichtpraktikum Psychologie am AKH Wien, Abteilung für Klinische Psychologie und Psychologische Diagnostik

Sprachkenntnisse

Muttersprache	Deutsch
weitere Sprachen	Englisch (fließend), Portugiesisch (Grundkenntnisse), Italienisch (Grundkenntnisse), Spanisch (Grundkenntnisse)

Sonstiges

2003	Herausgabe einer CD („Momente“) gemeinsam mit Ute Weitensfelder
2002	Herausgabe eines Buches mit Lyrik und Kurzprosa („Sonnenregen“, Verlag Johannes Heyn, Klagenfurt)
1999	Herausgabe eines Lyrikbandes („...und jeder Stern ein Traum von Dir“, Verlag Johannes Heyn, Klagenfurt)