



universität  
wien

# Diplomarbeit

Titel der Arbeit

Über den Einfluss der inhaltshomogenen  
Itemblockbildung auf die Dimensionalität eines  
Persönlichkeitsfragebogens

Verfasser

Norbert Dörrer

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juli 2009

Studienkennzahl: 298

Studienrichtung: Psychologie

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



## Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b><i>Einleitung</i></b> .....   | <b>5</b>  |
| 1.1      | <b>Problematik von Persönlichkeitsfragebogen</b> .....                     | 5         |
| 1.2      | <b>Reihenfolge von Items in Persönlichkeitsfragebogen</b> .....            | 7         |
| 1.3      | <b>Ziel der Untersuchung</b> .....   | 8         |
| 1.4      | <b>Effekte des Itemkontextes</b> .....                                     | 9         |
| 1.5      | <b>Prozesse der Beantwortung von Persönlichkeitsfragebogen-Items</b> ..... | 11        |
|          | 1.5.1 Das Modell von Rogers (1974) .....                                   | 11        |
|          | 1.5.2 Das Modell von Tourangeau und Rasinski (1988) .....                  | 12        |
| 1.6      | <b>Literatur zu Effekten des Itemkontextes</b> .....                       | 14        |
| 1.7      | <b>Zusammenfassung und Überleitung</b> .....                               | 22        |
| <br>     |  |           |
| <b>2</b> | <b><i>Methode</i></b> .....  | <b>23</b> |
| 2.1      | <b>Dichotom-logistisches Testmodell von Rasch (1960)</b> .....             | 23        |
| 2.2      | <b>Untersuchungsplan und Hypothesen</b> .....                              | 25        |
| 2.3      | <b>Untersuchungsmaterial</b> .....   | 27        |
|          | 2.3.1 Wiener Studieneignungs-Persönlichkeitsinventar (WSP) .....           | 27        |
|          | 2.3.2 Programmierung des WSP als Online-Self-Assessment .....              | 28        |
|          | 2.3.3 Unvollständiges Design, Bearbeitungsdauer und Randomisierung .....   | 30        |
|          | 2.3.4 Die Homepage „Besser-Studieren.at“ .....                             | 31        |
|          | 2.3.5 Zusätzliche Materialien: Poster, Flyer .....                         | 33        |
| 2.4      | <b>Untersuchungsteilnehmer</b> .....                                       | 34        |
|          | 2.4.1 Rekrutierung der Untersuchungsteilnehmer .....                       | 34        |
|          | 2.4.2 Deskriptive Beschreibung der Stichprobe .....                        | 35        |
| 2.5      | <b>Untersuchungsdurchführung und Computerprogramme</b> .....               | 38        |
| <br>     |  |           |
| <b>3</b> | <b><i>Ergebnisse</i></b> .....   | <b>41</b> |
| 3.1      | <b>Analyse der beiden Teilstichproben</b> .....                            | 41        |
|          | 3.1.1 Skala „Organisation“ .....   | 42        |
|          | 3.1.2 Skala „Arbeitsstil, Arbeitsverhalten“ .....                          | 44        |
|          | 3.1.3 Skala „Motivation“ .....   | 45        |
|          | 3.1.4 Skala „Belastbarkeit“ .....  | 46        |
|          | 3.1.5 Skala „Soziale Kompetenz“ .....                                      | 47        |
|          | 3.1.6 Skala „Emotionale Komponenten“ .....                                 | 48        |
|          | 3.1.7 Skala „Selbstbild“ .....   | 49        |

|   |           |
|---|-----------|
| <b>3.2 Analyse der Gesamtstichprobe .....</b>                       | <b>51</b> |
| 3.2.1 Modelltest in der Skala „Organisation“ .....                  | 52        |
| 3.2.2 Modelltest in der Skala „Arbeitsstil, Arbeitsverhalten“ ..... | 53        |
| 3.2.3 Modelltest in der Skala „Motivation“ .....                    | 54        |
| 3.2.4 Modelltest in der Skala „Belastbarkeit“ .....                 | 55        |
| 3.2.5 Modelltest in der Skala „Soziale Kompetenz“ .....             | 56        |
| 3.2.6 Modelltest in der Skala „Emotionale Komponenten“ .....        | 57        |
| 3.2.7 Modelltest in der Skala „Selbstbild“ .....                    | 58        |
| <b>3.3 Zusammenfassung der Ergebnisse .....</b>                     | <b>59</b> |
| <br>  |           |
| <b>4 Diskussion .....</b>   | <b>61</b> |
| <b>4.1 Interpretation .....</b>                                     | <b>61</b> |
| <b>4.2 Konklusion.....</b>  | <b>63</b> |
| <b>4.3 Kritik.....</b>  | <b>63</b> |
| <b>4.4 Aussicht.....</b>  | <b>65</b> |
| <br>  |           |
| <b>5 Zusammenfassung.....</b>                                       | <b>67</b> |
| <br>  |           |
| <b>6 Literaturverzeichnis.....</b>                                  | <b>69</b> |
| <br>  |           |
| <b>7 Tabellenverzeichnis .....</b>                                  | <b>75</b> |
| <br>  |           |
| <b>8 Abbildungsverzeichnis.....</b>                                 | <b>77</b> |
| <br>  |           |
| <b>9 Anhang.....</b>  | <b>79</b> |
| <b>9.1 Skalen und Subskalen des WSP.....</b>                        | <b>79</b> |
| <b>9.2 Skalenstruktur der zwei Bedingungen des WSP.....</b>         | <b>80</b> |
| <b>9.3 Instruktion des WSP.....</b>                                 | <b>81</b> |
| <b>9.4 Unvollständiges Design.....</b>                              | <b>83</b> |
| <b>9.5 Homepage Besser-Studieren.at.....</b>                        | <b>84</b> |
| <b>9.6 Abstract.....</b>  | <b>87</b> |
| <b>9.7 Lebenslauf.....</b>  | <b>89</b> |

# 1 Einleitung

## 1.1 Problematik von Persönlichkeitsfragebogen

Viele Fragestellungen der psychologischen Diagnostik machen es erforderlich, Persönlichkeitseigenschaften der zu testenden Personen genauer zu untersuchen. Die dominierenden psychologisch-diagnostischen Verfahren der Persönlichkeitsdiagnostik stellen dabei Persönlichkeits-Fragebogenbatterien dar (vgl. Kubinger, 2006). Obwohl die psychologische Forschung psychologisch-diagnostische Verfahren entwickelt hat um entsprechende Eigenschaftsausprägungen objektiv festzustellen (sog. Objektive Persönlichkeitstests), gehören Selbstauskünfte über individuelles Erleben und Verhalten mittels Persönlichkeits-Fragebogenbatterien zum Standardrepertoire der Persönlichkeitsdiagnostik.

Dabei ist dieses Vorgehen nicht unproblematisch und kann in mehrerer Hinsicht kritisiert werden (vgl. Kubinger, 2002, 2006). Besonders in Bewerbungssituationen führt die hohe Augenscheinvalidität und Durchschaubarkeit von Persönlichkeitsfragebogen zu einer bewusst verfälschten Selbstdarstellung (faking). Damit einhergehend gilt es, solche Verfahren generell hinsichtlich ihrer Zumutbarkeit zu hinterfragen. Diesbezüglich zu nennende Ansätze können das sogenannte Forced-choice-Format oder Fragen zum Privatbereich sein. Insbesondere aber (damit zusammenhängende) methodologische Probleme, wie sie Reliabilität, Validität und Skalierung betreffen sind Gegenstand zahlreicher empirischer Forschungsarbeiten.

Auch Schwarz (1999) weist darauf hin, dass Selbstauskünfte von Testpersonen eine sehr problematische Informationsquelle sind, und führt an, dass bereits geringe Änderungen in der Formulierung von Fragen, im Antwortformat oder in der Reihenfolge der Fragen bedeutsame Veränderungen im Antwortverhalten bewirken können. So stellen Schwarz, Strack und Mai (1991, zitiert nach Schwarz, 1999) fest: „Whether we conclude that marital satisfaction is a major or a minor contributor to general life-satisfaction depends on the order in

which both questions are asked, with correlations ranging from 0,18 to 0,67 as a function of question order and introduction.“ (S. 93).

Bei Fragebogenerhebungen scheinen also abgesehen von den Charakteristika der zu testenden Personen ganz besonders bestimmte Charakteristika der gestellten Fragen verantwortlich für Variationen beim Antwortverhalten zu sein. Krampen (1993) nennt einige Möglichkeiten der Beeinflussung des Beantwortungsprozesses: „Neben optisch-gestalterischen Aspekten, die das Layout und die Lesbarkeit betreffen (vgl. etwa Champion & Sear, 1968/69), den sprachlichen Formulierungen (vgl. etwa Löhr & Angleitner, 1980), der Polung der Items und den vorangegangenen Antwortalternativen oder -skalen (vgl. etwa Wegener, Faulbaum & Maag, 1982) ist hier vor allem auch die bislang in der Forschung weitgehend vernachlässigte Frage nach der Anordnung der Items zu nennen.“ (S. 98).

Becker (2003) führt vier Arten von Fehlerquellen bei der Diagnostik mittels Persönlichkeitsfragebogen an, nämlich Fehlerquellen im Medium Sprache (wie die Mehrdeutigkeit sprachlicher Ausdrücke), Fehlerquellen auf Seiten der Testperson (wie eine zu geringe Motivation), Fehlerquellen in der Untersuchungssituationen (wie die bereits erwähnten Charakteristika von Bewerbungssituationen) und schließlich Fehlerquellen im Fragebogen selbst (wie mangelhafte Verrechnungsfairness aufgrund ungenügender psychometrischer Eigenschaften des Fragebogens).

In der vorliegenden Diplomarbeit möchte ich mich in Anbetracht obiger Feststellungen mit einer speziellen Fehlerquelle im Fragebogen selbst beschäftigen, nämlich dem Itemkontext. Nach Tourangeau und Rasinski (1988) bildet sämtliches Material, das mit den Items eines bestimmten Fragebogens in Zusammenhang steht (Instruktion, Beispielitems, bisherige Items usw.), den diesbezüglichen Kontext im Fragebogen. Aber speziell die einem bestimmten Item vorangehenden Items werden in der Fachliteratur mit dem Terminus „Itemkontext“ bezeichnet. Insofern wird offensichtlich der Itemkontext durch die Reihenfolge der Items eines Persönlichkeitsfragebogens determiniert.

## 1.2 Reihenfolge von Items in Persönlichkeitsfragebogen

Üblicherweise werden die Items (multidimensionaler) Persönlichkeitsfragebogen in zufälliger beziehungsweise gleichverteilter Reihenfolge angeordnet. Mit gleichverteilter Reihenfolge ist hier gemeint, dass sich die den verschiedenen Skalen inhaltlich zugeordneten Items bei der Fragebogenbearbeitung derart abwechseln, dass die inhaltliche Distanz nacheinander präsentierter Items maximal wird. Besonders bei der Konstruktion multidimensionaler Persönlichkeits-Fragebogenbatterien ist die Gleichverteilung der Items einer Skala über den gesamten Fragebogen hinweg sehr häufig.

Diese Praxis soll die Intransparenz der zu messen beabsichtigten Eigenschaften gewährleisten (und also die Augenscheinvalidität verringern), sodass die zu testenden Personen ihre Ergebnisse nicht beziehungsweise schwerer in eine bestimmte Richtung beeinflussen können (Sparfeldt, Schilling, Rost & Thiel, 2006). Diese Beeinflussung kann bewusst oder unbewusst erfolgen und beispielsweise dazu führen sich anderen Personen gegenüber als besonders gut darzustellen (faking good und soziale Erwünschtheit). Auch das Gegenteil ist möglich, wenn sich Personen absichtlich besonders schlecht darstellen (faking bad [vgl. Franke, 2002])

Anger (1969, zitiert nach Ortner, 2004) erwähnt in diesem Zusammenhang auch einen Löschungseffekt als positive Auswirkung der maximalen inhaltlichen Distanz, insofern nämlich Fragen mit neuem Inhalt Effekte wie Priming minimieren können. Insgesamt wird daher in manchen einschlägigen Lehrbüchern zur Test- bzw. Fragebogenkonstruktion aus obigen Gründen eher eine gleichverteilte Itemreihenfolge bei multidimensionalen Persönlichkeitsfragebogen empfohlen (vgl. etwa Rost, 2004, Moosbrugger & Kelava, 2007 oder Mummendey, 2007)

Eine diesem Ansatz gegenüberstehende Vorgehensweise bei der Konstruktion von Persönlichkeitsfragebogen stellt die Zusammenfassung der zu einer Skala gehörenden Items zu inhaltshomogenen Blöcken dar, wodurch die inhaltliche Nähe nacheinander präsentierter Items maximal wird. Diese

Itemreihenfolge wird zum Beispiel dann realisiert, wenn einzelne Skalen aus multidimensionalen Persönlichkeitsfragebogen extrahiert werden, um sie separat an zu testenden Personen vorzugeben. Aber auch theoretische Überlegungen führen gelegentlich zur inhaltshomogenen Itemblockbildung, da diese nach Meinung mancher Autoren günstigere Auswirkungen auf die Motivation der Testpersonen und damit die Eindimensionalität des Antwortprozesses hätte (vgl. Schriesheim und DeNisi, 1980).

### **1.3 Ziel der Untersuchung**

In der einschlägigen Fachliteratur werden Vor- und Nachteile dieser beiden Vorgehensweisen bei der Fragebogenkonstruktion diskutiert. Entsprechende empirische Untersuchungen werden im Anschluss vorgestellt. Das Ziel dieser Diplomarbeit war es im Zuge einer empirischen Studie mit experimentellem Design die Auswirkungen verschiedener Itemreihenfolgen auf die testtheoretische Qualität eines (a-priori dimensionierten, multidimensionalen) Persönlichkeitsfragebogens zu untersuchen.

Dabei wurden in Anlehnung an die Praxis der Fragebogenkonstruktion zwei Bedingungen der Itemreihenfolge realisiert: einerseits die maximale inhaltliche Distanz der Items und andererseits die maximale inhaltlicher Nähe der Items. Zur Diskussion standen die Auswirkungen des so beeinflussten Itemkontextes auf die Dimensionalität der Skalen dieses psychologisch-diagnostischen Verfahrens im Sinne der Konformität der Items des Persönlichkeitsfragebogens mit dem dichotom-logistischen Testmodell von Rasch (1960).

Gleichzeitig sollte der im Zuge dieser Diplomarbeit erhobene Datensatz auch der abschließenden Testkonstruktion (im Sinne von Skalierung und Normierung) des hier herangezogenen Persönlichkeitsfragebogens dienen. Die durchgeführten Modelltests sollten also abgesehen von der Forschungsfrage im obigen Sinne, Aufschluss über die Verrechnungsfairness dieses psychologisch-diagnostischen Verfahrens geben.

## 1.4 Effekte des Itemkontextes

Phänomene, die mit der Itemreihenfolge in Zusammenhang stehen, werden in der Literatur „Itemkontext-Effekte“ oder einfach nur „Kontexteffekte“ beziehungsweise nach Rost und Hoberg (1997) auch „Reihenfolgeeffekte“, „Ausstrahlungseffekte“, „Platzierungseffekte“ oder „Sequenzeffekte“ genannt. Die Beschäftigung mit diesem Thema scheint deswegen notwendig, da wiederholt Einflüsse des Kontextes auf teststatistische Parameter (wie innere Konsistenz, diskriminante beziehungsweise konvergente Validität, die Faktorenstruktur oder die Konformität mit dem Rasch-Modell) von Persönlichkeitsfragebogen festgestellt werden konnten.

Ortner (2004) führt einige Erklärungsansätze für das in Abhängigkeit von der Itemreihenfolge veränderte Antwortverhalten an. Eine Ursache könnte sein, dass die zu testenden Personen zu Beginn der Befragung ihre Aufmerksamkeit verstärkt auf deren Selbst fokussieren können. Dahingegen könnte die Müdigkeit der zu testenden Personen gegen Ende der Befragung diesen Effekt minimieren und zu bestimmten Antworttendenzen führen, wie einem rigiden Antwortmuster.

Ortner (2004) verweist auch auf mögliche Halo-Effekte und Effekte des Primings, die entweder durch die Untersuchungssituation begünstigt werden können oder durch die Präsentation bestimmter Items vor anderen. Auch Jonkisz, Moosbrugger und Brandt (2007) machen in ihrem Beitrag auf sogenannte „Aktualisierungseffekte“ aufmerksam. Gemeint ist damit die Aktivierung von Kognitionen durch Items, welche die Interpretation nachfolgender Items beeinflussen können.

Krahé und Hermann (2003) untersuchten in diesem Sinne in einer Studie mit experimentellem Design die Beeinflussung der Urteilsbildung in der Beantwortung des NEO-FFI (Borkenau & Ostendorf, 1993) durch situative Manipulationen, die die gezielte Verfügbarkeit kognitiver Konzepte (hier der Dimension „Gewissenhaftigkeit“) erhöhen. Die Ergebnisse deuten für die Autoren in die Richtung, „dass die gedankliche Beschäftigung mit dem Persönlichkeitsprofil einer als sehr gewissenhaft beschriebenen Person die

Gewissenhaftigkeitsscores der Probanden gegenüber einer Kontrollbedingung erhöhen.“ (S. 113).

Aber nicht nur die Verfügbarkeit bestimmter kognitiver Konzepte könne nach Ortner (2004) die Beurteilungen der zu testenden Personen beeinflussen, sondern auch andere Charakteristika der vorgegebenen Items. Besonders wenn sehr „schwierige“ Items (bei denen die Wahrscheinlichkeit zuzustimmen nur sehr gering ist) gleich zu Beginn der Befragung präsentiert werden, kann das bei den zu testenden Personen zu der Erwartungshaltung führen, dass die Fragen generell sehr extrem formuliert sind beziehungsweise überhaupt unzumutbar sind. Diese Einschätzung könne nach Ortner (2004) zu einer Beeinflussung des Antwortverhaltens führen.

Um Effekte im Zusammenhang mit dem Itemkontext besser verstehen zu können, erweist es sich also als sinnvoll, zunächst jene psychologischen Prozesse genauer zu betrachten, die der Beantwortung von Persönlichkeitsfragebogen-Items zugrunde liegen. Die Grundidee der Klassischen Testtheorie ist, dass die Antworten einer Person auf Persönlichkeitsfragebogen-Items deren wahre Eigenschaftsausprägung, verfälscht durch einen Messfehler widerspiegeln. Aber was genau geschieht psychologisch, wenn Personen Items von Persönlichkeitsfragebogen bearbeiten? Entsprechende Erkenntnisse könnten dazu beitragen, den Einfluss bestimmter Charakteristika der gestellten Fragen auf das Antwortverhalten der Personen zu erklären.

Daher sind im Folgenden zwei psychologische Modelle dargestellt, die zwar selbst nicht mehr zur aktuellen Fachliteratur zählen, aber in dieser an mehreren Stellen zitiert werden. Da sie wesentlich zum Verständnis von Itemkontexteffekten beitragen können, erscheint mir die Auseinandersetzung damit an dieser Stelle als sinnvoll.

## **1.5 Prozesse der Beantwortung von Persönlichkeitsfragebogen-Items**

### **1.5.1 Das Modell von Rogers (1974)**

Nach Rogers (1974) löst die Konfrontation der Person mit einem Persönlichkeitsfragebogen einen Prozess aus, der durch die Interaktion zwischen Charakteristika der einzelnen Fragen und dem Selbstkonzept der Person bestimmt wird. Rogers (1974) beschreibt diesen Prozess folgendermaßen: „The time required for a ‘True’ or ‘False’ response to a personality item was thought to be divided into stages representing: (1) the translation of the item into cognitively functional units (Stimulus encoding), (2) the transformation of these units into useful working internal language (Stimulus comprehension) and (3) the use of this internal language to decide upon the appropriate response (Binary Decision).” (S. 129).

Hinsichtlich des Problems von Kontexteffekten ist besonders auf den dynamischen Charakter dieses Modells hinzuweisen. Schon Rogers (1974) stellt fest, dass Charakteristika der präsentierten Items und der Untersuchungssituation einen Einfluss darauf haben können, welche Elemente des Selbstkonzeptes als Vergleichsgröße herangezogen werden. Die Instruktion eines Persönlichkeitsfragebogens löst beispielsweise dann die Tendenz sich gut zu präsentieren aus, wenn die Testperson dadurch eher jene Teile des Selbstbildes heranzieht, von denen sie glaubt, dass sie von ihr erwartet werden. Entscheidend sind also die Erwartungen hinsichtlich der Untersuchungsintention, welche die Person auf Basis oben genannter Charakteristika bildet.

Bei der Beantwortung von ähnlichen Items kann es somit durch die Heranziehung unterschiedlicher Vergleichsgrößen zu Variationen kommen. Knowles (1988) weist mit Bezug auf Rogers (1974) ebenfalls darauf hin, dass der Prozess der Fragenbeantwortung nicht inert ist: „Answering one item leaves a trace on the content of the stored self-concept, on the ease of processing subsequent information, and on the answers to subsequent items.“ (S. 312).

### 1.5.2 Das Modell von Tourangeau und Rasinski (1988)

Auch Tourangeau und Rasinski (1988), die sich in ihrer Arbeit speziell mit Verfahren zur Messung von Einstellungen beschäftigen, postulieren für die Beantwortung entsprechender Fragen (z.B. zum Thema „Schwangerschaftsabbruch“) einen psychologischen Prozess. In ihrer Arbeit stellen die Autoren aber nicht nur ein Modell des Beantwortungsprozesses vor, sondern zeigen mögliche Einflüsse von Kontexteffekten auf jede ihrer postulierten Stufen auf. Den Itemkontext definieren sie als „the preceding items in the questionnaire“ (S. 301).

Der Prozess beginnt mit der Interpretation der gestellten Frage (erste Stufe). Die Personen müssen sich auf dieser Stufe damit auseinandersetzen, auf welche Eigenschaft das Item abzielt und ob sie überhaupt über die relevante Persönlichkeitsstruktur verfügen. Wie Rogers (1974) argumentieren auch Tourangeau und Rasinski (1988), dass bereits bearbeitete Items einen Bezugsrahmen darstellen, welcher die Basis für die Interpretation der Inhalte kommender Items bildet. So war die Zustimmung der Probanden und Probandinnen zu einem bestimmten Gesetzesentwurf stärker, wenn diesem Item ein Block von Items zum Thema „Inflation“ voranging. Der Effekt verschwand dementsprechend, wenn diese Items nicht zuvor geblockt präsentiert wurden.

Nachdem die Frage interpretiert wurde, müssen im Sinne der Urteilsbildung entsprechende Informationen aus dem Gedächtnis beziehungsweise dem Selbstkonzept abgerufen werden (zweite Stufe). Welche Informationen dabei zur Urteilsbildung herangezogen werden, hängt auch von den Eigenschaften der gestellten Fragen ab. Durch die Konfrontation mit dem interpretierten Iteminhalt ruft die Person nämlich eher Informationen aus dem Gedächtnis ab, die mit diesem Aspekt in Zusammenhang stehen (priming). Beispielsweise wurde von den Probandinnen und Probanden die Frage zum persönlichen Spendenverhalten weniger oft zustimmend beantwortet, wenn direkt davor ein inhaltshomogener Itemblock zur Verantwortung des Staates für bestimmte soziale Services wie Spitäler und Infrastruktur präsentiert wurde. Die Itemblockbildung hätte nach Meinung der Autoren zu einem kognitiven Priming hinsichtlich der Verantwortung des Staates bei der sozialen Fürsorge geführt.

Nachdem bestimmte Informationen des Selbstkonzeptes abgerufen wurden, werden diese auf der dritten Stufe im Sinne der Frage beurteilt. Die Autoren zitieren in diesem Zusammenhang eine Theorie von Anderson (1974, 1981), nach welcher die einzelnen Dimensionen eines Urteils nach bestimmten Regeln zu einem Gesamturteil integriert werden. Auf dieser dritten Stufe könnte der Kontext der Items als Normstandard zur Beurteilung eines aktuellen Items herangezogen werden. So gesehen können bereits bearbeiteten Items als ein Anker für spätere Beurteilungen dienen.

In der abschließenden Phase ihres Modells müssen die zu testenden Personen ihr Urteil schließlich abgeben. Zwei Komponenten sehen Tourangeau und Rasinski (1988) in dieser Phase als relevant an, nämlich die Projizierung der Beurteilung auf das vorhandene Antwortformat einerseits und die kognitive Nachbearbeitung der gewählten Antwortalternative andererseits. Vor allem diese Editierung der Antwort kann durch den Kontext ausgelöst werden. Die bisher gewählten Antworten können beispielsweise bei der Person einen Druck erzeugen, stimmige Antworten auf Items zu geben, von denen sie vermuten, dass sie das gleiche Merkmal erfassen (nach Jonkisz et al., 2007 auch „Konsistenzeffekt“ genannt). Beziehungsweise kann die Person allzu extreme Gesamtergebnisse in dieser Phase verhindern, indem sie auf spätere Items inkonsistente Antworten gibt (oder die Mittelkategorie wählt).

## 1.6 Literatur zu Effekten des Itemkontextes

Die Bildung von inhaltshomogenen Itemblocks stellt eine Besonderheit der Itemreihenfolge bei multidimensionalen Persönlichkeits-Fragebogenbatterien dar. Schon Schriesheim und DeNisi (1980) weisen darauf hin, dass sich zu testende Personen gelegentlich über jene Items „beschweren“, die sich im Fragebogen zu wiederholen scheinen. Das Entdecken solcher lediglich unterschiedlich formulierten Items (gleichen Inhalts), könne bei den zu testenden Personen aber den Eindruck erwecken, von den Testleitern kontrolliert zu werden. Die Zusammenfassung dieser zu einer Skala gehörenden Items zu inhaltshomogenen Blöcken würde nach Schriesheim und DeNisi (1980) beziehungsweise Schriesheim, Kopelman und Solomon (1989) die zu testenden Personen davon überzeugen können, dass es sich bei ähnlichen Items nicht um einen Trick der Testleiter handelt. So könnten Vertrauen und Motivation (als den Voraussetzungen für valide Antworten) gesteigert werden. Zusätzlich würde eine solche offensichtliche Unterteilung des Fragebogens in Blöcke (mit zusätzlicher Beschriftung dieser Blöcke nach ihrer zugrundeliegenden latenten Dimensionen) die Monotonie bei der Bearbeitung des Fragebogens durchbrechen, was sich günstig auf die Aufmerksamkeit und Motivation der zu testenden Personen auswirken würde.

In einer ersten empirischen Studie untersuchten Schriesheim und DeNisi (1980) die Auswirkungen eines solchen Vorgehens auf die Validität eines Fragebogens zum Führungsverhalten, und stellten eine verringerte konvergente und diskriminante Validität (im Sinne von Campbell & Fiske, 1959) unter der Versuchsbedingung mit inhaltshomogenen Itemblocks fest. Diese ungünstigen Auswirkungen führten die Autoren darauf zurück, dass Items bei der geblockten Versuchsbedingung in der Tendenz milder beurteilt wurden. Ihre Ergebnisse interpretieren die Autoren so, dass die zu testenden Personen aufgrund der vorgefundenen Skalendefinition vorschnell ein Urteil über die Dimension gebildet hätten, ohne sich genauer mit den Items auseinanderzusetzen. Die erhöhte Augenscheinvalidität könnte so den ungünstigen Effekt nach sich ziehen, dass der

Prozess der Itembearbeitung durch den Einsatz impliziter Theorien über die gefragten Dimensionen abkürzt werden würde.

In einer Reihe von Studien replizierten Schriesheim et al. (1989) ihre Ergebnisse und mussten schließlich erneut feststellen, dass ein geblocktes Fragebogenformat keinen Nutzen zu haben scheint: „Grouping items does not substantially and consistently improve any psychometric property and may, in some instances, impair measurement quality.“ (S. 506). Dort konnten die Autoren hinsichtlich der teststatistischen Parameter der internen Konsistenz, der Retest-Reliabilität und der diskriminanten und konvergenten Validität keinen Unterschied zur Standardreihenfolge des „Job Diagnostic Survey“ (Hackman & Oldham, 1975) und eines Lebenszufriedenheits-Fragebogens feststellen.

Aber nicht nur theoretische Überlegungen führen in der Praxis zur inhaltshomogenen Itemblockbildung. Krampen, Hense und Schneider (1992) beklagen in ihrer Arbeit die Fragmentierung differential- und persönlichkeitspsychologischer Ansätze in der Forschung durch eine zunehmende Konzentration auf lediglich einige wenige „moderne“ Konstrukte. Auch die ihrer Meinung nach in verschiedenen Forschungs- und Anwendungsbereichen der Psychologie nahezu üblich gewordene Praxis, „Subskalen aus umfassenden Persönlichkeitsfragebogen zu extrahieren und die Items dieser Subskalen dann als mehr oder weniger inhaltshomogene Itemblocks isoliert zur Beantwortung vorzugeben.“ (S. 229), sehen sie im Kontext dieser Fragmentierung. Zur Rechtfertigung solcher Subskalenextraktionen bei entsprechenden psychologisch-diagnostischen beziehungsweise empirischen Untersuchungen werden ökonomische Gründe angeführt, da eben nur jenes in Frage stehende Persönlichkeitskonstrukt interessieren würde.

Unter dem Rückgriff auf ökonomische Begründungen würden nach Krampen et al. (1992) allerdings kaum die methodischen Aspekte eines solchen Vorgehens reflektiert werden. So vermuten die Autoren, dass die inhaltliche Homogenisierung nicht nur Tendenzen zur sozialen Erwünschtheit verstärken würde, sondern auch, dass die häufige Wiederholung ähnlicher Frageninhalte und die thematische Perserveration bei den zu testenden Personen zu Reaktanz und

folglich Widerspruch führen würde. Gleichzeitig könnte die inhaltliche Homogenisierung auch zu einer Homogenisierung der Antworten führen und damit zu einer Überschätzung der internen Konsistenz beitragen (Krampen et al., 1992).

In zwei darauf aufbauenden experimentellen Studien konnten sie einige Effekte der Itemblockbildung (welche nach Krampen et al. [1992] der angesprochenen Extraktion von Subskalen aus umfassenden Persönlichkeitsfragebogen entspricht) auf teststatistische Parameter des FPI-A (Halbform A des Freiburger Persönlichkeitsinventares von Fahrenberg, Hampel & Selg, 1984) feststellen. Diese Effekte folgten aber kaum einer allgemeinen Richtung. Zudem kommen die Autoren in den beiden Studien zu inkonsistenten Ergebnissen, wie der Vergleich der Skalenhomogenitäten zeigt. Lediglich bei zwei von den neun Skalen des FPI-A ergaben sich unter der Blockbildung in beiden Studien übereinstimmend geringere interne Konsistenzen. Bei anderen zwei Skalen ergab sich kein statistisch bedeutsamer Unterschied der  $\alpha$ -Koeffizienten. Und in den übrigen fünf Skalen zeigten sich zwar statistisch signifikante, allerdings gegenläufige Veränderungen der internen Konsistenzen. Aufgrund dieser Befunde konnten die Autoren weder die Hypothese einer durchgängigen Homogenisierung des Antwortverhaltens noch die der durchgängigen Reaktanzerhöhung bestätigen. In einer weiteren experimentellen Studie konnte Krampen (1993) dieses Ergebnis bei einem anderen psychologisch-diagnostischen Verfahren, nämlich dem „Fragebogen zu Kompetenz- und Kontrollüberzeugungen“ (FKK; Krampen, 1991) replizieren. Dort waren die Effekte allerdings auch wieder inkonsistent.

Interessant erweist sich die Übereinstimmung der Auswertung von einigen Zusatzfragen bei beiden Arbeiten. Diese Zusatzfragen bezogen sich auf begleitende kognitive und motivationale Prozesse bei der Fragebogenbeantwortung der homogenen Itemblocks: „Die geschätzte Bearbeitungszeit für den FPI-A nimmt bei der Itemblockbildung in der Tendenz zu, die Langeweile bei der Fragebogenbearbeitung wird als höher beurteilt, und es treten eher Entscheidungsprobleme im Antwortverhalten sowie Überlegungen zur Messabsicht bei der Fragebogenbearbeitung auf.“ (Krampen et al., 1992, S. 244).

Zusammenfassend beurteilen die Autoren die diskutierte Subskalenextraktion als einen methodischen Kunstfehler, durch welchen dem Fragebogen Gewalt angetan werden würde und schlagen anstatt dieser für sie dadurch obsolet gewordenen Möglichkeit der ökonomischeren Gestaltung des psychologisch-diagnostischen Prozesses vor, entweder auf spezifische Persönlichkeitsfragebogen auszuweichen (optimale Lösung) oder die Skalen durch Itemselektionen (etwa nach Trennschärfekoeffizienten, Schwierigkeitsindizes oder Faktorladungen) zu verkürzen (suboptimale Lösung).

In sehr enger Anlehnung an diese Untersuchungen konnten Franke und Stäcker (1995) die Ergebnisse von Krampen (1993) beziehungsweise Krampen et al. (1992) teilweise replizieren (hier für die „Symptom-Check-Liste“ [SCL-90-R] von Derogatis, 1986). Die Autoren realisierten zur Überprüfung des von ihnen als „Krampen-Paradigma“ (S. 354) bezeichneten Phänomens (verändertes Antwortverhalten in Abhängigkeit von der inhaltshomogenen Itemblockung) bei zwei Studien unterschiedliche Versuchsbedingungen mit inhaltshomogener Itemblockbildung, nämlich durch die von Derogatis (1986) selbst vorgelegte Skalenstruktur einerseits (Version B) und durch eine Blockbildung nach vorheriger explorativer Faktorenanalyse andererseits (Version C). Ziel ihrer Untersuchungen war demnach die Überprüfung der SCL-90-R auf Äquivalenz mit diesen zwei Fragebogenversionen. Tatsächlich finden die Autoren signifikante Unterschiede im Antwortverhalten zwischen den drei Versionen (Standardreihenfolge A, Version B und Version C): „Die nach eigener Analyse geblockte Darbietung der Items führt somit zu z.T. signifikant höherem Antwortverhalten und damit u.U. psychologisch gesehen zu größerer Zustimmung; die Darbietung der Items nach den Skalen des Originalautors führt zu z.T. signifikant niedrigerem Antwortverhalten und damit psychologisch betrachtet u.U. zu größerer Ablehnung.“ (S. 363). Wie in diesem Zitat angedeutet, kommen die Autoren zu ähnlich inkonsistenten Ergebnissen wie bereits Krampen et al. (1992) zuvor, indem nämlich keine der signifikanten Mittelwertsdifferenzen der einen Studie mit dem Mittelwerts muster der anderen Studie übereinstimmten. Dieses inkonsistente Befundmuster ergab sich auch beim Vergleich der internen Konsistenzkoeffizienten.

Rost und Hoberg (1997) sprechen daher in Anbetracht der Ergebnisse von Krampen et al. (1992) und Franke und Stäcker (1995) von einem Ergebnischaos und stellen fest, „dass zwar wiederholt Kontexteffekte durch Blockbildung bei der Beantwortung von Fragebogen belegt worden sind, diese aber kaum einer einheitlichen Tendenz zu folgen scheinen.“ (S. 101). Die Autoren widmen diesem Problem daher eine weitere empirische Studie und wollen dem Einfluss von unterschiedlichen Itemreihenfolgen auf Konstruktvalidität und psychometrische Kennwerte (wie Homogenitätskoeffizienten, Mittelwerte und Streuungen) von Persönlichkeitsskalen am Beispiel des Selbstkonzeptes schulischer Leistungen und Fähigkeiten nachgehen. Im Gegensatz zu Franke und Stäcker (1995) und Krampen et al. (1992) beeinflusste die Blockbildung in ihrer Arbeit die Konstruktvalidität (hier im Sinne der faktoriellen Struktur) nicht: „Die Skaleninterkorrelationen und, daraus folgend, die Struktur der beiden Fragebogenversionen unterscheiden sich nicht.“ (S. 109). Auch die Skalenstreuung wurde durch die unterschiedliche Itemanordnung nicht beeinflusst. Bei zwei Skalen zeigten sich signifikante Effekte der Itemblockung. So führte die Blockbildung beim Selbstkonzept in Deutsch mit einer Differenz von -0,23 und beim Selbstkonzept in Erdkunde mit einer Differenz von -0,33 zu statistisch signifikant höheren Skalenmittelwerten. Außerdem kam es zu einer geringen Erhöhung der Homogenitäten. Letztere Effekte sind aber nach Rost und Hoberg (1997) von geringer Relevanz für die diagnostische Praxis (Anhebung von  $\alpha = 0,86$  auf  $\alpha = 0,91$  beziehungsweise von  $\alpha = 0,9$  auf  $\alpha = 0,93$ ), so dass sie abschließend hinsichtlich der kritisierten Subskalenextraktionen wenig Anlass zu besonderer Besorgnis sehen.

Sparfeldt et al. (2006) üben in ihrer Arbeit Kritik an bisherigen Untersuchungen zum Reihenfolgeeffekt. Besonders an der Tatsache, dass selten multivariate Analyseverfahren (hier im Sinne der Konfirmatorischen Faktorenanalyse) herangezogen würden, nehmen sie Anstoß. Gleichzeitig würden in empirischen Studien zu Effekten der Itemblockbildung sehr unterschiedliche Formen dieser inhaltshomogenen Itemblockung realisiert werden. Als extreme Form einer Zusammenfassung von zusammengehörigen Dimensionen in Blöcke präsentieren sie in ihrer Arbeit die Gestaltung eines Fragebogens zur Erfassung

schulischer Selbstkonzepte (hier von Mathematik, Physik, Deutsch und Englisch mit Hilfe des „Differentiellen Selbstkonzept Gitters“ [DISK] von Rost & Sparfeldt, 2002) als einem „Gitter“ beziehungsweise Raster (grid), bei welchem jeder einzelnen Selbstkonzeptskala eine eigene Spalte vorbehalten ist. Die für alle Skalen gleich formulierten Items sind dabei in die Zeilen des Gitters innerhalb einer eigenen Spalte eingetragen (z.B.: „I feel good about my course work in [...]“). Hinsichtlich des Einflusses des Kontextes vermuten die Autoren: „If different academic self-concept scales are presented in a grid, this could, for example, affect the internal comparison process and result in different path coefficients. So far, there is no information about the effects of such extreme item blocking.“ (S. 964).

In einer empirischen Studie mit experimentellem Design gaben Sparfeldt et al. (2006) den Personen der Versuchsgruppe den Fragebogen in der oben beschriebenen Form vor, während die Items des DISK in der Kontrollgruppe konventionell in zufälliger Reihenfolge als Liste angeordnet wurden. Obwohl sie aus theoretischen Überlegungen heraus bei der psychologischen Diagnostik des Selbstkonzeptes mit Einflüssen des Itemkontextes auf das Antwortverhalten gerechnet hätten, konnten sie keine wesentlichen Unterschiede hinsichtlich der faktoriellen Struktur, der Skalenmittelwerte und sonstiger psychometrischer Parameter feststellen. In beiden Gruppen konnte die postulierte, vier Skalen umfassende Struktur bestätigt werden, wobei sich die faktorielle Struktur beider Gruppen sehr ähnlich war. Zudem erschienen selbst die größten Mittelwertsdifferenzen den Autoren aufgrund der geringen Effektgröße vernachlässigbar. Obwohl von diesen Blickwinkeln aus betrachtet keines der untersuchten Fragebogenformate überlegen zu sein scheint, stellten Sparfeldt et al. (2006) Vorteile des Grid-Formates in einer anderen Hinsicht fest. So benötigten erstens die Personen der Versuchsgruppe für die Bearbeitung des gesamten Fragebogengitters nur halb so viel Zeit und zweitens empfanden die Personen der Kontrollgruppe die Bearbeitung der Fragebogenliste als langweiliger. „Hence, using a grid presentation format might be suggested to raise participants compliance.“ (S. 972).

Die Extraktion ganzer Skalen multidimensionaler Persönlichkeitsfragebogen aus ökonomischen Überlegungen und die Itemblockung aus theoretischen Überlegungen heraus stellen aber nicht die einzigen Möglichkeiten dar, wie in der psychologischen Diagnostik auf die Reihenfolge der Items psychologisch-diagnostischer Verfahren Einfluss genommen wird. Ortner (2004) führt eine weitere Möglichkeit an. Durch computerisiertes adaptives Testen nämlich wären unterschiedliche Itemreihenfolgen bei den zu testenden Personen unumgänglich, da jeder einzelnen Person nur jene Items aus dem Itempool vorgegeben werden, die der individuellen Fähigkeit adäquat sind und somit maximalen Informationsgewinn versprechen (Kubinger, 2006). Untersuchungen dazu, wie sich solche „zufälligen“ Änderungen der Itemreihenfolge auf testtheoretische Kennwerte von multidimensionalen Persönlichkeitsfragebogen auswirken, wären daher von großem Interesse.

Ortner (2004) führte eine entsprechende Untersuchung durch, insbesondere mit dem Ziel, die Itemhomogenität von zwei sich nur in der Itemreihenfolge unterscheidenden Versionen eines Persönlichkeitsfragebogens (hier einer modifizierten Fassung des „Eysenck Personality Profiler“ [EPP-D] von Eysenck, Wilson & Jackson, 1998) anhand eines Modells der sogenannten „Item Response Theory“ (IRT; vgl. Kubinger, 1989 und Rost, 2004) zu untersuchen. Die Notwendigkeit der Heranziehung der IRT (anstatt der Klassischen Testtheorie [vgl. wieder Rost, 2004]) ergibt sich in diesem Zusammenhang aus einigen wesentlichen testtheoretischen Grundannahmen. So wird üblicherweise die Summe der gelösten Items als das Testergebnis für die Eigenschaftsausprägung (beziehungsweise Leistungsfähigkeit) einer zu testenden Person herangezogen. Damit aber dieser Verrechnungsmodus fair ist, muss zunächst der Nachweis erbracht werden, dass für den konkreten Itempool das dichotome-logistische Testmodell von Rasch (1960) – auch einfach nur Rasch-Modell genannt – gilt (z.B. Kubinger, 2006). Zudem muss gesichert sein, dass die Lösung eines Items (abgesehen vom Zufall) nur von der Fähigkeit der zu testenden Person einerseits und der Schwierigkeit des Items andererseits abhängt, nicht aber davon, welche anderen Items die Person schon gelöst hat (z.B. Kubinger, 2006). Auch diese Voraussetzung der „lokalen stochastischen Unabhängigkeit“ als einer der

prüfbareren Implikationen des Rasch-Modells, macht für Ortner (2004) den Rückgriff auf die IRT notwendig.

Im Sinne obiger Frage untersuchte die Autorin in ihrer experimentellen Studie die Auswirkung der Itemreihenfolge auf die Schätzungen der Itemparameter der Items des EPP-D. Während die Kontrollgruppe dabei den Fragebogen in Standardreihenfolge bearbeiten musste, wurde der Versuchsgruppe der EPP-D mit genau umgekehrter Itemreihenfolge präsentiert. Bei drei der sieben Skalen unterschieden sich die im Teilungskriterium *Testversion* mit Hilfe des Likelihood-Ratio-Test (Anderson, 1973) geschätzten Itemschwierigkeitsparameter voneinander (allerdings nicht im Teilungskriterium *Score*). Bei zweien dieser Skalen konnte durch anschließenden Itemausschluss noch eine Passung mit dem Rasch-Modell erreicht werden („hypochondrial“ und „depressed“), bei der dritten Skala allerdings nicht („dogmatic“). Graphische Modellkontrollen bestätigten schließlich diese Befunde, sodass für diese drei Skalen durch die Umkehrung der Itemreihenfolge die Gültigkeit des Rasch-Modells nicht mehr vorausgesetzt werden könne. Folglich stellt die Autorin fest: „... that a simple change to the item order leads to changes in item difficulties in three of seven scales. In this case, items in different positions of a test showed a change in their item difficulty.“ (S. 475).

Zusätzlich evaluierte Ortner (2004) die Äquivalenz beider Versionen durch einen Vergleich der Scoremittelwerte. Dabei konnte sie aber keine signifikanten Mittelwertsunterschiede feststellen, so dass ihr solche Analysen für die Kontrolle metrischer Probleme infolge des Itemkontextes bei dieser Untersuchung ungeeignet erschienen. Hinsichtlich des computerisierten adaptiven Testens von Persönlichkeitseigenschaften stellt die Autorin schließlich fest: „A fair measurement of a person’s tendency in answering different items does not seem reasonable at this point in time.“ (S. 475).

## 1.7 Zusammenfassung und Überleitung

Es gibt verschiedene Möglichkeiten, die Items von Persönlichkeitsfragebogen vorzugeben: zufällig, gleichverteilt, adaptiv oder inhaltlich geblockt. Bei entsprechenden psychologisch-diagnostischen Verfahren wird hauptsächlich eine zufällige oder gleichverteilte Reihenfolge realisiert. Von einer inhaltshomogenen Itemblockbildung wird gelegentlich abgeraten (vor allem für psychologisch-diagnostische Zwecke der Bewerberauswahl).

Wiederholt konnten Zusammenhänge zwischen teststatistischen Parametern und dem aus der Reihenfolge resultierendem Kontext der Items bei Persönlichkeitsfragebogen festgestellt werden. Entsprechende empirische Studien und theoretische Überlegungen wurden kurz vorgestellt. Zu kritisieren bleibt, dass der Großteil der Studien keine Verfahren heranzieht, die sich probabilistischer Testmodelle bedienen, sondern auf die Methoden der klassischen Testtheorie vertraut. Dabei müssen diese Ansätze grundsätzlich kritisiert werden, da sie alle extrem stichprobenabhängig sind (Kubinger, 2006).

In diesem Sinne widmet sich die vorliegende empirische Studie der Äquivalenzprüfung zweier Versionen eines a-priori dimensionierten, multidimensionalen Persönlichkeitsfragebogens. Hier aber nicht mittels der zuvor kritisierten Verfahren, sondern durch die Überprüfung des Itempools auf die Konformität mit dem Rasch-Modell. Graphische Modellkontrollen beziehungsweise Likelihood-Ratio-Tests sollten darüber Aufschluss geben, ob die Schwierigkeit der Items (die sogenannten Itemparameter) vom Fragebogenformat (ferner von den aus diesem Format resultierenden kognitiven Prozessen wie Priming und Durchschauung der Messintention) beeinflusst werden.

Hinsichtlich der möglichen Beeinflussung des Beantwortungsprozesses durch den Itemkontext wird hier folgende Hypothese aufgestellt und vertreten: Während die mit der inhaltshomogenen Itemblockung einhergehende höhere Augenscheinvalidität einen eindimensionalen Antwortprozess fördert, schadet die gleichverteilte Itemreihenfolge diesem insofern, als durch das permanente wiedererkennen ähnlicher Items (gleichen Inhalts) Reaktanz ausgelöst wird.

## 2 Methode

Bevor näher auf die empirische Untersuchung des Reihenfolgeeffektes eingegangen wird, möchte ich an dieser Stelle kurz auf das den Analysen zugrundeliegende testtheoretische Modell eingehen und die sich daraus ableitbaren Modelltests kurz vorstellen. Das soll jene Vorzüge verdeutlichen, die die Probabilistische Testtheorie (im Gegensatz zur Klassischen Testtheorie) bei der empirischen Untersuchung des Reihenfolgeeffektes bietet.

### 2.1 Dichotom-logistisches Testmodell von Rasch (1960)

Das der sogenannten „Item Response Theory“ zuzuordnende Testmodell von Rasch (1960) – auch 1-PL Modell genannt – beschreibt die Lösungswahrscheinlichkeit eines Items als (vom Zufall abgesehen) abhängig von einem Personenparameter (welcher der Fähigkeit der Person entspricht) einerseits und einem Itemparameter (welcher der Schwierigkeit des Items entspricht) andererseits. Darüber hinausgehende Beeinflussungen der Lösungswahrscheinlichkeit (wie durch bisher gelöste Items) würden eine Verletzung dieser sogenannten „lokalen stochastischen Unabhängigkeit“ bedeuten und gegen die Gültigkeit des Rasch-Modells sprechen (z.B. Kubinger, 2006).

Da das Rasch-Modell den wissenschaftstheoretischen Ansatz der sogenannten „spezifisch objektiven Vergleiche“ erfüllt und sich so gesehen als stichprobenunabhängig herausstellt, lässt sich auch ein besonderer Modelltest ableiten. Dabei wird überprüft, ob sich die Differenzen der Itemparameter in verschiedenen Teilstichproben (z.B. Männer und Frauen, leistungsfähige und leistungsschwache Personen, ältere und jüngere Menschen etc.) unterscheiden. Interferenzstatistisch erfolgt der Modelltest des Rasch-Modells sinngemäß über einen Likelihood-Ratio-Test (z.B. Kubinger, 1989).

Stellt sich bei diesem Modelltest heraus, dass das Rasch-Modell für die Items eines bestimmten Tests nicht gilt, dann ist nicht von „lokaler stochastischer

Unabhängigkeit“ und Stichprobenunabhängigkeit auszugehen, und der Test ist notwendigerweise nicht verrechnungsfair. Erweisen sich die Items dabei allerdings als modellkonform, dann kann davon ausgegangen werden, dass sie fair und eindimensional dasselbe messen (z.B. Kubinger, 2006).

Eine weitere Möglichkeit auf die Geltung des Rasch-Modells zu prüfen besteht darin, in einem rechtwinkligen Koordinatensystem die pro Item (anhand der zwei Teilstichproben) geschätzten Parameter gegeneinander aufzutragen. Im theoretischen Idealfall ergibt sich dabei eine Graphik mit einer durch den Ursprung gehenden 45°-Geraden. Sollten dahingegen einzelne oder alle Punkte von der 45°-Geraden deutlich abweichen, „so gilt für diese die Stichprobenunabhängigkeit nicht, also ist die gegebene Verrechnung auch nicht fair.“ (Kubinger, 2006; S. 87).

Diese graphische Modellkontrolle dient vor allem auch dazu, jene auffälligen (nicht dem Rasch-Modell entsprechenden) Items a-posteriori aus dem Itempool auszuschließen, damit gegebenenfalls Konformität des Tests (oder der Skala) mit dem Rasch-Modell erreicht werden kann. Eine weitere Möglichkeit hierzu stellt der z-Test nach Fischer und Scheiblechner dar (z.B. Glas & Verhelst, 1995). Gerade für die Testkonstruktion stellen diese beiden Verfahren (in Kombination) eine wesentliche Entscheidungshilfe dar, um den schrittweisen Ausschluss von auffälligen Items zu rechtfertigen. Denn das Ziel ist stets, möglichst nur jene Items in das psychologisch-diagnostische Verfahren aufzunehmen, die fair und informativ sind.

Die Vorzüge der Anwendung dieser Methoden zur Untersuchung des Itemkontexteffektes liegen also offensichtlich darin, dass damit die wesentliche Grundannahme des Modells (die lokale stochastische Unabhängigkeit) als Voraussetzung für Eindimensionalität geprüft werden kann. Sollte sich bei den Modelltests herausstellen, dass den Lösungen der Items in Abhängigkeit der Fragebogenversion nicht nur Personenfähigkeit und Itemschwierigkeit zugrunde liegen, stellt das die Bestätigung der Vermutung dar, dass der Itemkontext zu Prozessen anregt, die die Eindimensionalität der Befragung stören (beispielsweise Reaktanz).

## 2.2 Untersuchungsplan und Hypothesen

Im Sinne eines experimentellen Designs wurden zwei Bedingungen des später genauer beschriebenen Persönlichkeitsfragebogens realisiert, wobei die Zuteilung der Personen zu den Bedingungen (computergestützt) randomisiert erfolgte. Den Personen der *Kontrollgruppe* wurden dabei in Anlehnung an die übliche Praxis der Befragung mit Persönlichkeitsfragebogen die Items einzelner Subskalen über den gesamten Fragebogen möglichst gleichverteilt präsentiert (Bedingung „Randomized“). Die in dieser Bedingung angestrebte maximale inhaltliche Distanz der Items wurde weiters dadurch erhöht, dass die inhaltlich einer Skala zugeteilten Subskalen möglichst weit voneinander im Fragebogen auseinander lagen. Die so resultierende Skalenstruktur und die sich daraus ergebende Itemreihenfolge ist im Anhang dargestellt (siehe Kapitel 9.2 „Skalenstruktur der zwei Bedingungen des WSP“ auf Seite 80).

Dahingegen wurde in der *Versuchsgruppe* die maximale inhaltliche Nähe nacheinander präsentierter Items einzelner Subskalen realisiert (Bedingung „Blocked“). Hier wurden nicht nur die Items der Subskalen zu inhaltshomogenen Blöcken gestaltet, sondern zusätzlich auch die inhaltlich einer Skala zugeteilten Subskalen nacheinander angeordnet. Die daraus resultierende Skalenstruktur und Itemreihenfolge ist ebenfalls im Anhang abgebildet (siehe ebenfalls Kapitel 9.2 „Skalenstruktur der zwei Bedingungen des WSP“ auf Seite 80).

Als *unabhängige Variable* ist somit der Itemkontext, hier im Sinne der Itemreihenfolge definiert. Zur Diskussion stehen bei dieser Untersuchung die Auswirkungen des Itemkontextes auf die Eindimensionalität der Skalen, hier im Sinne der Gültigkeit des dichotomen-logistischen Testmodells von Rasch (1960). Die *abhängige Variable* stellen somit die geschätzten Itemparameter dar. Da die Befragung computerisiert erfolgte, konnten negative Einflüsse auf das Gütekriterium der Objektivität als *Störvariablen* ausgeschlossen werden. Zusätzlich wurde versucht, Versuchleitereffekte bei der Rekrutierung der Teilnehmerinnen und Teilnehmer durch Standardisierung dieser Situationen zu minimieren. Auf darüber hinausgehende eventuelle Störvariablen, wie

Bedingungen der individuellen Testsituation konnte kein Einfluss genommen werden.

Für beide Gruppen getrennt durchgeführte Modelltests (Likelihood-Ratio-Tests nach Anderson [1974] und graphische Modellkontrollen) sollen darüber Aufschluss geben, in welcher Teilstichprobe (also „Randomized“ oder „Blocked“) eher Konformität mit dem Rasch-Modell vorliegt beziehungsweise durch weniger Itemausschließungen erreicht werden kann. Für die dieser Gruppe zugrundeliegende Testversion könnte dann gefolgert werden, dass deren Itemreihenfolge einen günstigeren Einfluss auf den Bearbeitungsprozess (im Sinne von Eindimensionalität) hat.

Die in diesem Sinne formulierten Hypothesen lauten:

H0-1: Die beiden Teilstichproben unterscheiden sich nicht hinsichtlich der Konformität mit dem Rasch-Modell.

H1-1: Die beiden Teilstichproben unterscheiden sich hinsichtlich der Konformität mit dem Rasch-Modell.

Zudem sollte die Berechnung von Likelihood-Ratio-Tests in der Gesamtstichprobe mit dem Teilungskriterium *Testversion* wenigstens in manchen Skalen zu signifikanten Ergebnissen führen. Denn das würde ausdrücken, dass sich die auf Basis der zwei Bedingungen geschätzten Itemparameter voneinander unterscheiden, was ein weiterer Hinweis auf die Abhängigkeit der Itemschwierigkeit vom Itemkontext wäre.

Die in diesem Sinne formulierten Hypothesen lauten:

H0-2: Die auf Basis der beiden Teilstichproben geschätzten Itemparameter unterscheiden sich nicht im Teilungskriterium *Testversion*.

H1-2: Die auf Basis der beiden Teilstichproben geschätzten Itemparameter unterscheiden sich im Teilungskriterium *Testversion*.

## 2.3 Untersuchungsmaterial

### 2.3.1 Wiener Studieneignungs-Persönlichkeitsinventar (WSP)

Das zur Überprüfung herangezogene psychologisch-diagnostische Verfahren basiert auf dem von Khorramdel et al. (2008) formulierten Anforderungsprofil für die universitäre Hochschulausbildung in Wien. Das Anforderungsprofil resultierte aus Anforderungsanalysen für acht Studienrichtungen an Wiener Universitäten. Nach der Methode der Critical Incident Technique (Flanagan, 1954) wurden dabei zunächst Fragebogenerhebungen und Experteninterviews mit Stelleninhaberinnen und Stelleninhabern an drei Wiener Universitäten durchgeführt. Die Auswertung erfolgte nach der qualitativen Inhaltsanalyse nach Mayring (2005).

Aus dem Anforderungsprofil wurden nach Festlegung eines "Cut-off"-Wertes und Reduktion kognitiver Leistungsmerkmale 7 Skalen beziehungsweise 22 Subskalen identifiziert, welche die für Studienerfolg relevanten Persönlichkeitseigenschaften beziehungsweise Verhaltensweisen repräsentieren. Die Skala „Organisation“ setzt sich dabei aus folgenden fünf Subskalen zusammen: *Zeitmanagement*, *Prioritäten Setzen*, *Planen*, *Organisationstalent* und *Informiertheit*. Die Skala „Arbeitsstil, Arbeitsverhalten“ besteht aus folgenden vier Subskalen: *Gewissenhaftigkeit*, *Genaues Arbeiten*, *Eigeninitiative* und *Selbstständiges Arbeiten*. Die Skala „Motivation“ ist zusammengesetzt aus folgenden sechs Subskalen: *Durchhaltevermögen*, *Lernbereitschaft*, *Engagement*, *Zielorientierung*, *Zielsetzung* und *Erfolgszuversicht versus Misserfolgsängstlichkeit*. Die Skala „Belastbarkeit“ wird repräsentiert durch die Subskala *Frustrationstoleranz*. Eine weitere Skala stellt „Soziale Kompetenz“ dar, welche sich aus folgenden drei Subskalen zusammensetzt: *Kommunikationsfähigkeit*, *Kontaktfähigkeit* und *Teamfähigkeit*. Weiters wurde die Skala „Emotionale Komponenten“ durch folgende zwei Subskalen berücksichtigt: *Distanzierungsfähigkeit* und *Offenheit*. Schließlich wird die Skala „Selbstbild“ über die Subskala *Selbstbewusstsein* erfasst. Eine Übersicht dieser Skalen und Subskalen inklusive der genauen Anzahl der darin enthaltenen Items ist im

Anhang abgebildet (siehe Kapitel 9.1 „Skalen und Subskalen des WSP“ auf Seite 79)

Aus den erhobenen Daten bei der Anforderungsanalyse generierte Maurer (2009) entlang eines eigens entwickelten Regelkataloges zur Itementwicklung für die oben angeführten Subskalen einen auf die Zielgruppe spezifisch zugeschnittenen Itempool, dessen Aufbereitung zum „Wiener Studieneignungs-Persönlichkeitsinventar“ (WSP) – inklusive Instruktion (siehe unten) – führte. Die Items des WSP sind dabei als Statements formuliert, auf welche die zu testenden Personen mittels eines 4-kategorialen Antwortformates zustimmend oder ablehnend reagieren können. Eine Frage lautet beispielsweise: *„Ich führe einen Kalender, um wichtige Termine zu notieren.“* Zur Beantwortung dieser Statements stehen folgende Antwortmöglichkeiten zur Verfügung: *„ja“*, *„eher ja“*, *„eher nein“* und *„nein“*.

Der WSP umfasst in seiner von Maurer (2009) vorgelegten Form 350 Items. Dass entsprechend viele Items konstruiert wurden (mit zwischen 9 und 25 Items je Subskala), lässt sich auf die Erfahrung zurückführen, nach welcher Modelltests im Zuge der Skalierung meist ein Ausscheiden von mehr oder weniger vielen auffälligen Items notwendig machen. Bei der Itemgenerierung wurde daher in strittigen Fällen darauf Wert gelegt, mehrere unterschiedlich formulierte Versionen eines Items zur Verfügung zu stellen. Im Zuge der Skalierung sollte dann durch Modellkontrollen festgestellt werden können, welche Version besser geeignet ist.

### **2.3.2 Programmierung des WSP als Online-Self-Assessment**

Die Entscheidung den WSP computerisiert im Sinne eines Online-Self-Assessment vorzugeben, basiert auf Überlegungen zu allen generellen Vorteilen von Computerverfahren versus Papier-Bleistift-Verfahren einerseits (vgl. Kubinger, 2006) und spezifischen Erfordernissen andererseits. Ein wesentlicher Grund für diese Entscheidung sind die günstigen Auswirkungen eines solchen

Vorgehens auf die Testleiterunabhängigkeit und die Verrechnungssicherheit als den Voraussetzungen für das Hauptgütekriterium der Objektivität. Weiters können dadurch (eine gewissenhafte Programmierung vorausgesetzt) sehr einfach Auswertungs- bzw. Datenübertragungsfehler verhindert werden. Zusätzlich kann die Zuteilung zu den zwei Fragebogenversionen sehr einfach randomisiert gestaltet werden.

Wie später noch verdeutlicht werden wird, war es ein wesentliches Anliegen, den Teilnehmern<sup>1</sup> unmittelbar nach der Befragung ihre individuellen Ergebnisse rückzumelden. Allein deshalb war eine computerisierte Vorgabe (über das Internet mit eigener Website) notwendig. Ein weiterer wesentlicher Grund ergab sich aus der Absicht, den WSP nach erfolgreicher testtheoretischer Absicherung und Normierung in ein bereits bestehendes Online-Self-Assessment (zur Studienwahlberatung) zu integrieren. Um außerdem einer ansonsten notwendigen Äquivalenzüberprüfung zwischen Papier-Bleistift-Version und Computerversion zuvorzukommen, wurde der WSP daher gleich für die Vorgabe am Computer und über das Internet programmiert.

Vor der Bearbeitung des eigentlichen Persönlichkeitsfragebogens wurde eine Instruktion zur Verfügung gestellt (siehe Kapitel 9.3 „Instruktion des WSP“ auf Seite 81). Zudem waren einige deskriptive Angaben zu machen, nämlich bezüglich Alter, Geschlecht, Muttersprache, Berufstätigkeit und Studienfach beziehungsweise Studienrichtung. Der Wortlaut der Instruktion war an entsprechende Formulierungen gängiger Persönlichkeits-Fragebogen angelehnt. Im Zuge dieser Instruktion wurde jedem Teilnehmer ein eigener zufällig generierter zwölfstelliger Zahlen-Buchstaben-Code präsentiert, mit Hilfe dessen später die eigenen Ergebnisse eingesehen werden konnten. Außerdem beinhaltete

---

<sup>1</sup> Im Folgenden wird durchgehend von „den (Untersuchungs-)Teilnehmern“, „den Probanden“ etc. die Rede sein wenn alle befragten Personen – also teilgenommene Frauen und Männer – gemeint sind. Dieses Vorgehen soll eine bessere Lesbarkeit des Textes ermöglichen.

die Instruktion ein Beispielitem, das die Teilnehmer mit der Beantwortung der einzelnen Fragen vertraut machen sollte.

Im Persönlichkeitsfragebogen selbst wurde immer nur ein Item nach dem anderen präsentiert. Hatten sich Teilnehmer eine abgegebene Antwort noch einmal anders überlegt, so hatten sie die Möglichkeit das zuletzt beantwortete Item auszubessern. Zur zeitlichen Orientierung stand eine Prozentleiste (prozentueller Anteil der bisher bearbeiteten Items) zur Verfügung, die den Fortschritt im WSP anzeigte. Ein eigener Link ermöglichte es, jederzeit die Instruktion in einem zusätzlichen Fenster nachzulesen.

An dieser Stelle sei Frau Mag. Khorramdel gedankt, welche durch ihr engagiertes Feedback wesentlich zu der raschen Umsetzung des WSP als einem Online-Fragebogen beitrug. Mit freundlicher Unterstützung stand sie mir bei der Erstellung dieser wissenschaftlichen Abschlussarbeit zur Seite.

### **2.3.3 Unvollständiges Design, Bearbeitungsdauer und Randomisierung**

Aufgrund der hohen Itemanzahl und der damit verbundenen langen Bearbeitungszeit des Fragebogens erschien ein sogenanntes unvollständiges Design für die Fragebogenvorgabe sinnvoll. Dadurch wurde es möglich, den Personen nicht alle Items der Subskalen vorzugeben, sondern nur einen Teil davon. Die daraus resultierende Reduzierung der Bearbeitungsdauer von ursprünglich etwa einer Stunde auf etwa eine halbe bis dreiviertel Stunde sollte dazu beitragen die Motivation und Aufmerksamkeit der Teilnehmer aufrechtzuerhalten beziehungsweise reaktantes Antwortverhalten und die Abbruchquote zu minimieren.

Aber nicht alle Subskalen wurden unvollständig vorgegeben. Subskalen mit weniger als 10 Items wurden vollständig vorgegeben. Weiters wurden unterschiedlich dichte „Verlinkungen“ der Items gebildet. Bei Subskalen mit einer ursprünglichen Itemanzahl zwischen 10 und 15 Items wurde darauf geachtet, dass jedes Item in drei Testheften enthalten ist. Bei Subskalen mit einer Itemanzahl

größer als 15 Items erschien es ausreichend, dass jedes Item in nur zwei Testheften enthalten ist. Im Anhang ist für zwei Subskalen die Bildung des unvollständigen Designs exemplarisch dargestellt (siehe Kapitel 9.4 „Unvollständiges Design“ auf Seite 83).

Die Konstruktion des unvollständigen Designs führte zur Bildung von vier Testheften je Bedingung. Somit standen insgesamt acht Testhefte zur Verfügung. In vieren waren die Items des WSP mit maximaler inhaltlicher Distanz gereiht, in den anderen vieren mit maximaler inhaltlicher Nähe. Die Itemanzahl der einzelnen Testhefte war dabei leicht unterschiedlich, mit 218 Items im ersten beziehungsweise fünften Testheft, 221 Items im zweiten beziehungsweise sechsten Testheft, 215 Items im dritten beziehungsweise siebten Testheft und 217 Items im vierten beziehungsweise achten Testheft.

Die Zuteilung zu einem dieser acht Testhefte erfolgte randomisiert, wobei die Wahrscheinlichkeit für die Vorgabe eines bestimmten Testheftes umso geringer war, je öfter es bereits zuvor bearbeitet wurde. Diese Maßnahme ermöglichte, dass alle Testhefte möglichst gleich häufig bearbeitet wurden.

#### **2.3.4 Die Homepage „Besser-Studieren.at“**

Der Zugang zum computerisierten WSP wurde über eine eigens erstellte Homepage eingerichtet, welche über den Link [www.besser-studieren.at](http://www.besser-studieren.at) erreichbar war. Dieses Vorgehen sollte eine Reihe von für die Untersuchung vorteilhaften Voraussetzungen erfüllen. So war es ein wesentliches Anliegen, bei den potentiellen Teilnehmern Interesse und Neugier für die Themen Studienerfolg beziehungsweise studienrelevante Persönlichkeitseigenschaften und Verhaltensweisen zu wecken. In diesem Sinne wurde auf der Homepage auch ein Diskussionsforum angeboten, das dem Erfahrungs- und Meinungsaustausch der Studierenden hinsichtlich obiger Themen dienen sollte. Besucher konnten dort Tipps von Studienkollegen finden beziehungsweise selber dokumentieren, beispielsweise wie sie sich auf schwierige Prüfungen vorbereiten, wie sie mit

Prüfungsstress und sonstigen studienrelevanten Emotionen umgehen (Stichwort Frustrationstoleranz) oder mit welchen Arbeits- und Organisationstechniken sie ihren Studienalltag erleichtern (Stichwort Mindmap, Zeitmanagement etc.).

Die Website sollte die Funktion einer Plattform zum Themenbereich „Besser Studieren“ erfüllen. Die zwei Säulen dieser Plattform waren dabei erstens das Diskussionsforum, das dem Austausch studienrelevanter Erfahrungen diene und zweitens der Persönlichkeitsfragebogen selbst. Vor allem die Bearbeitung des computerisierten WSP bedeutete für die Teilnehmer einen erheblichen Gewinn. Im Sinne eines Self-Assessment konnten sie jene Anforderungen an sich selbst überprüfen, die durch die Studie von Khorramdel et al. (2008) als studienrelevant identifiziert wurden. Dadurch erhielten sie auch einen guten allgemeinen Überblick über diese relevanten Anforderungen.

Zusätzlich erhielten sie mit ihren Ergebnissen automatisch Literatur- und Softwareempfehlungen rückgemeldet. Bei den Literaturempfehlungen handelte es sich um vom Autor recherchierte populäre Literatur zu Themenbereichen wie Zeitmanagement- und Organisationstechniken, Selbstmanagement, Softskills, Lern- und Arbeitstechniken für das Studium, Motivation und Zielorientierung. Bei den Softwareempfehlungen handelte es sich um ebenfalls vom Autor recherchierte (über das Internet gratis beziehbare) Software zur Erstellung von Mindmaps, zur Erstellung von To-Do-Listen und einem Tool zum Zeitmanagement. Außerdem wurde jeder Teilnehmer auf die zwei Einrichtungen „Test- und Beratungsstelle“ ([www.testzentrum.at](http://www.testzentrum.at)) und „Psychologische Studentenberatung“ ([www.studentenberatung.at](http://www.studentenberatung.at)) verwiesen. All diese Empfehlungen waren nach Eingabe des individuellen Personencodes direkt über die Homepage einzusehen.

Im Sinne der Idee des „informed consent“ wurden relevante Informationen zu Fragebogen, Bearbeitung des Fragebogens und Diskussionsforum auf der Homepage angeboten. Es wurde genau informiert, von wem der Persönlichkeitsfragebogen entwickelt wurde, dass es sich dabei um ein psychologisch-diagnostisches Verfahren in der Erprobungsphase handelt, dass die Teilnahme völlig anonym erfolgt und etwa dreißig Minuten Zeit in Anspruch nähme, welche Vorteile die Bearbeitung böte, und dass die Ergebnisse im

Rahmen der Diplomarbeit des Autors verwendet werden. Die Texte der Homepage sind im Anhang wiedergegeben (siehe Kapitel 9.5 „Homepage Besser-Studieren.at“ auf Seite 84).

Das Layout der Homepage wurde absichtlich sehr einfach gehalten und sollte in der optischen Gestaltung an eine Schreibtafel erinnern. Ebenso einfach war die logische Struktur der Homepage, mit nur sehr wenigen Verzweigungen. Hauptsächlich mussten sich die Besucher nur zwischen dem Link „Zum Fragebogen“ und dem Link „Zum Forum“ entscheiden.

### **2.3.5 Zusätzliche Materialien: Poster, Flyer**

Für „Marketingzwecke“ wurden als zusätzliche Materialien Poster und Flyer eingesetzt. Die grundlegende Idee der Gestaltung des Layouts war dabei, dieses maximal simpel zu halten. Somit beschränkte sich die Botschaft im Wesentlichen auf den Internetlink „Besser-Studieren.at“. Durch die Kombination eines neonfarbenen Papiers mit dem eingängigen und kurzen Text wurde versucht, die (erfahrungsgemäß sehr schwer zugängliche) Aufmerksamkeit der Personen zu erreichen.

Inwieweit das bei den Postern, welche regelmäßig in den Gängen der Universität Wien, der Technischen Universität Wien und der Wirtschaftsuniversität Wien affiziert wurden, gelungen ist, kann nicht abgeschätzt werden. Die Flyer waren dahingegen in erster Linie als Erinnerungstütze für jene Personen gedacht, welche im direkten Gespräch auf das Angebot hingewiesen wurden. Zudem wurden täglich Flyer in den Computerhörsälen des Zentralen Informatikdienstes der Universität Wien und der Technischen Universität Wien sowie in diversen Hörsälen dieser Universitäten ausgelegt.

## **2.4 Untersuchungsteilnehmer**

### **2.4.1 Rekrutierung der Untersuchungsteilnehmer**

Im Sinne der Zielgruppe des nach Khorramdel et al. (2008) beziehungsweise Maurer (2009) entwickelten Persönlichkeitsfragebogens, der sich einem förderungs-diagnostischen Ansatz folgend an Studierende richtet, waren eben solche auch das Ziel der Datenerhebung dieser Studie. Dabei wurden verschiedene Möglichkeiten zur Rekrutierung von Teilnehmern genutzt. Insgesamt wurde dabei bedacht, möglichst viele Variablen auszuschalten, die in dieser Phase zu einem Stichprobenbias beitragen können. Der Kontakt zwischen Versuchsleiter und potentiellm Versuchsteilnehmer wurde daher entweder standardisiert gestaltet oder gänzlich unterbunden.

Letzteres Vorgehen konnte durch den Aushang von Postern in den Universitätsgebäuden, das Auslegen von Flugzetteln in den Computersälen der Zentralen Informatikdienste, den Versand von Massenmails durch dieselben und Foreneinträge in einschlägigen Internet-Diskussionsforen für Studierende erreicht werden. Besonders das tägliche Auslegen von Flugzetteln vor die einzelnen PCs der Computersäle der Zentralen Informatikdienste (im sogenannten Neuen Institutsgebäude und am Universitätscampus der Universität Wien beziehungsweise im Universitätsgebäude der Technischen Universität Wien) erwies sich dabei als effektiv.

Zusätzlich wurden persönliche Kontakte mit potentiellen Teilnehmern gesucht. Dazu wurden Personen vor und in den Universitätsgebäuden direkt angesprochen und über das Angebot kurz informiert. Zusätzlich erhielt jede angesprochene Person einen Flugzettel. Zur Standardisierung dieser Situationen des persönlichen Kontaktes wurde zuvor ein Text formuliert und eingeübt, der beim Ansprechen von potentiellen Teilnehmern „vorgetragen“ wurde. Ebenso wurden Antworten auf häufige Fragen möglicher Probanden zuvor überlegt, sodass in diesen Situationen der Teilnehmerrekrutierung immer ein maximal ähnlicher Eindruck der Person des Versuchsleiters erstens und der Versuchsteilnahme zweitens erreicht werden konnte. Bei der Rekrutierung von

Versuchspersonen waren Studienkollegen behilflich, welche zum Erwerb eines Zeugnisses für die Lehrveranstaltung „Forschungsseminar“ notwendige Hilfestunden absolvieren mussten. Namentlich waren Frau Kathrin Gruber, Frau Marianne Liedlbauer, Frau Nina Teubenbacher, Frau Sabine Weinberger und Herr Robin Gleeson beteiligt. Diese unterstützend mitwirkenden Personen wurden über die Formulierungen genau instruiert bevor sie mit Personen in Kontakt traten.

Da jedoch Prozesse wie persönliche Sympathie und Antipathie unbewusst und daher unkontrollierbar mit Sicherheit eine Rolle spielen, können gewisse Effekte auf die Stichprobe (z.B. die unbewusste Selektion von anzusprechenden Personen) nicht völlig ausgeschlossen werden. Als eine vorteilhaftere Gestaltung der Rekrutierung von Versuchsteilnehmern wurden aber auch zusätzlich Personengruppen in Vorlesungen und Seminaren an Wiener Universitäten kurz über das Angebot informiert. Hierzu wurden zunächst jene Stelleninhaberinnen und Stelleninhaber um Unterstützung gebeten, welche schon bei Khorramdel et al. (2008) für die Teilnahme am Experteninterview bereit waren, so dass nach deren Zusage Studierende in deren Lehrveranstaltungen gesammelt auf das Angebot hingewiesen werden konnten. Der hierzu zuvor überlegte und formulierte Text wurde dann in allen diesen Situationen gleichlautend vorgetragen.

#### **2.4.2 Deskriptive Beschreibung der Stichprobe**

Die erhaltene Pseudozufallsstichprobe umfasste schließlich insgesamt 721 Studierende von Wiener Universitäten. 201 Datensätze mussten allerdings ausgeschlossen werden, da diese Studierende den Persönlichkeitsfragebogen (aus nicht näher bekannten Gründen) nicht abgeschlossen hatten. Für die Analysen standen daher Daten von 520 Studierenden zur Verfügung. Davon bearbeiteten 259 Personen die vier Testhefte „Randomized“ und 261 Personen die vier Testhefte „Blocked“.

Für die deskriptive Beschreibung der Stichprobe wurden folgende Variablen erhoben: Geschlecht, Alter, Semester, Muttersprache, Berufstätigkeit und

Studienrichtung. Lediglich 10% der Teilnehmer gaben eine andere Muttersprache als Deutsch an. Tabelle 1 gibt die deskriptiven Statistiken zu Geschlechterverteilung, Alter und Semesterstatus der Personen in den zwei Teilstichproben wieder. Tabelle 2 zeigt die verschiedenen Studienfächer und Studienrichtungen der in der Stichprobe enthaltenen Personen.

Betrachtet man Tabelle 1, so fällt ein gewisser Frauenüberhang auf (mit nämlich insgesamt 343 Frauen und 177 Männern), der in der Zusammensetzung der Stichprobe hinsichtlich der Studienrichtungen begründet liegen könnte. Nach Tabelle 2 sind nämlich eher Studienrichtungen repräsentiert, die einen hohen Frauenanteil aufweisen. Hier sei darauf hingewiesen, dass dies keineswegs beabsichtigt war. Eventuell war die Bereitschaft von Teilnehmerinnen der entsprechenden Studienrichtungen (insbesondere Psychologie) größer, das Angebot wahrzunehmen.

***Tabelle 1.** Deskriptive Statistiken und Häufigkeiten zu Geschlechterverteilung, Alter und Semesterstatus der Personen in den zwei Bedingungen*

| <b>Bedingung</b> | <b>Geschlecht</b> | <b>n</b> | <b>%</b> |
|------------------|-------------------|----------|----------|
| Randomized       | männlich          | 86       | 33,2     |
|                  | weiblich          | 173      | 66,8     |
|                  | insgesamt         | 259      | 100      |
| Blocked          | männlich          | 91       | 34,9     |
|                  | weiblich          | 170      | 65,1     |
|                  | insgesamt         | 261      | 100      |

| <b>Bedingung</b> | <b>Variable</b> | <b>Min.</b> | <b>Max.</b> | <b>MW</b> | <b>SD</b> |
|------------------|-----------------|-------------|-------------|-----------|-----------|
| Randomized       | Alter           | 18          | 51          | 23,25     | 4,296     |
|                  | Semester        | 1           | 26          | 6,24      | 3,942     |
| Blocked          | Alter           | 18          | 50          | 23,36     | 4,383     |
|                  | Semester        | 1           | 24          | 6,21      | 4,423     |

*Tabelle 2. Häufigkeiten der Studienfächer und Studienrichtungen in der Stichprobe*

| <b>Studienfächer und Studienrichtungen</b>   | <b>n</b> | <b>%</b> |
|--|----------|----------|
| Human- und Sozialwissenschaften              | 143      | 27,5     |
| Soziologie                                   | 27       | 5,2      |
| Psychologie                                  | 80       | 15,4     |
| Sonstige                                     | 36       |          |
| Geisteswissenschaften                        | 71       | 13,7     |
| Wirtschaftswissenschaften                    | 71       | 13,7     |
| Betriebswirtschaftslehre                     | 40       | 7,7      |
| Volkswirtschaftslehre                        | 14       | 2,7      |
| Sonstige                                     | 17       |          |
| Rechtswissenschaften                         | 67       | 12,9     |
| Rechtswissenschaften                         | 46       | 8,8      |
| Sonstige                                     | 21       |          |
| Ingenieurwissenschaften und Informatik       | 60       | 11,5     |
| Architektur                                  | 20       | 3,8      |
| Maschinenbau                                 | 7        | 1,3      |
| Sonstige                                     | 33       |          |
| Medien- und Kommunikationswissenschaften     | 48       | 9,2      |
| Publizistik und Kommunikationswissenschaften | 44       | 8,5      |
| Sonstige                                     | 4        |          |
| Mathematik und Naturwissenschaften           | 23       | 4,4      |
| Human- und Zahnmedizin                       | 7        | 1,3      |
| Kunst und Kunstwissenschaften                | 4        | ,8       |
| Ernährungswissenschaften                     | 2        | ,4       |
| Veterinärmedizin                             | 1        | ,2       |
| Agrar- und Forstwissenschaften               | 1        | ,2       |
| Sonstige                                     | 22       | 4,2      |

## 2.5 Untersuchungsdurchführung und Computerprogramme

Auf die Computerisierung des für die Online-Vorgabe aufbereiteten Persönlichkeitsfragebogens WSP (inklusive Recherche der Literatur- und Softwareempfehlungen) folgte die Konzipierung und Programmierung der Homepage „Besser-Studieren.at“. Danach wurden Flyer beziehungsweise Poster gestaltet und produziert. Nach erfolgreicher Kontaktaufnahme mit Stelleninhaberinnen und Stelleninhabern konnten erste Teilnehmer in Vorlesungen angeworben werden. In der Phase der Datenerhebung wurden zwecks Anwerbung von Teilnehmern zusätzlich Plakate ausgehängt, Flyer ausgelegt und Personen angesprochen. Die Datenerhebung erfolgte wie oben dargestellt computerisiert über die Bearbeitung des WSP durch die Untersuchungsteilnehmer.

Nach Abschluss der Datenerhebung wurden erste deskriptive Analysen vorgenommen. Dazu diente das Computerprogramm „SPSS Statistics 17.0“. Darüber hinausgehende Analysen wurden mit Hilfe eines für „R for Windows, Version 2.8.1“ programmierten Softwarepackages vorgenommen (extended Rasch modelling von Mair & Hatzinger, 2007; vgl. Poinstingl, Mair & Hatzinger, 2007), das speziell für die Anwendung des Rasch-Modells entwickelt wurde.

Da die Analysen nach dem Rasch-Modell einen dichotomen Datensatz voraussetzen, wurden nach Abschluss der Datenerhebung die ursprünglich 4-kategoriellen Antworten der Teilnehmer dichotomisiert, sodass „eher ja“ und „ja“ beziehungsweise „eher nein“ und „nein“ zu „eher ja“ beziehungsweise „eher nein“ zusammengefasst wurden.

In Anlehnung an die zuvor aufgestellten Untersuchungshypothesen wurden zwei verschiedene Ansätze der Auswertung realisiert. Die erste der oben formulierten Alternativhypothesen besagt nämlich, dass sich die zwei Teilstichproben in ihrer Konformität mit dem Rasch-Modell unterscheiden. Sollte der Itemkontext bei der Bedingung „Randomized“ also tatsächlich im Sinne der oben aufgestellten Hypothese zum psychologischen Beantwortungsprozess einen Einfluss auf die (Ein-) Dimensionalität der Items haben, so würden die Items der

Skalen und Subskalen aufgrund der verletzten lokalen stochastischen Unabhängigkeit eher nicht dem Rasch-Modell folgen. Gleichzeitig sollte sich der günstige Einfluss der Bedingung „Blocked“ auf die (Ein-) Dimensionalität dadurch ausdrücken, dass die Items der Skalen und Subskalen Rasch-Modell-Konformität aufweisen.

Modelltests in diesem Sinne wurden separat in beiden Teilstichproben für die folgenden Teilungskriterien durchgeführt: *Score*, *Geschlecht*, *Alter* und *Semester*. Die Wahl dieser Teilungskriterien ergibt sich aus der Zielgruppe des WSP. Da dieses psychologisch-diagnostische Verfahren Studienanfängern als auch fortgeschrittenen Studierenden dienen soll, erschien das Teilungskriterium *Semester* notwendig. Dabei wurde die Einteilung in Studierende bis inklusive 4. Semester und Studierende ab dem 5. Semester getroffen. Die Gruppenbildung im Teilungskriterium *Alter* erfolgte derart, dass in beiden Gruppen gleich viele Personen enthalten sind. In diesem Fall also mit einem Split beim 22. Lebensjahr (= Median der Altersverteilung). Das Teilungskriterium *Score* wurde durch einen Split beim Median der Rohscores einer Skala realisiert und soll dazu dienen, leistungsstärkere mit leistungsschwächeren Personen zu vergleichen.

Als Modelltests dienten Likelihood-Ratio-Tests (Anderson, 1974) und Graphische Modellkontrollen. Items, die sowohl in der graphischen Modellkontrolle als auch im z-Test nach Fischer und Scheiblechner (gelegentlich auch „Waldtest“ genannt) bei Teilungskriterien auffällig waren, wurden im Falle signifikanter Likelihood-Ratio-Tests solange schrittweise ausgeschlossen, bis in diesen ein nicht signifikantes Ergebnis erreicht werden konnte. Für eine genauere Auseinandersetzung mit diesen Verfahren siehe zum Beispiel Kubinger (1989) und Fischer und Molenaar (1995).

Die zweite der oben formulierten Alternativhypothesen besagt einen Unterschied der im Teilungskriterium *Testversion* geschätzten Itemparameter. Durch einen solchen Split werden die auf Basis der beiden Teilstichproben geschätzten Itemparameter einander direkt gegenübergestellt. Übt der Itemkontext einen Einfluss auf die Schwierigkeiten der Items der Skalen aus, so sollten das Likelihood-Ratio-Tests und graphische Modellkontrollen anzeigen. Modelltests in

diesem Sinne wurden also in einer zweiten Auswertung unter Heranziehung der gesamten Stichprobe für das Teilungskriterium *Testversion* durchgeführt.

Es erschien sinnvoll zunächst alle Items der inhaltlich einer Skala zugeordneten Subskalen zusammenzufassen und einer gemeinsamen Analyse zu unterziehen. Sollte sich nämlich dabei herausstellen, dass Rasch-Modell-Konformität bereits auf der übergeordneten Ebene der Skalen (z.B. „Motivation“ inklusive aller dazugehörenden Subskalen) vorliegt, so dürfte aufgrund dieses Nachweises der Eindimensionalität nicht in Subskalen differenziert werden. Solange nicht mehr als etwa 15% der Items einer Skala bei diesen Analysen ausgeschlossen werden müssen, erscheint daher eine Analyse auf Subskalenebene unangebracht.

## 3 Ergebnisse

### 3.1 Analyse der beiden Teilstichproben

Im Folgenden sind die für beide Teilstichproben getrennt gerechneten Modelltests in den Teilungskriterien *Score*, *Geschlecht*, *Alter* und *Semester* wiedergegeben. Wie zuvor erwähnt wurde mit der Analyse der gesamten Skalen begonnen. Im Falle von Rasch-Modell-Konformität auf Ebene der Skalen wurde auf die Analyse der Subskalen verzichtet, da deren Verrechnung in diesem Fall ohnehin nicht zulässig erscheint.

Wesentlich beim Vergleich der beiden Teilstichproben war auch die Beobachtung, wie viele Items ausgeschlossen werden müssen, um Rasch-Modell-Konformität zu erreichen. Daher sei hier darauf hingewiesen, dass schrittweise nur jene Items ausgeschlossen wurden, die in der graphischen Modellkontrolle und dem z-Test nach Fischer und Scheiblechner am stärksten auffällig waren. Sobald bei den Likelihood-Ratio-Tests in allen Teilungskriterien ein nicht signifikantes Ergebnis erreicht wurde, wurden keine weiteren Items ausgeschlossen (auch wenn weiterhin einzelne Items nach obigen Methoden auffällig waren).

### 3.1.1 Skala „Organisation“

#### 3.1.1.1 Teilstichprobe „Randomized“

Siehe Tabelle 3 für die Ergebnisse des Modelltests der Skala „Organisation“ in der Teilstichprobe „Randomized“. Durch Ausschluss von 7 Items konnte im Teilungskriterium *Score* ein nicht signifikantes Ergebnis erreicht werden. In Anbetracht der insgesamt 78 Items dieser Skala erscheint ein Differenzieren in inhaltlich verschiedene Subskalen daher nicht zulässig zu sein.

**Tabelle 3.** Modelltest für die Skala „Organisation“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items    | Teilungskriterien |            |        |          |
|---|-------------------|------------|--------|----------|
|   | Score             | Geschlecht | Alter  | Semester |
| Anderson $\chi^2 =$                                   | <b>167,55</b>     | 91,79      | 91,93  | 91,68    |
| Degrees of Freedom =                                  | <b>72</b>         | 77         | 77     | 77       |
| $\chi^2$ at alpha (1 percent) =                       | <b>101,62</b>     | 107,58     | 107,58 | 107,58   |
| Likelihood-Ratio-Tests nach<br>Ausschluss von 7 Items | Teilungskriterien |            |        |          |
|   | Score             | Geschlecht | Alter  | Semester |
| Anderson $\chi^2 =$                                   | 93,68             | 88         | 84,34  | 68,02    |
| Degrees of Freedom =                                  | 67                | 70         | 70     | 70       |
| $\chi^2$ at alpha (1 percent) =                       | 95,63             | 99,23      | 99,23  | 99,23    |

### 3.1.1.2 Teilstichprobe „Blocked“

Siehe Tabelle 4 für die Ergebnisse des Modelltests der Skala „Organisation“ in der Teilstichprobe „Blocked“. Durch Ausschluss von 13 Items konnte in den Teilungskriterien *Score* und *Geschlecht* ein nicht signifikantes Ergebnis erreicht werden. Ein Differenzieren in Subskalen erscheint daher unzulässig zu sein.

**Tabelle 4.** Modelltest für die Skala „Organisation“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor      | Teilungskriterien |               |        |          |
|---------------------------------|-------------------|---------------|--------|----------|
| Ausschluss von Items            | Score             | Geschlecht    | Alter  | Semester |
| Anderson $\chi^2$ =             | <b>118,6</b>      | <b>132,03</b> | 64,89  | 93,63    |
| Degrees of Freedom =            | <b>74</b>         | <b>77</b>     | 77     | 77       |
| $\chi^2$ at alpha (1 percent) = | <b>104,01</b>     | <b>107,58</b> | 107,58 | 107,58   |
| Likelihood-Ratio-Tests nach     | Teilungskriterien |               |        |          |
| Ausschluss von 13 Items         | Score             | Geschlecht    | Alter  | Semester |
| Anderson $\chi^2$ =             | 84,61             | 72,56         | 59,12  | 54,55    |
| Degrees of Freedom =            | 63                | 64            | 64     | 64       |
| $\chi^2$ at alpha (1 percent) = | 90,80             | 92,01         | 92,01  | 92,01    |

### 3.1.2 Skala „Arbeitsstil, Arbeitsverhalten“

#### 3.1.2.1 Teilstichprobe „Randomized“

Siehe Tabelle 5 für die Ergebnisse des Modelltests der Skala „Arbeitsstil, Arbeitsverhalten“ in der Teilstichprobe „Randomized“. Ein Ausschluss von Items war nicht erforderlich. Aufgrund der Eindimensionalität dieser Skala, wird „Arbeitsstil, Arbeitsverhalten“ nicht weiter in Subskalen differenziert.

**Tabelle 5.** Modelltest für die Skala „Arbeitsstil, Arbeitsverhalten“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2$ =                                | 89,38             | 70,33      | 60,56 | 77,69    |
| Degrees of Freedom =                               | 65                | 70         | 69    | 71       |
| $\chi^2$ at alpha (1 percent) =                    | 93,22             | 99,23      | 98,03 | 100,43   |

#### 3.1.2.2 Teilstichprobe „Blocked“

Siehe Tabelle 6 für die Ergebnisse des Modelltests der Skala „Arbeitsstil, Arbeitsverhalten“ in der Teilstichprobe „Blocked“. Durch Ausschluss von 2 Items konnte im Teilungskriterium *Alter* ein nicht signifikantes Ergebnis erreicht werden. Da diese Skala insgesamt 72 Items umfasst, scheint auch hier angesichts der lediglich 2 auszuschließenden Items ein Differenzieren in Subskalen unzulässig zu sein.

**Tabelle 6.** Modelltest für die Skala „Arbeitsstil, Arbeitsverhalten“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |               |          |
|--|-------------------|------------|---------------|----------|
|  | Score             | Geschlecht | <b>Alter</b>  | Semester |
| Anderson $\chi^2$ =                                | 90,99             | 95,07      | <b>113,72</b> | 86,96    |
| Degrees of Freedom =                               | 64                | 71         | <b>69</b>     | 71       |
| $\chi^2$ at alpha (1 percent) =                    | 92,01             | 100,43     | <b>98,03</b>  | 100,43   |

| Likelihood-Ratio-Tests nach<br>Ausschluss von 2 Items | Teilungskriterien |            |       |          |
|---|-------------------|------------|-------|----------|
|   | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2$ =                                   | 62,42             | 91,1       | 92,64 | 78,44    |
| Degrees of Freedom =                                  | 65                | 69         | 67    | 68       |
| $\chi^2$ at alpha (1 percent) =                       | 93,22             | 98,03      | 95,63 | 96,83    |

### 3.1.3 Skala „Motivation“

#### 3.1.3.1 Teilstichprobe „Randomized“

Siehe Tabelle 7 für die Ergebnisse des Modelltests der Skala „Motivation“ in der Teilstichprobe „Randomized“. Ein Ausschluss von Items war nicht erforderlich. Daher wird die Skala „Motivation“ auch nicht weiter in Subskalen differenziert.

**Tabelle 7.** Modelltest für die Skala „Motivation“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |        |          |
|--|-------------------|------------|--------|----------|
|  | Score             | Geschlecht | Alter  | Semester |
| Anderson $\chi^2 =$                                | 107,13            | 115,63     | 114,71 | 83,36    |
| Degrees of Freedom =                               | 84                | 91         | 91     | 92       |
| $\chi^2$ at alpha (1 percent) =                    | 115,88            | 124,12     | 124,12 | 125,29   |

#### 3.1.3.2 Teilstichprobe „Blocked“

Siehe Tabelle 8 für die Ergebnisse des Modelltests der Skala „Motivation“ in der Teilstichprobe „Blocked“. Durch Ausschluss von 4 Items konnte in den Teilungskriterien *Score* und *Geschlecht* ein nicht signifikantes Ergebnis erreicht werden. Aufgrund der insgesamt 93 Items dieser Skala wird daher nicht weiter in Subskalen differenziert.

**Tabelle 8.** Modelltest für die Skala „Motivation“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |               |        |          |
|--|-------------------|---------------|--------|----------|
|  | Score             | Geschlecht    | Alter  | Semester |
| Anderson $\chi^2 =$                                | <b>114,12</b>     | <b>134</b>    | 96,89  | 74,1     |
| Degrees of Freedom =                               | <b>86</b>         | <b>90</b>     | 92     | 92       |
| $\chi^2$ at alpha (1 percent) =                    | <b>118,24</b>     | <b>122,94</b> | 125,29 | 125,29   |

| Likelihood-Ratio-Tests nach<br>Ausschluss von 4 Items | Teilungskriterien |            |        |          |
|---|-------------------|------------|--------|----------|
|   | Score             | Geschlecht | Alter  | Semester |
| Anderson $\chi^2 =$                                   | 96,46             | 117,1      | 84,94  | 79,8     |
| Degrees of Freedom =                                  | 83                | 86         | 88     | 88       |
| $\chi^2$ at alpha (1 percent) =                       | 114,69            | 118,24     | 120,59 | 120,59   |

### 3.1.4 Skala „Belastbarkeit“

#### 3.1.4.1 Teilstichprobe „Randomized“

Siehe Tabelle 9 für die Ergebnisse des Modelltests der Skala „Belastbarkeit“ (welche der Subskala „Frustrationstoleranz“ entspricht) in der Teilstichprobe „Randomized“. Ein Ausschluss von Items auf Basis dieser Befunde ist nicht erforderlich.

**Tabelle 9.** Modelltest für die Skala „Belastbarkeit“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2$ =                                | 26,07             | 15,14      | 12,13 | 22,04    |
| Degrees of Freedom =                               | 15                | 19         | 20    | 20       |
| $\chi^2$ at alpha (1 percent) =                    | 29,14             | 34,81      | 36,19 | 36,19    |

#### 3.1.4.2 Teilstichprobe „Blocked“

Siehe Tabelle 10 für die Ergebnisse des Modelltests der Skala „Belastbarkeit“ (welche der Subskala „Frustrationstoleranz“ entspricht) in der Teilstichprobe „Blocked“. Ein Ausschluss von Items war demnach nicht erforderlich.

**Tabelle 10.** Modelltest für die Skala „Belastbarkeit“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2$ =                                | 25,27             | 23,53      | 26,52 | 23,21    |
| Degrees of Freedom =                               | 14                | 18         | 19    | 18       |
| $\chi^2$ at alpha (1 percent) =                    | 27,69             | 33,41      | 34,81 | 33,41    |

### 3.1.5 Skala „Soziale Kompetenz“

#### 3.1.5.1 Teilstichprobe „Randomized“

Siehe Tabelle 11 für die Ergebnisse des Modelltests der Skala „Soziale Kompetenz“ in der Teilstichprobe „Randomized“. Ein Ausschluss von Items war nicht erforderlich, weswegen nicht weiter in Subskalen differenziert wurde.

**Tabelle 11.** Modelltest für die Skala „Soziale Kompetenz“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2 =$                                | 31,73             | 34,44      | 49,4  | 24,42    |
| Degrees of Freedom =                               | 33                | 35         | 35    | 35       |
| $\chi^2$ at alpha (1 percent) =                    | 53,49             | 56,06      | 56,06 | 56,06    |

#### 3.1.5.2 Teilstichprobe „Blocked“

Siehe Tabelle 12 für die Ergebnisse des Modelltests der Skala „Soziale Kompetenz“ in der Teilstichprobe „Blocked“. Durch Ausschluss von 3 Items konnte im Teilungskriterium *Score* ein nicht signifikantes Ergebnis erreicht werden. Aufgrund der insgesamt 36 Items dieser Skala scheint ein Differenzieren in Subskalen also unzulässig.

**Tabelle 12.** Modelltest für die Skala „Soziale Kompetenz“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items    | Teilungskriterien |            |       |          |
|---|-------------------|------------|-------|----------|
|   | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2 =$                                   | <b>65,71</b>      | 43,65      | 37,57 | 41,28    |
| Degrees of Freedom =                                  | <b>34</b>         | 35         | 35    | 35       |
| $\chi^2$ at alpha (1 percent) =                       | <b>54,78</b>      | 56,06      | 56,06 | 56,06    |
| Likelihood-Ratio-Tests nach<br>Ausschluss von 3 Items | Teilungskriterien |            |       |          |
|   | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2 =$                                   | 35,71             | 42,86      | 38,74 | 48,08    |
| Degrees of Freedom =                                  | 30                | 32         | 32    | 32       |
| $\chi^2$ at alpha (1 percent) =                       | 49,59             | 52,19      | 52,19 | 52,19    |

### 3.1.6 Skala „Emotionale Komponenten“

#### 3.1.6.1 Teilstichprobe „Randomized“

Siehe Tabelle 13 für die Ergebnisse des Modelltests der Skala „Emotionale Komponenten“ in der Teilstichprobe „Randomized“. Ein Ausschluss von Items war nicht erforderlich, was für die Eindimensionalität dieser 32 Items umfassenden Skala spricht. Es wird daher nicht weiter in Subskalen differenziert.

**Tabelle 13.** Modelltest für die Skala „Emotionale Komponenten“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2 =$                                | 34,249            | 20,91      | 21,87 | 26,38    |
| Degrees of Freedom =                               | 28                | 31         | 31    | 31       |
| $\chi^2$ at alpha (1 percent) =                    | 46,96             | 50,89      | 50,89 | 50,89    |

#### 3.1.6.2 Teilstichprobe „Blocked“

Siehe Tabelle 14 für die Ergebnisse des Modelltests der Skala „Emotionale Komponenten“ in der Teilstichprobe „Blocked“. Ein Ausschluss von Items war nicht erforderlich, weswegen ein Differenzieren in Subskalen auch hier unzulässig erscheint.

**Tabelle 14.** Modelltest für die Skala „Emotionale Komponenten“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |        |          |
|--|-------------------|------------|--------|----------|
|  | Score             | Geschlecht | Alter  | Semester |
| Anderson $\chi^2 =$                                | 13,88             | 42,04      | 19,946 | 34,62    |
| Degrees of Freedom =                               | 21                | 31         | 30     | 31       |
| $\chi^2$ at alpha (1 percent) =                    | 37,57             | 50,89      | 49,59  | 50,89    |

### 3.1.7 Skala „Selbstbild“

#### 3.1.7.1 Teilstichprobe „Randomized“

Siehe Tabelle 15 für die Ergebnisse des Modelltests der Skala „Selbstbild“ (welche der Subskala „Selbstbewusstsein“ entspricht) in der Teilstichprobe „Randomized“. Ein Ausschließen von Items aus dieser 18 Items umfassenden Skala war nicht erforderlich.

**Tabelle 15.** Modelltest für die Skala „Selbstbild“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2 =$                                | 19,28             | 13,82      | 7,64  | 25,36    |
| Degrees of Freedom =                               | 16                | 17         | 17    | 17       |
| $\chi^2$ at alpha (1 percent) =                    | 30,58             | 31,99      | 31,99 | 31,99    |

#### 3.1.7.2 Teilstichprobe „Blocked“

Siehe Tabelle 16 für die Ergebnisse des Modelltests der Skala „Selbstbild“ (welche der Subskala „Selbstbewusstsein“ entspricht) in der Teilstichprobe „Blocked“. Ein Ausschluss von Items war nicht erforderlich.

**Tabelle 16.** Modelltest für die Skala „Selbstbild“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.

| Likelihood-Ratio-Tests vor<br>Ausschluss von Items | Teilungskriterien |            |       |          |
|--|-------------------|------------|-------|----------|
|  | Score             | Geschlecht | Alter | Semester |
| Anderson $\chi^2 =$                                | 16,89             | 24,36      | 27    | 25,76    |
| Degrees of Freedom =                               | 16                | 17         | 17    | 17       |
| $\chi^2$ at alpha (1 percent) =                    | 30,58             | 31,99      | 31,99 | 31,99    |



### **3.2 Analyse der Gesamtstichprobe**

Im Folgenden sind die Ergebnisse der Likelihood-Ratio-Tests angeführt wie auch die graphischen Modellkontrollen für das Teilungskriterium *Testversion* abgebildet.

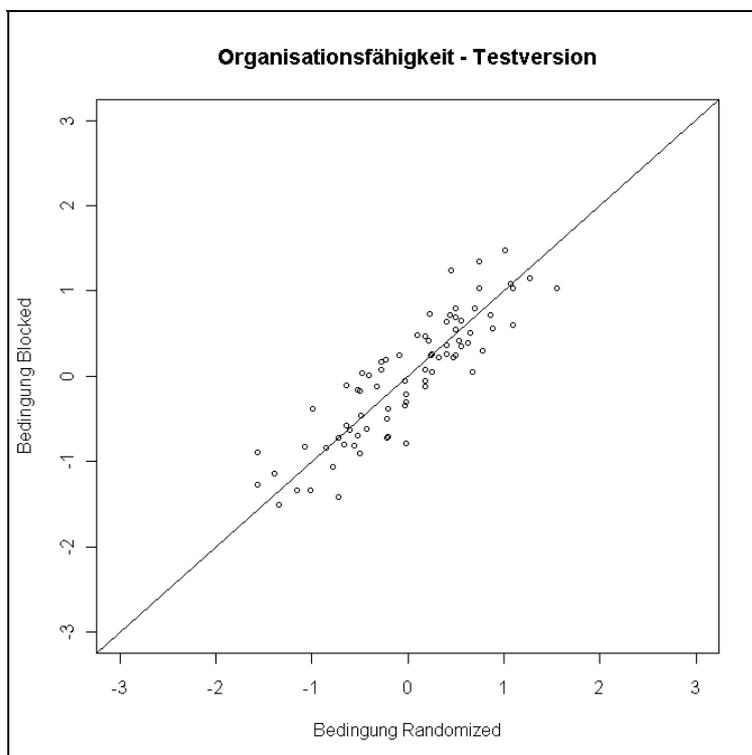
Auf die genauere Analyse der Subskalen wurde hier verzichtet, da sich bei den ersten Analysen sowohl in der Teilstichprobe „Randomized“ als auch in der Teilstichprobe „Blocked“ herausgestellt hatte, dass bereits auf Skalenebene Eindimensionalität gegeben ist. Die Verrechnung der Subskalen scheint somit auf Basis dieser Ergebnisse unzulässig zu sein.

### 3.2.1 Modelltest in der Skala „Organisation“

Die Itemparameter der Skala „Organisation“ unterscheiden sich nicht in den beiden Teilstichproben (siehe das nicht signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 17). Ein Ausschluss von Items ist daher nicht erforderlich. Dieser Befund drückt sich auch in der Graphischen Modellkontrolle aus (siehe Abbildung 1).

**Tabelle 17.** Nicht signifikanter Modelltest für das Teilungskriterium *Testversion* der Skala „Organisation“ als Ergebnisausdruck aus dem Programm eRm.

|  | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|--|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items | 83,61             | 77                 | 107,58                        |



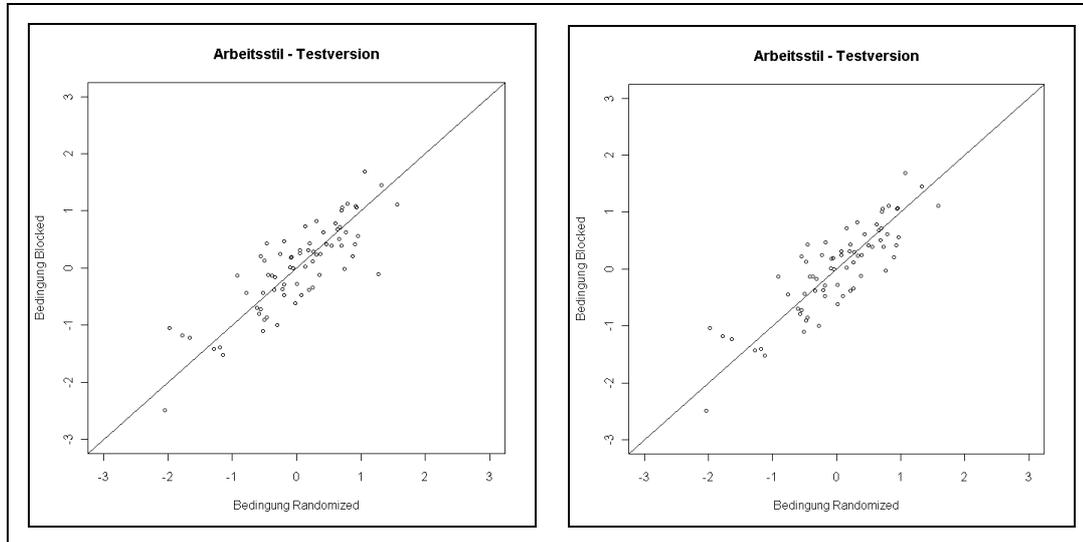
**Abbildung 1.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Organisation“.

### 3.2.2 Modelltest in der Skala „Arbeitsstil, Arbeitsverhalten“

Die Itemparameter der Skala „Arbeitsstil, Arbeitsverhalten“ unterscheiden sich in den beiden Teilstichproben (siehe das signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 18). Durch Ausschluss eines Items konnte allerdings für das Teilungskriterium *Testversion* ein nicht signifikantes Ergebnis erreicht werden (siehe Tabelle 18 und Abbildung 2).

**Tabelle 18.** Modelltest für das Teilungskriterium *Testversion* der Skala „Arbeitsstil, Arbeitsverhalten“ als Ergebnisausdruck aus dem Programm eRm.

|  | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|--|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items   | 107,87            | 71                 | 100,43                        |
| Likelihood-Ratio-Test nach Ausschluss von 1 Item | 82,77             | 70                 | 99,23                         |



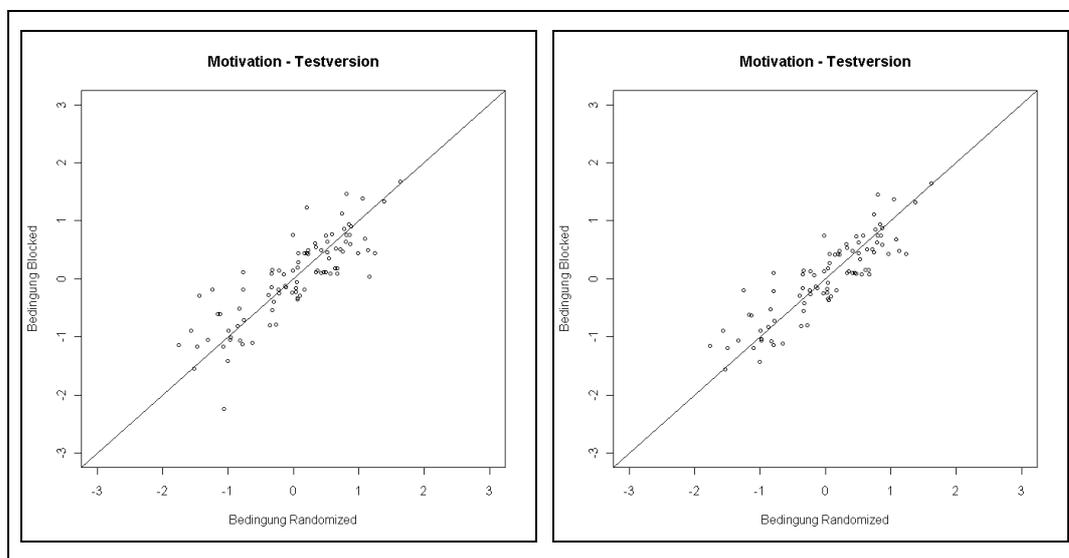
**Abbildung 2.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Arbeitsstil, Arbeitsverhalten“ vor dem Ausschluss von Items (links) und nach Ausschluss von einem Item (rechts).

### 3.2.3 Modelltest in der Skala „Motivation“

Die Itemparameter der Skala „Motivation“ unterscheiden sich in den beiden Teilstichproben (siehe das signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 19). Durch Ausschluss von 4 Items konnte allerdings für das Teilungskriterium *Testversion* ein nicht signifikantes Ergebnis erreicht werden (siehe Tabelle 19 und Abbildung 3).

**Tabelle 19.** Modelltest für das Teilungskriterium *Testversion* der Skala „Motivation“ als Ergebnisausdruck aus dem Programm eRm.

|   | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|---|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items    | 131,52            | 92                 | 125,29                        |
| Likelihood-Ratio-Test nach Ausschluss von 4 Items | 96,88             | 88                 | 120,59                        |



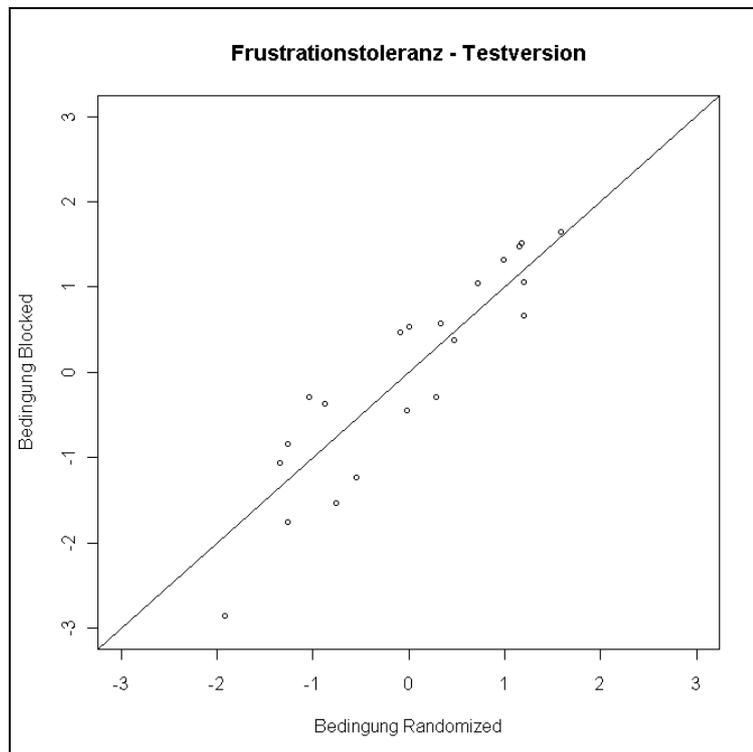
**Abbildung 3.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Motivation“ vor dem Ausschluss von Items (links) und nach Ausschluss von 4 Items (rechts).

### 3.2.4 Modelltest in der Skala „Belastbarkeit“

Die Itemparameter der Skala „Belastbarkeit“ (die der Subskala „Frustrationstoleranz“ entspricht) unterscheiden sich nicht in den beiden Teilstichproben (siehe das nicht signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 20). Dieser Befund drückt sich auch in der Graphischen Modellkontrolle aus (siehe Abbildung 4).

**Tabelle 20.** Nicht signifikanter Modelltest für das Teilungskriterium *Testversion* der Skala „Belastbarkeit“ als Ergebnisausdruck aus dem Programm eRm.

|  | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|--|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items | 23,2              | 20                 | 36,19                         |



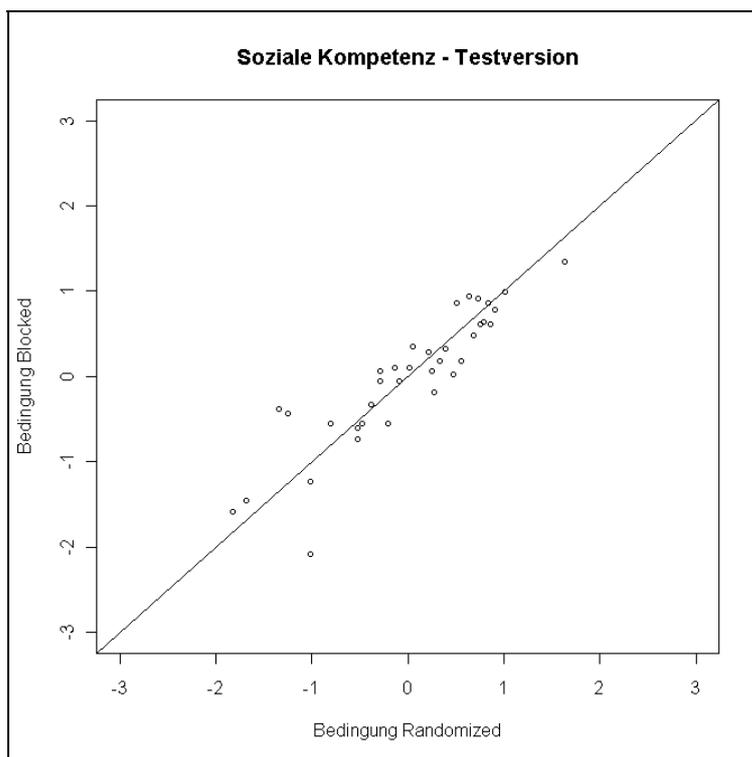
**Abbildung 4.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Belastbarkeit“.

### 3.2.5 Modelltest in der Skala „Soziale Kompetenz“

Die Itemparameter der Skala „Soziale Kompetenz“ unterscheiden sich nicht in den beiden Teilstichproben (siehe das nicht signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 21). Dieser Befund drückt sich auch in der Graphischen Modellkontrolle aus (siehe Abbildung 5).

**Tabelle 21.** Nicht signifikanter Modelltest für das Teilungskriterium *Testversion* der Skala „Soziale Kompetenz“ als Ergebnisausdruck aus dem Programm eRm.

|  | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|--|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items | 42,11             | 35                 | 56,06                         |



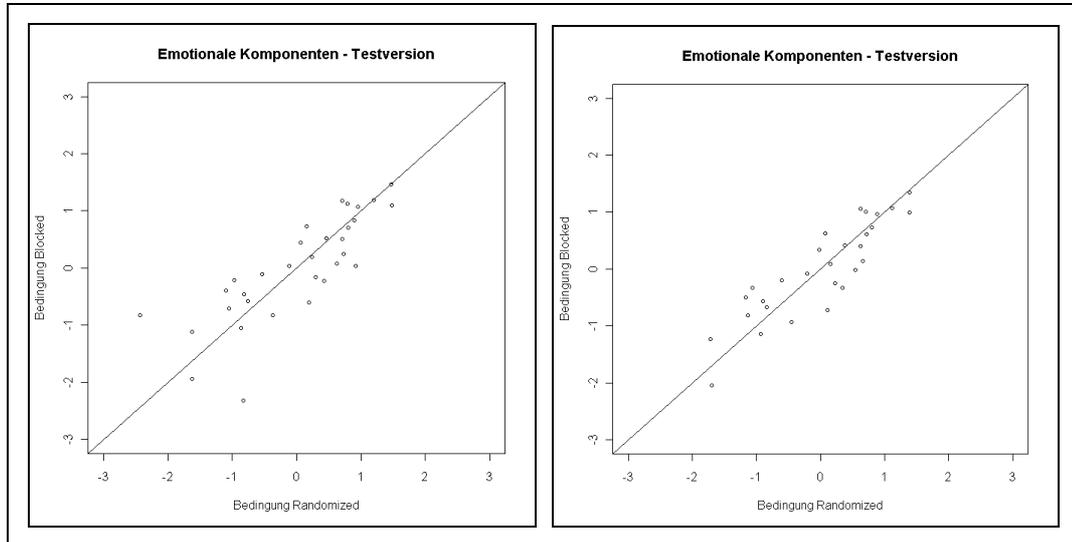
**Abbildung 5.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Soziale Kompetenz“.

### 3.2.6 Modelltest in der Skala „Emotionale Komponenten“

Die Itemparameter der Skala „Emotionale Komponenten“ unterscheiden sich in den beiden Teilstichproben (siehe das signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 22). Durch Ausschluss von 3 Items konnte allerdings für das Teilungskriterium *Testversion* ein nicht signifikantes Ergebnis erreicht werden (siehe Tabelle 22 und Abbildung 6).

**Tabelle 22.** Modelltest für das Teilungskriterium *Testversion* der Skala „Emotionale Komponenten“ als Ergebnisausdruck aus dem Programm eRm.

|   | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|---|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items    | 52,89             | 31                 | 50,89                         |
| Likelihood-Ratio-Test nach Ausschluss von 3 Items | 36,56             | 28                 | 46,96                         |



**Abbildung 6.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Emotionale Komponenten“ vor dem Ausschluss von Items (links) und nach Ausschluss von 3 Items (rechts).

### 3.2.7 Modelltest in der Skala „Selbstbild“

Die Itemparameter der Skala „Selbstbild“ (welche der Subskala „Selbstbewusstsein“ entspricht) unterscheiden sich nicht in den beiden Teilstichproben (siehe das nicht signifikante Ergebnis des Likelihood-Ratio-Tests in Tabelle 23). Dieser Befund drückt sich auch in der Graphischen Modellkontrolle aus (siehe Abbildung 7).

**Tabelle 23.** Nicht signifikanter Modelltest für das Teilungskriterium *Testversion* der Skala „Selbstbild“ als Ergebnisausdruck aus dem Programm eRm.

|  | Anderson $\chi^2$ | Degrees of Freedom | $\chi^2$ at alpha (1 percent) |
|--|-------------------|--------------------|-------------------------------|
| Likelihood-Ratio-Test vor Ausschluss von Items | 22,65             | 17                 | 31,99                         |



**Abbildung 7.** Graphische Modellkontrolle für das Teilungskriterium *Testversion* der Skala „Selbstbild“.

### 3.3 Zusammenfassung der Ergebnisse

Ein erstes wesentliches Ergebnis ist, dass sich bereits auf der Ebene der Skalen die Konformität der Items mit dem Rasch-Modell herausgestellt hat (und zwar in beiden Teilstichproben). Selbst im Falle von signifikanten Likelihood-Ratio-Tests bei einigen Teilungskriterien auf dieser Ebene konnte durch Ausschließen von Items in einem relativ geringen Umfang noch Eindimensionalität der Skalen erreicht werden. Deswegen wurde bei den weiteren Analysen auf eine Differenzierung in Subskalen verzichtet.

Allerdings zeigten sich Unterschiede der beiden Teilstichproben in einem anderen und für die Fragestellung relevanteren Zusammenhang. So ergaben für beide Bedingungen getrennt durchgeführte Modelltests in der Teilstichprobe „Blocked“ bei den drei Skalen „Arbeitsstil, Arbeitsverhalten“ (für das Teilungskriterium *Alter*), „Motivation“ (für das Teilungskriterien *Score* und *Geschlecht*) und „Soziale Kompetenz“ (für das Teilungskriterium *Score*) signifikante Likelihood-Ratio-Tests. Durch einen Ausschluss von Items konnte dort allerdings eine Passung mit dem Rasch-Modell erreicht werden (siehe Tabellen 6, 8 und 12). Gar keine Items mussten jedoch bei diesen Skalen in der Bedingung „Randomized“ ausgeschlossen werden, da Likelihood-Ratio-Tests dort keine signifikanten Ergebnisse angezeigt hatten (siehe Tabellen 5, 7 und 11).

Bei diesem Vergleich erwiesen sich die Items der Skala „Organisation“ in beiden Teilstichproben (vor Itemausschluss) als nicht mit dem Rasch-Modell konform (siehe Tabellen 3 und 4). So war in der Teilstichprobe „Randomized“ der Likelihood-Ratio-Test für das Teilungskriterium *Score* signifikant. In der Teilstichprobe „Blocked“ waren hier die Likelihood-Ratio-Tests für die Teilungskriterien *Score* und *Geschlecht* signifikant. Um die Eindimensionalität der Skala „Organisation“ zu erreichen, mussten also in beiden Bedingungen Items ausgeschlossen werden. Dieses Ausscheiden von Items führte zwar in beiden Teilstichproben zu einer Modellpassung, allerdings mussten in der Teilstichprobe „Randomized“ lediglich 7 Items ausgeschieden werden, während in der Teilstichprobe „Blocked“ 13 Items ausgeschieden werden mussten.

Die Skalen „Belastbarkeit“ (siehe Tabellen 9 und 10), „Emotionale Komponenten“ (siehe Tabellen 13 und 14) und „Selbstbild“ (siehe Tabellen 15 und 16) erwiesen sich in beiden Teilstichproben noch vor dem Ausschluss von Items als mit dem Rasch-Modell konform. Dort ließ sich also mit dieser Methode kein Unterschied zwischen den zwei Bedingungen feststellen.

Eine weitere Untersuchung des Datensatzes hinsichtlich des zur Frage stehenden Itemkontexteffektes sah die Analyse der Gesamtstichprobe im Teilungskriterium *Testversion* vor. Durch einen solchen Split können die auf Basis der zwei Teilstichproben geschätzten Itemparameter direkt gegenübergestellt werden. Eventuelle Unterschiede können so direkt auf die den Bedingungen zugrundeliegenden, verschiedenen Itemreihenfolgen zurückgeführt werden. Da die Verrechnung der Subskalen aufgrund obiger Befunde unzulässig erscheint, wurde auch bei dieser Untersuchung nicht weiter in Subskalen differenziert, sondern die gesamten Skalen zur Analyse herangezogen.

Die Untersuchung der Gesamtstichprobe in diesem Sinne ergab signifikante Likelihood-Ratio-Tests in den drei Skalen „Arbeitsstil, Arbeitsverhalten“ (siehe Tabelle 18 beziehungsweise Abbildung 2), „Motivation“ (siehe Tabelle 19 beziehungsweise Abbildung 3) und „Emotionale Komponenten“ (siehe Tabelle 22 beziehungsweise Abbildung 6). Erst durch den Ausschluss von einigen auffälligen Items konnte für diese Skalen Eindimensionalität im Sinne der Konformität mit dem Rasch-Modell erreicht werden. In den weiteren Skalen („Organisation“, „Belastbarkeit“, „Soziale Kompetenz“ und „Selbstbild“) konnten keine auffälligen Likelihood-Ratio-Tests nachgewiesen werden.

## 4 Diskussion

### 4.1 Interpretation

Ein erstes Ergebnis bezieht sich auf die Struktur des WSP, welcher aus Skalen besteht, die weiter in Subskalen unterteilt sind. Gegen diese durch inhaltliche Überlegungen entstandene Unterteilung scheint allerdings die Skalierung des Persönlichkeitsfragebogens zu sprechen. Denn bereits auf der Ebene der Skalen liegt nach obigen Befunden Eindimensionalität der Items vor (und das unabhängig von der Fragebogenform). Die Verrechnung der einzelnen Subskalen zu inhaltlich verschiedenen Persönlichkeitsdimensionen scheint somit unzulässig zu sein.

Zur Überprüfung der Forschungsfrage wurde der Datensatz (unter Anwendung des dichotomen-logistischen Testmodells von Rasch [1960]) hinsichtlich jenes Einflusses untersucht, den der Itemkontext auf die Bearbeitung eines Persönlichkeitsfragebogens haben könnte. Dazu wurden zwei verschiedene Ansätze gewählt. Erstens wurde jede Teilstichprobe für sich untersucht. Dabei wurde überprüft, in welcher von diesen eher Konformität mit dem Rasch-Modell besteht beziehungsweise erreicht werden kann. Zweitens wurden die beiden Teilstichproben im Teilungskriterium *Testversion* direkt gegenübergestellt. Diese Analyse der Gesamtstichprobe sollte anzeigen, ob sich die geschätzten Itemparameter beider Testversionen unterscheiden.

Die resultierenden Ergebnisse stützen die zuvor formulierte Alternativhypothese, dass sich die zwei Testversionen hinsichtlich ihrer Konformität mit dem Rasch-Modell unterscheiden (H1-1). So waren die in der Teilstichprobe „Randomized“ durchgeführten Likelihood-Ratio-Tests seltener signifikant beziehungsweise mussten in der einzigen dort auffälligen Skala weniger Items ausgeschieden werden. Dieser Befund spricht dafür, dass der Prozess der Bearbeitung des Fragebogens unter der Bedingung „gleichverteilte Itemreihenfolge“ eindimensionaler ist als unter der Bedingung „inhaltlich geblockte Itemreihenfolge“. Diese Bedingung desselben Persönlichkeits-

fragebogens führte nämlich öfter zu signifikanten Likelihood-Ratio-Tests beziehungsweise mussten dort wesentlich mehr auffällige Items ausgeschlossen werden.

Auch die zweite Alternativhypothese (H1-2) kann durch obige Befunde gestützt werden. So zeigten sich im Teilungskriterium *Testversion* abhängig von der Itemreihenfolge unterschiedliche Itemschwierigkeiten (Itemparameter) in den beiden Teilstichproben. Likelihood-Ratio-Tests kamen in drei Skalen zu signifikanten Ergebnissen. Auch wenn durch entsprechende Itemausschlüsse noch eine Konformität dieser Skalen mit dem Rasch-Modell erreicht werden konnte, spricht dieser Befund dafür, dass der Itemkontext einen Einfluss auf die Schwierigkeiten der Items hatte, je nachdem, ob die inhaltlich zusammengehörenden Items gleichverteilt über den Fragebogen hinweg angeordnet oder in inhaltshomogenen Blöcken nacheinander präsentiert wurden.

Nicht gestützt werden konnte allerdings die Hypothese, dass diese Blockung der inhaltlich zusammengehörenden Items und die Vorgabe dieser Blöcke im WSP günstige Auswirkungen auf die Bearbeitung dieses Persönlichkeitsfragebogens haben. Eher das Gegenteil musste festgestellt werden. Wie oben bereits dargestellt, war die Modellkonformität unter dieser Bedingung öfters nur durch Ausschlüsse von auffälligen Items zu erreichen. Testtheoretisch kann daraus gefolgert werden, dass die Lösungswahrscheinlichkeit der Items im Falle der inhaltlichen Blockung nicht nur von Zufall, Itemschwierigkeit und Fähigkeit der Person abhängig ist, sondern noch weitere Dimensionen eine Rolle spielen. Eventuell führte die Aneinanderreihung von ähnlichen Items zu Motivationsverlust in Folge der Monotonie. Dahingegen könnte die Aufmerksamkeit der Teilnehmer in der Bedingung „Randomized“ durch die dort wesentlich abwechslungsreichere Itemreihenfolge aufrecht erhalten geblieben sein.

Unterstützung für eine solche Erklärung lässt sich aus manchen Arbeiten der psychologischen Fachliteratur ableiten (siehe Krampen et al. [1992]). Diese Autoren stellten im Zuge der Untersuchung des Einflusses der inhaltshomogenen Itemblockung auf begleitende kognitive und motivationale Prozesse eine dabei erhöhte (geschätzte) Bearbeitungszeit, erhöhte Langeweile bei der Fragebogen-

bearbeitung, sowie Entscheidungsprobleme im Antwortverhalten und intensivere Überlegungen zur Messabsicht des Fragebogens fest.

## **4.2 Konklusion**

In dieser Studie wurde ein weiteres Mal der Nachweis erbracht, dass sich der durch die Itemreihenfolge manipulierte Itemkontext auf die Bearbeitung eines Persönlichkeitsfragebogens auswirkt. Entgegen der plausibel klingenden Vermutung, dass die erhöhte Augenscheinvalidität günstige Auswirkungen auf die Eindimensionalität des psychologisch-diagnostischen Verfahrens hat, stellte sich die Bedingung mit einer Gleichverteilung der inhaltlich zusammengehörenden Items (und somit verringerten Augenscheinvalidität) als eher mit dem Rasch-Modell konform heraus. Eventuell trug die abwechslungsreichere Gestaltung des Fragebogens zu einer Aufrechterhaltung von Aufmerksamkeit und Motivation bei.

## **4.3 Kritik**

Ein erster wesentlicher Kritikpunkt bezieht sich auf den Fragebogen selbst. Dieser beinhaltet in seiner ursprünglich von Maurer (2009) vorgelegten Form nämlich etliche Fragen, die in ihrer Formulierung fast ident sind. Zwei Items aus der Skala „Soziale Kompetenz“ können hierfür als Beispiel dienen: „Es fällt mir leicht, mit anderen Studenten in ein Gespräch zu kommen“ und „Es fällt mir leicht, mit fremden (unbekannten) Personen ein Gespräch zu beginnen“. Entsprechende Items wurden bei der Fragebogenkonstruktion generiert, da erfahrungsgemäß etliche Items im Zuge der Skalierung ausgeschlossen werden müssen. Durch das zur Verfügung stellen von möglichst vielen Items kann erreicht werden, dass schlussendlich nur noch die besten Items in das psychologisch-diagnostische Verfahren aufgenommen werden.

Ein Nachteil, der sich jedoch durch die Vorgabe eines solchen erweiterten Itempools ergeben kann, ist die erhöhte Bearbeitungsdauer. Motivationsverlust und Ermüdung können die Folgen sein. Vor allem in Kombination mit dem Entdecken der sehr ähnlichen Items kann reaktantes Antwortverhalten dadurch gefördert werden. Besonders nämlich in der Bedingung „Blocked“, in welcher die Augenscheinvalidität ohnehin stark erhöht ist und Items mit ähnlichem Inhalt nacheinander präsentiert werden, stechen diese fast identen Items besonders ins Auge.

Es wäre daher von Vorteil, künftige Untersuchungen mit einem auf die inhaltlich und testtheoretisch besten Items reduzierten Itempool durchzuführen. Ein solcher hätte zudem den Vorzug, dass Rasch-Modell-Konformität beziehungsweise Eindimensionalität bereits nachgewiesen wäre und nicht erst im Zuge der Analyse der Forschungsfrage erbracht werden muss. Bei Verletzungen der Eindimensionalität bestünde dann weniger Unsicherheit darüber, ob diese durch den Einfluss des Itemkontextes oder die (bisher unbekannt) testtheoretische Qualität des Verfahrens verursacht wurde.

Zudem könnte dann aufgrund der geringeren Bearbeitungsdauer ein vollständiges Design verwendet werden, bei dem jedes einzelne Item von jeder Person bearbeitet wird. In Kombination mit einer größeren Stichprobe, bei der insbesondere mehr Personen auch von technischen Studiengängen erfasst werden sollten, würde ein vollständiges Design wesentlich mehr Daten liefern.

Um mehr Hinweise auf die psychologischen Ursachen einer eventuellen Beeinflussung des Beantwortungsprozesses zu erhalten, empfiehlt es sich einige diesbezügliche Fragen zusätzlich zu stellen, ganz so, wie das bereits in einigen anderen empirischen Studien (vgl. oben Krampen et al, 1992 und Sparfeldt et al, 2006) realisiert wurde. Entsprechende Fragen sollten sich auf begleitende kognitive und motivationale Prozesse beziehen und könnten darüber Aufschluss geben, inwieweit die inhaltshomogene Itemblockbildung tatsächlich Motivations- und Aufmerksamkeitsverlust fördert. Aufgrund des ohnehin sehr umfangreichen Itempools mit entsprechend langer Bearbeitungsdauer wurde bei dieser Studie auf eine zusätzliche Belastung der Teilnehmer verzichtet.

#### **4.4 Aussicht**

Der im Zuge dieser empirischen Studie erhobene Datensatz soll der Normierung und Skalierung im Sinne der abschließenden Testkonstruktion dienen. Ziel ist es, ein praxistaugliches psychologisch-diagnostisches Verfahren im Sinne eines Self-Assessment zur Verfügung zu stellen. Erste Hinweise auf Itemkontexteffekte und die testtheoretische Qualität konnten im Zuge dieser Untersuchung bereits erbracht werden.

Für die Auswahl der endgültigen Items des WSP sollten schrittweise all jene Items ausgeschlossen werden, die sehr ähnlich formuliert sind und daher überflüssig scheinen. Die Auswahl kann sich an den Graphischen Modellkontrollen orientieren und/oder an den Ergebnissen des z-Test nach Fischer und Scheiblechner. Vor allem aber auch inhaltliche Überlegungen dürfen dabei nicht vernachlässigt werden.

In Anbetracht der Ergebnisse dieser Studie hinsichtlich des ungünstigen Einflusses der inhaltshomogenen Itemblockbildung auf die Skalierung des WSP empfiehlt sich für die endgültige Fragebogenform die gleichverteilte Itemreihenfolge. Daher sollte die (vorläufige) Normierung nur anhand jener Personen geschehen, die auch dieser Bedingung ausgesetzt waren.

Für die weitere Zukunft empfiehlt sich aber eine neuerliche Datenerhebung (mit vollständigem Design), die insbesondere auch mehr Personen von technischen Studiengängen beinhaltet, sodass eine studienrichtungsspezifische Normierung möglich wird. Würden dann genügend Daten vorliegen, sollte in entsprechenden Teilungskriterien auch untersucht werden, ob die Schwierigkeiten der einzelnen Items für Personen unterschiedlicher Studienrichtungen verschieden sind.



## 5 Zusammenfassung

In der Praxis der Testkonstruktion gibt es verschiedene Möglichkeiten, Items von Persönlichkeitsfragebogen anzuordnen (zufällig, adaptiv, gleichverteilt und geblockt). In der psychologischen Fachliteratur werden positive und negative Auswirkungen auf teststatistische Parameter der so gestalteten Fragebogen ausgeführt. Entsprechende empirische Studien wurden kurz vorgestellt. Besonders die Zusammenfassung mancher Items zu inhaltshomogenen Blöcken hat zu fachlichen Diskussionen geführt. So könnte die inhaltshomogene Itemblockbildung zu mehr Vertrauen seitens der zu testenden Personen führen, indem die erhöhte Augenscheinvalidität keine Zweifel über die der Befragung zugrundeliegenden Persönlichkeitseigenschaften aufkommen lässt. Ehrliche und valide Antworten könnten so gefördert werden.

Bei dieser empirischen Studie mit experimentellem Design wurden daher zwei Bedingungen eines Persönlichkeitsfragebogens realisiert. Einerseits eine Bedingung mit inhaltshomogener Blockbildung („Blocked“) und andererseits eine Bedingung mit der gleichmäßigen Verteilung der inhaltlich den verschiedenen Skalen zugeordneten Items über den gesamten Fragebogen hinweg („Randomized“). Untersucht wurde, welche Auswirkungen die so beeinflusste Itemreihenfolge auf die Dimensionalität des psychologisch-diagnostischen Verfahrens (im Sinne der Konformität mit dem dichotomen-logistischen Testmodell von Rasch [1960]) hatte.

Die Datenerhebung erfolgte computerisiert über das Internet, und die Zuteilung zu den beiden Versionen des Fragebogens randomisiert. Insgesamt standen danach Datensätze von 520 Studierenden unterschiedlichster Studienrichtungen zur Verfügung.

Für beide Teilstichproben wurden getrennt Likelihood-Ratio-Tests gerechnet, um zu überprüfen, in welcher das Rasch-Modell gilt beziehungsweise durch Ausschlüsse auffälliger Items eher erreicht werden kann. Gültigkeit des Rasch-Modells bedeutet, dass die Beantwortung der Items nur von der Fähigkeit der Person und der Schwierigkeit des Items (und vom Zufall) abhängt, nicht aber

von weiteren Variablen beeinflusst wird. Eindimensionalität in diesem Sinne konnte eher für die Teilstichprobe „Randomized“ festgestellt werden, da in dieser seltener Likelihood-Ratio-Tests signifikant waren beziehungsweise weniger auffällige Items ausgeschlossen werden mussten.

In einer weiteren Auswertung wurden die beiden Teilstichproben im Teilungskriterium *Testversion* direkt gegenübergestellt. Dabei konnte festgestellt werden, dass sich die Itemparameter in manchen Skalen bei den zwei Bedingungen unterscheiden. Auch dieser Befund spricht also für die Beeinflussung des Antwortverhaltens durch die Reihenfolge der Items.

Gleichzeitig allerdings sprechen die Ergebnisse gegen die Hypothese, dass die Bedingung „Blocked“ günstige Auswirkungen auf die Dimensionalität des Persönlichkeitsfragebogens hat. Denn wie erwähnt, war es nicht diese Bedingung, sondern „Randomized“ in welcher Likelihood-Ratio-Tests eher Eindimensionalität anzeigten. Das anfangs erwähnte, plausibel klingende Argument für die inhaltshomogene Itemblockbildung konnte in dieser Studie empirisch nicht bekräftigt werden. Eventuell ist die aus dieser Bedingung resultierende Eintönigkeit doch zu gravierend.

## 6 Literaturverzeichnis

- Anderson, N. (1974). Cognitive algebra. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1–101). New York: Academic Press.
- Anderson, N. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Anger, H. (1969). Befragung und Erhebung. [Questioning and Survey]. In C. F. Graumann (Ed.), *Sozialpsychologie* (Vol. 7/1). Göttingen: Hogrefe.
- Becker, P. (2003). *Persönlichkeitsfragebogen*. In Kubinger, K.D. und Jäger, R.S. (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 332–337). Weinheim, Basel, Berlin: Beltz PVU.
- Borkenau, P. & Ostendorf, F. (1993). Das NEO-Fünf-Faktoren-Inventar (NEO-FFI). Göttingen: Hogrefe.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Champion, D.J. & Sear, A.M. (1968/69). Questionnaire response rate: A methodological analysis. *Social Forces*, 47, 335–339.
- Derogatis, L.R. (1986). SCL-90-R. Self-Report Symptom Inventory. In Collegium Internationale Psychiatriae Sclerum (Hrsg.), *Internationale Skalen für Psychiatrie*. Weinheim: Beltz.
- Eysenk, H.J., Wilson, C.D. & Jackson, C.J. (1998). Eysenck Personality Profiler (EPP-D). Frankfurt/M.: Swets.
- Fahrenberg, J., Hampel, R. & Selg, H. (1984). *Das Freiburger Persönlichkeitsinventar* (revidierte Aufl.). Göttingen: Hogrefe.
- Fischer, G.H. und Molenaar, J.W. (Eds.), (1995). *Rasch Models. Foundations, Recent Developments and Applications*. New York: Springer.

- Flanagan, J.C. (1954). The critical incident technique. *Psychological-Bulletin*, 51(4), 327–358.
- Franke, G.H. (2002). Faking bad in personality inventories: Consequences for the clinical context. *Psychologische Beiträge*, 44(1), 50–61.
- Franke, G.H. & Stäcker, K.H. (1995). Reliabilität und Validität der Symptom-Check-Liste (SCL-90-R; Derogatis, 1986) bei Standardreihenfolge versus inhaltshomogener Itemblockbildung. *Diagnostica*, 41(4), 349–373.
- Glas, A.W. & Verhelst, N.D. (1995). Testing the Rasch-Model. In G. H. Fischer und I. W. Molenaar (Eds.), *Rasch models* (S. 69–95). New York: Springer.
- Hackman, J.R. & Oldham, G.R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, 60, 159–179.
- Jonkisz, E., Moosbrugger, H. & Brandt, H. (2007). Planung und Entwicklung von psychologischen Tests und Fragebogen. In Moosbrugger, H. & Kelava, A. (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 27 – 71). Heidelberg: Springer.
- Khorramdel, L., Maurer, M. & Kubinger, K.D. (2008). *A requirement analysis of study specific demands – What requirements of ability and personality do students need to be successful?* 29th International Congress of Psychology, Berlin, 20th-25th July 2008 (Poster).
- Knowles, E.S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, 55(2), 312–320.
- Krahé, B. & Herrmann J. (2003). Verfälschungstendenzen im NEO-FFI: Eine experimentelle Überprüfung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(2), 105–117.
- Krampen, G. (1991). *Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK)*. Göttingen: Hogrefe.

- Krampen, G. (1993). Effekte von Bewerbungsinstruktionen und Subskalenextraktion in der Fragebogendiagnostik. *Diagnostica*, 39(2), 97–108.
- Krampen, G., Hense, H. & Schneider, J.F. (1992). Reliabilität und Validität von Fragebogenskalen bei Standardreihenfolge versus inhaltshomogener Blockbildung ihrer Items. *Zeitschrift für experimentelle und angewandte Psychologie*, 39, 229-248.
- Kubinger, K.D. (Hrsg.), (1989). *Moderne Testtheorie. Ein Abriss samt neuesten Beiträgen* (2. Aufl.). Weinheim, Basel: Beltz.
- Kubinger, K.D. (Edt.), (2002). Personality questionnaires: Some critical points of view [Themenheft]. *Psychologische Beiträge*, 44(1).
- Kubinger, K.D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Löhr, F.J. & Angleitner, A. (1980). Eine Untersuchung zu sprachlichen Formulierungen der Items in deutschen Persönlichkeitsfragebogen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 217–235.
- Mair, P. & Hatzinger, R. (2007). Extended Rasch Modelling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 20(9).
- Maurer, M. (2009). *Wiener Studieneignungs-Persönlichkeitsinventar (WSP): Formulierung eines Regelkataloges zur Itemgenerierung und dessen Anwendung zur Erstellung eines Itempools zur Erfassung studienrelevanter Eigenschaften*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Mayring, P. (Hrsg.), (2005). *Die Praxis der qualitativen Inhaltsanalyse*. Weinheim: Beltz.
- Moosbrugger, H & Kelava, A. (Hrsg.), (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.

- Mummendey, H.D. (2007). Die Fragebogen-Methode (5. Aufl.). Göttingen, Niedersachs: Hogrefe.
- Ortner, T. M. (2004). On changing the position of items in personality questionnaires: Analysing effects of item sequence using IRT. *Psychology Science*, 46(4), 466–476.
- Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modelling): Anwendung des Rasch-Modells (1-PL Modell)*. Lengrich, Wien: Papst Science Publishers.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Rogers, T.B. (1974). An Analysis of two central stages underlying responding to personality items: the self-referent decision and response selection. *Journal of Research in Personality*, 8, 128–138.
- Rost, D.H. & Hoberg, K. (1997). Itempositionsveränderungen in Persönlichkeitsfragebogen: Methodischer Kunstfehler oder tolerierbare Praxis? *Diagnostica*, 43(2), 97-112.
- Rost, D.H. & Sparfeldt, J.R. (2002). Facetten des schulischen Selbstkonzepts. Ein Verfahren zur Messung des differentiellen Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter). *Diagnostica*, 48, 130–140.
- Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion (2. Aufl.). Bern: Verlag Hans Huber.
- Schriesheim, C.A. & DeNisi, A.S. (1980). Item presentation as an influence on questionnaire validity: A field experiment. *Educational and Psychological Measurement*, 40(1), 175–182.

- 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(3), 407-508.
- Schwarz, N., Strack, F. & Mai, H.P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, 55, 3–23.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105.
- Sparfeldt, J.R., Schilling, S.R., Rost D.H. & Thiel A. (2006). Blocked Versus Randomized Format of Questionnaires: A Confirmatory Multigroup Analysis. *Educational and Psychological Measurement*, 66(6), 961–974.
- Tourangeau, R., & Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3), 299–314.
- Wegener, B., Faulbaum, F. & Maag, G. (1982). Die Wirkung von Antwortvorgaben bei Kategorienskalen. *ZUMA-Nachrichten*, 10, 3–20.



## 7 Tabellenverzeichnis

|   |    |
|---|----|
| <i>Tabelle 1. Deskriptive Statistiken und Häufigkeiten zu Geschlechterverteilung, Alter und Semesterstatus der Personen in den zwei Bedingungen .....</i> | 36 |
| <i>Tabelle 2. Häufigkeiten der Studienfächer und Studienrichtungen in der Stichprobe .....</i>  | 37 |
| <i>Tabelle 3. Modelltest für die Skala „Organisation“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm. ....</i>                 | 42 |
| <i>Tabelle 4. Modelltest für die Skala „Organisation“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm. ....</i>                    | 43 |
| <i>Tabelle 5. Modelltest für die Skala „Arbeitsstil, Arbeitsverhalten“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm. .</i>   | 44 |
| <i>Tabelle 6. Modelltest für die Skala „Arbeitsstil, Arbeitsverhalten“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm. ....</i>   | 44 |
| <i>Tabelle 7. Modelltest für die Skala „Motivation“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm. ....</i>                   | 45 |
| <i>Tabelle 8. Modelltest für die Skala „Motivation“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm. ....</i>                      | 45 |
| <i>Tabelle 9. Modelltest für die Skala „Belastbarkeit“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm. ....</i>                | 46 |
| <i>Tabelle 10. Modelltest für die Skala „Belastbarkeit“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm. ....</i>                  | 46 |
| <i>Tabelle 11. Modelltest für die Skala „Soziale Kompetenz“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm. ....</i>           | 47 |
| <i>Tabelle 12. Modelltest für die Skala „Soziale Kompetenz“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm. ....</i>              | 47 |

|   |    |
|---|----|
| <b>Tabelle 13.</b> Modelltest für die Skala „Emotionale Komponenten“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm. .                     | 48 |
| <b>Tabelle 14.</b> Modelltest für die Skala „Emotionale Komponenten“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm. ....                     | 48 |
| <b>Tabelle 15.</b> Modelltest für die Skala „Selbstbild“ der Teilstichprobe „Randomized“ als Ergebnisausdruck aus dem Programm eRm.....                               | 49 |
| <b>Tabelle 16.</b> Modelltest für die Skala „Selbstbild“ der Teilstichprobe „Blocked“ als Ergebnisausdruck aus dem Programm eRm.....                                  | 49 |
| <b>Tabelle 17.</b> Nicht signifikanter Modelltest für das Teilungskriterium Testversion der Skala „Organisation“ als Ergebnisausdruck aus dem Programm eRm.....       | 52 |
| <b>Tabelle 18.</b> Modelltest für das Teilungskriterium Testversion der Skala „Arbeitsstil, Arbeitsverhalten“ als Ergebnisausdruck aus dem Programm eRm.              | 53 |
| <b>Tabelle 19.</b> Modelltest für das Teilungskriterium Testversion der Skala „Motivation“ als Ergebnisausdruck aus dem Programm eRm.....                             | 54 |
| <b>Tabelle 20.</b> Nicht signifikanter Modelltest für das Teilungskriterium Testversion der Skala „Belastbarkeit“ als Ergebnisausdruck aus dem Programm eRm.....      | 55 |
| <b>Tabelle 21.</b> Nicht signifikanter Modelltest für das Teilungskriterium Testversion der Skala „Soziale Kompetenz“ als Ergebnisausdruck aus dem Programm eRm. .... | 56 |
| <b>Tabelle 22.</b> Modelltest für das Teilungskriterium Testversion der Skala „Emotionale Komponenten“ als Ergebnisausdruck aus dem Programm eRm....                  | 57 |
| <b>Tabelle 23.</b> Nicht signifikanter Modelltest für das Teilungskriterium Testversion der Skala „Selbstbild“ als Ergebnisausdruck aus dem Programm eRm. ....        | 58 |

## 8 Abbildungsverzeichnis

|   |    |
|---|----|
| <b>Abbildung 1.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Organisation“ .....  | 52 |
| <b>Abbildung 2.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Arbeitsstil, Arbeitsverhalten“ vor dem Ausschluss von Items (links) und nach Ausschluss von einem Item (rechts)..... | 53 |
| <b>Abbildung 3.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Motivation“ vor dem Ausschluss von Items (links) und nach Ausschluss von 4 Item (rechts). .....                      | 54 |
| <b>Abbildung 4.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Belastbarkeit“ .....   | 55 |
| <b>Abbildung 5.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Soziale Kompetenz“ .....   | 56 |
| <b>Abbildung 6.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Emotionale Komponenten“ vor dem Ausschluss von Items (links) und nach Ausschluss von 3 Item (rechts).....            | 57 |
| <b>Abbildung 7.</b> Graphische Modellkontrolle für das Teilungskriterium <i>Testversion</i> der Skala „Selbstbild“ .....  | 58 |



## 9 Anhang

### 9.1 Skalen und Subskalen des WSP

|   |                          |          |
|---|--------------------------|----------|
| <b>Organisation<br/>(78 Items)</b>                      | Zeitmanagement           | 24 Items |
|   | Organisationstalent      | 15 Items |
|   | Informiertheit           | 9 Items  |
|   | Prioritäten setzen       | 13 Items |
|   | Planen                   | 17 Items |
| <b>Arbeitsstil,<br/>Arbeitsverhalten<br/>(72 Items)</b> | Eigeninitiative          | 16 Items |
|   | Selbstständiges Arbeiten | 12 Items |
|   | Genaueres Arbeiten       | 19 Items |
|   | Gewissenhaftigkeit       | 25 Items |
| <b>Motivation<br/>(93 Items)</b>                        | Zielsetzung              | 12 Items |
|   | Zielorientierung         | 17 Items |
|   | Erfolgszuversicht        | 17 Items |
|   | Lernbereitschaft         | 16 Items |
|   | Durchhaltevermögen       | 12 Items |
|   | Engagement               | 18 Items |
| <b>Belastbarkeit<br/>(21 Items)</b>                     | Frustrationstoleranz     | 21 Items |
| <b>Soziale Kompetenz<br/>(36 Items)</b>                 | Kommunikationsfähigkeit  | 9 Items  |
|   | Kontaktfähigkeit         | 9 Items  |
|   | Teamfähigkeit            | 18 Items |
| <b>Emotionale<br/>Komponenten<br/>(32 Items)</b>        | Distanzierungsfähigkeit  | 15 Items |
|   | Offenheit                | 17 Items |
| <b>Selbstbild<br/>(18 Items)</b>                        | Selbstbewusstsein        | 18 Items |

## 9.2 Skalenstruktur der zwei Bedingungen des WSP

| <b>Bedingung „Randomized“<sup>2</sup></b> |                               | <b>Bedingung „Blocked“<sup>3</sup></b> |                               |
|---|-------------------------------|--|-------------------------------|
| <b>M<sub>1</sub></b>                      | Zielsetzung                   | <b>SB</b>                              | Selbstbewusstsein             |
| <b>O<sub>1</sub></b>                      | Informiertheit                | <b>B</b>                               | Frustrationstoleranz          |
| <b>A<sub>1</sub></b>                      | Eigeninitiative               | <b>E<sub>1</sub></b>                   | Distanzierungsfähigkeit       |
| <b>E<sub>1</sub></b>                      | Distanzierungsfähigkeit       | <b>E<sub>2</sub></b>                   | Offenheit                     |
| <b>S<sub>1</sub></b>                      | Kontaktfähigkeit              | <b>S<sub>1</sub></b>                   | Kontaktfähigkeit              |
| <b>M<sub>2</sub></b>                      | Erfolgszuversicht             | <b>S<sub>2</sub></b>                   | Kommunikationsfähigkeit       |
| <b>O<sub>2</sub></b>                      | Zeitmanagement                | <b>S<sub>3T</sub></b>                  | <i>Teamorientierung</i>       |
| <b>A<sub>2</sub></b>                      | genaues Arbeiten              | <b>S<sub>3KB</sub></b>                 | <i>Kompromissbereitschaft</i> |
| <b>E<sub>2</sub></b>                      | Offenheit                     | <b>S<sub>3KO</sub></b>                 | <i>Kollegialität</i>          |
| <b>B</b>                                  | Frustrationstoleranz          | <b>M<sub>1</sub></b>                   | Zielsetzung                   |
| <b>S<sub>2</sub></b>                      | Kommunikationsfähigkeit       | <b>M<sub>3</sub></b>                   | Zielorientierung              |
| <b>M<sub>3</sub></b>                      | Zielorientierung              | <b>M<sub>2</sub></b>                   | Erfolgszuversicht             |
| <b>O<sub>3</sub></b>                      | Prioritäten setzen            | <b>M<sub>5</sub></b>                   | Lernbereitschaft              |
| <b>A<sub>3</sub></b>                      | Gewissenhaftigkeit            | <b>M<sub>4</sub></b>                   | Durchhaltevermögen            |
| <b>S<sub>3KB</sub></b>                    | <i>Kompromissbereitschaft</i> | <b>M<sub>6</sub></b>                   | Engagement                    |
| <b>M<sub>4</sub></b>                      | Durchhaltevermögen            | <b>A<sub>1</sub></b>                   | Eigeninitiative               |
| <b>O<sub>4</sub></b>                      | Planen                        | <b>A<sub>4</sub></b>                   | selbstständiges Arbeiten      |
| <b>A<sub>4</sub></b>                      | selbstständiges Arbeiten      | <b>A<sub>2</sub></b>                   | genaues Arbeiten              |
| <b>S<sub>3T</sub></b>                     | <i>Teamorientierung</i>       | <b>A<sub>3</sub></b>                   | Gewissenhaftigkeit            |
| <b>M<sub>5</sub></b>                      | Lernbereitschaft              | <b>O<sub>2</sub></b>                   | Zeitmanagement                |
| <b>O<sub>5</sub></b>                      | Organisationstalent           | <b>O<sub>5</sub></b>                   | Organisationstalent           |
| <b>SB</b>                                 | Selbstbewusstsein             | <b>O<sub>1</sub></b>                   | Informiertheit                |
| <b>M<sub>6</sub></b>                      | Engagement                    | <b>O<sub>3</sub></b>                   | Prioritäten setzen            |
| <b>S<sub>3KO</sub></b>                    | <i>Kollegialität</i>          | <b>O<sub>4</sub></b>                   | Planen                        |

<sup>2</sup> Auf ein Item „Zielsetzung“ folgt eines aus „Informiertheit“, darauf eines aus „Eigeninitiative“...

<sup>3</sup> Auf die Items „Selbstbewusstsein“ folgen die Items „Frustrationstoleranz“ usw.

### 9.3 Instruktion des WSP

#### Erste Seite:

##### Willkommen!

Mit diesem Fragebogen werden Eigenschaften und Verhaltensweisen erfasst, die für die erfolgreiche Bewältigung des Studiums besonders relevant sind.

Auf den nächsten Seiten finden Sie insgesamt 217<sup>4</sup> Aussagen über typische Situationen und Verhaltensweisen im Studienalltag.

Lesen Sie bitte jede dieser Aussagen aufmerksam durch und geben Sie an, inwiefern diese **grundsätzlich** auf Sie persönlich zutreffen oder nicht.

Sollten Situationen beschrieben sein, die Sie noch nicht persönlich erlebt haben, dann schätzen Sie bitte ein, wie Sie sich in solchen Situationen **grundsätzlich** am ehesten verhalten würden.

<sup>3</sup>Einige der Aussagen sind sich in der Formulierung ähnlich. Jede der 217<sup>4</sup> Aussagen unterscheidet sich jedoch inhaltlich von den anderen, bitte lassen Sie sich nicht davon irritieren.

#### Zweite Seite:

Ihre individuellen Ergebnisse können Sie mit einem Ihnen zufällig zugeteilten Personencode nach der Testung anonym abrufen. Ihre Daten werden selbstverständlich anonym und vertraulich behandelt und ausschließlich für Forschungszwecke genutzt.

Ihr Personencode lautet: -----<sup>5</sup>

**Bitte notieren Sie sich unbedingt diesen Code! Sie benötigen diesen um später ihre Ergebnisse einzusehen!**

---

<sup>4</sup> Hier erscheint die Itemanzahl jenes Testheftes, welches zufällig zugeteilt wurde.

<sup>5</sup> Hier erscheint der zufällig generierte, zwölfstellige Zahlen-Buchstaben-Code.

**Dritte Seite:**

Zur Beurteilung der Aussagen steht Ihnen eine vierfach abgestufte Skala zur Verfügung. Klicken Sie bitte immer die für Sie am ehesten zutreffende Antwort an.

**Durch das Anklicken der Antwort wird automatisch die nächste Frage angezeigt!**

Probieren Sie das bitte anhand des folgenden Beispiels aus!

|   |
|---|
| <b>Ich führe einen Kalender, um mir wichtige Termine zu notieren.</b> |
|---|

trifft zu:

|                        |
|------------------------|
| <b>ja</b> <sup>6</sup> |
|------------------------|

|                             |
|-----------------------------|
| <b>eher ja</b> <sup>6</sup> |
|-----------------------------|

|                               |
|-------------------------------|
| <b>eher nein</b> <sup>6</sup> |
|-------------------------------|

|                          |
|--------------------------|
| <b>nein</b> <sup>6</sup> |
|--------------------------|

**Vierte Seite:**

**Genau so geht's!**

Sie werden bei der Befragung die Möglichkeit haben über einen entsprechenden Link zu der zuletzt beantworteten Aussage zurückzukehren.

**Bedenken Sie, dass es keine richtigen oder falschen Antworten gibt. Antworten Sie so, wie es für Sie persönlich am ehesten zutrifft. Nur dann können Sie von den Ergebnissen des Fragebogens profitieren.**

Klicken Sie bitte auf *Weiter*, damit Ihre Befragung beginnt.

---

<sup>6</sup> Durch ein Anklicken dieser Antwort wurde die nächste Seite der Instruktion angezeigt.

## 9.4 Unvollständiges Design

| <b>Subskala Selbstbewusstsein</b> |              |              |              |              |
|-----------------------------------|--------------|--------------|--------------|--------------|
|                                   | Testheft I   | Testheft II  | Testheft III | Testheft IV  |
| Item 392                          | enthalten    | enthalten    |              |              |
| Item 398                          | enthalten    | enthalten    |              |              |
| Item 404                          | enthalten    | enthalten    |              |              |
| Item 395                          |              | enthalten    | enthalten    |              |
| Item 407                          |              | enthalten    | enthalten    |              |
| Item 401                          |              | enthalten    | enthalten    |              |
| Item 405                          |              |              | enthalten    | enthalten    |
| Item 399                          |              |              | enthalten    | enthalten    |
| Item 393                          |              |              | enthalten    | enthalten    |
| Item 394                          | enthalten    |              |              | enthalten    |
| Item 400                          | enthalten    |              |              | enthalten    |
| Item 406                          | enthalten    |              |              | enthalten    |
| Item 396                          |              | enthalten    |              | enthalten    |
| Item 402                          |              | enthalten    |              | enthalten    |
| Item 408                          |              | enthalten    |              | enthalten    |
| Item 397                          | enthalten    |              | enthalten    |              |
| Item 403                          | enthalten    |              | enthalten    |              |
| Item 409                          | enthalten    |              | enthalten    |              |
| $\Sigma = 18$                     | $\Sigma = 9$ | $\Sigma = 9$ | $\Sigma = 9$ | $\Sigma = 9$ |

| <b>Subskala Distanzierungsfähigkeit</b> |               |               |               |               |
|---|---------------|---------------|---------------|---------------|
|   | Testheft I    | Testheft II   | Testheft III  | Testheft IV   |
| Item 360                                | enthalten     | enthalten     | enthalten     |               |
| Item 364                                | enthalten     | enthalten     | enthalten     |               |
| Item 368                                | enthalten     | enthalten     | enthalten     |               |
| Item 372                                | enthalten     | enthalten     | enthalten     |               |
| Item 361                                |               | enthalten     | enthalten     | enthalten     |
| Item 365                                |               | enthalten     | enthalten     | enthalten     |
| Item 369                                |               | enthalten     | enthalten     | enthalten     |
| Item 373                                |               | enthalten     | enthalten     | enthalten     |
| Item 362                                | enthalten     | enthalten     |               | enthalten     |
| Item 366                                | enthalten     | enthalten     |               | enthalten     |
| Item 370                                | enthalten     | enthalten     |               | enthalten     |
| Item 374                                | enthalten     | enthalten     |               | enthalten     |
| Item 363                                | enthalten     |               | enthalten     | enthalten     |
| Item 367                                | enthalten     |               | enthalten     | enthalten     |
| Item 371                                | enthalten     |               | enthalten     | enthalten     |
| $\Sigma = 15$                           | $\Sigma = 11$ | $\Sigma = 12$ | $\Sigma = 11$ | $\Sigma = 11$ |

## 9.5 Homepage Besser-Studieren.at

### **WAS... kennzeichnet erfolgreiche Studierende?**

Organisatorische Fähigkeiten, Arbeitsstil, Belastbarkeit, (Leistungs-)Motivation, Soziale Kompetenzen und einige andere Eigenschaften konnten in einer Studie der Test- und Beratungsstelle der Fakultät für Psychologie an der Universität Wien mittels Experteninterviews als studienrelevante Anforderungen identifiziert werden.

**Der auf Basis dieser Eigenschaften konstruierte Fragebogen lässt dich eigene Stärken und auch Schwächen erkennen und hilft dir dabei deine Leistung zu verbessern!**

### **WIE... kann dir diese Homepage helfen besser zu studieren?**

Teste dich mit dem kostenlosen und anonymen Fragebogen selbst und finde so deine Stärken und Schwächen in Bezug auf studienrelevante Fähigkeiten und Eigenschaften. Profitiere aus einem ausführlichem Feedback (schriftliche Ergebnismeldung) nach der Fragebogenbearbeitung und zusätzlichen Tipps und Links, die dir gemeinsam mit deinem individuellen Ergebnis rückgemeldet werden.

Erhalte außerdem im Forum Tipps von KollegInnen, wie du das Studium effizienter bewältigen kannst. Beteilige Dich an Diskussionen zum Thema Studienerfolg und tausche dich über Strategien aus, die helfen, den Studienalltag zu erleichtern.

### **WER... ist für diese Website verantwortlich?**

Diese Website wurde im Rahmen meiner Diplomarbeit erstellt und die hier erhobenen Testergebnisse stellen die Grundlage meiner Arbeit dar. Mein Name ist Norbert Dörrer. Ich bin Diplomand an der Fakultät für Psychologie und werde bei meiner Arbeit von Herrn Univ.Prof. Dr. Mag. Klaus Kubinger und Frau Mag. Lale Khorramdel betreut.

Für Feedback und Fragen bin ich unter folgender E-mail Adresse erreichbar:  
[besser-studieren@gmx.at](mailto:besser-studieren@gmx.at)

## ZUM FRAGEBOGEN...

### **Wie verhalten sich besonders erfolgreiche Studierende im Studienalltag? Welche Persönlichkeitseigenschaften sind dabei hilfreich?**

Aus einer Umfrage unter UniversitätsprofessorInnen und Studierenden aus 8 verschiedenen Studienrichtungen (darunter human- und sozialwissenschaftliche, technische, wirtschaftswissenschaftliche, rechtswissenschaftliche) gingen mehrere solcher Eigenschaften hervor! Der zur Verfügung gestellte Fragebogen wurde an der Test- und Beratungsstelle des Arbeitsbereiches Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien im Rahmen der Diplomarbeit von Frau Bakk.Phil. Martina Maurer entwickelt

Hier hast du jetzt die Möglichkeit deine eigenen Stärken und Schwächen kennen zu lernen!

### **Was sind deine Vorteile?**

Bitte nimm dir **rund 30 Minuten Zeit** und fülle den Fragebogen aus. Dabei wirst du mit Situationen des Studienalltags konfrontiert, wie du sie wahrscheinlich schon erlebt hast. Schätze ein, wie du dich in diesen Situationen grundsätzlich verhältst!

Danach kannst du jederzeit **dein individuelles Testergebnis und darauf abgestimmte konkrete Empfehlungen** abrufen. Das sind Literaturtipps, weiterführende Websites aber auch studienrelevante Softwareempfehlungen, die dir dabei helfen sollen, deine Leistungen in bestimmten Bereichen des Studienalltags zu steigern.

Sobald meine Datenerhebung abgeschlossen ist (voraussichtlich Mai 2009), hast du zudem die Möglichkeit herauszufinden, wo im **Vergleich zu den anderen Teilnehmern** deine ganz persönlichen Stärken liegen. Hierzu wird dir der Prozentsatz jener Personen rückgemeldet, die deines oder ein niedrigeres Ergebnis erreicht haben.

## WEITER ZUM FRAGEBOGEN<sup>7</sup>...

## ZU MEINEN ERGEBNISSEN<sup>8</sup>...

---

<sup>7</sup> Dieser Link führt direkt zu einem zufällig zugeteilten Testheft des WSP.

<sup>8</sup> Dieser Link führt nach Eingabe des Personencodes zu der individuellen Ergebnisseite.

## **ZUM FORUM...**

### **Was bringt dir das Diskussionsforum?**

#### **Beteilige dich an Diskussionen zum Thema Studienerfolg!**

Was bedeutet Erfolg im Studium? Ist es der schnelle Abschluss oder ein toller Notenschnitt? Was macht Studienerfolg nach Meinung der Studierenden aus? Prüfungsvorbereitung aus Mitschriften und Skripten oder die intensive Auseinandersetzung mit dem Stoffgebiet durch das Lesen zusätzlicher (nicht geforderter) Literatur?

#### **Lerne die Meinungen von Lehrveranstaltungsleitern und Praktikern aus der Wirtschaft kennen zum Thema Studienerfolg!**

Wie definieren sie Studienerfolg? Und was sollten Studierende ihrer Meinung nach beachten, um effizienter und besser zu studieren?

#### **Profitiere aus den Erfahrungen anderer Studierenden!**

Was würden sie tun, wenn sie noch einmal zu studieren anfangen würden? Was können sie dir raten, damit du nicht die gleichen Fehler machen musst wie sie?

#### **Tausche dich über Strategien aus, die den Studienalltag erleichtern helfen!**

Wie motivierst du dich vor und für schwierige Prüfungen? Wie gehst du mit negativen Emotionen wie Prüfungsangst um? Was hilft dir bei Lernblockaden? Was kannst du anderen Studierenden mit ähnlichen Problemen raten?

#### **Zusätzlich gibt es einen eigenen Bereich zum Thema „Schreiben wissenschaftlicher Abschlussarbeiten“!**

Welche Software kann beim erstellen und formatieren wissenschaftlicher (Abschluss-) Arbeiten verwendet werden? Wie können sie einem das Arbeiten erleichtern?

## 9.6 Abstract

Aufgabe dieser Diplomarbeit war es, den Einfluss des Itemkontextes auf die Bearbeitung eines multidimensionalen Persönlichkeitsfragebogens zu überprüfen. Da der Itemkontext maßgeblich durch die Reihenfolge der Items determiniert wird, wurden in einem experimentellen Design zwei Fragebogenversionen desselben Verfahrens realisiert. Da insbesondere die Auswirkungen der inhaltshomogenen Itemblockbildung interessierten, wurde die Versuchsgruppe mit dem solcher Art gestalteten Fragebogen konfrontiert. Dahingegen wurde der Kontrollgruppe derselbe Fragebogen mit gleichverteilter Reihenfolge präsentiert. In beiden Gruppen wurden Modelltests hinsichtlich des dichotom-logistischen Testmodells von Rasch durchgeführt. Zudem wurden die beiden Gruppen im Teilkriterium *Testversion* gegenübergestellt. Es ergaben sich verschiedene Itemschwierigkeitsparameter in drei der sieben Skalen in Abhängigkeit von der Itemreihenfolge. Die Befunde sprechen weiters für eine Überlegenheit der gleichverteilten Reihenfolge gegenüber der inhaltshomogenen Itemblockbildung.

This thesis deals with the influence of the item context on the processing of a personality questionnaire. Two experimental groups were formed to assess the impact of item order: One group consisting of the items of the same scales blocked together and the other group consisting of the items in a conventional order (uniformly distributed). By applying the dichotomous Rasch model for both groups, advantages of the conventional order could be shown. By applying this test model with the splitting criterion *testversion* for the whole sample, various difficulties for items had been caused by the item order in three of the seven scales.



## 9.7 Lebenslauf

**Norbert Dörrer**, geboren am 14. August 1982 in Wien

Haslingergasse 72 / 2, 1170 Wien

Tel.: 0660 / 5551585, E-Mail: [n.doerrerr@gmx.at](mailto:n.doerrerr@gmx.at)

|                       |   |
|-----------------------|---|
| 1992 – 2000           | Bundesgymnasium Waidhofen an der Thaya  |
| Juni 2000             | Matura am Bundesgymnasium Waidhofen an der Thaya  |
| Oktober 2000          | Beginn des Diplomstudiums Psychologie an der Universität Wien                                   |
| März – September 2005 | Präsenzdienst beim Heeressportwissenschaftlichen Dienst am Heeressport-Zentrum in Wien          |
| Juni und Juli 2007    | 6-Wochen-Pflichtpraktikum am Institut für Entwicklungspsychologie und Psychologische Diagnostik |