



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Effekte bei der Veränderung der Itempositionen anhand
des Anstrengungsvermeidungstest
und B5PO

Verfasserin

Elke Eisenhauer

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, im Dezember 2008

Studienkennzahl: A 298

Studienrichtung: Psychologie

Betreuer: Univ. Prof. Dr. Mag. Klaus D. Kubinger

Danksagung

An dieser Stelle möchte ich gerne die Möglichkeit nutzen mich bei den Menschen zu bedanken, die mich auf dem Weg zu dieser Diplomarbeit begleitet haben.

An erster Stelle möchte ich meiner Familie danken, besonderes meinen Großeltern, die mich in vielfältiger Weise unterstützt haben. Ein weiteres großes Dankeschön richtet sich an meine Eltern, die mir dieses Studium ermöglichten und mir auf meinem gesamten Lebensweg in reichhaltiger Weise geholfen haben. Danke, an meine Schwester, die mich mit ihrer Hilfsbereitschaft und ihren wunderbaren Fähigkeiten in jeder Form während des Studiums und vor allem bei der Erstellung dieser Diplomarbeit begleitet hat.

Auch meinem Freund Wolfgang, der für technische Fragen immer zu Verfügung stand und mich mit seiner liebevollen Art immer wieder aufs Neue motivieren konnte, möchte ich an dieser Stelle danken. Meinen „Mädels“ möchte ich danken, dass sie immer ein offenes Ohr für mich hatten und für ihre unermüdliche Unterstützung während des gesamten Studiums.

Ein weiteres großes Dankeschön gilt Herrn Univ.-Prof. Dr. Mag. Klaus D. Kubinger für die Betreuung und Unterstützung während dieser Diplomarbeit, sowie den Mitarbeitern des Arbeitsbereichs für Psychologische Diagnostik.

Die rasche Akquirierung meiner Versuchspersonen haben Herr Hofrat DI Gottfried Kellner, Direktor der Höheren Bundeslehr- und Forschungsanstalt für Gartenbau in Schönbrunn/ Wien und Herr DI Alois Rosenberger, Direktor der Höheren Bundeslehr- und Forschungsanstalt für Landwirtschaft, Landtechnik und Lebensmitteltechnologie in Wieselburg/ Niederösterreich, sowie Herr AV DI Martin Kerschbaumer durch ihre unkomplizierte und rasche Unterstützung möglich gemacht. Auch die Schüler, die an der Untersuchung teilgenommen haben, trugen wesentlich zur Entstehung dieser Arbeit bei.

Vielen Dank.

Abstract

In dieser Diplomarbeit werden die Effekte, die aufgrund einer Itempositionsveränderung in zwei Persönlichkeitsfragebogen entstehen können, untersucht. Eine weitere Analyse beschäftigt sich mit den möglichen Einflüssen, die bei einer variierten Bearbeitungsreihenfolge von Fragebögen zu beobachten sind. Zur Untersuchung dieser Fragestellungen werden 151 Schüler/innen der Sekundarstufe in vier Versuchsgruppen randomisiert zugeteilt. Den Versuchspersonen werden das Big Five Plus One Persönlichkeitsinventar (B5PO, Holocher-Ertl, Kubinger und Menghin, 2003) und der Anstrengungsvermeidungstest (AVT, Rollet & Bartram, 1998) in ihrer originalen Itemreihenfolge, in der umgekehrten Itemreihenfolge und in nach ihren Subskalen geordneten Itemblöcken, vorgegeben. Die erste Versuchsgruppe die als Kontrollgruppe dient bearbeitet die Originalversion, die zweite Versuchsgruppe bearbeitet die umgekehrte Abfolge der Items und der dritten Gruppe wird die geblockte Form der Items vorgegeben. Zur Prüfung von Effekten aufgrund der Bearbeitungsreihenfolge, bearbeitet die Kontrollgruppe zuerst den AVT und anschließend den B5PO. Einer weiteren Versuchsgruppe werden die Fragebögen in deren umgedrehter Reihenfolge vorgelegt. Die Ergebnisse zeigen, dass eine Veränderung der Itempositionen keine signifikanten Effekte auf die Mittelwerte der Versuchspersonen hat. Weder eine umgekehrte Reihenfolge der Items noch die Items die nach ihren Skalen vorgegeben wurden, haben einen Einfluss auf die Ergebnisse der Versuchspersonen. Die Analyse der Bearbeitungsreihenfolge kommt ebenfalls zu keinem signifikanten Mittelwertsunterschied in den untersuchten Skalen.

The paper at hand examines the effects caused by changes in the itempositioning in two personalityquestionnaires. Further analysis deals with possible impacts stemming from varying respondencesequences. The investigation is based on samples from 151 students who were allocated to four respondergroups by randomization. Subjects were confronted with two personalityquestionnaires, the Big Five Plus One personalitytest (B5PO) and the so-called „Anstrengungsvermeidungstest“ (AVT, testing effort avoidance). Both tests were incorporated in the study in three constitutions differentiated by itempositions: beside original itempositions reversed itempositions and blocked items sorted by subscales were used. Concerning effects due to a variation in respondencesequences two respondergroups tested the original items, one facing the ATV first followed by the B5PO, the other one vice versa. Group three answered the questionnaires with reverse itempositions, group four dealt with blocked items sorted by subscales. The multivariate Analysis concerning itemposition and respondencesequence showed no significant differences in the mean values.

Inhaltsverzeichnis

I. Einleitung	15
II. Theoretischer Teil	17
1. Bisherige Forschungsergebnisse	17
2. Kognitiven Aufgaben eines Befragten	20
2.1. Die kognitiven Prozesse	21
2.1.1. Interpretation der Frage	22
2.1.2. Abruf der relevanten Gedächtnisinhalte	23
2.1.3. Bildung des Urteils	24
2.1.4. Formatierung des Urteils	24
2.1.5. Editierung des Urteils	25
3. Kontexteffekte	27
3.1. Arten von Kontexteffekten bei Befragungen	27
3.1.1. Kontexteffekte bei der Interpretation der Frage	29
3.1.2. Kontexteffekte beim Abrufen relevanter Gedächtnisinhalte	29
3.1.3. Kontexteffekte bei der Urteilsbildung	30
3.1.4. Kontexteffekte bei der Formatierung des Urteils	30
3.1.5. Kontexteffekte bei der Editierung des Urteils	31

III. Empirischer Teil	33
4. Ziel der Untersuchung	33
4.1. Hypothesen	33
4.1.1. Hypothesen bezüglich der Skalen des AVT	34
4.1.2. Hypothesen bezüglich der Skalen des B5PO	34
5. Methode	35
5.1. Versuchsplan	35
5.2. Erhebungsinstrumente	37
5.2.1. Big Five Plus One Persönlichkeitsinventar (B5PO)	37
5.2.2. Anstrengungsvermeidungstest (AVT)	38
5.3. Durchführung des Versuchs	39
5.3.1. Vorbereitungen	39
5.3.2. Ablauf	39
5.4. Stichprobe	40
5.4.1. Versuchsgruppen	40
6. Ergebnisse	44
6.1. Auswertung	44
6.1.1. Prüfung der Voraussetzungen	44
6.2. Ergebnisdarstellung	44
6.2.1. Ergebnisse der multivariaten Varianzanalyse	45
6.2.2. Ergebnisse der Hypothesen bezüglich der Skalen des AVT	46
6.2.3. Ergebnisse der Hypothesen bezüglich der Skalen des B5PO	47

6.3.	Zusammenfassung der Ergebnisse	51
7.	Diskussion	52
8.	Zusammenfassung	55
9.	Literaturverzeichnis	58
9.1.	Abbildungsverzeichnis	62
9.2.	Tabellenverzeichnis	63

Lebenslauf

I. Einleitung

„The whole is more than the sum of its parts“

(Aristoteles, griechischer Philosoph, 384 - 322 v. Chr.)

Dieses Zitat von Aristoteles beschreibt sehr gut, die Thematik dieser Diplomarbeit. Überträgt man dieses Zitat auf die Persönlichkeitsfragebögen, mit der die Psychologische Diagnostik arbeitet, erklärt dies, ob die Ergebnisse gleich sind, wenn die Befragten den ganzen Fragebogen in seiner Originalversion bearbeiten, oder sie die einzelnen Items in anderer Reihenfolge bzw. nach Skalen extrahiert bearbeiten.

In dieser Untersuchung soll überprüft werden, ob die in der Praxis gängige Methode, der Subskalen-Extraktion aus Persönlichkeitsfragebögen eine zulässige Methode ist. Diese Vorgehensweise wird seit Jahren unter dem Deckmäntelchen der Ökonomie durchgeführt. Die methodischen Aspekte aber werden kaum reflektiert (Krampen, Hense & Schneider, 1992; Rost & Hoberg, 1997). Krampen et al. (1992) stellen fest, dass Deusinger (1986) die Vorgehensweise der Subskalen-Extraktion mit den „Frankfurter Selbstkonzeptskalen“ (FSKN) die sowohl als Fragebogengesamtversion als auch als Einzelversionen der FSKN-Subskalen¹ vorgegeben werden können, geradezu propagiert, obwohl Deusingers Untersuchungsbefunde noch Zweifel offen lassen.

Auch aufgrund meiner persönlichen Erfahrung in der Psychologischen Diagnostik, die ich an der Universität Wien und in meinen Praktika sammeln durfte, kann ich diese Methode nachvollziehen. Aus der Sicht des Diagnostikers wie aber auch der Testpersonen, Patienten und Klienten ist die höchstmögliche Ökonomie anzustreben. Es stellt sich aber dennoch die Frage, ob es sich bei dieser „praktikablen“ Vorgehensweise um einen methodischen Kunstfehler oder um tolerierbare Praxis handelt (Rost & Hoberg, 1997).

Eine weitere Analyse dieser Diplomarbeit dient der Feststellung von Effekten auf die Ergebnisse von Fragebögen deren Vorgabenreihenfolge variiert wird. In der

¹ FSKN-Subskalen Deusinger (1986): FSAL Leistungsfähigkeit, FSAP Problembewältigung, FSVE Verhaltens- und Entscheidungssicherheit, FSSW Selbstwertschätzung, FSEG Empfindlichkeit und Gestimmtheit, FSST Standfestigkeit, FSKU Soziale Kontakt- und Umgangsfähigkeit, FSWA Wertschätzung durch andere, FSIA Irritierbarkeit durch andere, FSGA Gefühle und Beziehungen zu anderen.

psychologisch diagnostischen Praxis ist es der Regelfall, dass die Klienten mehrere Fragebögen in einer Sitzung bearbeiten. Ob die Reihenfolge in der die Fragebögen bearbeitet werden, einen Effekt auf die Ergebnisse hat, soll empirisch überprüft werden. Auf diese möglichen Beeinflussungen beim Einsatz mehrerer Fragebögen, weisen Dalbert et al. (1991) hin. In ihrer Arbeit beschreiben sie, dass die Bearbeitung von Bewältigungsfragebögen und die intensive Auseinandersetzung mit Erinnerungen die diese Fragebögen mit sich ziehen, Effekte auf die Ergebnisse eines nachfolgenden Befindlichkeitsfragebogen haben können. So könnten die Antworten mit günstigen Bewältigungsstrategien zu einer positiven Befindlichkeitsangabe führen und bei ungünstigen Bewältigungsstrategien zu einer negativen. Die Ergebnisse von Dalbert et al. (1991) bestätigen diese Annahmen. Sie zeigen, dass die Bearbeitungsreihenfolge der Fragebögen, einen Einfluss auf die Bewältigungs- und Befindlichkeitsangaben hat. Lazarus-Mainka (1987) weißt ebenfalls darauf hin, dass die Bearbeitung eines Angstfragebogens eine Auswirkung auf die Ergebnisse eines nachfolgenden Fragebogens mit sich ziehen könnte.

Auf diese Problemstellungen soll im Zuge der Untersuchung anhand des Big Five Plus One Persönlichkeitsinventar (Holocher-Ertl, Kubinger und Menghin, 2003) und des Anstrengungsvermeidungstests (Rollet & Bartram, 1998) eingegangen werden.

Im anschließenden Theorieteil werden bisherige Forschungsergebnisse besprochen. Sparfeldt et al. (2006) deuten mit dem Begriff des Item-Kontexteffektes darauf hin, dass die psychometrischen Eigenschaften eines Items oder einer Skala durch die Darstellung oder durch eine geblockte bzw. randomisierte Itemposition verändert werden. Aufgrund dessen, sollen die kognitiven Prozesse, die ein Befragter bei der Beantwortung von Fragen durchläuft und die Auswirkungen von Kontexteffekten auf diese Prozesse, in Abschnitt 3 diskutiert werden.

II. Theoretischer Teil

1. Bisherige Forschungsergebnisse

An dieser Stelle soll auf einige der zahlreiche empirische Untersuchungen zu Reihenfolgeeffekten eingegangen werden.

Die Extraktion von Subskalen aus Persönlichkeitsfragebögen wird von Krampen, Hense und Schneider (1992) als „methodischer Kunstfehler“ bezeichnet. Zu diesem Schluss kommen die Autoren, aufgrund ihrer Studie in der sie Reliabilität und Validität eines Fragebogens mit dessen Standardreihenfolge der Items und inhaltshomogener Item-Blockbildung vergleichen. In zwei Doppelstudien (Gesamt-N=213) wurden anhand der Halbform A des „Freiburger-Persönlichkeitsinventar“ (FPI-A, Fahrenberg, Selg & Hampel, 1973) dessen originale Itempositionen mit einer inhaltshomogenen Item-Blockbildung verglichen. Die Autoren kamen zu signifikanten Effekten bezüglich der Skalenmittelwerte, Skalen-Interkorrelation und der internen Konsistenz der Skalen. Die Studien zeigten Veränderungen der Reliabilitäts- und Validitätsparameter und lassen daher Zweifel an der Zuverlässigkeit der Methode der Subskalen-Extraktion. Eine weitere Studie der inhaltshomogenen Item-Blockbildung des Freiburger-Persönlichkeitsinventars (FPI-A1, Fahrenberg, Hampel & Selg, 1984) von Frank (1998) lieferte keine signifikanten Effekte.

Rost und Hoberg (1997) überprüfen ebenfalls die Auswirkungen von Itempositionsveränderungen. In ihrer Studie bearbeiten Schüler (N=600) die in zwei Stichproben zugeteilt wurden 18 Zielitems (je sechs identisch formulierte Items zum spezifischen schulischen Selbstkonzept bezüglich der Fächer Mathematik, Erdkunde und Deutsch) die zufällig unter 107 Distraktoritems gemischt wurden. Zusätzlich wurden weitere sechs Zielitems die sich auf das generelle schulische Selbstkonzept beziehen unter diese Items gemischt. Die Itemblockbildung in der Versuchsgruppe wurde dadurch gewährleistet, dass die jeweils drei identisch formulierten Items bezüglich der drei Schulfächer untereinander angeordnet wurden und zusätzlich mit einem grauen Hintergrund unterlegt wurden. So ergaben sich insgesamt sechs Blöcke á drei Items. Im Gegensatz zu Krampen et al. (1992) kamen Rost und Hoberg (1997)

zu dem Ergebnis, dass sich nur geringe Auswirkungen der Blockbildung zeigten. Konstruktvalidität und Skalenstreuung wurden nicht beeinflusst. Bei zwei Skalen kam es unter Itemblockbildung zu Erhöhung der Mittelwerte und der Homogenität. Die Autoren sprechen jedoch von einer geringen praktischen Relevanz der Ergebnisse.

Zu einem ähnlichen Ergebnis kommen Sparfeldt et al. (2006) bei ihrem Vergleich des Differentiellen Selbstkonzept Gitter (DISK-Gitter, Rost & Sparfeldt, 2002, Schilling, Sparfeldt, Rost & Nickels, 2005) in einem geblockten Itemformat und einem randomisierten herkömmlichen Itemformat. Das Blockformat wurde der Versuchsgruppe besonders hervorstechend präsentiert. Die Items wurden in einer Tabelle bzw. in einem Gitter dargestellt. In den Reihen der Tabelle standen acht Fragen, in den Spalten die vier Schulfächer (Mathematik, Physik, Englisch und Deutsch) deren schulisches Selbstkonzept erhoben werden soll. So ergaben sich jeweils acht identisch formulierte Fragen zu den vier Schulfächern. Der Kontrollgruppe wurden die gleichen Fragen in einem herkömmlichen randomisierten Format vorgegeben. Eine Faktorenanalyse bestätigte ähnliche Faktorenstrukturen wie z.B. Faktoren-Ladung, Faktoren-Varianz und Faktoren-Kovarianz. Die Skalenmittelwerte und Homogenität waren in beiden Formaten sehr ähnlich.

Weitere Studien (Schriesheim, Kopelmann & Solomon, 1989; Frank & Stäcker, 1995; Frank, 1998) die sich mit der inhaltshomogenen Itemblockbildung beschäftigen kamen zu folgenden Ergebnissen. Schriesheim et al. (1989) überprüften die Effekte der inhaltshomogenen Blockbildung anhand eines Fragebogens zu Lebenszufriedenheit und den „Job Diagnostik Survey“ (Hackman & Oldham, 1975). Sie zeigten, dass die Blockbildung zu Unterschieden in den Validitäts- und Reliabilitätskennwerten führte und stellten aufgrund dieser Ergebnisse die Methode der Itemblockbildung in Frage.

Franke und Stäcker (1995) stellten ebenfalls signifikante Effekte bei ihrer Untersuchung der Symptom-Checkliste-90-R (SCL-90-R, Derogatis, 1986), fest. Beim Vergleich der Ergebnisse bei Vorgabe der Standartreihenfolge der Items und einer Itemblockbildung, konnten signifikante Unterschiede bezüglich der Reliabilität, der Mittelwerte und der Interkorrelationen der Subskalen ermittelt werden. Auch sie kamen zu dem Schluss, dass die Subskalen-Extraktion eine zweifelhafte

Vorgehensweise darstellt. Franke (1996) untersuchte die SCL-90-R hinsichtlich der Fragestellung, ob die Vorgabe einer zufälligen Itemreihenfolge zu anderen Ergebnissen führt als die originale Itemreihenfolge. Im Gegensatz zur inhaltshomogenen Itemblockbildung (Franke & Stäcker, 1995) konnten bei einer zufälligen Itemreihenfolge keine Unterschiede in den Mittelwerten ermittelt werden.

Eine weitere Studie die sich mit den Auswirkungen einer Itempositionsveränderung beschäftigt, stammt von Ortner (2004). Sie vergleicht anhand des EPP-D (Eysenk, Wilson & Jackson, 1998) die Effekte der Positionsveränderungen. Zu diesem Zweck wurden zwei Versuchsgruppen (n= 168) gebildet. Eine Versuchsgruppe bearbeitete die Rasch-Modell konformen Items des EPP-D in seiner Originalform. Die zweite Versuchsgruppe beantwortete die identischen Items in umgekehrter Reihenfolge. Die Autorin konnte keine signifikanten Unterschiede der Mittelwerte feststellen. Weitere Analysen bezüglich der IRT (Item-Response-Theory) kamen zu dem Ergebnis, dass sich in drei von sieben Skalen eine unterschiedliche Itemschwierigkeit aufgrund der Itempositionsveränderung ergab. Zusammenfassend weist Ortner (2004) darauf hin, dass die klassische Testtheorie für Studien der Effekte bei Itempositionsveränderungen nicht ausreicht.

2. Kognitive Aufgaben eines Befragten

Bei der Beantwortung eines Items in einem Fragebogen, durchlaufen die Befragten einige kognitive Prozesse (Starck, 1994; Trost, 1997). Diese beginnen mit dem Verstehen der Frage, einer Formulierung der Antwort und enden mit dem Schreiben der Antwort oder, wie es meistens der Fall ist, mit dem Ankreuzen eines Kästchens. Diese Prozesse stehen unter dem Einfluss von Kontexteffekten. Diese Itemkontexteffekte führen nach Hartig und Hinrichs (2002) mit zunehmender Anzahl der zuvor bearbeiteten Items zu einer größeren Trennschärfe der Items und die Itemmittelwerte sinken. Der Erklärungsansatz der Autoren für diese Auswirkungen ist, dass im Laufe der Bearbeitung eines Fragebogens die den Items zugrunde liegenden Konstrukte erschlossen werden und sich dadurch die Bedeutung der Items verändern und sich diese gegenseitig verstärken. Die Autoren nehmen an, dass eine Verringerung dieser Kontexteffekte, durch eine Erläuterung der zu erfassenden Konstrukte vor der Bearbeitung des Fragebogens und durch eine Wiederholungsmessung zu erreichen ist. Diese Annahme wurde von ihnen mittels des State-trait-Heiterkeits-Inventars überprüft. An dieser Studie nahmen insgesamt 524 Versuchspersonen teil, die zu zwei Zeitpunkten die geblockten Items des Inventars bearbeiteten. Die Hälfte dieser Versuchspersonen erhielt vor der Bearbeitung des Fragebogens Informationen bezüglich der erfassten Konstrukte dieses Verfahrens. Wie von Hartig und Hinrichs erwartet, konnten die allgemeinen Itemkontexteffekte auf die Itemtrennschärfe und die Itemmittelwerte festgestellt werden. Die Messwiederholung führte dazu, dass die Skalenreliabilität und Faktorenladung vom ersten zum zweiten Messpunkt anstiegen. Die Erläuterung der Konstrukte, die die Versuchspersonen vor der Bearbeitung des Fragebogens erhielten, führte zu keinen Veränderungen der Kontexteffekte.

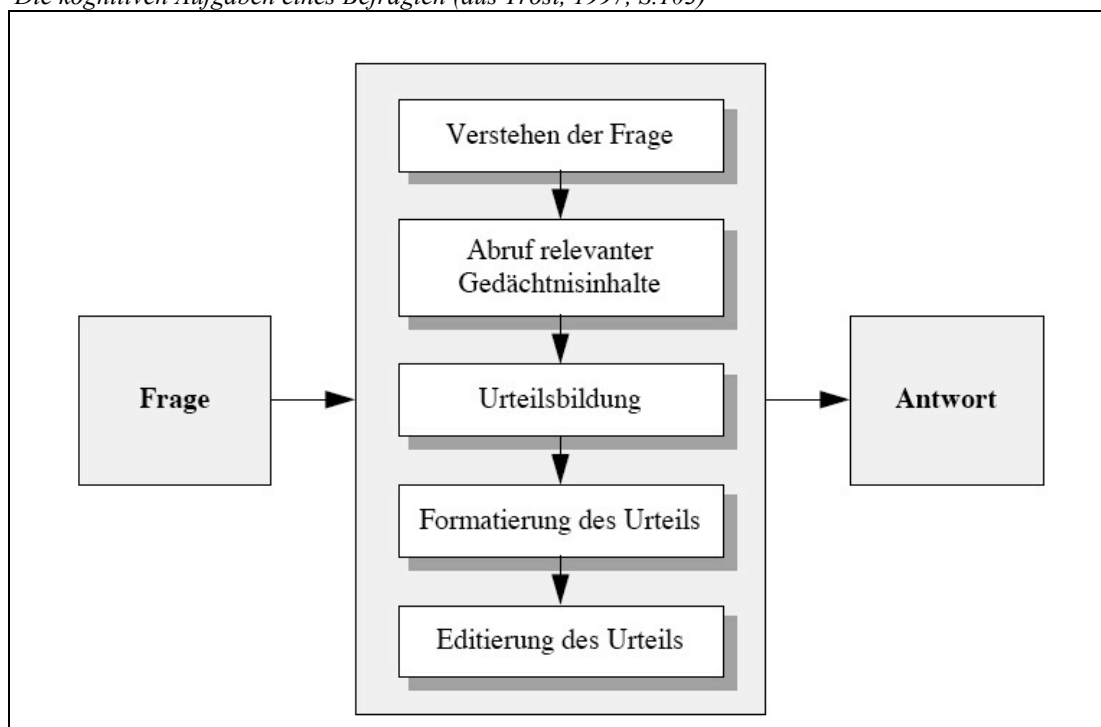
Diese Kontexteffekte könnten auch einen Einfluss bei einer Veränderung der Itempositionen auf die Ergebnisse haben. Aus diesem Grund soll auf eine Beschreibung dieser kognitiven Prozesse und ihrer Kontexteffekte, in dieser Arbeit nicht verzichtet werden. Einen sehr detaillierten Überblick bezüglich dieses Themas findet man bei Trost (1997), auf dessen Ausführung ebenfalls eingegangen werden soll.

2.1. Die kognitiven Prozesse

Trost (1997) beschreibt folgendes Prozessmodell bei der Beantwortung von Fragen: Zunächst muss der Befragte die Frage verstehen, anschließend folgt das Erinnern an für die Frage relevante Gedächtnisinhalte. Kann der Befragte auf kein Urteil aus der Vergangenheit zurückgreifen, kommt es zu einer Urteilsbildung aufgrund der relevanten Erinnerungen die er im Bezug auf die gestellte Frage abrufen konnte. Als nächsten Schritt kommt es zur Formatierung des Urteils. Der Befragte hat die Aufgabe, sein Urteil der vorgegeben Antwortkategorie anzupassen. Der letzte Prozess in der Beantwortung von Fragen besteht darin, dass Urteil zu editieren. Hier muss der Befragte entscheiden, ob er das tatsächliche Urteil bekannt gibt oder eine in eine bestimmte Richtung korrigierte Antwort auf die Frage gibt. Dieses Prozessmodell wird in Abbildung 1 dargestellt.

Abbildung 1

Die kognitiven Aufgaben eines Befragten (aus Trost, 1997, S.103)



2.1.1. Interpretation der Frage

Nach Trost (1997) besteht der erste Schritt, den ein Befragter zu machen hat, in der Interpretation der gestellten Frage. Eine semantische Interpretation der gesamten Frage, ergibt sich aus dem Verständnis der einzelnen Begriffe, aus denen sich die Frage zusammensetzt. Zusätzlich muss sich der Befragte eine Vorstellung davon machen, was der Untersucher mit dieser Frage in Erfahrung bringen will. Trost (1997) geht davon aus, dass einer täglichen Konversation dieser Prozess gleichermaßen zugrunde liegt. Die Frage „Wie geht’s?“ wird kaum zu Verständnisproblemen führen. Im Gegensatz zur pragmatischen Interpretation dieser Frage. Während einige Personen mit „Danke, gut!“ antworten werden, können andere Personen eine detaillierte Beschreibung ihres Befindens geben. Diese pragmatische Interpretation ist aber grundsätzlich nur bei offenen Fragen von Relevanz.

Natürlich spielt nach Trost (1997) auch die semantische Interpretation eine wichtige Rolle. Bei der Beantwortung von Fragen bezüglich der demographischen Daten kommt es zu keinen Verständnisproblemen. Dies ändert sich jedoch beispielsweise bei Fragen nach „interner oder externer Kundenzufriedenheit“. Dieser Faktor ist bei Fragebogen-Erhebungen problematisch, da die Interpretation der Frage letztendlich dem Befragten überlassen wird. Eine Möglichkeit Verständnisfragen zu stellen ist bei dieser Art der Erhebung kaum möglich.

Eine weitere Annahme, die die Interpretation einer Frage beeinflusst, beschreibt die kognitive Psychologie. Der Befragte legt eine Frage in die Richtung jener Gedächtnisinhalte aus, die ihm in dieser Situation am ehesten in den Sinn kommen. Diese Inhalte sind entweder solche, die ihn in der gegenwärtigen Situation am meisten beschäftigen, solche die in der Vergangenheit eine wichtige Rolle gespielt haben oder Inhalte die aus sehr prägnanten Ereignissen bestehen (Trost, 1997). Trost (1997) gibt in seiner Arbeit folgendes Beispiel, wenn die Frage „Wie stehen Sie zum Thema Golf?“ gestellt wird, wird die Interpretation eines Golf-Spielers eine andere sein als die eines Golf-Fahrers.

2.1.2. Abruf der relevanten Gedächtnisinhalte

Die nächste kognitive Aufgabe eines Befragten besteht nach Trost (1997) darin, relevante Informationen bezüglich der gestellten Frage aus dem Gedächtnis abzurufen.

Diese „Suche“ nach Erinnerungen wird durch einige Faktoren erschwert. Erstens wird beispielsweise bei der Frage „Wie zufrieden sind Sie mit Ihrem bisherigen Studienverlauf?“ eine Vielzahl von einzelnen Merkmalen aufgerufen. Zweitens bleibt beim Ausfüllen eines Fragebogens oftmals zu wenig Zeit, um sich mit dieser Vielzahl von Merkmalen intensiv auseinanderzusetzen. Der dritte Faktor bezieht sich auf die menschliche Informationsverarbeitungskapazität die von Natur aus eingeschränkt ist. Aus diesem Grund ist es dem Befragten nicht möglich alle Gedächtnisinhalte z.B. des Studienverlaufs abzurufen. Wie ist es aber trotz dieser Faktoren möglich, dass Menschen Fragen beantworten können?

Die kognitivpsychologische Forschung hat folgende Antwort: Der Mensch unternimmt immer den geringsten kognitiven Aufwand um eine Entscheidung zu treffen (Taylor, 1981; Leyens & Codol, 1990, zitiert nach Trost, 1997). Für die Beantwortung von Fragen bedeutet dies, dass der Befragte nur so lange in seinem Gedächtnis „sucht“ bis er für ihn genug zufrieden stellende Informationen gefunden hat.

Nach Trost (1997) spielt hier die Verfügbarkeit von Gedächtnisinhalten eine wichtige Rolle. Diese Verfügbarkeit wird erleichtert, wenn Merkmale, an die kurze Zeit vor der Befragung gedacht wurde, erinnert werden müssen. Dies könnte z.B. durch die vorhergehende Frage bewirkt werden (Strack, 1994) oder sich der Befragte in der Vergangenheit oft mit diesen Merkmalen beschäftigen musste. Auch Merkmale mit denen sich der Befragte intensiv beschäftigte oder Merkmale die mit einem prägnanten Erlebnis in Verbindung stehen, sind leichter verfügbar. Es können zwei Arten von verfügbaren Gedächtnisinhalten unterschieden werden. Erstens die chronischen Gedächtnisinhalte, die vom Befragten ohne größere kognitive Anstrengung abgerufen werden können und zweitens die temporären Inhalte die ein intensives Nachdenken verlangen, damit sie dem Befragten bewusst werden.

2.1.3. Bildung des Urteils

Der dritte kognitive Prozess bei der Beantwortung von Fragen dient der Urteilsbildung. Oft können Befragte Urteile abrufen, die sie schon zu einem früheren Zeitpunkt gebildet haben (Strack, 1994). Dieses Urteil kann, auf Basis der bereits bestehenden Meinung bezüglich einer berühmten Persönlichkeit z.B. den Bundespräsidenten, gebildet werden.

Nach Strack (1994), bildet sich eine Person eine Meinung über eine bestimmte Thematik aus verschiedenen Gründen. Die Priorität eines Sachverhalts spielt hierbei eine wichtige Rolle. Wichtige Sachverhalte führen zu einer stärkeren Meinungsbildung als nicht so wichtige. Sachverhalte, die eine Handlung erfordern, benötigen ebenfalls eine Meinungsbildung. Zusammenfassend basieren Urteile nicht unbedingt auf urteilsrelevanten Informationen, sondern auf Urteilen die schon früher getroffen wurden.

Falls zur aktuellen Frage noch keine bestehende Meinungen bzw. Urteile existieren, werden von dem Befragten die für das Urteil relevanten Informationen aus dem Gedächtnis abgerufen. In diesem Fall ist die Verfügbarkeit der Informationen wie schon in den beiden ersten kognitiven Prozessen (vgl. Abschnitt 2.1.1. und 2.1.2.) von zentraler Bedeutung.

2.1.4. Formatierung des Urteils

Nach Strack (1994) ist es bei standardisierten Befragungen in der Regel so, dass die Antworten in einem vorgegebenen Format abgegeben werden müssen. Dies bedeutet für den Befragten, dass er sein Urteil in ein vorgegebenes Antwortformat übersetzen muss. Diese Übersetzung erfordert kognitive Operationen, die Kontexteffekten ausgesetzt sind.

Besonders die Interpretation der vorgegeben Antwortkategorien wird von Kontexteffekten beeinflusst. Dies zeigen Schwarz et al. (1985) in ihrer Untersuchung, in der die Versuchspersonen anhand von zwei Arten von Antwortkategorien Angaben bezüglich ihres täglichen Fernsehkonsums machen mussten. Der Hälfte der insgesamt 132 Versuchspersonen wurde eine Antwortskala

mit hohen Häufigkeiten vorgegeben. Die zweite Versuchsgruppe gab ihren täglichen Fernsehkonsum anhand einer Antwortskala mit eher niedrigen Häufigkeiten an. Die Ergebnisse dieser Untersuchung sind der Tabelle 1 zu entnehmen.

Tabelle 1

Angaben bezüglich des täglichen Fernsehkonsum mit niedrigen oder hohen Antwortvorgaben (aus Schwarz et al., 1985)

Niedrige Antwortvorgaben		Hohe Antwortvorgaben	
bis 1/2 Stunde	7,4% (5)	bis 2 1/2 Stunden	62,5% (40)
1/2 bis 1 Stunde	17,7% (12)	2 1/2 bis 3 Stunden	23,4% (15)
1 bis 1 1/2 Stunde	26,5% (18)	3 bis 3 1/2 Stunden	7,8% (5)
1 1/2 bis 2 Stunde	14,7% (10)	3 1/2 bis 4 Stunden	4,7% (3)
2 bis 2 1/2 Stunden	17,7% (12)	4 bis 4 1/2 Stunden	1,6% (1)
mehr als 2 1/2 Stunden	16,2% (11)	mehr als 4 1/2 Stunden	0,0% (0)
	100% (68)		100% (64)

Aufgrund dieser Ergebnisse vermuten Schwarz et al. (1985), dass die Befragten vom Mittelpunkt der Antwortskala auf den Mittelwert der Häufigkeitsverteilung der Population schließen. Deshalb wird die eigene Angabe des täglichen Fernsehkonsums im Verhältnis zum Mittelwert geschätzt.

Ein weiterer Einflussfaktor ist nach Strack (1994) die Verankerung der Urteilsskala. Diese Verankerung bedeutet, dass Befragte ihre Angabe eines Skalenwertes bezüglich der Ausprägung eines Merkmals von der Breite der vorgegebenen Antwortskala, bestimmt wird.

2.1.5. Editierung des Urteils

Ob ein Befragter aber sein tatsächliches Urteil bekannt gibt oder gezielt eine andere Antwort wählt, wird von Trost (1997) als „Editierung des Urteils“ bezeichnet. Diese Vorgehensweise wird ein Befragter dann wählen, wenn seine Antwort positive oder aber auch negative Konsequenzen nach sich ziehen kann.

Das sehr häufig diskutierte Verhalten der sozialen Erwünschtheit, soll an dieser Stelle erwähnt werden. Hier sind die Befragten motiviert sich in der

Befragungssituation positiver darzustellen, als dies bei ihrer tatsächlichen Antwort der Fall wäre. Trost (1997) weist darauf hin, dass diese Art der Antworttendenzen bei persönlichen Interviews häufiger zu finden ist, als bei anonymen schriftlichen Befragungen.

Auf die Kontexteffekte die mit der Editierung des Urteils in Verbindung stehen wird in Abschnitt 3.5. genauer eingegangen.

3. Kontexteffekte

In dieser Diplomarbeit soll überprüft werden, ob Effekte bei der Veränderung der Itempositionen in einem Fragebogen zu finden sind. Aber aus welchem Grund sollten die Ergebnisse einer Versuchsperson die einen Fragebogen mit seinen originalen Itempositionen bearbeitet andere sein, als wenn dieselbe Person den identischen Fragebogen mit veränderten Position der Items oder Items die nach den Subskalen geordnet wurden, bearbeitet. In der Fragebogenkonstruktion weist Anger (1969) auf die Vorteile einer zufällig gemischten Itemvorgabe in Fragebögen hin. Der so genannte „Löscheffekt“ der nach Anger (1969) mit dieser Methode entsteht, minimiert mögliche Ausstrahlungseffekte der vorangegangenen Frage. Anger (1969) warnt aber auch vor einer übertriebenen Anwendung dieser Vorgehensweise wie folgt.

Man wird aber fragen müssen, ob allzu bunte Fragen-Potpourris – von der Zahnpasta zur Kirche, von der Büchsenmilch zum Sex oder zur Politik – nicht das Vertrauen der Befragten in die Ernsthaftigkeit der Untersuchung erschüttern, ganz abgesehen von den erheblichen Umstellungsschwierigkeiten, die jeder extrem sprunghafte Themenwechsel mit sich bringt. (S. 587)

In der Regel wird diese Vorgehensweise in der Fragebogenkonstruktion angewendet, um mögliche Ausstrahlungseffekte zu verhindern. Die am Anfang dieser Arbeit schon erwähnte Subskalen-Extraktion, die aus ökonomischen Gründen in der Praxis häufig eingesetzt wird, wäre nach diesen Annahmen eine bedenkliche Vorgehensweise. Durch die inhaltshomogene Vorgabe der Items könnten mögliche Kontexteffekte auftreten.

Im folgenden Abschnitt dieser Arbeit, soll auf die Arten von Kontexteffekten und deren Wirkung auf die fünf kognitiven Prozesse eines Befragten, eingegangen werden.

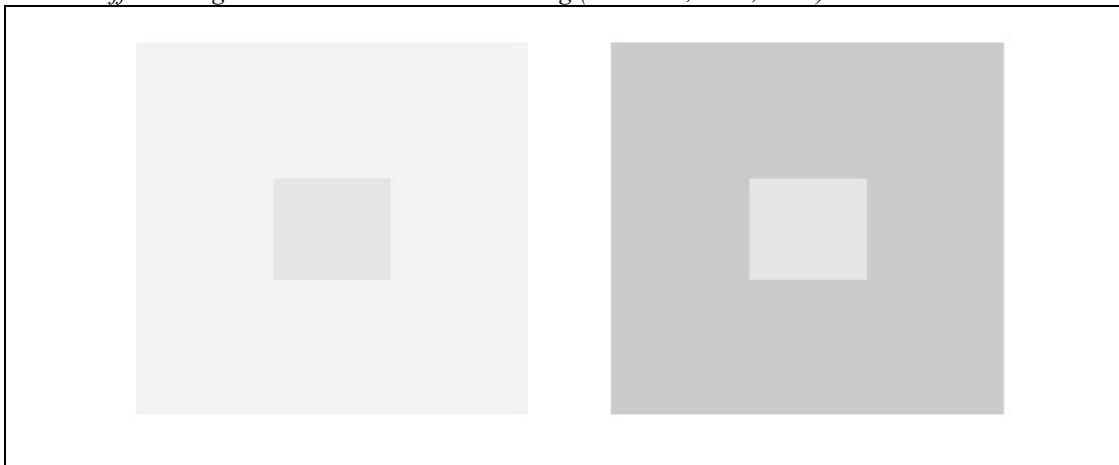
3.1. Arten von Kontexteffekten bei Befragungen

Die Wahrnehmungspsychologie beschäftigt sich schon seit langer Zeit mit den Einflüssen von Kontexteffekten. Ein besonders eindrucksvolles Beispiel der

Kontextabhängigkeit zeigt die Abbildung 2. Bei der jeweiligen Darbietung dieser zwei Quadrate würde man nach Trost (1997) bezüglich der Frage nach der Intensität der Farbe des inneren Quadrates, völlig unterschiedliche Antworten erhalten. Obwohl objektiv gesehen die beiden inneren Quadrate die identische Intensität aufweisen.

Abbildung 2

Kontexteffekt bezüglich der Intensitätswahrnehmung (aus Trost, 1997, S. 47)



Überträgt man dieses Beispiel der Wahrnehmungspsychologie auf Befragungssituationen können folgende Arten von Kontexteinflüssen entstehen.

Beeinflusst die Beantwortung der ersten Frage inhaltlich die Antwort der zweiten Frage, so spricht man von einem „Konsistenzeffekt. Dieser Effekt tritt nach Strack (1994) vor allem dann auf, wenn zuerst eine spezielle Frage gestellt wird und anschließend die Antwort auf eine allgemeine Frage gegeben wird. Anger (1969) meint, der „Konsistenzeffekt“ impliziert die Bemühung der Befragten, keine widersprüchlichen Antworten zugeben.

Der so genannte „Kontrasteffekt“ spiegelt den gegengerichteten Einfluss von Fragen wieder. Wenn die Antwort der zweiten Frage sich im starken Ausmaß von der Antwort der ersten Frage unterscheidet, spricht man nach Strack (1994) von einem „Kontrasteffekt“.

Ein weiterer sehr oft diskutierter Effekt betreffend der Reihenfolge von Fragen ist der „Halo-Effekt“. Hier kommt es dazu, dass positive oder negative Emotionen von einer Frage ausgelöst werden und diese Gefühle die nachfolgenden

Urteile der Befragten beeinflussen. Ein besonders drastisches Beispiel beschreibt Schumann (2006) anhand der Frage „Sind Sie für oder gegen die Todesstrafe?“. Wäre die befragte Person vorher mit Fragen zu einem Mord an einem Taxifahrer konfrontiert worden oder mit Fragen bezüglich eines abscheulichen Kindesmissbrauchs, würde dies das Urteil der Person beeinflussen.

Auf Wirkungen, dieser Kontexteffekte auf die kognitiven Prozesse die ein Befragter durchläuft, soll anschließend eingegangen werden.

3.1.1. Kontexteffekte bei der Interpretation der Frage

Eine vorausgehende Frage kann, nach Trost (1997), die Interpretation einer Frage beeinflussen. Der Inhalt eines Items kann auf die Erfassung der nächsten Items einwirken. Die Frage „Mögen Sie Kohl?“ wird nach Strack (1994) in eine andere Richtung interpretiert, wenn die vorausgehende Frage „Mögen Sie Karotten?“ wäre, oder wenn die vorausgehende Frage „Mögen Sie Lafontaine?“ lauten würde.

Um mögliche Kontexteffekte beim Verstehen einer Frage zu minimieren, ist eine eindeutige Formulierung unumgänglich (Trost, 1997). Im Bezug auf das oben angeführte Beispiel wäre folgende Formulierung geeigneter „Mögen Sie den ehemaligen deutschen Bundeskanzler Helmut Kohl?“ .

3.1.2. Kontexteffekte beim Abrufen relevanter Gedächtnisinhalte

Bei der Verfügbarkeit relevanter Gedächtnisinhalte ist die vorausgehende Frage von zentraler Bedeutung. Besonders die Art der aufgerufenen Gedächtnisinhalte, d.h. ob es sich um chronische oder temporäre Informationen (vgl. Abschnitt 2.1.2.) handelt, spielt nach Trost (1997) eine wichtige Rolle. Diese Annahme überprüfen Schwarz, Strack und Mai (1991) in ihrer Arbeit, an der 456 Erwachsene (18 Jahre oder älter) teilnahmen. Die Versuchspersonen wurden nach ihrer generellen Lebenszufriedenheit befragt und bezüglich der Zufriedenheit mit der Ehe. Die Reihenfolge, in der diese Fragen präsentiert wurden wurde in den jeweiligen Gruppen verändert. Folgende Ergebnisse konnten Schwarz et al. (1991) feststellen. Wurde zuerst nach der generellen Lebenszufriedenheit gefragt und anschließend

nach der Zufriedenheit in der Ehe, kam es zu einer Korrelation von $r = 0,32$. Bei der Befragung anhand der umgekehrten Reihenfolge der Fragen konnte eine Korrelation von $r = 0,67$ ermittelt werden. Diese Ergebnisse werden wie folgt interpretiert. Wird zuerst die Frage nach der Zufriedenheit in der Ehe gestellt, so werden die Gedächtnisinhalte auf diesen Aspekt gelenkt und Gefühle aktiviert, die wiederum die Beantwortung der zweiten Fragen beeinflussen. Die Frage nach der Zufriedenheit in der Ehe verlangt nach dem Abruf vieler verschiedener Merkmale. Diese große Anzahl an Informationen führt dazu, dass der Befragte auf temporär verfügbare Gedächtnisinhalte zurückgreift (Trost, 1997). Die Frage, bezüglich der generellen Lebenszufriedenheit wird aufgrund chronisch verfügbarer Informationen beantwortet. Zusammenfassend ist nach Trost (1997) zuzusagen, dass Fragen die hauptsächlich durch den Abruf chronischer Gedächtnisinhalte zu beantworten sind, eher zu geringeren Kontexteinflüssen führen, als dies bei Fragen zu temporär verfügbaren Inhalten der Fall ist.

3.1.3. Kontexteffekte bei der Urteilsbildung

Trost (1997) beschreibt in seiner Arbeit die Annahme mehrerer Untersuchungen, dass ein Urteil das zur Beantwortung einer Frage gebildet wurde, die Urteilsbildung der nachkommenden Fragen mit ähnlichen Sachverhalten beeinflusst.

Die Kontexteffekte kommen hier besonders bei Zufriedenheitsurteilen zu tragen, da die Urteile zur eigenen Zufriedenheit Vergleichsstandards verlangen. Das Niveau dieser normativen Standards ist situationsabhängig und kann aus diesem Grund vom Fragekontext beeinflusst werden (Trost, 1997, Strack, 1994).

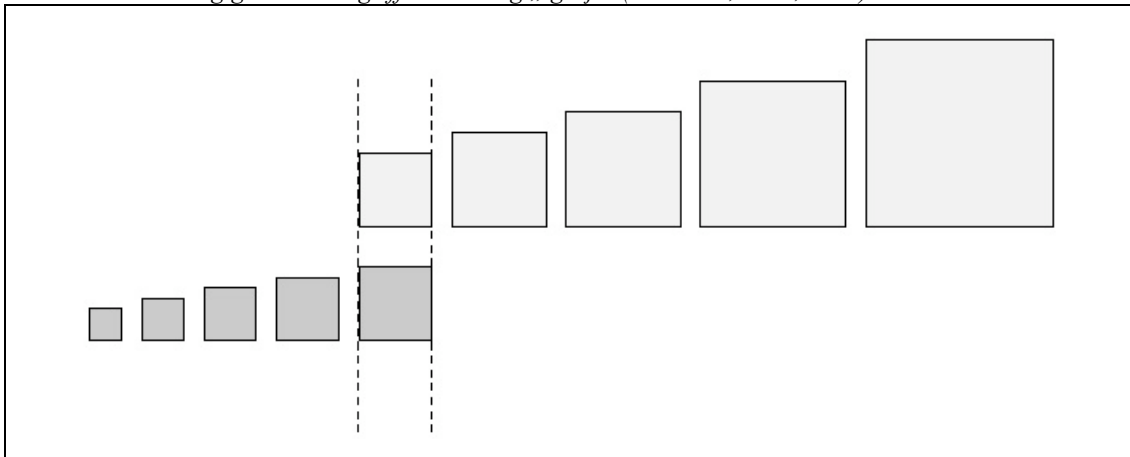
3.1.4. Kontexteffekte bei der Formatierung des Urteils

Bei der Übersetzung des gebildeten Urteils in die vorgegebenen Antwortkategorien, können ebenfalls Kontexteinflüsse auf die subjektiven Bedeutungen der Antwortkategorien einwirken. Eine Studie die diese Kontexteffekte untersucht wurde bereits in Abschnitt 2.1.4 erläutert.

Trost (1997) demonstriert in seiner Arbeit ein sehr anschauliches Beispiel, dass hier in Abbildung 3 dargestellt wird.

Abbildung 3

Die Kontextabhängigkeit der Begriffsbedeutung „ groß“ (aus Trost, 1997, S. 50)



Bei der Beurteilung anhand einer Antwortskala mit den Endkategorien „groß“ und „klein“ würde sich die Begriffsbedeutung des Wortes „ groß“ je nach Darbietung der oberen oder unteren Quadratreihe verändern.

3.1.5. Kontexteffekte bei der Editierung des Urteils

Wie schon in Abschnitt 2.1.5. erwähnt, dient dieser letzte kognitive Prozess den Befragten dazu, ihre tatsächlichen Antworten noch abzuändern ins besondere bei der Erwartung von Konsequenzen. Eine zentrale Bedeutung in dieser Thematik hat die oft diskutierte soziale Erwünschtheit. Die kognitive Verfügbarkeit dieser Antworttendenz kann durch vorausgehende Fragen erhöht werden (Trost, 1997).

Folgendes fiktives Szenario von Trost (1997) soll diese Kontexteffekte bei der Editierung eines Urteils verdeutlichen. Studenten derselben Lehrveranstaltung sollen gemeinsam an einem Teamprojekt arbeiten. Den Studenten, werden diesbezüglich folgende Fragen nacheinander gestellt:

Inwieweit ist eine erfolgreiche Projektarbeit von dem Engagement jedes Studenten abhängig? (Frage A)

Inwiefern sind Sie bereit mehr Leistung zu erbringen als man von Ihnen erwartet? (Frage B)

Der Frage A wird mit hoher Wahrscheinlichkeit von einem Großteil der Befragten zugestimmt. Dadurch, dass das Engagement ein sozial erwünschtes Verhalten darstellt, wird auch dieses bei einer Mehrheit der Studenten bewusst. Nach der Annahme von Trost (1997) wäre nun zu erwarten, dass die Frage B eher sozial erwünscht beantwortet wird, als dies beim Fehlen der Frage A der Fall wäre.

III. Empirischer Teil

4. Ziel der Untersuchung

In dieser Untersuchung soll erhoben werden, ob bei einer Veränderung der Itempositionen, Effekte auf die Ergebnisse von Persönlichkeitsfragebögen festzustellen sind. Zusätzlich soll analysiert werden, ob sich die Ergebnisse verändern, wenn die Bearbeitungsreihenfolge von Fragebögen variiert wird. Sind die Ergebnisse andere, wenn zuerst Fragebogen A bearbeitet und anschließend Fragebogen B oder wenn Fragebogen B vor Fragebogen A bearbeitet wird.

Um diese Einflüsse feststellen zu können, wurden folgende Fragebögen verwendet:

- Big Five Plus One Persönlichkeitsinventar (B5PO, Holocher-Ertl, Kubinger und Menghin, 2003)
- Anstrengungsvermeidungstest (AVT, Rollet & Bartram, 1998)

4.1. Hypothesen

Für die statistische Analyse dieser Diplomarbeit wurden insgesamt sechs Hypothesen formuliert. Es wurde geprüft, ob Unterschiede bei den Ergebnissen auf die Skalen (Pflichteifer und Anstrengungsvermeidung) des AVT und der Skalen (Extraversion, Offenheit, Empathie, Gewissenhaftigkeit, Verträglichkeit und Emotionale Kontrolle) des B5PO aufgrund der Itempositionsveränderungen festzustellen sind. Die Manipulation der Itempositionen erfolgte durch die Darbietung der originalen Itemreihenfolge in deren umgekehrter Reihenfolge. Eine zweite manipulierte Vorgabeform, wurde durch die Darbietung der Items die nach ihren Subskalen geordnet wurden, erreicht (siehe Tabelle 2). Daher konnten für die statistische Analyse für den AVT zwei Hypothesen und für den B5PO zwei Hypothesen formuliert werden.

Für die Analyse von Effekten bezüglich der Bearbeitungsreihenfolge der beiden Persönlichkeitsfragebögen konnten weitere zwei Hypothesen formuliert werden.

4.1.1. Hypothesen bezüglich der Skalen des AVT

- | | |
|------------------|---|
| H ₀₋₁ | Es gibt keinen Unterschied bei den Ergebnissen der Skalen ² des AVT zwischen originaler Itempositionierung und umgekehrter Itempositionierung. |
| H ₀₋₂ | Es gibt keinen Unterschied bei den Ergebnissen der Skalen ² des AVT zwischen originaler Itempositionierung und randomisierter Itempositionierung |
| H ₀₋₃ | Es gibt keinen Unterschied bei den Ergebnissen der Skalen ² des AVT bezüglich der Bearbeitungsreihenfolge der Fragebögen. |

4.1.2. Hypothesen bezüglich der Skalen des B5PO

- | | |
|------------------|---|
| H ₀₋₄ | Es gibt keinen Unterschied bei den Ergebnissen der Skalen ³ des B5PO zwischen originaler Itempositionierung und umgekehrter Itempositionierung. |
| H ₀₋₅ | Es gibt keinen Unterschied bei den Ergebnissen der Skalen ³ des B5PO zwischen originaler Itempositionierung und randomisierter Itempositionierung. |
| H ₀₋₆ | Es gibt keinen Unterschied bei den Ergebnissen der Skalen ³ Des B5PO bezüglich der Bearbeitungsreihenfolge der Fragebögen. |

² AVT-Skalen: Anstrengungsvermeidung, Pflichteifer

³ B5PO-Skalen: Extraversion, Offenheit, Empathie, Gewissenhaftigkeit, Verträglichkeit, Emotionale Kontrolle

5. Methode

5.1. Versuchsplan

Für die vorliegende Untersuchung wurde folgendes Versuchsplan umgesetzt:

Um festzustellen, ob es bei einer Veränderung der Itempositionen in einem Fragebogen Effekte gibt, wurden vier Gruppen gebildet. Diese vier Versuchsgruppen werden von Schüler/innen der 9. und 10. Schulstufe aus der Höheren Bundeslehr- und Forschungsanstalt für Gartenbau in Schönbrunn/ Wien und der Höheren Bundeslehr- und Forschungsanstalt für Landwirtschaft, Landtechnik und Lebensmitteltechnologie in Wieselburg/ Niederösterreich gebildet (genaue Beschreibung der Stichprobe - siehe Abschnitt 5.4.).

Für diese Erhebung wurden vier Testformen in Papier/Bleistift Form erstellt. Aus diesen Testformen ergeben sich vier Testhefte die mit vier verschiedenen farbigen Deckblättern den Versuchspersonen vorgegeben wurden um ein „Abschreiben“ als wenig erfolgreich zu indizieren. Die Manipulation der Testhefte erfolgte nach Tabelle 2. Insgesamt wurden pro Testform 40 Testhefte erstellt, so ergab sich eine Gesamtanzahl von 160 Testheften die vorbereitet wurden.

Die Testform I beinhaltet die Original Itemreihenfolge des B5PO und des AVT. Betreffend der Bearbeitungsreihenfolge der Fragebögen wurde in dieser Testform zuerst der AVT und anschließend der B5PO vorgegeben.

In Testform II wurden ebenfalls die Original Itemreihenfolgen der Fragebögen angewandt aber die Reihenfolge der Fragebögen wurde vertauscht. Das heißt, es wurde zuerst der B5PO und anschließend der AVT von den Versuchspersonen bearbeitet.

In Testform III wurden die Itempositionen umgedreht, das heißt dass das erste Item aus der Originalreihenfolge nun an der letzten Stelle zu finden war und das letzte Item der Originalreihenfolge an die erste Stelle gerückt war.

Die Testform IV wurde ebenfalls wieder mit der Bearbeitungsreihenfolge der Fragebögen wie in Testform I und Testform III vorgegeben. Die Itempositionen in der letzten Testform wurden dahingehend modifiziert, dass die Items nach Skalen geordnet vorgegeben wurden.

Tabelle 2

Überblick der dargebotenen Testhefte

Testheft I	Testheft II	Testheft III	Testheft IV
AVT B5PO Originale Item- Positionen	B5PO AVT Originale Item- positionen	AVT B5PO Umgekehrte Item- positionen	AVT B5PO Nach Skalen geordnete Item- positionen

Die für diese Untersuchung verwendeten Variablen sind den Tabellen 3 und 4 zu entnehmen.

Tabelle 3

Unabhängige Variablen (UV) Testheft

UV ₁	Original Itempositionierung
UV ₂	Umgekehrte Itempositionierung
UV ₃	Nach Skalen geordneter Itempositionierung
UV ₄	Bearbeitungsreihenfolge

Tabelle 4

Abhängige Variablen (AV)

AV ₁	AVT-Skala: Anstrengungsvermeidung
AV ₂	AVT-Skala: Pflichteifer
AV ₃	B5PO-Skala: Extraversion
AV ₄	B5PO-Skala: Offenheit
AV ₅	B5PO-Skala: Empathie
AV ₆	B5PO-Skala: Gewissenhaftigkeit
AV ₇	B5PO-Skala: Verträglichkeit
AV ₈	B5PO-Skala: Emotionale Kontrolle

5.2. Erhebungsinstrumente

5.2.1. Big Five Plus One Persönlichkeitsinventar [B5PO] (Holocher-Ertl, Kubinger & Menghin, 2003)

Bei diesem Rasch-Modell konformen Persönlichkeitsinventar handelt es sich um ein computergestütztes Verfahren zur Erfassung der Big-Five-Dimensionen Extraversion, Emotionale Kontrolle, Gewissenhaftigkeit, Offenheit und Verträglichkeit plus einer zusätzlichen Dimension Empathie. Die Selbsteinschätzung der Versuchspersonen bezüglich dieser relativ stabilen Persönlichkeitseigenschaften erfolgt über die sog. Analogskalen der insgesamt 66 Items des B5PO. Die Items des B5PO bestehen jeweils aus einem bipolaren Eigenschaftswort-Paar

Holocher-Ertl, Kubinger und Menghin (2003) geben an, dass die Objektivität aufgrund der standardisierten Computervorgabe des B5PO gegeben ist. Im Sinne einer inneren Konsistenz ist die Reliabilität aufgrund der Geltung des Rasch-Modells gegeben und die Konstruktvalidität kann aufgrund der Bestätigung des Big-Five-Persönlichkeitsmodell angenommen werden.

Bezüglich der Auswertung werden die Items dichotomisiert verrechnet, wobei sich aber bei den einzelnen Skalen unterschiedliche Aufteilungen der Analogskalen ergeben.

Für die Skalen Extraversion, Offenheit und Empathie wurde nach Holocher-Ertl, Kubinger und Menghin (2003) eine Dichotomisierung von 20:80 gewählt. Für die Interpretation, zum Beispiel der Dimension Extraversion bedeutet dies, dass eine Testperson die ihre Antworten in diesem 20%-Bereich angibt als „extravertiert“ zu verrechnen ist. Bezüglich der Skala Gewissenhaftigkeit konnte eine Dichotomisierung von 30:70 festgelegt werden. Bei den Skalen Verträglichkeit und Emotionale Kontrolle konnte festgestellt werden, dass die gesamte Breite der Analogskalen genutzt wurde. Daher erfolgte eine Dichotomisierung von 50:50 bei der Skala Verträglichkeit und eine 60:40 für die Skala Emotionale Kontrolle.

Aus organisatorischen Gründen wurde der B5PO bei dieser Untersuchung in Papier/Bleistift-Version vorgegeben, die hier verwendete Version inklusive der Instruktion für die Versuchspersonen, wurde von Wolfbauer (2006) übernommen.

Aus diesem Grund erfolgte die Auswertung mittels Schablonen die für die verschiedenen Testhefte jeweils erstellt wurden.

Um die dichotomisierte Auswertung ökonomisch und fehlerfrei durchführen zu können, wurden die Balkenunterteilungen wie folgt berechnet. Da die gesamte Balkenbreite 5,92 cm betrug, ergaben sich folgende Unterteilungen, die in Tabelle 5 angegeben werden.

Tabelle 5

Aufteilung der Analogskalen

Skalen	Aufteilung der Analogskalen in %	Aufteilung der Analogskalen in cm
Extraversion, Offenheit und Empathie	20:80	1,18:4,74
Gewissenhaftigkeit	30:70	1,78:4,14
Verträglichkeit	50:50	2,96:2,96
Emotionale Kontrolle	60:40	3,53:2,37

5.2.2. Anstrengungsvermeidungstest [AVT] (Rollett & Bartram,1998)

Beim AVT handelt es sich um einen speziellen Persönlichkeitsfragebogen, dessen Ziel es ist schulbezogene Anstrengungsvermeidung zu erfassen. Der Fragebogen kann sowohl als Papier-Bleistift- Version und als Computer-Version vorgegeben werden. Die 41 Items erfassen die Skalen Anstrengungsvermeidung und Pflichteifer und benötigen ca. 15-20 min Bearbeitungszeit. Der AVT kann auch als Gruppenverfahren eingesetzt werden und verfügt über eine standardisierte Instruktion. Das Verfahren verfügt über geschlechtsspezifischen Prozentnormen und Standardwertnormen für Schüler/innen der 5.-9. Schulstufe aus Hauptschulen und weiterführenden Schulen.

Von den insgesamt 41 Items des AVT sind 30 Items Rasch-Modell konform. Bei der vorliegenden Testung wurden die Items, die nicht Rasch-Modell konform sind, ausgeschieden. Da diese nach Angaben von Rollett und Bartram (1998) für Diskussionen und als Anknüpfungspunkte in der Therapie gedacht sind und aus diesem Grund auch gesondert ausgewertet werden.

Es wurde die originale Instruktion von Rollett und Bartram (1998) für diese Untersuchung übernommen.

Bezüglich der Testgütekriterien sind folgende Angaben im Manual des Fragebogens zu finden. Die Autoren geben an, durch die Anwendung der Auswertungsschablonen bzw. des Auswertungsprogramms maximale Objektivität zu erreichen. Die Reliabilität für die Anstrengungsvermeidungsskala ist mit $r_{tt} = 0,80$ und für die Pflichteiferskala mit $r_{tt} = 0,69$ angegeben. Aussagen bezüglich der Validität konnten mittels umfangreichen Untersuchungen getroffen werden. Diese Untersuchungen ergaben, dass der AVT mit seiner Messung der Anstrengungsvermeidungstendenzen in der Lage ist, schulischen Lernerfolg zu prognostizieren (vgl. auch Rollett & Bartram, 1998, S. 17-19).

Die Auswertung des AVT erfolgte, analog zum B5PO mittels für jedes Testheft angefertigten Schablonen um eine ökonomische und fehlerfreie Auswertung der Items zu gewährleisten. Da die Items des AVT mit „stimmt“ oder „stimmt nicht“ zu beantworten sind, konnte die Auswertung relativ schnell durchgeführt werden.

5.3. Durchführung des Versuchs

5.3.1. Vorbereitungen

Die Vorbereitungen dieser Testung begannen im Februar 2008 mit der Erstellung der vier Testhefte. Nach der Fertigstellung der Testhefte im März 2008 und der Zusagen der Direktoren der beiden Schulen, konnte der erste Testtag, an dem zwei Schulklassen teilnahmen, am 09. April 2008 in der Höheren Bundeslehr- und Forschungsanstalt für Gartenbau in Schönbrunn stattfinden. Die zweite Testung in der Höheren Bundeslehr- und Forschungsanstalt für Landwirtschaft, Landtechnik und Lebensmitteltechnologie in Wieselburg wurde am 30. Mai 2008 in drei Schulklassen, durchgeführt.

5.3.2. Ablauf

Die Versuchspersonen wurden in fünf Testterminen getestet, diese fanden jeweils vormittags statt. Es wurde darauf geachtet, dass eine ganze Unterrichtsstunde pro

Testung zu Verfügung stand. Somit konnte verhindert werden, dass die Schüler durch die Pausenglocke oder Pausenlärm gestört wurden oder die Testung von den Versuchspersonen abgebrochen wurde.

Um eine randomisierte Zuweisung der Versuchspersonen zu gewährleisten, wurden die Testhefte nach folgender Prozedur zugeteilt. Die Testhefte wurden nach der Sitzreihenfolge der Versuchspersonen ausgeteilt, dies bedeutet, dass der ersten Versuchsperson das Testheft eins zugewiesen wurde, der zweiten Versuchsperson das Testheft zwei usw., so dass alle vier Testhefte in einem Raum gleichzeitig bearbeitet wurden.

Anschließend an die Vorstellung der Testleiter wurde eine kurze mündliche Einführung gegeben. Hier wurde besonders darauf hingewiesen, dass es sich hierbei um eine anonyme Erhebung handelt. Im Bezug auf den Inhalt der Testung wurde lediglich bekannt gegeben, dass die Testhefte zwei Persönlichkeitsfragebögen beinhalten und diese ehrlich und vollständig von den Versuchspersonen zu beantworten sind. Nach dem Hinweis, dass sich die jeweiligen Instruktionen zu den Fragebögen in den Testheften befinden, wurden die Testungen gestartet.

Die Versuchspersonen arbeiteten während der gesamten Testung sehr konzentriert und ruhig. Es gab keine finanzielle Vergütung o.ä. für die Untersuchungsteilnahme.

5.4. Stichprobe

An der Untersuchung nahmen insgesamt 151 Personen teil. Die Testungen fanden nach der Genehmigung der Schulleiter der beiden Schulen an zwei Tagen statt. Die Verteilungen der demographischen Merkmale (Alter, Geschlecht) der Versuchspersonen, werden nachfolgend beschrieben.

5.4.1. Versuchsgruppen

Die Versuchspersonen wurden randomisiert zu vier Versuchsgruppen zugeteilt, davon dient eine Gruppe als Kontrollgruppe. Diese Zuteilung erfolgte, durch die zufällige Zuordnung der vier Testhefte. Die Kontrollgruppe wird von jenen

Versuchspersonen gebildet, denen das Testheft I zugeteilt wurde. Die drei Versuchsgruppen ergaben sich durch die Zuteilung der Testhefte II, III und IV. Die Verteilung der Versuchspersonen sind der Tabelle 6 zu entnehmen.

Tabelle 6

Verteilung der Versuchsgruppen

	Häufigkeit	Prozent
KG: Testform I	38	25,2
VG ₁ : Testform II	38	25,2
VG ₂ : Testform III	37	24,5
VG ₃ : Testform IV	38	25,2

Die Versuchspersonen (n = 151) waren Schüler und Schülerinnen des ersten und zweiten Jahrgangs der beiden weiterführenden Schulen. Das Alter der Versuchspersonen lag zwischen 14 und 18 Jahren.

Das Durchschnittsalter der Schülerinnen und Schüler betrug 15,03 Jahre (SD = 0,739). Tabelle 7 zeigt, die Verteilung der Versuchspersonen nach ihrem Alter für die Versuchsgruppen und die Gesamtstichprobe ist der Tabelle 8 zu entnehmen.

Tabelle 7

Verteilung der Versuchspersonen nach dem Alter für die Versuchsgruppen

Alter	KG: Testheft I		VG ₁ : Testheft II		VG ₂ : Testheft III		VG ₃ : Testheft IV	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
14	11	28,9	6	15,8	8	21,6	5	13,2
15	21	55,3	24	63,2	22	59,5	27	71,1
16	6	15,8	6	15,8	4	10,8	6	15,8
17	-	-	1	2,6	2	5,4	-	-
18	-	-	1	2,6	1	2,7	-	-
Gesamt	38	100,0	38	100,0	37	100,0	38	100,0

Tabelle 8

*Verteilung der Versuchspersonen nach dem Alter
für die Gesamtstichprobe*

Alter	Gesamtstichprobe	
	Häufigkeit	Prozent
14	30	19,9
15	94	62,3
16	22	14,6
17	3	2,0
18	2	1,3
Gesamt	151	100,0

Insgesamt nahmen 40 weibliche und 111 männliche Versuchspersonen an der Untersuchung teil (siehe Tabelle 9). Die Verteilung des Geschlechts auf die Versuchsgruppen ist in Tabelle 10 zu finden. Die ungleiche Verteilung des Geschlechts ist darauf zurückzuführen, dass von den insgesamt fünf getesteten Klassen, drei Klassen mit dem Unterrichtsschwerpunkt Landwirtschaft und Landtechnik untersucht wurden. Dass dieser Schulzweig überdurchschnittlich von männlichen Schülern besucht wird, wurde auch von der Schulleitung bestätigt.

Tabelle 9

Verteilung der Versuchspersonen nach dem Geschlecht für die Versuchsgruppen

	KG: Testheft I		VG₁: Testheft II		VG₂: Testheft III		VG₃: Testheft IV	
	Häufig- keit	Prozent	Häufig- keit	Prozent	Häufig- keit	Prozent	Häufig- keit	Prozent
weiblich	9	23,7	12	31,6	10	27,0	9	23,7
männlich	29	76,3	26	68,4	27	73,0	29	76,3
Gesamt	38	100,0	38	100,0	37	100,0	38	100,0

Tabelle 10

*Verteilung der Versuchspersonen nach dem Geschlecht
für die Gesamtstichprobe*

Gesamtstichprobe		
	Häufigkeiten	Prozent
weiblich	40	26,5
männlich	111	73,5
Gesamt	151	100,0

6. Ergebnisse

6.1. Auswertung

Zur Prüfung der in dieser Untersuchung aufgestellten Unterschiedshypothesen wurde eine Multivariate Varianzanalyse gewählt.

Bei signifikanten Ergebnissen kam der Student-Newman-Keuls-Test zum Einsatz, dieser Post-hoc-Test ist in der Lage zu zeigen, zwischen welchen der vier Versuchsgruppen ein signifikanter Unterschied besteht.

Für die statistische Analyse der Untersuchung wurde das Programm „Statistical Package for Social Scientists“ (SPSS for Windows 15.0, deutsche Version) angewendet.

6.1.1. Prüfung der Voraussetzungen

Zu den Voraussetzungen einer multivariaten Varianzanalyse zählt die multivariate Normalverteilung der Variablen und die Homogenität der Varianz-Kovarianz-Matrix.

Aufgrund der kleinen Stichprobe erfolgte eine optische Prüfung der univariaten Normalverteilung der Variablen mittels Histogramm.

Zur Überprüfung der Homogenität der Varianz-Kovarianz-Matrix wurde der Box-M-Tests eingesetzt, er kam zu keinem signifikanten Ergebnis.

6.2. Ergebnisdarstellung

Der Tabelle 11 sind die gesamten Mittelwerte und Standardabweichungen der vier Versuchsgruppen zu entnehmen.

Tabelle 11

Die Mittelwerte (Standardabweichungen) der vier Versuchsgruppen der abhängigen Variablen (Pflichteifer, Anstrengungsvermeidung, Extraversion, Offenheit, Empathie, Gewissenhaftigkeit, Verträglichkeit, Emotional Kontrolle)

	KG:	VG₁:	VG₂:	VG₃:	Insgesamt
	Testform I	Testform II	Testform III	Testform IV	
	n = 38	n = 38	n = 37	n = 38	
Pflichteifer	6,57 (2,06)	5,70 (2,25)	6,33 (1,89)	5,89 (1,52)	6,12 (1,96)
Anstrengungs- vermeidung	6,16 (3,23)	6,59 (3,55)	7,25 (4,50)	5,82 (2,98)	6,45 (3,60)
Extraversion	4,16 (2,56)	4,30 (3,03)	4,47 (2,66)	3,37 (2,22)	4,07 (2,64)
Offenheit	5,08 (2,30)	5,22 (2,96)	5,97 (2,62)	5,45 (2,81)	5,43 (2,68)
Empathie	4,22 (2,22)	4,49 (2,288)	5,08 (2,84)	4,00 (2,87)	4,44 (2,72)
Gewissen- haftigkeit	9,49 (3,88)	6,92 (4,40)	10,33 (3,71)	9,18 (4,09)	8,97 (4,19)
Verträglich- keit	6,00 (2,05)	5,73 (2,00)	6,61 (1,77)	6,34 (1,90)	6,17 (1,95)
Emotional Kontrolle	4,51 (1,62)	4,14 (1,45)	4,31 (1,56)	4,24 (1,60)	4,30 (1,55)

6.2.1. Ergebnisse der multivariaten Varianzanalyse

Wie Tabelle 12 zu entnehmen ist, führte die multivariate Varianzanalyse zu keinem signifikanten Ergebnis. Es können alle Nullhypothesen (H_{0-1} , H_{0-2} , H_{0-3} , H_{0-4} , H_{0-5} , H_{0-6}) beibehalten werden. Dennoch soll im folgenden Abschnitt auf die Mittelwertsdifferenzen der vier Testformen eingegangen werden.

Tabelle 12

Multivariate Tests

Effekt	Wert	F	Hypothese df	Fehler df	Signifikanz
Testheft					
Wilks-Lambda	0,773	1,538	24,00	397,94	0,052

Für alle Hypothesenprüfungen wurde das Risiko 1. Art auf $\alpha = 0,05$ festgelegt.

6.2.2. Ergebnisse der Hypothesen bezüglich der Skalen des AVT

a) Anstrengungsvermeidung

Wie aus Tabelle 13 zu entnehmen, erreicht die Versuchsgruppe die das Testheft III bearbeitete einen höheren Mittelwert in der Skala Anstrengungsvermeidung, als dies bei den anderen Versuchsgruppen der Fall war.

Tabelle 13

Ergebnisse der Skala Anstrengungsvermeidung. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
VG ₂ : Testform III	37	7,25	4,50
VG ₁ : Testform II	37	6,59	3,55
KG: Testform I	38	6,16	3,23
VG ₃ : Testform IV	38	5,82	2,98
Gesamt	150	6,45	3,60

b) Pflichteifer

Im Bezug auf die Skala Pflichteifer sind nur sehr geringe Mittelwertsunterschiede festzustellen, wie aus Tabelle 14 ersichtlich. Hier konnte die KG den höchsten Mittelwert erreichen. Die Position der Items als auch die Bearbeitungsreihenfolge hatte keinen signifikanten Effekt auf die Ergebnisse der Versuchspersonen.

Tabelle 14

Ergebnisse der Skala Pflichteifer. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
KG: Testform I	38	6,57	2,06
VG ₂ : Testform III	37	6,33	1,89
VG ₃ : Testform IV	38	5,89	1,52
VG ₁ : Testform II	38	5,70	2,25
Gesamt	151	6,12	1,96

6.2.3. Ergebnisse der Hypothesen bezüglich der Skalen des B5PO

a) Extraversion

Der Tabelle 15 ist zu entnehmen, dass in der Skala Extraversion nur sehr geringe Abweichungen der Mittelwerte analysiert werden konnten. Hier kam die VG₂ zum höchsten Ergebnis.

Tabelle 15

Ergebnisse der Skala Extraversion. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
VG ₂ : Testform III	37	4,47	2,66
VG ₁ : Testform II	38	4,30	3,03
KG: Testform I	38	4,16	2,56
VG ₃ : Testform IV	38	3,37	2,22
Gesamt	151	4,07	2,64

b) Offenheit

Der Vergleich der Mittelwerte der Versuchsgruppen in der Skala Offenheit kommt ebenfalls nur zu einem sehr kleinen Effekt, vergleiche Tabelle 16. Weder die Itemposition noch die Bearbeitungsreihenfolge der Fragebögen hatten einen signifikanten Einfluss auf die Ergebnisse.

Tabelle 16

Ergebnisse der Skala Offenheit. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
VG ₂ : Testform III	37	5,97	2,62
VG ₃ : Testform IV	38	5,45	2,81
VG ₁ : Testform II	38	5,22	2,96
KG: Testform I	38	5,08	2,30
Gesamt	151	5,43	2,68

c) Empathie

In der Skale Empathie erreichten die Versuchspersonen denen das Testheft III vorgegeben wurde den höchsten Mittelwert (Tabelle 17). Dennoch hatte die umgekehrte Itemreihenfolge nur einen kleinen, statistisch unbedeutsamen Effekt auf die Ergebnisse.

Tabelle 17

Ergebnisse der Skala Empathie. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
VG ₂ : Testform III	37	5,08	2,84
VG ₁ : Testform II	38	4,49	2,88
KG: Testform I	38	4,22	2,22
VG ₃ : Testform IV	38	4,00	2,87
Gesamt	151	4,44	2,72

d) Gewissenhaftigkeit

Wie Tabelle 18 zeigt, können in der Skala Gewissenhaftigkeit deutlich höhere Mittelwertsunterschiede zwischen den Versuchsgruppen festgestellt werden. Dennoch sind diese Unterschiede aus statistischer Sicht unbedeutsam. Diese Resultate zeigen, dass die Bearbeitungsreihenfolge der Fragebögen einen geringen Einfluss auf das Ergebnis in der Skala Gewissenhaftigkeit hat. Die Versuchsgruppe in der die Fragebögen in der manipulierten Bearbeitungsreihenfolge (Testform II) vorgegeben wurde, erreichte den niedrigsten Mittelwert (siehe Tabelle 18). Die graphische Darstellung in Abbildung 4 soll diesen Unterschied zeigen.

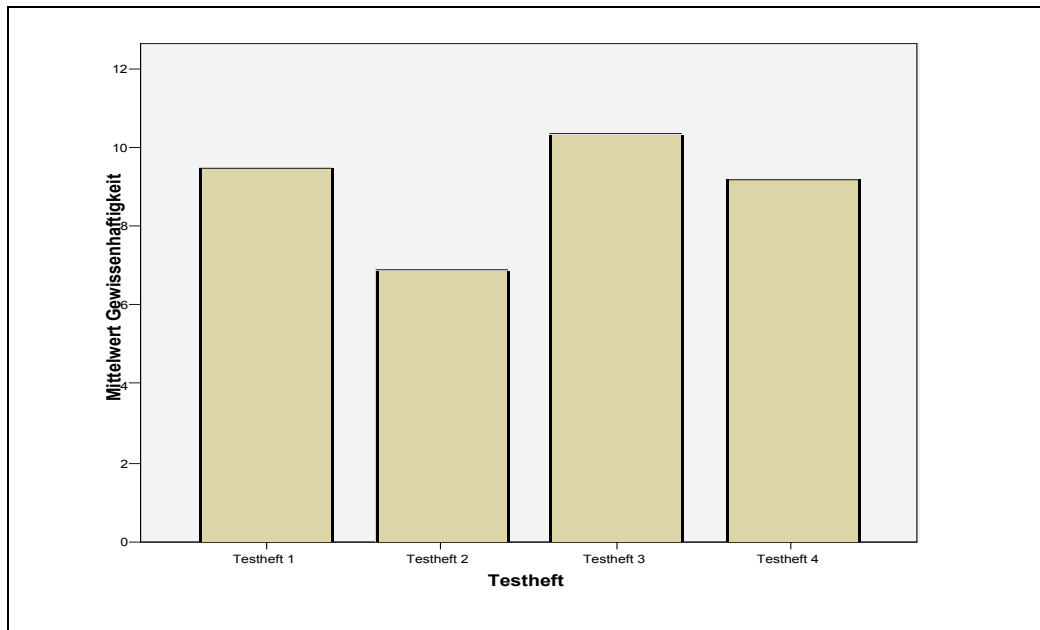
Tabelle 18

Ergebnisse der Skala Gewissenhaftigkeit. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
VG ₂ : Testform III	36	10,33	3,71
KG: Testform I	37	9,49	3,88
VG ₃ : Testform IV	38	9,18	4,09
VG ₁ : Testform II	37	6,92	4,40
Gesamt	149	8,97	4,19

Abbildung 4

Mittelwerte der vier Versuchsgruppen in der Skala Gewissenhaftigkeit



e) Verträglichkeit

Die Mittelwerte bezüglich der Skala Verträglichkeit weisen nur geringe Differenzen auf (Tabelle 19). Den höchsten Mittelwert erreichte die Versuchsgruppe in der die umgekehrte originale Itemreihenfolge des AVT und B5PO bearbeitet wurde. Der niedrigste Mittelwert konnte in der VG₁ ermittelt werden.

Tabelle 19

Ergebnisse der Skala Verträglichkeit. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
VG ₂ : Testform III	37	6,61	1,77
VG ₃ : Testform IV	38	6,34	1,90
KG: Testform I	38	6,00	2,05
VG ₁ : Testform II	38	5,73	2,00
Gesamt	151	6,17	1,95

f) Emotionale Kontrolle

Wie Tabelle 20 zu entnehmen, konnten die durchgeführten Manipulationen der Fragebögen einen geringen Unterschied in den Mittelwerten der Versuchsgruppen erzeugen. In der Skala Emotionale Kontrolle erreichte die KG den höchsten Mittelwert.

Tabelle 20

Ergebnisse der Skala Emotionale Kontrolle. Mittelwerte und Standardabweichungen

Versuchsgruppen	n	Mittelwerte	Standardabweichung
KG: Testform I	38	4,51	1,62
VG ₂ : Testform III	37	4,31	1,56
VG ₃ : Testform IV	38	4,24	1,60
VG ₁ : Testform II	38	4,14	1,45
Gesamt	151	4,30	1,55

6.3. Zusammenfassung der Ergebnisse

Insgesamt wurden sechs Hypothesen formuliert. Wobei vier Hypothesen sich auf die Positionsveränderungen der Items in den Skalen der Fragebögen AVT und B5PO beziehen. Nach der statistischen Analyse ist festzustellen, dass alle vier Nullhypothesen beibehalten werden können. Die Positionen der Items hatten keinen Einfluss auf die Ergebnisse.

Um Unterschiede in den Ergebnissen der Versuchspersonen bezüglich der Bearbeitungsreihenfolge der Fragebögen AVT und B5PO zu untersuchen, wurden zwei Hypothesen formuliert. Hier zeigten, die Ergebnisse der statistischen Analyse, dass beide Nullhypothesen beibehalten werden können.

7. Diskussion

Das Hauptanliegen dieser Diplomarbeit galt der Analyse von Effekten auf die Ergebnisse von Persönlichkeitsfragebögen aufgrund einer Positionsveränderung ihrer Items. Können Effekte festgestellt werden, wenn die originale Reihenfolge der Items umgekehrt vorgegeben wird bzw. die Items nach ihren Skalen geordnet beantwortet werden? Zur statistischen Analyse der Mittelwertsunterschiede der vier Versuchsgruppen wurde eine multivariate Varianzanalyse angewendet.

In der vorliegenden Untersuchung konnten keine Effekte auf die Mittelwerte aufgrund einer Itempositionsveränderung festgestellt werden. Weder die Vorgabe der Items in einem skalenhomogenen Block, noch eine umgekehrte Reihenfolge der originalen Itemreihenfolge beeinflusste die Ergebnisse der Versuchspersonen. Studien von Rost und Hoberg (1997), Sparfeldt et al. (2006) und Franke (1998) bezüglich der Überprüfung von Itemblockbildungen, kamen ebenfalls zu diesem Ergebnis. Rost und Hoberg (1997) stellten zwar unter der Bedingung der Blockbildung Mittelwerts- und Homogenitätsunterschiede fest, sprechen aber von einer geringen praktischen Relevanz. Bei dem Vorgabemodus, in dem die originale Itemreihenfolge in umgekehrter Reihenfolge vorgegeben wurde, konnte auch Ortner (2004) keinen Einfluss auf die Mittelwerte feststellen.

An dieser Stelle soll darauf hingewiesen werden, dass die Ergebnisse dieser Untersuchung sich ausschließlich auf die zwei untersuchten Fragebögen, den Anstrengungsvermeidungstest (AVT) und den Big Five Plus One (B5PO), beziehen. Da folgende Studien zeigten, dass die Veränderungen von Itempositionen in anderen Fragebögen, Effekte mit sich ziehen können: Schriesheim et al. (1989) konnten unter Itemblockbildungen Unterschiede bei den Validitäts- und Reliabilitätskennwerten feststellen. Franke und Stäcker (1995) und Franke (1996) fanden ebenfalls signifikante Effekte beim Vergleich von originalen Itemreihenfolgen und Itemblockbildungen.

Zusammenfassend bleibt zu diskutieren, ob sich veränderte Itempositionen bzw. Itemblockbildungen nicht in jedem Fragebogen anders verhalten könnten. Natürlich wäre die Untersuchung jedes einzelnen Fragebogens, der in der psychologisch-diagnostischen Praxis eingesetzt wird, hinsichtlich der besprochenen

Thematik aus ökonomischen Gründen sehr schwierig, wenn nicht sogar unmöglich. Die Fragebögen bei denen in der Praxis besonders häufig eine Subskalen-Extraktion angewendet wird, sollten hier dennoch einer Prüfung unterzogen werden.

Eine weitere statistische Untersuchung in dieser Diplomarbeit befasste sich mit den Einflüssen, die bei einer variierenden Bearbeitungsreihenfolge der zwei Fragebögen (AVT, B5PO) entstehen könnten. Die durchgeführte multivariate Varianzanalyse kam zu keinem signifikanten Ergebnis. Bei einer Analyse der Mittelwerte konnten lediglich in der Skala Gewissenhaftigkeit Differenzen festgestellt werden. Obwohl diese Mittelwertsunterschiede aus statistischer Sicht unbedeutsam sind, soll dennoch kurz auf diese Effekte eingegangen werden. Eine mögliche Interpretation wäre, dass die Versuchspersonen der Testhefte I, III und IV zuerst durch die Bearbeitung des AVT, Fragen zu ihren Anstrengungsvermeidungstendenzen beantworten mussten. Und aufgrund dieser intensiven Reflektierung bezüglich ihres Pflichtverhaltens und dem Umgang mit Anstrengungsvermeidung, schätzten sich diese Versuchspersonen gewissenhafter ein. In diesem Fall wäre es möglich, dass der so genannte „Konsistenzeffekt“ eine Rolle spielt. Nach Strack (1994), tritt dieser Effekt auf Itemebene dann auf, wenn zuerst eine spezielle Frage und anschließend eine allgemeine Frage beantwortet wird. Dieser spezielle Kontexteffekt könnte auch auf Fragebogenebene eine Rolle spielen. Anscheinend kommt es durch die Beantwortung eines speziellen Persönlichkeitsfragebogens (AVT) zu einer Beeinflussung der Antworten auf eine, wie in diesem Fall, anschließend bearbeitete Persönlichkeits-Fragebogenbatterie (B5PO).

Die Studien von Dalbert et al. (1991) und Lazarus-Mainka (1987) im Bezug auf diese Thematik führten zu signifikanten Ergebnissen. Die Untersuchungen dieser Autoren beziehen sich allerdings auf Fragebögen die das Angsterleben bzw. Angstbewältigung erfassen. Die Vergleichbarkeit dieser Studien mit den Ergebnissen der vorliegenden Studie wird daher in Frage gestellt.

Besonders in der psychologisch-diagnostischen Praxis ist es in der Regel so, dass mehrere Fragebögen direkt hintereinander bearbeitet werden und sich so gegenseitig beeinflussen könnten. Diese Kontexteffekte die durch die Bearbeitung

eines Fragebogens hervorgerufen werden, könnten von zentraler Bedeutung sein und sollten in Zukunft noch genauer untersucht werden.

Ein weiterer Anstoß für zukünftige Untersuchungen dieser Thematik, mit besonderem Blick auf die Analogskalen des B5PO, soll hier gegeben werden, nämlich die Analyse der Effekte aufgrund einer Itempositionsveränderung auf Item-Niveau. Durch eine genaue Vermessung der einzelnen Analogskalen-Antworten könnten Unterschiede bezüglich jedes einzelnen Items festgestellt werden. So könnten eventuelle einzelne Itemreihenfolgeeffekte untersucht werden.

Bei der Durchführung dieser vorliegenden Arbeit traten keine größeren Probleme auf. Lediglich bei der ersten Testung der Versuchspersonen kam es zu Verwirrungen aufgrund eines Rechtschreibfehlers bezüglich eines Items des B5PO. Dieses Problem konnte sehr schnell gelöst werden und die Richtigstellung wurde an die Versuchspersonen dieser ersten Testgruppe sofort weiter gegeben. Den weiteren Versuchspersonen wurde selbstverständlich eine korrigierte Fassung des B5PO vorgelegt. Auswirkungen auf die Ergebnisse können fast gänzlich ausgeschlossen werden.

Weiters soll noch erwähnt werden, dass der B5PO grundsätzlich computerunterstützt ausgewertet wird. Da eine Testung des B5PO aus organisatorischen Gründen nur mittels einer Papier/Bleistift Version möglich war, musste auch die Auswertung mittels eigens angefertigter Schablonen erfolgen. Die Erstellung der Schablonen, wie auch die Auswertung anhand dieser, wurde mit der größtmöglichen Präzision durchgeführt. Dennoch soll hier darauf hingewiesen werden, dass Auswertungsfehler nicht mit 100%iger Sicherheit ausgeschlossen werden können.

8. Zusammenfassung

Die in der Praxis häufig angewandte Methode der Subskalen-Extraktion wurde in zahlreichen Studien untersucht. Die Ergebnisse einiger Untersuchungen zeigten, dass diese Vorgehensweise als methodischer Kunstfehler zu bewerten ist. Es konnten signifikante Veränderungen bezüglich der Mittelwerte, der Validitäts- und Reliabilitätsparameter unter der Bedingung der Subskalen-Extraktion und der daraus resultierenden Itemblockbildung (Schriesheim et al., 1989; Krampen et al., 1992; Franke & Stäcker, 1995) gefunden werden.

Im Gegensatz zu diesen Studien kommen Rost und Hoberg (1997) aufgrund ihrer Ergebnisse zu dem Schluss, dass es sich bei der Itemblockbildung um eine tolerierbare, praktische Vorgehensweise handelt. Dies bestätigen auch Sparfeldt et al. (2006) und Franke (1998). Bei der Vorgabe einer umgekehrten Original – Itemreihenfolge von Items konnte Ortner (2004) keine signifikanten Veränderungen der Mittelwerte feststellen. Franke (1996) fand beim Vergleich der originalen und der zufälligen Itemreihenfolge des SCL-90-R ebenfalls keine Veränderungen der Mittelwerte.

Die Zielsetzung dieser Diplomarbeit gilt nun der Untersuchung von Effekten die bei der Veränderung der Itemreihenfolge entstehen können, anhand des Big Five Plus One Persönlichkeitsinventar (B5PO, Holocher-Ertl, et al., 2003) und des Anstrengungsvermeidungstest (AVT, Rollet & Bartram, 1998). Eine weitere Analyse dient dazu, festzustellen ob die Reihenfolge in der die Fragebögen bearbeitet werden einen Effekt auf die Mittelwerte hat.

An der Studie nahmen insgesamt 151 Schüler/-innen aus zwei berufsbildenden höheren Schulen mit einem Durchschnittsalter von 15,03 Jahren freiwillig teil.

Die Versuchspersonen wurden vier Versuchsgruppen randomisiert zugeteilt. Versuchsgruppe I bearbeitete die Fragebögen mit ihrer originalen Itemreihenfolge. Hier wurde zuerst der AVT und anschließend der B5PO bearbeitet. Der zweiten Versuchsgruppe wurden die Fragebögen in umgekehrter Reihenfolge vorgegeben. Die Items der Fragebögen in umgekehrter Reihenfolge bearbeitete die dritte Versuchsgruppe. Ihre persönlichen Angaben machten die Versuchspersonen der

vierten Versuchsgruppe anhand einer, nach Subskalen geblockten, Itemvorgabe der beiden Fragebögen.

Die Ergebnisse einer multivariaten Varianzanalyse zeigten, dass die Itemreihenfolge keine signifikanten Effekte auf die Mittelwerte der Skalen hatte. Weder eine Bearbeitung der umgekehrten Reihenfolge der Items, noch eine Vorgabe der Items nach ihren Subskalen geordnet, führte zu signifikanten Unterschieden der Mittelwerte der Versuchspersonen. Diese Ergebnisse sind konform mit den Resultaten einer langen Reihe von wissenschaftlichen Untersuchungen (Rost & Hoberg, 1997; Sparfeldt et al., 2006; Franke, 1996, 1998; Ortner, 2004). Dennoch kann nicht ausgeschlossen werden, dass Veränderungen der Itempositionen in anderen Fragebögen zu Effekten führen. Aus diesem Grund werden weitere Untersuchungen dieser Thematik in Zukunft unumgänglich sein.

Bei der Bearbeitungsreihenfolge der Fragebögen konnten ebenfalls keine signifikanten Effekte anhand der multivariaten Varianzanalyse festgestellt werden. In der Skala Gewissenhaftigkeit konnte aufgrund einer optischen Prüfung die höchsten Mittelwertsunterschiede im Vergleich zu den anderen Skalen festgestellt werden. Die jedoch auf Grund des nicht signifikanten Ergebnisses als statistisch unbedeutsam eingestuft werden können.

Dennoch sind weitere Untersuchungen von Effekten die bei der Bearbeitung von mehreren Fragebögen entstehen könnten unumgänglich. Dies zeigen die Studien von Dalbert et al. (1991) und Lazarus-Mainka (1987) die zu signifikanten Ergebnissen führten. Der von Strack (1994) auf Itemebene festgestellter „Konsistenzeffekt“ könnte auch auf Fragebogenebene eine Rolle spielen. Wenn Personen dazu tendieren keine widersprüchlichen Antworten in einem Fragebogen zu geben (Anger, 1969), liegt die Vermutung nahe, dass sich die Personen bemühen bei der Bearbeitung von mehreren Fragebögen ebenfalls konsistente Antworten zu geben.

Ein weiterer sehr interessanter Punkt in der zukünftigen Forschung wäre die Untersuchung der Reihenfolgeeffekte im Bezug auf objektive Persönlichkeitstests. Die Untersuchung von Baldinger (2006) zu dieser Thematik, kam zu einem signifikanten niedrigeren Ergebnis in der Skala „Entschlussfreudigkeit“, wenn die Versuchspersonen als erstes Verfahren die objektiven Persönlichkeitsbatterie

„Arbeitshaltungen“ (Kubinger & Ebenhöf 1996) und anschließend den herkömmlichen Persönlichkeitsfragebogen BIP (Hossiep, Paschen & Mühlhaus, 2003) bearbeiteten.

Zusammenfassend bleibt zu sagen, dass trotz der nicht signifikanten Ergebnisse der vorliegenden Arbeit weitere Untersuchungen von Reihenfolgeeffekten notwendig sind.

9. Literaturverzeichnis

Anger, H. (1969). Befragung und Erhebung. In: Graumann, C.F. (Hrsg.), *Handbuch der Psychologie*, 7, (S.567-618). Göttingen: Hogrefe.

Baldinger, D. (2006). *Der Einfluss sozial erwünschten Verhaltens auf das Ergebnis objektiver Persönlichkeitstests*. Unveröffentlichte Diplomarbeit, Universität Wien.

Dalbert, C., Strack, K., Boesken, W. H., & Schwenkmezger P. (1991). Beeinflusst die Auseinandersetzung mit Bewältigungsverhalten das Angsterleben? *Zeitschrift für klinische Psychologie*, 20, 65-74.

Deusinger, I.M. (1986). *Die Frankfurter Selbstkonzeptskalen (FSKN)*. Göttingen: Hogrefe.

Eysenck, H.-J., Wilson C. D., & Jackson, C.J. (1998). *Eysenck Personality Profiler (EPP-D)*. Frankfurt/M.: Swets

Franke, G. H., & Stäcker, K. H. (1995) Reliabilität und Validität der Symptom-Checkliste (SCL-90-R; Derogatis, 1986) bei Standardreihenfolge und inhaltshomogener Itemblockbildung. *Diagnostica*, 41, 349-373.

Franke, G. H. (1996). Effekte der Typographie und Itempositionierung in der Fragenbogendiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 17, 187-200.

Franke, G. H. (1998). Auswirkung der PC-gestützten Anwendung des Freiburger Persönlichkeitsinventars unter besonderer Berücksichtigung der Itemreihenfolge: Eine experimentelle Studie. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 19, 187-199.

Hartig, J. & Hinrichs, J. (2002, September). *Kontexteffekte bei der Beantwortung von Persönlichkeitsfragebögen: Eine Überprüfung der Bedeutungsänderungshypothese*. Vortrag auf dem 43. Kongress der Deutschen Gesellschaft für Psychologie, Berlin.

Holocher-Ertl, S., Kubinger K.D. & Menghin S. (2005). *Big Five Plus One Persönlichkeitsinventar(B5PO)* . Test: Software und Manual. Mödling: Wiener Testsystem/ Schuhfried.

Hossiep, R. & Paschen, M. (2003, unter Mitarbeit von O. Mühlhaus). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)*. (2. Aufl.). Göttingen: Hogrefe.

Krampen, G., Hense, H. & Schneider J.F. (1992). Reliabilität und Validität von Fragenbogenskalen bei Standartreihenfolge versus inhaltshomogener Blockbildung ihrer Items. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 39, 229-248.

Kubinger K.D. & Ebenhöf, J. (1996) *Arbeitshaltungen – Manual*. Frankfurt/M.: Swets.

Lazarus-Mainka, G. (1987). Ängstlichkeit und die konnotative Bedeutung von Angst- bzw. Glückwörtern. *Psychologische Beiträge*, 29, 132-139.

Leyens, J.-P. & Codol, J.-P. (1990). Soziale Informationsverarbeitung. In W. Stroebe, M. Hewstone, J.-P. Codol & G. M. Stephenson (Hrsg.), *Sozialpsychologie. Eine Einführung*. (S. 89-111). Berlin u.a.: Springer.

Ortner, T. M. (2004). On changing the position of items in personality questionnaires: Analyzing effects of item sequence using IRT. *Psychology Science*, 46, 466–476.

Rasch, D. & Kubinger, K.D. (2006). *Statistik für das Psychologiestudium – Mit Softwareunterstützung zur Planung und Auswertung von Untersuchungen sowie zu sequentiellen Verfahren*. Heidelberg: Spectrum.

Rollett, B. & Bartram, M. (1998). *Anstrengungsvermeidungstest (AVT)*. (3., überarbeitete Auflage). Göttingen: Hogrefe.

Rost, D. H., & Sparfeldt, J. R. (2002). Facetten des schulischen Selbstkonzepts. Ein Verfahren zur Messung des differentiellen Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter) [Facets of academic self-concept. Development of a self-concept grid: Psychometric properties and some validity data]. *Diagnostica*, 48, 130-140.

Rost, D.H., & Hoberg, K. (1997). Itempositionsveränderungen in Persönlichkeitsfragebögen: Methodischer Kunstfehler oder tolerierbare Praxis? *Diagnostica*, 43, 97-112.

Schilling, S. R., Sparfeldt, J. R., Rost, D. H., & Nickels, G. (2005). Schulische Selbstkonzepte: Zur Validität einer erweiterten Version des Differentiellen Selbstkonzept Gitters (DISK-Gitter) [Facets of academic self-concept: Validity of the Differential Self-Concept Grid (DISC-Grid)]. *Diagnostica*, 51, 21–28.

Schriesheim, C. A., Kopelman, R. E., & Solomon, E. (1989). The effect of grouped versus randomized questionnaire format on scale reliability and validity : A three-study investigation. *Educational and Psychological Measurement*, 49, 487-508.

Schumann, S. (2006). *Repräsentative Umfrage: Praxisorientierte Einführung in empirische Methoden und Statistische Analyseverfahren*. (4. überarbeitete Auflage). München u.a.: Oldenbourg Wissenschafts Verlag.

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments. *Public Opinion Quarterly*, 49, 388-395.

Sparfeldt, J.R., Schilling, S.R., Rost, D.H., & Thiel, A. (2006). Blocked versus Randomized Format Questionnaires. A confirmatory multigroup Analysis. *Educational and Psychological Measurement*, 66, 961-974.

Strack, F. (1994). Zur Psychologie der standardisierten Befragung: Kognitive und kommunikative Prozesse. In: Albert D., Pawlik K., Stapf K.H. & Stroebe W. (1994). (Hrsg.), *Lehr- und Forschungstexte Psychologie*, 48. Berlin u.a.: Springer Verlag.

Taylor, S. E. (1981). The interface of cognitive and social psychology. In J. Harvey (Ed.), *Cognition, social behavior, and the environment*. (S. 189-211). Hillsdale, NJ: Erlbaum.

Trost, A. (1997). Das Antwortverhalten befragter Mitarbeiter. Eine kognitionspsychologische Perspektive. In: Bungard W. & Jöns I. (1997). (Hrsg.), *Mitarbeiterbefragung – Ein Instrument des Innovations- und Qualitätsmanagements*. (S. 38-57). Weinheim: Psychologie Verlags Union.

Wolfsbauer, Ch. (2006). *Persönlichkeitsspezifisches Lösungsverhalten bei einem Leistungstest mit speziellen Antwortmöglichkeiten*. Unveröffentlichte Diplomarbeit, Universität Wien.

9.1. Abbildungsverzeichnis

Abbildung 1: Die kognitiven Aufgaben eines Befragten	21
Abbildung 2: Kontexteffekte bezüglich der Intensitätswahrnehmung.....	28
Abbildung 3: Die Kontextabhängigkeit der Begriffsbedeutung "groß"	31
Abbildung 4: Mittelwerte der vier Versuchsgruppen in der Skala Gewissenhaftigkeit	49

9.2. Tabellenverzeichnis

Tabelle 1: Angaben bezüglich des täglichen Fernsehkonsum mit niedrigen oder hohen Antwortvorgaben (aus Schwarz et al., 1985)	25
Tabelle 2: Überblick der dargebotenen Testhefte	36
Tabelle 3: Unabhängige Variablen (UV) Testheft	36
Tabelle 4: Abhängige Variablen (AV)	36
Tabelle 5: Aufteilung der Analogskalen.....	38
Tabelle 6: Verteilung der Versuchsgruppen.....	41
Tabelle 7: Verteilung der Versuchspersonen nach dem Alter für die Versuchsgruppen.....	41
Tabelle 8: Verteilung der Versuchspersonen nach dem Alter	42
Tabelle 9: Verteilung der Versuchspersonen nach dem Geschlecht für die Versuchsgruppen	42
Tabelle 10: Verteilung der Versuchspersonen nach dem Geschlecht für die Gesamtstichprobe	43
Tabelle 11: Die Mittelwerte (Standardabweichungen) der vier Versuchsgruppen der abhängigen Variablen (Pflichteifer, Anstrengungsvermeidung, Extraversion, Offenheit, Empathie, Gewissenhaftigkeit, Verträglichkeit, Emotional Kontrolle).....	45
Tabelle 12: Multivariate Tests	45
Tabelle 13: Ergebnisse der Skala Anstrengungsvermeidung. Mittelwerte und Standardabweichungen	46
Tabelle 14: Ergebnisse der Skala Pflichteifer. Mittelwerte und Standardabweichungen.....	46
Tabelle 15: Ergebnisse der Skala Extraversion. Mittelwerte und Standardabweichungen	47
Tabelle 16: Ergebnisse der Skala Offenheit. Mittelwerte und Standardabweichungen	47
Tabelle 17: Ergebnisse der Skala Empathie. Mittelwerte und Standardabweichungen	48
Tabelle 18: Ergebnisse der Skala Gewissenhaftigkeit. Mittelwerte und Standardabweichungen	48
Tabelle 19: Ergebnisse der Skala Verträglichkeit. Mittelwerte und Standardabweichungen ...	50
Tabelle 20: Ergebnisse der Skala Emotionale Kontrolle. Mittelwerte und Standardabweichungen	50

LEBENS LAUF

ELKE EISEHAUER

Geboren	am 07. Oktober.1980 in Horn/ NÖ
Wohnhaft	in Fuglau 33, 3591 Altenburg
Eltern	Franz und Christa Eisenhauer
Geschwister	Schwester Karin

AUSBILDUNG

1987-1991	Volksschule Altenburg/ NÖ
1991-1995	Hauptschule Horn/ NÖ
1996-2000	Höhere Lehranstalt für Tourismus und Umwelt in Retz/ NÖ
Seit WS 2001	Hochschulstudium Psychologie an der Universität Wien
Jänner 2005	erste Diplomprüfung bestanden

PRAKTIKUM

Juli/ August 2005	Psychologisches Praktikum im Psychosomatikzentrum Eggenburg
-------------------	--

BERUFSLAUFBAHN

2000-2001	Angestellte im Kursbüro des Stift Geras/ NÖ
2001	Angestellte in der Stiftstaverne in Altenburg/ NÖ
2002-2003	geringfügig als Thekenkraft im Cafe-Restaurant Landtmann in Wien beschäftigt