



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Untersuchung des Einflusses unterschiedlicher
Itemreihenfolgen im Wiener Studieneignungs-
Persönlichkeitsinventar (WSP) mittels LLTM: eine
experimentelle Variation.“

verfasst von

Julia Kober

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Dezember 2013

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Mag. Dr. Lale Khorramdel

DANKSAGUNGEN

Für die Unterstützung bei der Erstellung dieser Diplomarbeit möchte ich mich bei Mag. Dr. Lale Khorramdel für ihre engagierte Betreuung und hilfreichen Anregungen bedanken.

Auch möchte ich mich herzlich bei Mag. Dr. Christine Hohensinn für ihre freundliche Unterstützung und konstruktive Kritik bedanken.

Mein ganz besonderer Dank gilt MMag. Daniel Weigl, der mir beim Schreiben dieser Diplomarbeit stets helfend zur Seite stand.

Auch möchte ich mich bei allen bedanken, die mir mit alltäglicher motivationaler Unterstützung in der Phase des Schreibens dieser Diplomarbeit zur Seite standen.

Inhaltsverzeichnis

1	Einleitung	7
2	Theorie	8
2.1	Kontexteffekte	9
2.2	Zwei Modelle zur Entstehung von Kontexteffekten	12
2.3	Zur Itemreihenfolge in Persönlichkeitsfragebögen	15
3	Methodik.....	19
3.1	Die verwendeten statistischen Verfahren	19
3.1.1	Das dichotom logistische Modell von Rasch (1 PL-Modell)	19
3.1.2	Der Martin-Löf-Test.....	21
3.1.3	Das Linear Logistische Testmodell von Fischer (LLTM)	22
3.2	Untersuchungsplan und Hypothesen	24
3.2.1	Aufgestellte Hypothesen	27
3.2.2	Postulierte Modelle	28
3.3	Untersuchungsmaterial	31
3.3.1	Das Wiener Studieneignungs-Persönlichkeitsinventar (WSP)	31
3.3.2	Das Antwortformat des WSP	35
3.3.3	Die Entwicklung des Wiener Studieneignungs-Persönlichkeitsinventars (WSP).....	37
3.4	UntersuchungsteilnehmerInnen.....	40
3.4.1	Rekrutierung der UntersuchungsteilnehmerInnen.....	40
3.4.2	Deskriptive Beschreibung der Stichprobe	41
3.5	Durchführung der Untersuchung	47
3.5.1	Die Homepage www.hochschultest.at	47
3.5.2	Demografische Variablen	48
3.5.3	Darbietung der Items des WSP	49
3.5.4	Ergebnisrückmeldung an die TeilnehmerInnen	50

4	Ergebnisse	51
4.1	Prüfung der Abbruchraten der Versionen	52
4.2	Analyse der Homogenität der Subskalen innerhalb der Skalen des WSP mittels Martin-Löf-Tests	53
4.3	Analyse auf Rasch-Modell-Konformität	66
4.3.1	Die Skala „Organisation“	67
4.3.2	Die Skala „Arbeitsstil und Arbeitsverhalten“	68
4.3.3	Die Skala „Belastbarkeit“	70
4.3.4	Die Skala „Distanzierungsfähigkeit“	72
4.3.5	Die Skala „Offenheit“	73
4.3.6	Die Skala „Motivation“	75
4.3.7	Die Skala „Kontaktfähigkeit“	77
4.3.8	Die Skala „Teamfähigkeit“	78
4.3.9	Die Skala „Kommunikationsfähigkeit“	79
4.3.10	Die Skala „Selbstbild“	81
4.4	Analysen mittels Linear Logistischem Testmodell.....	83
4.4.1	Ergebnisse des Rasch-Modells	85
4.4.2	Ergebnisse des quasi-saturierten Modells	85
4.4.3	Ergebnisse des Itemkontextmodells	86
4.5	Modellvergleiche	86
4.5.1	Modellvergleich des quasi-saturierten Modells mit dem Itemkontextmodell..	87
4.5.2	Modellvergleich des Itemkontextmodell mit dem Rasch-Modell.....	88
4.5.3	Gegenüberstellung der drei Modelle	89
5	Diskussion (und Ausblick).....	91
6	Zusammenfassung	95
7	Literaturverzeichnis	100

8	Tabellenverzeichnis.....	103
9	Abbildungsverzeichnis	108
10	Anhang.....	109
10.1	Itemreihenfolgen der drei Fragebogenversionen	109
10.2	Der Text der Homepage <i>www.hochschultest.at</i>	111
10.3	Instruktion des webbasierten WSP	113
10.4	Rückmeldung der Ergebnisse an die TeilnehmerInnen.....	113
10.5	Liste der teilnehmenden Fachhochschulen Österreichs	115
10.6	Liste der studentischen universitätsnahen Foren, auf denen die Teilnahme an der Studie beworben wurde	116
11	Abstract	117
12	Curriculum Vitae.....	118

1 Einleitung

Nach Kubinger können Verfahren der psychologischen Diagnostik inhaltlich den Bereichen Leistungsdiagnostik, Persönlichkeitsdiagnostik und Diagnostik ‚hybrider‘ Eigenschaften zugeordnet werden (Kubinger, 2006: S. 183). In der Leistungsdiagnostik werden vorwiegend kognitive Fähigkeiten erfasst, wobei die diesbezüglich verwendeten Verfahren sich in Intelligenz-Testbatterien und spezielle Leistungstests gliedern.

Der offensichtlichste Unterschied zwischen Leistungs- und Persönlichkeitsverfahren besteht darin, dass für die Aufgaben bei Intelligenztests richtige und falsche Antwortmöglichkeiten definiert werden. Bezüglich der Abfolge der Items bestehen ebenfalls Unterschiede. So werden die Items von Intelligenz- und Leistungstests mehrheitlich in inhaltlichen Blöcken präsentiert, wobei z. B. beim IST 2000R die Items der Intelligenzdimension ‚verbale Intelligenz‘ als Untertests geblockt, gefolgt von den Items der Subtests der ‚numerischen Intelligenz‘ vorgegeben werden. Zur Messung von Persönlichkeitseigenschaften werden neben wenigen objektiven Persönlichkeitstests hauptsächlich Persönlichkeitsfragebögen zur Selbst- bzw. Fremdbeurteilung verwendet. Die übliche Vorgabemodalität dieser Fragebögen zur Persönlichkeitsdiagnostik besteht aus einer durchmischten Itemreihenfolge (Ortner, 2004: S. 467), welche häufig auch als randomisiert bezeichnet wird (selbst wenn die Zuordnung der Items häufig nicht randomisiert im eigentlichen Sinn erfolgt). Hierfür werden die Items einer Skala oder auch deren Subskalen mit inhaltlichem Abstand voneinander vorgegeben, sodass auf ein Item einer Skala jeweils ein Item einer anderen Skala folgt (die Wahl der Abfolge der Skalen ist dabei häufig nicht randomisiert, die Zuordnung der einzelnen Items ist es meist).

Auch der für diese Studie verwendete Fragebogen, das Wiener Studieneignungs-Persönlichkeitsinventar (WSP), wurde bisher mit dieser Art der Itemreihung vorgegeben. In der klinisch-diagnostischen Praxis kommt es allerdings öfter vor, dass zur Beantwortung von spezifischen Fragestellungen einzelne Skalen in geblockter Reihenfolge vorgegeben werden (ebd.: S. 467), wobei die entsprechenden Manuale oft keine Informationen darüber liefern, ob dies einen Einfluss auf die Skalierung und Validität der Messung hat. Auch für wissenschaftliche Studien werden die interessierenden Skalen oder Konstrukte häufig aus den entsprechenden Fragebögen extrahiert, wodurch deren ursprüngliche Fragereihenfolge verändert wird.

Die vorliegende Studie soll untersuchen, ob unterschiedliche Itemreihenfolgen einen Einfluss auf die Skalierung und somit auf die interne Validität des WSP haben. Dabei wird im speziellen der Einfluss einer durchmischten (bzw. randomisierten), einer skalenweise geblockten und einer skalen- und subskalenweise geblockten Itemreihenfolge auf das Antwortverhalten von StudentInnen beim WSP untersucht. Dies soll nicht nur Klarheit darüber schaffen, welche Vorgabevariante für das WSP geeigneter wäre sondern darüberhinaus auch beantworten, ob man die Skalen des WSP auch einzeln in geblockter Reihenfolge vorgeben könnte (je nachdem welche Skala für die jeweilige Fragestellung von Interesse ist) ohne die Qualität des Verfahrens dadurch zu beeinträchtigen. Da es sich hierbei um Positions- und Reihenfolge handelt möchte ich mit dem nächsten Abschnitt den Forschungsstand zu dieser Thematik, sofern er für die vorliegende Arbeit relevant erscheint, umreißen.

Da bei der Test- oder Fragebogenkonstruktion mögliche Positions- und Reihenfolgeeffekte bedacht werden sollten um die lokale stochastische Unabhängigkeit der Itemantworten¹ nicht zu beeinflussen, widmet sich der nächste Abschnitt dieser Thematik. Dabei können zwei Arten von Effekten unterschieden werden: Positionseffekte, welche die Beeinflussung des Antwortverhaltens der Items in Zusammenhang mit deren Stelle im Test oder Fragebogen beschreiben, und Reihenfolgeeffekte, welche einen Effekt auf die Itemantworten der vorhergehenden Aufgaben postulieren.

2 Theorie

Dieses Kapitel soll einen Überblick über Kontext- und Reihenfolgeeffekte in Einstellungsfragebögen liefern. Hierfür wird zuerst ein Überblick über die unterschiedlichen Arten von Kontexteffekten und ihre möglichen Einflüsse auf die Beantwortung von Einstellungsfragen gegeben. Im Anschluss daran werden zwei Modelle skizziert, die das Zustandekommen von Kontexteffekten in Einstellungsfragebögen erklären und damit erheblich zum Verständnis ihrer etwaigen Entstehungsbedingungen beitragen. Schließlich werden die Auswirkungen von Kontexteffekten bei unterschiedlicher Reihung der Fragen, den (Item)-Reihenfolgeeffekten erläutert.

¹ Unabhängigkeit der Lösungswahrscheinlichkeit des derzeitigen Items von zuvor bearbeiteten Aufgaben.

2.1 Kontexteffekte

An erster Stelle soll der der Thematik zugrundeliegende Bereich der Einstellung und ihrer psychometrischen Erfassung behandelt werden. Obwohl Einstellungen zu einem gewissen Thema Resultate der bisherigen Beschäftigung damit und der diesbezüglich gemachten Erfahrungen sind, basieren die Urteile bei Einstellungsfragen oft auf kurzfristigen Entscheidungen. Bei der Beantwortung einer Frage in einem Fragebogen beeinflusst nicht nur der thematische Inhalt, die Formulierung, die Polung und das Layout des Items (der entsprechenden Einstellungsfrage) das Einstellungsurteil sondern auch das Antwortformat bzw. die zur Verfügung stehenden Antwortmöglichkeiten und deren Formulierungen². Somit sind nach Schwarz (1999: S. 94-96) zum Verständnis der Inhalte der Items zwei ineinander verflochtenen Prozesse notwendig: Das rein sprachliche und das pragmatische Verständnis - Zweiteres kann Schlussfolgerungen über den Untersuchungsgegenstand, die Absichten der ForscherInnen oder auch deren erkenntnistheoretischen Zugang beinhalten. Anschließend müssen entsprechende Gedächtnisinhalte aktiviert werden, welche unter Einbezug von Aspekten sozialer Erwünschtheit und Selbstdarstellung zur Urteilsfindung herangezogen werden, schließlich muss dieses Einstellungsurteil noch in Einklang mit den vorgegebenen Antwortmöglichkeiten gebracht werden (ebd.: S. 97).

Darüber hinaus haben auch die vor dem betreffenden Item bearbeiteten Fragen – der Itemkontext – einen Einfluss auf das Einstellungsurteil. Dafür werden in der entsprechenden Literatur die Begriffe Kontexteffekt, (Item-) Reihenfolgeeffekt, Sequenzeffekt oder Positionseffekt verwendet. Die vorhergehenden Fragen fungieren dabei z. B. als impliziter Standard für das vorzunehmende Einstellungsurteil oder können die Zugänglichkeit (Steinberg, 1994: S. 341) zu bestimmten Überlegungen im Gedächtnis erleichtern.

Die vorliegende Arbeit beschäftigt sich mit solchen Reihenfolgeeffekten, welche von Rost (2004: S. 70) als „Beeinflussung der Itemantwort dadurch, welche anderen Items zuvor bearbeitet wurden“, definiert werden. Nach Tourangeau, Rips und Rasinski (2000: S. 228) treten solche Kontexteffekte auf, wenn ein konzeptioneller Zusammenhang zwischen den Themengebieten der aktuellen und jenen der bereits bearbeiteten Fragen besteht. Demnach

² In weiteren Definitionsansätzen kann unter Itemkontext darüber hinaus auch die Situation gemeint sein, in welcher der Fragebogen vorgegeben wird, z. B. die Einflüsse der TestleiterInnen und weiterer Anwesenden sowie die zum Zeitpunkt der Beantwortung vorherrschende Meinung zu dem Thema.

sollten thematisch ähnliche Fragen, welche hintereinander vorgegeben werden, zu konsistenteren Einstellungsurteilen führen. Ebenso können – vor allem bei wenig vertrauten Themengebieten – die der gerade zu bearbeitenden Frage vorausgehenden Items als thematisch einem Untersuchungsgegenstand angehörig angesehen werden, was Schlussfolgerungen oder daraus resultierende Erwartungen für die vorliegende Einstellungsfrage evoziert. Sofern die Fragen einer gewissen Thematik nach durchmischt über den ganzen Fragebogen verteilt werden und die Frageabfolge nacheinander unterschiedliche Inhalte vorsieht, sollten also weniger Überlegungen zum derzeitigen Fragengegenstand aktiviert sein – was weniger konsistente Einstellungsurteile nach sich ziehen könnte.

Dies wurde z. B. bei Einstellungsfragen zum Thema Abtreibung untersucht. Unter anderem konnte festgestellt werden, dass die Antwort auf die Frage nach dem Grad der Zustimmung zu legaler Abtreibung für den Fall, dass die Frau verheiratet war und keine Kinder mehr wollte, abhängig von der zuvor gestellten Frage war (Schuman, 1992; Schuman & Presser, 1981, zit. nach Tourangeau, Rips & Rasinski, 2000: S. 197). Wenn dieser Frage eine nach dem Grad der Zustimmung zu Abtreibung beim Vorliegen einer hohen Wahrscheinlichkeit für Geburtsdefekte des Kindes vorrausging, so sank der angegebene Zustimmungsggrad bei der Folgefrage. Auch konnte von McGuire (1960, zit. nach Tourangeau, Rips & Rasinski, 2000: S. 213) gezeigt werden, dass nacheinander vorgegebene Items in einem Fragebogen zu konsistenteren Antworten der Befragten führen, da – so folgert der Autor – dadurch der Fokus auf ihren logischen Zusammenhang gelenkt wird³.

Kontexteffekte können nach *unconditional* und *conditional context effects* unterschieden werden, wobei bei Ersteren der thematische Inhalt die nachfolgende Itembeantwortung beeinflusst, während bei Zweiteren auch von einem Einfluss der Antworten auf die folgenden Fragen ausgegangen werden kann (Smith, 1986, zit. nach Tourangeau, Rips & Rasinski, 2000: S. 199). Weiters können Kontexteffekte auch nach der Richtung der Beeinflussung der Antworten unterteilt werden: *Consistency effects* beschreiben Tendenzen, welche der Thematik zuvor bearbeiteter Fragen entsprechen, *contrast effects* bezeichnen gegenläufige Neigungen (Tourangeau, Rips & Rasinski, 2000: S. 201). Ein weiterer Aspekt von

³ T. W. Smith (1983, zit. nach Tourangeau, Rips & Rasinski, 2000: S. 213-214) konnte zeigen, dass dieser Effekt nicht nur bei logischen sondern auch bei thematischen Zusammenhängen besteht.

Kontexteffekten ist die veränderte Zugänglichkeit von Gedanken zum Thema der derzeitigen Frage aufgrund thematisch ähnlicher vorhergehender Items, welche *retrieval-based assimilation effects* genannt werden und unter anderem dazu führen können, dass Fragen schneller beantwortet werden (ebd.: S. 206). Anderson (1983) zeigte, dass diese *assimilation effects* nur eine gewisse Zeit andauern indem er thematisch ähnliche Fragen einerseits in Blöcken und andererseits zwischen unzusammenhängenden Items vorgab. Auch Tourangeau, Rips und Rasinski (2000: S. 207) konnten zeigen, dass der Assimilationseffekt stärker war, je ähnlicher die Themen der davor bearbeiteten Fragen waren. Dabei ist zu beachten, dass zuvor beantwortete, ähnliche Fragestellungen nicht immer zu Assimilationseffekten führen, sondern auch ignoriert werden und so zu subtraktiven Kontexteffekten oder Kontrasteffekten führen können. Bei der Fragebogenbeantwortung können Kontexteffekte auch die Übertragung der Einstellungsurteile auf die Antwortmöglichkeiten beeinflussen, dies kann auch dadurch bedingt sein, dass die Antworten zuerst redigiert werden, woraus wieder Assimilations- und Kontrasteffekte resultieren können (ebd.: S. 213). Gut überlegte Einstellungen, die bereits eine hohe Zugänglichkeit aufweisen, dürften generell weniger von Kontexteffekten beeinflussbar sein (ebd.: S. 217). Tourangeau (1992, zit. nach Tourangeau, Rips & Rasinski, 2000: S. 217) argumentiert, dass dies allerdings auch erst die Voraussetzung für das Zustandekommen gewisser Kontexteffekte sein könnte, da die Überlegungen zu den nacheinander in Fragen behandelten Inhalten im Gedächtnis vernetzt sein müssten, was wiederum eine gewisse Beschäftigung mit dem Gegenstand voraussetzen würde.

Im Hinblick auf die vermehrt computerisierte Vorgabe von Einstellungsfragebögen sind Kontexteffekte nach Ortner (2008: S. 255) besonders dann relevant, wenn entsprechende Persönlichkeitsinventare mittels *computerised adaptive testing* (CAT) gemessen werden, sodass die Itemreihenfolge von den individuellen Antwortpfaden der TeilnehmerInnen abhängen. Eventuell bestehende Itemkontexteffekte können so die Erfassung der zugrundeliegenden Persönlichkeitsstrukturen beeinflussen oder verfälschen.

Im folgenden Abschnitt wird auf zwei Modelle eingegangen, welche Erklärungsansätze für das Auftreten von Kontexteffekten in Einstellungsfragebögen liefern.

2.2 Zwei Modelle zur Entstehung von Kontexteffekten

Um generelle Voraussagen über die Stabilität von Antworten auf Einstellungsfragen basierend auf Veränderungen der Abruf- und Urteilsprozesse vornehmen zu können, stellten Tourangeau, Rips und Rasinski (2000: S. 179) das *Belief-sampling-model* auf. Einstellungen werden in diesem Modell als Ansammlung verschiedener Emotionen, Werte, Überzeugungen und Gedanken sowie als das Wissen um vorhergegangene Urteile zu einem gewissen Thema angesehen – welche Erwägungen daraus abgerufen werden ist abhängig von deren momentaner Zugänglichkeit. Diese Erwägungen, welche aus dem Langzeitgedächtnis abgerufen werden, und etwaige Diskrepanzen zwischen ihnen, werden in weiterer Folge für die Urteilsfindung mit den im Zuge der Befragung aufgetretenen neuen Argumenten, Urteilen, Einstellungen sowie mit Aspekten der sozialen Erwünschtheit und Selbstdarstellungsmotiven kombiniert. Dies wird als sukzessiver Anpassungsprozess verstanden, indem die Beantwortenden die betreffenden Einstellungen abrufen und für die vorliegende Frage daraus Schlussfolgerungen ableiten (ebd.: S. 180). Diesen Überlegungen zufolge postulierten Tourangeau, Rips und Rasinski das zu erstellende Einstellungsurteil J als Mittel der Einzelurteile s_i (ebd.):

$$J = \frac{\sum s_i}{n}$$

Um zu einem Einstellungsurteil zu gelangen, können die an diesem Prozess beteiligten Einstellungen auch ihrer Relevanz entsprechend gewichtet werden.

Die eben beschriebenen Annahmen weisen auf eine inhärente Instabilität von Antworten in Einstellungsfragebögen hin, da diese jeweils auf den zu diesem Zeitpunkt zugänglichen Erwägungen im Langzeitgedächtnis, die durch die Fragestellung aktiviert wurden, basieren. Aus diesem Modell leiten die Autoren ab, dass die Stabilität von Einstellungsurteilen einer zu verschiedenen Befragungszeitpunkten gestellten Frage von drei Faktoren abhängig ist – der Homogenität der dem Thema zugrundeliegenden Überlegungen, den Ähnlichkeiten der Überlegungen zu den gestellten Fragen, die zu jedem Befragungszeitpunkt angestellt werden, sowie den Beurteilungsprozessen um diese Informationen zu verarbeiten. Alle drei Faktoren können durch den Itemkontext beeinflusst werden (Tourangeau, Rips & Rasinski, 2000: S. 197).

Die zuvor beantworteten Fragen können einerseits zu Antworttendenzen *overall direction of answers* oder *directional context effects* – also gerichteten Kontexteffekten – führen (ebd.: S. 198) indem die Inhalte der vorhergehenden Fragen als Anker für die derzeitigen Einstellungsurteile fungieren. Diesbezüglich erwähnt Ortner, dass der Kontext, welcher, die vorhergehenden Fragen darstellen, die individuellen Anker verändern kann. Dies entspricht auch der Annahme Tourangeaus und Rasinskis (1998 zit. nach Ortner, 2008: S. 251), dass der jeweilige Kontext sowohl die Skalierung der Persönlichkeitseigenschaften, als auch die Auswahl der Informationen, die zur Beurteilung herangezogen werden, beeinflussen kann.

Zusammenfassend kann die Konsistenz der Antworten auf dieselbe Einstellungsfrage unter verschiedenen Bedingungen als von den folgenden drei Parametern abhängig beschrieben werden (Ortner, 2008: S. 225):

- Dem Grad der Konsistenz der Beurteilung der gleichen Überlegungen zu den verschiedenen Bedingungen,
- dem Grad der Konsistenz der potentiell relevanten Überlegungen und
- dem Überlappungsgrad dieser Überlegungen.

Somit postuliert das Modell, dass Kontexteffekte dann einen Einfluss auf das Antwortverhalten ausüben, wenn Überlegungen die Inhalte, aus denen die Einstellungsurteile resultieren, deren Wichtigkeit oder die Bewertung dieser Überlegungen beeinflussen (ebd.: S. 225).

Diese Annahmen sind jenen des *Inklusions-/Exklusions Modells* nach Schwarz und Bless (1992, zit. nach Tourangeau, Rips & Rasinski, 2000: S. 221-224) ähnlich, in dem die Autoren davon ausgehen, dass, um Einstellungsurteile zu einer gewissen Frage abgeben zu können, zuerst Repräsentationen des Frageninhaltes sowie eines Vergleichsmaßstabes gebildet werden müssen, welche anschließend zur Urteilsbildung herangezogen werden. Kontexteffekte können dementsprechend über die Zugänglichkeit relevanter Informationen die Repräsentation der Frage, des Vergleichsmaßstabs oder beider diesen Prozess beeinflussen (ebd.: S. 224). Nach Schwarz kann zwischen chronisch und temporär zugänglichen Informationen im Gedächtnis unterschieden werden, welche zur Bildung der Repräsentationen des Frageinhaltes sowie des Vergleichsstandards verwendet werden. Bereits bearbeitete Fragen haben Einfluss darauf, welche Informationen im Gedächtnis

temporär zugänglicher sind, was als Kontexteffekt verstanden werden kann – im Vergleich dazu sind chronisch zugängliche Aspekte eher kontextunabhängig (Schwarz, 1999: S. 100). Kontexteffekte können auch entstehen, wenn gewisse Normen der zwischenmenschlichen Konversation zum Tragen kommen, indem unter anderem Redundanz zu vermeiden versucht wird. In Einstellungsfragebögen kann dies beispielsweise dazu führen, dass zugängliche Informationen exkludiert werden, wenn diese in vorhergehenden Antworten schon mitgeteilt wurden (ebd.: S. 102). Damit können sowohl aus der Art des Umgangs mit den gerade zugänglichen Informationen und aus den sie begleitenden mentalen Interpretationsprozessen Assimilations- oder Kontrasteffekte resultieren.

Das Modell ermöglicht es abzuleiten, wann zugängliche Inhalte für die Repräsentation der derzeitigen Frage exkludiert werden, wofür es vier Bedingungen postuliert, von denen mindestens eine erfüllt sein muss (Schwarz & Bless, 1992 zit. nach Tourangeau, Rips & Rasinski, 2000: S. 224):

- Wirken sich die Aspekte auf die gestellte Frage aus?
- Werden gewisse Inhalte aufgrund von relevanten Einflüssen aktiviert?
- Sind die Überlegungen der gerade zu bearbeitenden Frage thematisch zugehörig?
- Soll die beantwortende Person diese verwenden?

Tourangeau, Rips und Rasinski gehen in ihrem *Belief-sampling-Modell* von der unterschiedlichen Zugänglichkeit von Einstellungsaspekten aus, welche in sukzessiven Anpassungsprozessen zu einem Einstellungsurteil führen. Auch im *Inklusions-/Exklusionsmodell* von Schwarz und Bless ist die Zugänglichkeit von Informationen, in Form von chronisch oder temporär zugänglichen Informationen, essentiell. Allerdings werden in diesem Modell zuerst Repräsentationen des Inhaltes einer Frage und des Vergleichsmaßstabes gebildet, welche dann für Urteile bei Einstellungsfragen herangezogen werden. Ähnlich dem *Inklusions-/Exklusionsmodell* von Schwarz und Bless (1992), in dem Inklusion zu Assimilationseffekten führt und Kontrasteffekte durch Exklusion verursacht werden, werden diese im *Belief-Sampling Modell* von Tourangeau, Rips und Rasinski (2000) durch Prozesse in den Abläufen der Abfrage, Urteilsbildung und Fragebeantwortung sowie mittels der vorgegebenen Antwortkategorien erklärt.

Der Überblick über die Erklärungsansätze in beiden Modelle sollte zum besseren Verständnis der Entstehungsbedingungen von Kontexteffekten in Einstellungsfragebögen beitragen. Der folgende Abschnitt behandelt nun das zentrale Anliegen dieser Arbeit: Den Spezialfall der Kontexteffekte in Form von Itemreihenfolgeeffekten.

2.3 Zur Itemreihenfolge in Persönlichkeitsfragebögen

Die Reihenfolge der Items in einem Persönlichkeitsfragebogen orientiert sich meist an den Annahmen des zugrundeliegenden Persönlichkeitskonzeptes und der von diesem vorgegebenen Struktur. Fragebögen, denen multidimensionale Persönlichkeitskonzepte zugrunde liegen, werden oft in durchmischter Itemreihenfolge vorgegeben, indem die Fragen einer Skala bzw. Subskala mit jenen anderer Skalen oder Subskalen durchmischt präsentiert werden. Dabei besteht zwischen den Fragen einer Skala eine maximale inhaltliche Distanz. Als Beispiel sei hier das NEO-Fünf-Faktoren-Inventar (NEO-FFI) genannt. Diese Vorgabemodalität soll ein mögliches Durchschauen der Untersuchungsabsicht seitens der Beantwortenden erschweren, wodurch die Augenscheinvalidität eines Fragebogens und somit Antworttendenzen der beantwortenden Personen reduziert werden sollen.

Solche bewusst oder unbewusst einseitig verzerrten Angaben können durch Tendenzen der zu testenden Person, sich besser oder schlechter darzustellen (*faking good* bzw. *faking bad*) auftreten. Dem könnte die Intention zugrunde liegen sozial erwünschter bzw. im Sinn einer modifizierten Selbstdarstellung zu antworten, weswegen es vorteilhaft erscheint „für eine Persönlichkeitsdiagnostik die Undurchschaubarkeit der Messintention seitens der Tp [sic!] anzustreben“ ist (Kubinger, 2006: S. 115).

Ebenso kann die Itemreihenfolge auch spezifischere Effekte provozieren, exemplarisch seien nachfolgend einige angeführt: Vorhergehende Fragen können *Halo-Effekte* bedingen, welche sich dadurch auszeichnen, dass von einer Eigenschaft auf andere Merkmale des zu beurteilenden Objektes (z. B. auf Persönlichkeitseigenschaften) geschlossen wird. Die Reihung von Fragen kann auch *Primacy-Recency-Effekte* provozieren, welche von Bortz und Döring (2006: S. 184). als Urteilsverzerrungen, die mit der sequenziellen Position der zu beurteilenden Objekte zusammenhängen, definiert werden. Ebenfalls können *Effekte der*

kognitiven Dissonanz nach Festinger (1957)⁴ von der Abfolge der Items ausgelöst werden, wobei unvereinbar scheinende Kognitionen (Einstellungen, Überlegungen, Gedanken, Wünsche etc.) einen inneren Konflikt hervorrufen und nachfolgende Einstellungsurteile verzerren können.

Allerdings kann bei der Fragebogenkonstruktion auch gezielt *Kontexteffekten* entgegengewirkt werden:

- Indem eine durchmischte Fragereihenfolge angestrebt wird (Ortner, 2004: S. 467), und somit thematisch ähnliche Fragen nicht nacheinander sondern zwischen thematisch andersartigen Items eingebettet präsentiert werden, wird eine ausreichende Zeitspanne gewährleistet, sodass beim nächsten entsprechenden Item thematisch ähnliche Assoziationen nicht mehr im Gedächtnis aktiviert sind.
- Darüber hinaus soll eine thematisch abwechselnde Itemvorgabe Monotonie verhindern und so dazu beitragen die Aufmerksamkeit der Beantwortenden aufrechtzuerhalten (Sparfeldt, Schilling, Rost & Thiel, 2006: S. 962).
- Zudem können *Reaktanzeffekte* seitens der zu testenden Personen aufgrund scheinbar redundanter Fragen durch eine thematisch abwechselnde Darbietung der Items vermieden werden.
- Ein weiterer Vorteil von einer durchmischten Itemreihenfolge ist, dass Müdigkeitseffekte gegen Ende von Fragebögen⁵ nicht das Antwortverhalten bei einzelnen Skalen betreffen, sondern über alle Skalen verteilt einen weniger extremen Einfluss auf das Gesamtergebnis ausüben und das erhobene Profil nicht beeinträchtigen.

Oft werden aufgrund von ökonomischen Überlegungen Skalen aus einer Fragebogenbatterie herausgenommen um spezifisch einzelne Aspekte des Persönlichkeitskonstruktes zu messen, indem die Items inhaltlich geblockt vorgegeben werden. Während bisher eher Argumente gegen eine solche Darbietung angeführt wurden, sprechen die nachfolgend genannten Argumente für die serielle Itemdarbietung von ähnlichen Iteminhalten. Im Anschluss daran

⁴ Siehe hierfür Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

⁵ Effekte, welche durch Müdigkeit gegen Ende oder „Warming-up“-Effekte am Anfang von Fragebögen werden Positionseffekte genannt, sie zeichnen sich nach Rost durch „die Veränderung der Schwierigkeit oder anderer Merkmale eines Items infolge seiner Platzierung im Test“ aus (Rost, 2004: S. 69).

werden einige Studien angeführt, im Rahmen derer Itemreihenfolgeeffekte untersucht wurden.

Die inhaltlich geblockte Darbietung von Fragen einer Skala oder Subskala kann durch die gebotene Transparenz nach Sparfeldt, Schilling, Rost und Thiel (2006: S. 962) das Vertrauen und die Auskunftsbereitschaft der Beantwortenden erhöhen, was zu einer gesteigerten Kooperation seitens der TeilnehmerInnen führen kann. Dieser Aspekt kann sich dementsprechend auch positiv auf die Motivation der beantwortenden Personen auswirken.

Nach Schriesheim und DeNisi (1980, zit. nach Sparfeldt, 2006: S. 962) könnte die Aneinanderreihung von Items eines Themengebietes auch die Eindimensionalität der Antworten fördern. Auch Steinberg (1994: S. 341) weist darauf hin, dass die wiederholte Beantwortung von Items, welche ein unidimensionales psychologisches Konstrukt repräsentieren, die Zugänglichkeit zu den relevanten eigenen Einstellungen oder Gefühlen erhöht und somit die Beziehung zwischen der Itembeantwortung und dem zugrundeliegenden Konstrukt gestärkt wird. Der Aufmerksamkeitsfokus wird dabei auf die zu messenden Aspekte gelenkt, somit können die Itemantworten durch Iteminterpretationsprozesse oder die Erleichterung des Abrufs von relevanten Emotionen oder Einstellungen verändert werden (ebd.). Diese Annahme wird von einigen Studien gestützt, die höhere Reliabilitäten (Korrelationen der einzelnen Items mit dem Gesamtscore) der Items gegen Ende von Persönlichkeitsfragebögen und bei jenen mit höherer Augenscheinvalidität fanden (Knowles, 1998: S. 316).

Die nachfolgend angeführten Studien sollen einen kurzen Überblick über die für diese Arbeit relevante bisherige Forschung zu Itemreihenfolgeeffekten geben:

Schriesheim und DeNisi (1980: S. 175) verglichen den Einfluss von geblockten und randomisierten Itemreihenfolgen auf die Validität eines Fragebogens zur Erfassung von Führungsverhalten⁶, dabei erwies sich die Bedingung mit randomisierter Fragenreihung als valider. Ortner (2004: S. 466, 475) untersuchte Unterschiede im Antwortverhalten von österreichischen Bundesheersoldaten bei zwei Versionen eines Persönlichkeitsfragebogens⁷ mit durchmischter Itemreihenfolge, wobei der Versuchsgruppe die Fragen in genau

⁶ Die resultierenden Daten des *Leader Behavior Description Questionnaire* wurden mittels *multitrait-multimethod* (MTMM) und Varianzanalysen ausgewertet.

⁷ Es wurde eine deutsche Version des *Eysenck Personality Profiler* (EPP-D) vorgegeben.

umgekehrter Reihenfolge als der Kontrollgruppe präsentiert wurden. Während keine signifikanten Mittelwertsunterschiede zwischen den Versionen gefunden wurden, zeigten Methoden der *Item Response Theory* Abweichungen der Itemschwierigkeiten bei einigen Skalen. In einer Studie von Krampen, Hense und Schneider (1992, zit. nach Sparfeldt et al., 2006: S. 963). wurde ein Persönlichkeitsfragebogen⁸ in geblockter und randomisierter Reihenfolge vorgegeben, es wurden nur wenige statistisch bedeutsame Unterschiede zwischen den Bedingungen gefunden. Sparfeldt et al. (2006: S. 967-971) untersuchten die Selbstkonzepte von SchülerInnen in Mathematik, Physik, Deutsch und Englisch anhand des *Differential Self-Concept* (DISC) Rasters. Beim Vergleich einer skalenweise geblockten Bedingung mit entsprechend ausformulierten und randomisierten Fragen für die Kontrollgruppe fanden sie keine praktisch bedeutsamen Unterschiede der Faktorenstruktur und der Skalenmittelwerte der beiden Fragebogenversionen⁹ (ebd.).

Da die oben genannten Studien sehr unterschiedliche Persönlichkeitskonstrukte umfassen, verschiedene Arten der Bildung inhaltlicher Blöcke (z. B. nach Skalen und/oder nach Subskalen, mit der oder ohne die Verwendung von Einleitungssätzen etc.) bzw. Arten der Randomisierung realisieren, sowie die verwendeten Fragebögen unterschiedliche Antwortformate aufweisen, sind die oben angeführten Ergebnisse schwer vergleichbar. Die Tatsache, dass Großteils Mittelwertvergleiche und nicht Methoden der *Item Response Theory* zur Datenanalyse verwendet wurden, beeinträchtigt zusätzlich deren Aussagekraft. Überdies weisen die gefundenen Effekte teilweise in unterschiedliche Richtungen. Angesichts dieser teils widersprüchlichen Ergebnisse könnten für diese Studie keine Annahmen über die zu erwartenden Effekte bei unterschiedlich geblockten sowie durchmischten Items beim Wiener Studieneignungs-Persönlichkeitsinventar (WSP) angenommen werden.

Das folgende Kapitel widmet sich den in dieser Studie verwendeten statistischen Verfahren und bietet eine theoretische Einführung in diese.

⁸ Hier wurde eine Version des *Freiburger Persönlichkeitsinventars* verwendet.

⁹ Dafür wurden *konfirmatorische Faktorenanalysen* und eine (*between-subjects*) MANOVA gerechnet.

3 Methodik

Mit dieser Arbeit soll untersucht werden, ob beim Persönlichkeitsfragebogen WSP ein Effekt durch die Variation der Reihenfolge der Itempräsentation auftritt. Das LLTM – als Verallgemeinerung des Rasch-Modells – ermöglicht es, Modelle, die einen solchen Effekt aufweisen, zu postulieren und dabei zu prüfen, ob diese die empirisch erhobenen Daten adäquat und sparsam beschreiben. Die Anwendung des LLTMs erfordert die Gültigkeit des Rasch-Modells für die interessierende Eigenschaft in der untersuchten Population.

Hinsichtlich der praktischen Vorgehensweise wurde die Gültigkeit des Rasch-Modells wurde mittels Likelihood-Ratio-Test nach Andersen geprüft. War diese nicht a priori gegeben, wurden mittels Itemfit-Charakteristiken und grafischer Modellkontrollen die für die Modellpassung ungünstigsten Items ermittelt und in Folge ausgeschlossen. Überdies wurden die Subskalen der Skalen mittels Martin-Löf-Test dahingehend geprüft, ob die Annahme der Eindimensionalität gerechtfertigt sei. Aufgrund dieser Ergebnisse wurde für diese Analyse die Skalenstruktur des WSP optimiert. Die verwendeten statistischen Verfahren und ihre theoretischen Grundlagen werden im folgenden Kapitel skizziert.

3.1 Die verwendeten statistischen Verfahren

3.1.1 Das dichotom logistische Modell von Rasch (1 PL-Modell)

Als zentrales Modell des Item-Response-Ansatzes unterscheidet das Rasch-Modell die Fähigkeitsdimension der Person von der Schwierigkeitsdimension des Items. Die Modellgleichung des dichotom¹⁰ logistischen Modells von Rasch (1980) bringt diese Parameter zusammen, indem die Wahrscheinlichkeit der Lösung eines Items i durch Person v als Funktion von θ_v , dem Fähigkeitsparameter der Person und β_i , dem Itemschwierigkeitsparameter des bearbeiteten Items, konzipiert wird:

$$p(x_{vi}) = \frac{\exp(x_{vi}(\theta_v - \beta_i))}{1 + \exp(\theta_v - \beta_i)}$$

(Rost, 2004: S. 119)

¹⁰ Wobei Erweiterungen des Rasch-Modells für Daten mit mehr als zwei Antwortkategorien existieren.

Der bei Gültigkeit des Rasch-Modells erwartete Zusammenhang zwischen den Antworten der Personen und dem postulierten psychischen Merkmal kann mittels einer Itemfunktion oder grafisch in einer Itemcharakteristischen Kurve (Abk. ICC: *item characteristic curve*) wiedergegeben werden (Rost, 2004: S. 96).

Aus der oben angeführten Modellgleichung kann die Wahrscheinlichkeit, dass eine Person v ein Item i löst, wie nachfolgend dargestellt formuliert werden, wobei θ_v den Fähigkeitsparameter der Person v und β_i den Schwierigkeitsparameter des Items i darstellt:

$$P(X_{iv} = 1 | \theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}.$$

Durch das Rasch-Modell kann die Wahrscheinlichkeit, dass eine Person v das Item i löst, abhängig von dem Personenparameter θ_v , welcher die Fähigkeit der Person v beschreibt, und dem Itemparameter β_i , der die Schwierigkeit des Items i postuliert, definiert werden (Kubinger, 2003: S. 416).

Bei der Analyse von Persönlichkeitsfragebögen können statt der Termini Lösungswahrscheinlichkeit und Personenparameter (oder Fähigkeitsdimension) die Begriffe Zustimmungswahrscheinlichkeit und Eigenschaftsdimension (Persönlichkeitseigenschaft) verwendet werden. Je größer also die Eigenschaftsdimension θ einer Person v ist, desto höher ist die Wahrscheinlichkeit, dass er oder sie die Frage (beziehungsweise das Item) i zustimmend beantwortet. „Die Antwortwahrscheinlichkeit hängt also nur von der Differenz zwischen dem Eigenschafts- und dem Itemparameter ab“, womit Vergleiche auf einem Absolutskalenniveau möglich werden (vgl. Fischer, 1974: S. 209).

Das Rasch-Modell basiert auf den Grundannahmen der *lokal stochastischen Unabhängigkeit*, der *erschöpfenden Statistik* und der *Eindimensionalität*. Wenn das Rasch-Modell für eine Population gilt, gibt die Summenstatistik eine *erschöpfende* oder *suffiziente Statistik* (die Anzahl aller gelösten oder zugestimmten Items) einer Person die gesamte Information bezüglich der gemessenen Eigenschaft der Person wieder (Rost, 2004: S. 114).

Aus der Grundannahme der *Eindimensionalität* ergibt sich, dass alle Items homogen sind, da sie dieselbe latente Eigenschaft messen. Nach Kubinger (2003) bedeutet die *lokale stochastische Unabhängigkeit*, dass die Beantwortung einer Aufgabe nur von der Fähigkeit (oder der Persönlichkeitseigenschaft) einer Person und der Itemschwierigkeit der jeweiligen

Aufgabe abhängt, jedoch nicht davon, welche Items die Person bereits bearbeitet hat oder noch bearbeiten wird. Nach Irtel (1995: S. 49) kann aus der *lokalen stochastischen Unabhängigkeit* auch gefolgert werden, „dass die Antwortvektoren unterschiedlicher Personen stochastisch unabhängig sind“ – dies bedeutet, dass die postulierte latente Dimension stichprobenunabhängig erfasst werden kann. Die Tatsache, dass bei einem Rasch-Modell-homogenen Datensatz der Vergleich von Personenparametern unabhängig von den verwendeten Items erfolgt sowie die Itemparameter unabhängig von den Personenparametern gegenübergestellt werden können, wird *spezifische Objektivität* des Rasch-Modells genannt.

Mittels Modelltests kann die Rasch-Modell-Konformität der Daten überprüft werden. Dies erfolgt, indem der Datensatz anhand mehrerer Teilungskriterien¹¹ in Teilstichproben aufgeteilt wird, für die jeweils getrennt voneinander Itemparameter berechnet werden. Falls eine oder mehrere der Teilstichproben nicht Rasch-Modell-konform sein sollten, können auffällige Items z. B. mittels des Likelihood-Ratio-Tests nach Andersen, der Itemfit-Statistiken oder grafischen Modelltests identifiziert werden, um diese aus dem Datensatz ausschließen zu können und somit Rasch-Modell-Konformität herzustellen.

3.1.2 Der Martin-Löf-Test

Um die Fragestellung, ob die Subskalen innerhalb einer übergeordneten Skala ein homogenes Konstrukt messen, zu beantworten, wurde der Martin-Löf-Test durchgeführt. Diesem hypothesengeleiteten Signifikanztest liegt die Annahme zugrunde, dass verschiedene Itemgruppen eines Tests oder Fragebogens unterschiedliche Persönlichkeitseigenschaften oder Fähigkeiten erfassen könnten. Um dies zu prüfen, vergleicht er die Summenscores (welche bei Rasch-Modell-skalierten Daten erschöpfende Statistiken darstellen) der Personen, welche die entsprechenden Itemgruppen bearbeitet haben.

So kann mit dem Martin-Löf-Test festgestellt werden, ob z. B. zwei Itemgruppen ($k = 1,2$) unidimensional im Sinn des Rasch-Modells sind, indem mittels eines Likelihood-Quotienten-Tests die beiden entstandenen Itemgruppen k_1 und k_2 miteinander verglichen werden. Die nachstehende Formel soll das dabei verwendete mathematische Prinzip verdeutlichen,

¹¹ Üblicherweise werden hierfür sowohl externe Kriterien wie z. B. das Alter als auch interne Kriterien wie z. B. der Mittelwert der erhobenen Werte herangezogen.

wobei r für die Itemantworten der Aufgabengruppen k_1 und k_2 steht und nr die Anzahl derjenigen Personen angibt, welche die entsprechenden Itemgruppen beantwortet haben.

$$LR_{ML} = 2 \ln \left(\frac{L_c}{L_c^{(1)} \cdot L_c^{(2)}} \cdot q \right) \quad q = \frac{\prod_r \left(\frac{n_r}{N} \right)^{nr}}{\prod_r \left(\frac{n_r}{N} \right)^{nr}}$$

$$LR_{ML} = 2 \left(\sum_r n_r \ln \left(\frac{n_r}{N} \right) - \sum_r n_r \ln \left(\frac{n_r}{N} \right) - \ln L_c + \ln L_c^{(1)} + \ln L_c^{(2)} \right)$$

LR_{ML} folgt einer asymptotischen χ^2 -Verteilung mit $df = k_1 k_2 - 1$.

Für die Berechnungen wurden die einzelnen Items einer Skala den jeweiligen Subskalen zugeordnet. Für die so entstandenen Itemgruppen kann mittels Martin-Löf-Test festgestellt werden, ob für sie innerhalb einer Skala Homogenität angenommen werden kann.

3.1.3 Das Linear Logistische Testmodell von Fischer (LLTM)

Als Spezialfall oder Erweiterung des Rasch-Modells postuliert das Linear Logistische Testmodell (in Folge: LLTM) nach Fischer (1973) die Itemschwierigkeit eines Items als Funktion additiver Elementarparameter. Als Basis- oder Elementarparameter können all jene interessierenden Bedingungen, welche potentiell Einfluss auf die Itemschwierigkeit haben, mit einbezogen werden. Dies ermöglicht es somit im Vergleich zum Rasch-Modell mit seinen jedem Item zugeordneten spezifischen Schwierigkeitsparametern, dass das Zustandekommen der empirischen Daten mit weniger Elementarparametern geschätzt werden kann, was zu einer Verdichtung der enthaltenen Information führt. Im LLTM sind also die Itemschwierigkeitsparameter des Rasch-Modells σ_i (wobei $i = 1, \dots, k$) weiter in eine Kombination von Elementarparametern η_l ($l = 1, \dots, p$) und deren Gewichten q_{il} aufgeteilt, was durch die nachstehende Formel nachvollziehbar ist:

$$\sigma_i = \sum_{l=1}^p q_{il} \eta_l$$

Die spezifische Zusammensetzung der Teiloperationen kann dabei Hypothesen darüber repräsentieren, welche kognitiven Lösungsschritte zur Lösung eines Items notwendig sind, (Kubinger, 2003: S. 421). Ebenso können aber auch Itemcharakteristika, die zur

Itemschwierigkeit beitragen, mitteinbezogen werden. Somit kann das LLTM z. B. für die Analyse von Itemkonstruktionsregeln, Übungs-oder Lerneffekten, Positionseffekten, Effekten des Antwortformats etc. eingesetzt werden. Um das LLTM auf einen Datensatz anzuwenden, muss für diesen die Gültigkeit des Rasch-Modells vorliegen.

Für den Fall, dass ein Item denselben Personen zu unterschiedlichen Zeitpunkten vorgegeben wird oder – wie in der vorliegenden Studie – ein Item in unterschiedlichen Testversionen mit verschiedener Itemreihenfolge präsentiert wird und dieses Item für weitere Analysen zu verschiedenen Zeitpunkten beziehungsweise in unterschiedlichen Versionen diskriminiert werden soll, kommen sogenannte *virtuelle Items* zum Einsatz. Diese sind in Folge durch den hochgestellten Buchstaben v gekennzeichnet. Dabei wird ein Item quasi als zwei oder mehrere Items dargestellt, wobei sich diese durch gewisse Spezifikationen, welche als Elementarparameter formuliert werden, voneinander unterscheiden, und sich dabei einen gemeinsamen grundlegenden Schwierigkeitsparameter teilen. Zum Beispiel können aus dem Item I_i mit dem Itemparameter β_i die virtuellen Items I_a^v , I_b^v und I_c^v abgeleitet werden, wobei für I_c^v folgendes definiert sein kann:

$$\beta_c^v = \beta_i + \varepsilon$$

Die Schätzung der Basisparameter η_j erfolgt, wie auch beim Rasch-Modell, mittels der Maximum-Likelihood-Methode, wobei die Likelihoodfunktion „die Wahrscheinlichkeit der beobachteten Testdaten unter der Bedingung des angenommenen Testmodells als Funktion der Modellparameter“ (Rost, 2004: S. 303) beschreibt. Das erhaltene Ergebnis gibt demnach die Wahrscheinlichkeit dafür an, dass für die erhobenen Daten das angenommene Testmodell gilt – je höher dieser Wert ist, desto besser erklärt das Modell das Zustandekommen der Daten.

Ein Vergleich konkurrierender Modelle erfolgt mithilfe eines Likelihood-Ratio-Tests, der es ermöglicht, die Likelihoodquotienten von zwei verschiedenen Testmodellen für denselben Datensatz zu vergleichen, was durch die nachstehende Formel deutlich wird.

$$LR = \frac{TM_0}{TM_1}$$

Nach Rost (2004) müssen drei Bedingungen erfüllt sein um einen entsprechenden Likelihood-Ratio-Test durchführen zu können: Jenes Modell, welches sich im Nenner

befindet (TM_0) darf nicht durch die Restriktion der Parameter des zweiten Modells darstellbar sein; das Modell im Zähler soll darüber hinaus nicht durch Null-Setzen von Parametern aus dem oben stehenden Modell hervorgehen und die Modellgültigkeit des allgemeineren Modells (TM_0) muss nachgewiesen sein. Die erhaltene Prüfstatistik ist χ^2 -verteilt, wobei sich deren Freiheitsgrade aus der Differenz zwischen der Parameteranzahl des Modells TM_0 und jener des Modells TM_1 ergeben.

Im Rahmen dieser Arbeit wird so, indem das Rasch-Modell als Spezialfall des LLTM, bei dem jedes Item einen eigenen Schwierigkeitsparameter aufweist, gegen die alternativ formulierten Linear Logistischen Testmodelle geprüft wird, das Modell, das die erhobenen Daten am effizientesten (unter Berücksichtigung der Parameteranzahl) beschreibt, ermittelt. Im Anschluss an eine solche Analyse können aus diesem Modell noch die nicht signifikanten Effektparameter entfernt werden um den Allgemeinheitsgrad des Modells zu erhöhen. Für letztgenanntes wird abschließend eine neuerliche Modellprüfung vorgenommen.

Eine notwendige Voraussetzung für die Durchführung dieser inferenzstatistischen Likelihoodquotiententests ist, dass die Prüfgröße $-2 \log L$ annähernd χ^2 -verteilt sein muss. Falls die Stichproben nicht ausreichend groß sein sollten, können, ebenfalls durch Maximieren der Likelihoodfunktion eines Modelles, *informationstheoretische Kriterien* verwendet werden, welche den Modellvergleich von Modellen erlauben, deren Geltung nicht belegt sein muss und deren Beziehung zueinander nicht hierarchisch aufgebaut ist (Rost, 2004: S. 339). Die nachstehende Formel des Akaike information criterion (AIC) verdeutlicht, dass dieses informationstheoretische Maß die Anzahl der benötigten Parameter n_p und den $-2 \log L$ Wert abwägt.

$$AIC = 2(n_p - \log(L_0))$$

Je kleiner dieser Informationsindex ist, desto besser beschreibt das Modell den Datensatz.

3.2 Untersuchungsplan und Hypothesen

Das Wiener Studieneignungs-Persönlichkeitsinventar gliedert sich in die folgenden sieben Persönlichkeitsdimensionen oder Skalen: „Organisation“, „Arbeitsstil und Arbeitsverhalten“, „Motivation“, „Belastbarkeit“, „Soziale Kompetenz“, „Emotionale Kompetenz“ und „Selbstbild“, welche weiter in 22 Subskalen aufgeteilt werden können (siehe Abschnitt 3.3.1 Das Wiener Studieneignungs-Persönlichkeitsinventar (WSP) auf S. 29f). Diese Untersuchung

behandelt die Frage, ob das Antwortverhalten der Testpersonen beim WSP durch die Reihenfolge der Darbietung der Items dieser Skalen beeinflusst wird. Das experimentelle Design dieser Studie besteht aus drei Bedingungen, in denen die verschiedene Abfolgen der Items und Subskalen realisiert wurden – die Zuteilung der teilnehmenden Personen zu diesen Bedingungen erfolgte randomisiert.

Die Bedingung der Fragebogenversion A mit durchmischter Itemreihenfolge entsprach dabei der in der Einleitung erwähnten und in der Praxis hauptsächlich verwendeten Itemreihenfolge bei Fragebögen: Dabei wechseln sich die einzelnen Fragen der sieben Skalen und 22 Subskalen mit dem größtmöglichen Abstand gleicher Skalen und Subskalen ab, wobei die Items innerhalb der jeweiligen Subskalen randomisiert platziert wurden.

In der Vorgabebedingung B wurden die Items skalenweise geblockt vorgegeben, wobei die Items der zugehörigen Subskalen innerhalb der Skala mit größtmöglichem Abstand alternieren. Alle Items einer Skala werden demnach gemeinsam vorgegeben, allerdings variiert die Itemabfolge nach der Subskalenzugehörigkeit. Welches Item einer Subskala vorgegeben wird, wurde durch Zufall festgelegt.

In Bedingung C wurden die Items nach Skalen und Subskalen geblockt gereiht. Diese Fragebogenversion stellt damit die inhaltshomogenste Wiedergabe der Items dar.

Die Itemreihenfolgen der drei in dieser Studie verwendeten Fragebogenversionen sind dem Anhang zu entnehmen (siehe Kapitel 10.1 *Itemreihenfolgen der drei Fragebogenversionen* auf Seite 109f). Die Skalenstruktur des WSP ist in der Tabelle 3 *Die Struktur des Wiener Studieneignungs-Persönlichkeitsinventars (WSP)* (des Kapitels 3.3.1 *Das Wiener Studieneignungs-Persönlichkeitsinventar (WSP)* auf S. 31f) dargestellt.

Zur Beantwortung der Fragestellung war die Analyse der erhobenen Daten mittels des Linear Logistischen Testmodells vorgesehen, wofür den drei Bedingungen noch ein gemeinsamer *Linkingteil* hinzugefügt werden musste (siehe hierfür Kapitel 3.1.3 *Das Linear Logistische Testmodell von Fischer (LLTM)*, S. 22f). Dies ist notwendig, da die drei Fragebogenversionen mit den verschiedenen Itemreihenfolgen, welche die Items der sieben Skalen beinhalten, auf der Ebene der virtuellen Items als voneinander unabhängig konstruiert wurden, sodass nur durch alle Versionen durchziehende Items auch eine gemeinsame Parameterschätzung vorgenommen werden kann. Der *Linkingteil* wurde als für jede Bedingung am Ende des Fragebogens in identischer, durchmischter Reihenfolge vorgegebener Itemgruppe realisiert.

Gemäß der Fragestellung, ob der Itemkontext¹² der Fragebogenversionen das Antwortverhalten der TeilnehmerInnen beeinflusst, werden *abhängige* und *unabhängige Variable* wie folgt definiert: Der Itemkontext stellt die *unabhängige Variable* dar, das Antwortverhalten der Personen die *abhängige Variable*. Das Zustandekommen etwaiger Änderungen des Antwortverhaltens kann über die, über den geteilten Schwierigkeitsparameter hinausgehenden Elementareffekte geprüft werden. Die konservative Grundannahme besteht darin, dass der Itemkontext keinen Einfluss auf das Antwortverhalten der Testpersonen aufweist. Das Untersuchungsdesign soll durch die nachfolgende Tabelle veranschaulicht werden.

Tabelle 1: Untersuchungsdesign

	BEDINGUNG A	BEDINGUNG B	BEDINGUNG C
	durchmischt	geblockt-randomisiert- innerhalb	geblockt-vollständig
124 Items	Items der 7 Skalen und 22 Subskalen wechseln sich mit dem größtmöglichen Abstand gleicher Skalen und Subskalen ab, die Items der jeweiligen Subskalen sind randomisiert zugeteilt worden.	Items werden skalenweise geblockt vorgegeben, wobei die Items der zugehörigen Subskalen sich innerhalb der Skala mit größtmöglichem Abstand abwechseln.	Items werden sowohl nach Skalen als auch nach Subskalen geblockt präsentiert.
29 Items	Linkingteil (durchmischt)		

Vollständiges Design mit randomisierter Zuteilung der TeilnehmerInnen zu den drei Bedingungen.

Die Itemvorgabe erfolgte in der Befragungssituation standardisiert und computerisiert um *Störvariablen* und Versuchsleitereffekte, welche Einbußen der Gütekriterien Objektivität, Validität etc. nach sich ziehen könnten, zu minimieren. Die eingangs beschriebene Forschungsfrage dieser Untersuchung, ob das Antwortverhalten beim WSP durch einen veränderten Itemkontext in Form einer nach Skalen geblockten oder durchmischten Itemreihung beeinflusst wird, wird im folgenden Abschnitt in operationale und statistische Hypothesen übergeführt.

¹² Der Itemkontext wurde hier durch die durchmischte Itemreihenfolge und die beiden auf unterschiedliche Arten (skalenweise) geblockten Darbietungen der Items in den drei Bedingungen operationalisiert.

3.2.1 Aufgestellte Hypothesen

Diese Untersuchung behandelt die Frage, ob die durch drei unterschiedliche Itemreihenfolgen realisierte Manipulation des Itemkontextes beim WSP das Antwortverhalten der TeilnehmerInnen beeinflusst. Dies wird in der ersten Hypothese formuliert und soll mittels LLTM-Analysen beantwortet werden. Um diese statistischen Verfahren für die sieben Skalen des WSP durchführen zu können, muss für die erhobenen Daten die Gültigkeit des Rasch-Modells anzunehmen sein. Hierfür muss die Voraussetzung der *Eindimensionalität*, welche für alle Items derselben Skala Homogenität verlangt (da auf eine latente Eigenschaft geschlossen wird) geprüft werden, was in der zweiten und dritten Hypothese ausformuliert wurde. Da in allen Versionen viele Fragebögen von den StudentInnen nicht vollständig bearbeitet wurden und die nicht kompletten Datensätze in die statistischen Analysen nicht einbezogen werden konnten, postuliert die vierte (post-hoc formulierte) Hypothese mögliche Unterschiede in den Bedingungen bezüglich der Abbruchraten, da auch diese mitunter einen wesentlichen Hinweis auf mögliche Itemkontexteffekte liefern könnten.

Die erste Hypothese thematisiert den Einfluss der drei Vorgabemodalitäten mit verschiedenen Itemreihenfolgen auf das Antwortverhalten der TeilnehmerInnen:

H_{0-1} : Die drei Teilstichproben unterscheiden sich nicht hinsichtlich des Antwortverhaltens.

H_{1-1} : Mindestens eine der drei Teilstichproben unterscheidet sich von den anderen hinsichtlich des Antwortverhaltens.

Bevor diese zentrale Hypothese der Studie einer Prüfung mittels LLTM unterzogen werden kann muss die Voraussetzung der Rasch-Modell-Konformität der Daten gewährleistet sein, diese wird in den beiden nachfolgenden Hypothesen näher ausgeführt. Die zweite Hypothese postuliert, dass die Subskalen innerhalb einer übergeordneten Skala ein homogenes Konstrukt messen bzw. die einzelnen Subskalen Bestandteile des Persönlichkeitskonstrukts der jeweiligen Skala sind. Dafür wurden die Items jeder Subskala, welche aus inhaltlichen Überlegungen gebildet wurden, zu einer eigenen Itemgruppe zusammengefasst, die anschließenden Analysen wurden mittels Martin-Löf-Tests durchgeführt. Das Verwerfen der H_{0-2} würde zu einer Anpassung der Skalenstruktur des WSP führen.

H₀₋₂: Die Subskalen messen skalenweise eindimensional.

H₁₋₂: Mindestens eine Subskala mindestens einer Skala misst nicht konform dem Konstrukt der Skala.

Für die jeweiligen Items der Skalen des WSP sollte nachfolgend festgestellt werden, ob Rasch-Modell-Konformität gegeben ist. Um dies zu prüfen wurden die Items des WSP mittels Likelihood-Ratio-Tests skalenweise auf ihre Eindimensionalität getestet. Dies formuliert die dritte Hypothese:

H₀₋₃: Die Skalen des WSP weisen Rasch-Modell-Konformität auf.

H₁₋₃: Mindestens eine der Skalen des WSP weist keine Rasch-Modell-Konformität auf.

Die vierte, post-hoc aufgestellte Hypothese thematisiert, ob sich die Abbruchraten in den drei Fragebogenversionen voneinander unterscheiden. Ein Signifikanztest wurde durchgeführt um diese Fragestellung zu beantworten.

H₀₋₄: Die Abbruchraten in den Versionen A, B und C sind gleichverteilt.

H₁₋₄: Die Abbruchraten in den Versionen A, B und C sind nicht gleichverteilt.

3.2.2 Postulierte Modelle

Das WSP ist hierarchisch auf zwei Ebenen gegliedert – auf Skalen- und Subskalenniveau. Dörner (2009: S. 61) prüfte in seiner Arbeit, ob die Items des Fragebogens in der von ihm erhobenen Stichprobe auf Skalenniveau Eindimensionalität im Sinne des Rasch-Modells aufweisen. Auch in der vorliegenden Studie erfolgte die Prüfung auf Rasch-Modell-Konformität sowie die Ergebnismeldung an die TeilnehmerInnen skalenweise. Die aufgrund inhaltlicher Ähnlichkeiten gebildeten Subskalen geben die inhaltliche Aufstellung der einzelnen Skalen wieder. Der *Itemkontext* wurde in dieser Studie über die Itemreihenfolgen der Fragebogenversionen realisiert. In der Bedingung A erfolgte die Itemvorgabe in einer durchmischten Präsentationsvariante, wie sie bei Einstellungsfragebögen üblich ist. Die beiden anderen Versionen des WSP geben dieselben Items in nach Skalen bzw. nach Skalen und Subskalen geblockter Reihenfolge wieder.

In der Literatur werden mehrere mögliche Ursachen angeführt, um potentielle Unterschiede zwischen einer randomisierten oder geblockten Vorgabe zu erklären: So vermeide eine geblockte Vorgabe Monotonie, verstärke jedoch soziale Erwünschtheit und weise eine hohe Augenscheinvalidität auf. Zugleich könnten Testpersonen dazu neigen nahe zusammenliegende inhaltlich ähnliche Items aufeinander abzustimmen um Dissonanz im Sinne von Festingers Dissonanztheorie zu vermeiden. Der Nachweis solcher Effekte setzt jedoch tatsächliche Unterschiede im Antwortverhalten der TeilnehmerInnen zwischen den verschiedenen Bedingungen voraus. Um die im vorigen Abschnitt aufgestellte Hypothese, ob sich die Teilstichproben A, B und C hinsichtlich des Antwortverhaltens unterscheiden, prüfen zu können, muss zumindest ein Modell aufgestellt werden, welches die Items nach Versionszugehörigkeit differenziert sowie eines, dass die Itemzugehörigkeit zu den Bedingungen nicht unterscheidet. Anschließend können die Modellpassungen der Modelle mit den Daten untereinander verglichen werden um herauszufinden, welches Modell besser zu den erhobenen Daten passt, bzw. die Daten sparsamer beschreibt.

Im Rasch-Modell werden Itemparameter für jedes Item definiert, der Itemkontext geht lediglich als Fehler in die Parameterschätzung ein. Das Modell Itemkontext erweitert die Inhaltsparameter des Rasch-Modells um die drei Versionsparameter, welche Effekte des Itemkontextes zwischen den drei unabhängigen Bedingungen erfassen sollen. Zusätzlich wurde ein quasi-saturiertes Modell postuliert, in dem für jede Kombination aus Inhaltsparameter und Darbietungsversion jeweils ein eigener Itemschwierigkeitsparameter geschätzt wird. Die folgende Tabelle listet die drei hypothetischen Modelle auf:

Tabelle 2: Mittels LLTM zu vergleichende, hypothetische Modelle

Quasi-saturiertes Modell¹³:	Inhaltsparameter (Root) * 3 Versionen
Modell Itemkontext:	Inhaltsparameter (Root) + 3 Versionsparameter
Rasch-Modell:	Itemparameter = Inhaltsparameter (Root)

Das quasi-saturierte Modell verabsolutiert die Zugehörigkeit eines Items zu einer der drei Versionen. Die Items der drei Bedingungen werden so behandelt, als würden alle drei

¹³ Das quasi-saturiertes Modell ist im statistischen Sinne als saturiert zu verstehen, indem es das Zustandekommen der Daten dadurch, dass seine Parameteranzahl der Anzahl an unabhängigen Daten entspricht, zu 100 Prozent erklären kann.

Fragebogenversionen verschiedene Items beinhalten. Das Modell Itemkontext definiert Inhaltsparameter für jedes Item, zusätzlich werden auch drei Versionsparameter für die Bedingungen A, B und C postuliert. Somit wird der Itemkontexteffekt, welcher als Vergleich von zwei unterschiedlich geblockten und einer durchmischten Itemreihenfolge operationalisiert ist, erfassbar. Das Rasch-Modell postuliert hingegen, dass der Schwierigkeitsparameter lediglich durch das zugrunde liegende Item bestimmt wird.

Diese aufgestellten Modelle müssen für die computerunterstützte Durchführung der Analysen mittels LLTM in Gewichtungen der Elementarparameter übertragen und festgelegt werden. Dafür werden diese prä-experimentellen Gewichtungen q_{ij} in eine Q-Matrix (oder auch Designmatrix) eingeschrieben. In die Zeilen werden Items eingetragen und die Itemkomponenten werden in den entsprechenden Spalten aufgelistet. Somit können aus einer Q-Matrix die Hypothesen über die Aufgabenstruktur des verwendeten Modells herausgelesen werden.

Diese Vorgehensweise soll durch die nachfolgende Abbildung 1 verdeutlicht werden. Es handelt sich um die Q-Matrix des Itemkontextmodells für die Skala „Belastbarkeit“ des WSP. Die virtuellen Items sind der Übersichtlichkeit wegen in jene der Versionen A, B und C aufgeschlüsselt, die Items des Linkingteils sind mit einem L gekennzeichnet. Die Itemkomponenten bestehen aus inhaltlichen Komponenten sowie Komponenten der Versionen. Die Inhaltskomponenten der Items wurden in diejenigen, die Teil der verschiedenen Versionen und solche, die Teil des Linkingblocks (wieder durch L verdeutlicht) sind, unterschieden. Da die Informationen der Inhaltskomponente 1 und der Versionskomponente A durch die jeweils restlichen Komponenten ersichtlich und somit redundant sind, wurden diese entfernt. (So ist zum Beispiel die Information der Zugehörigkeit eines Items zur Bedingung A bereits durch dessen Nicht-Zugehörigkeit zu den Versionen B und C gegeben.)

		Komponenten												
		2	3	4	5	6	7	8	9	L1	L2	L3	B	C
virtuelle Items	A1	0	0	0	0	0	0	0	0	0	0	0	0	0
	A2	1	0	0	0	0	0	0	0	0	0	0	0	0
	A3	0	1	0	0	0	0	0	0	0	0	0	0	0
	A4	0	0	1	0	0	0	0	0	0	0	0	0	0
	A5	0	0	0	1	0	0	0	0	0	0	0	0	0
	A6	0	0	0	0	1	0	0	0	0	0	0	0	0
	A7	0	0	0	0	0	1	0	0	0	0	0	0	0
	A9	0	0	0	0	0	0	0	1	0	0	0	0	0
	B1	0	0	0	0	0	0	0	0	0	0	0	1	0
	B2	1	0	0	0	0	0	0	0	0	0	0	1	0
	B3	0	1	0	0	0	0	0	0	0	0	0	1	0
	B5	0	0	0	1	0	0	0	0	0	0	0	1	0
	B6	0	0	0	0	1	0	0	0	0	0	0	1	0
	B8	0	0	0	0	0	0	1	0	0	0	0	1	0
	B9	0	0	0	0	0	0	0	1	0	0	0	1	0
	C1	0	0	0	0	0	0	0	0	0	0	0	0	1
	C2	1	0	0	0	0	0	0	0	0	0	0	0	1
	C3	0	1	0	0	0	0	0	0	0	0	0	0	1
	C4	0	0	1	0	0	0	0	0	0	0	0	0	1
	C5	0	0	0	1	0	0	0	0	0	0	0	0	1
	C6	0	0	0	0	1	0	0	0	0	0	0	0	1
	C7	0	0	0	0	0	1	0	0	0	0	0	0	1
	C8	0	0	0	0	0	0	1	0	0	0	0	0	1
	L1	0	0	0	0	0	0	0	0	1	0	0	0	0
	L2	0	0	0	0	0	0	0	0	0	1	0	0	0
	L3	0	0	0	0	0	0	0	0	0	0	1	0	0

Abb. 1: Exemplarische Darstellung der Q-Matrix der Skala „Belastbarkeit“ des Itemkontextmodells, wobei L1, L2 und L3 die Linkingitems dieser Skala sind.

3.3 Untersuchungsmaterial

3.3.1 Das Wiener Studieneignungs-Persönlichkeitsinventar (WSP)

Das WSP ist ein Selbstbeurteilungsfragebogen zur Erfassung studienrelevanter Persönlichkeitseigenschaften. Nach Khorramdel und Maurer (2012) wurden durch das Erfragen von studienrelevanten Verhaltensweisen und Einstellungen jene Persönlichkeitseigenschaften erfasst, welche ein erfolgreiches Studieren wahrscheinlich machen. Auf Basis einer Anforderungsanalyse (siehe Kapitel 3.3.3 *Die Entwicklung des Wiener Studieneignungs-Persönlichkeitsinventars (WSP)*, S. 37f) konnten die folgenden

sieben Persönlichkeitsdimensionen als für den Studienerfolg relevant identifiziert werden: „Organisation“, „Arbeitsstil und Arbeitsverhalten“, „Motivation“, „Belastbarkeit“, „Soziale Kompetenz“, „Emotionale Kompetenz“ und „Selbstbild“. Die 278 Items des Wiener Studieneignungs-Persönlichkeitsinventars können diesen sieben Persönlichkeitsdimensionen oder Skalen zugeordnet werden, wobei die Skalen „Organisation“, „Arbeitsstil und Arbeitsverhalten“, „Motivation“, „Soziale Kompetenz“ und „Emotionale Kompetenz“ mehrere Subskalen aufweisen. Maurer (2009: S. 71) beschreibt die Struktur des WSP folgendermaßen: „Das WSP ist multidimensional konzipiert, wobei die einzelnen Subtests unidimensional erstellt worden sind.“ Die sieben Skalen des WSP bestehen wiederum aus 22 Subskalen, welche in der folgenden Tabelle dargestellt sind:

Tabelle 3: Die Struktur des Wiener Studieneignungs-Persönlichkeitsinventars (WSP)

Skalen	Zugehörige Subskalen
Organisation	Zeitmanagement
	Prioritätensetzung
	Planen
	Organisieren
	Informiertheit
Arbeitsstil und Arbeitsverhalten	Gewissenhaftigkeit
	Genauer Arbeitsstil
	Eigeninitiative
	Selbstständiger Arbeitsstil
Motivation	Durchhaltevermögen
	Lernbereitschaft
	Engagement
	Zielorientierung
	Zielsetzung
Belastbarkeit	Erfolgszuversicht
	Frustrationstoleranz
Soziale Kompetenz	Kontaktfähigkeit
	Teamfähigkeit
	Kommunikationsfähigkeit
Emotionale Kompetenz	Distanzierungsfähigkeit
	Offenheit
Selbstbild	Selbstbewusstsein

Die Persönlichkeitsdimension „Organisation“ erfasst mit fünf Subskalen Fähigkeiten zum Zeitmanagement, zum Ordnen von Aufgaben nach deren Wichtigkeit, zur

Informationsbeschaffung, zum Planen und zum strukturierten Vorgehen. Khorramdel und Maurer (2012: S. 109) beschreiben diese Persönlichkeitsdimension folgendermaßen:

Personen mit hohen Werten beschreiben sich so, dass sie bei der Herangehensweise und Bearbeitung von Aufgaben und Tätigkeiten ein günstiges Zeitmanagement aufweisen (indem sie Zeitpläne erstellen, die Lernzeit einteilen, den zeitlichen Arbeitsaufwand gut einschätzen können), Prioritäten setzen können (indem sie die Aufgaben nach ihrer Wichtigkeit einteilen können und mit der Wichtigsten beginnen), sich die relevanten Informationen besorgen (indem sie sich über das Studienangebot und bestimmte Termine informieren), Arbeitsschritte planen (indem sie z. B. recherchieren, welche Arbeitsschritte nötig und wichtig sind) und diese organisieren können (indem sie mehrere Aufgaben oder Arbeitsschritte koordinieren und dabei strukturiert vorgehen).

Die Skala „Arbeitsstil und Arbeitsverhalten“ beinhaltet die Subskalen Selbstständigkeit, Gewissenhaftigkeit und Genauigkeit bei der Arbeitsverrichtung. Nach Khorramdel und Maurer (ebd.: S. 109) zeichnen sich Personen mit hohen Werten in dieser Skala durch...

einen gewissenhaften und genauen Arbeitsstil aus (indem sie sich an Geplantes halten, sich z. B. ausreichend auf Prüfungen vorbereiten und bei der Arbeit auf Vollständigkeit und Exaktheit Wert legen). Sie beschreiben sich als jemanden, der Eigeninitiative zeigt und selbständig arbeitet (indem sie z. B. eigenständig Lösungsvorschläge und Ideen erarbeiten und versuchen, diese selbständig ohne Hilfe oder Aufforderung von außen umzusetzen).

Mit der Skala „Motivation“, die sechs Subskalen beinhaltet, werden das Engagement und die Bereitschaft zum Lernen, die Zuversicht Vorgenommenes auch umsetzen zu können sowie die Fähigkeit zur Zielsetzung und Zielverfolgung erfragt. Die Skalenbeschreibung lautet nach Khorramdel und Maurer (ebd.: S. 109f) folgendermaßen:

Personen mit hohen Werten beschreiben sich so, dass sie sich regelmäßig Ziele setzen (z. B. in der Ausbildung), welche sie nicht aus den Augen verlieren und von denen sie sich nicht ablenken lassen. Sie geben an, hohes Engagement zu zeigen (indem sie sich z. B. über das vom Studium Geforderte hinaus mit Lerninhalten oder fachlichen Inhalten beschäftigen oder mehr Praxiserfahrung sammeln, als vom Studium

vorgeschrieben wird) und Zuversicht zu besitzen, Aufgaben erfolgreich bewältigen zu können (indem sie sich von anfänglichen Schwierigkeiten nicht abhalten lassen und zuversichtlich sind, das Studium erfolgreich abzuschließen; neue Aufgaben sehen sie als Herausforderung an). Weiters beschreiben sie für sich die Bereitschaft Neues zu (er)lernen (indem sie Verbesserungsvorschläge annehmen können und Kritik als Verbesserungspotenzial sehen) sowie Aufgaben und Ziele mit dem entsprechenden Durchhaltevermögen zu verfolgen (indem sie selbst bei großem Aufwand und Anstrengung oder weniger interessanten Inhalten nicht aufgeben).

Die Skala „Belastbarkeit“ erfasst die Tendenz, sich von Misserfolgen in der Zielerreichung nicht beeinträchtigen zu lassen. Nach Khorramdel und Maurer (ebd.: S. 110) beschreiben sich Personen, die hohe Werte in dieser Skala erzielen, als sehr frustrationstolerant, sie lassen sich in ihrer Leistungsbereitschaft von erlebten Misserfolgen (wie z. B. von schlechten Noten oder Kritik auch für Arbeiten, in die viel Zeit investiert wurde) nicht verunsichern oder demotivieren.

Die Skala „Soziale Kompetenz“ erfragt Kommunikationsfähigkeit, Kontaktfreudigkeit und die Fähigkeit zur Zusammenarbeit in Teams, welche mittels drei Subskalen abgefragt werden. Khorramdel und Maurer (ebd.: S. 110) fassen diese Persönlichkeitsdimension folgendermaßen zusammen:

Personen mit hohen Werten beschreiben für sich die Bereitschaft, mit anderen zu kommunizieren und die Regeln einer angemessenen Kommunikation zu beherrschen (indem sie einen freundlichen und respektvollen Umgang mit anderen pflegen ohne jemanden zu kränken, sowie andere ausreden lassen). Sie beschreiben sich des Weiteren so, dass sie keine Schwierigkeiten damit haben, neue Kontakte zu knüpfen und auf fremde Personen zugehen zu können. Sie geben außerdem an, dass es ihnen leicht fällt, mit anderen in einem Team zu arbeiten, und dass sie sich dabei als kollegial, hilfsbereit und kompromissbereit zeigen.

Die beiden Subskalen der Skala „Emotionale Kompetenz“ gliedern sich in Offenheit für Neues und die Fähigkeit zur Distanzierung. Nach Khorramdel und Maurer (ebd.: S. 109) beschreiben sich Personen mit hohen Werten in dieser Skala als offen gegenüber neuen Sichtweisen, Erfahrungen und Inhalten und vermögen es sich von der eigenen Arbeit zu

distanzieren. Sie geben z. B. an sachlich und objektiv an Themen herangehen zu können, die eigenen Emotionen kontrollieren zu können sowie Probleme aus der Arbeit nicht ins Privatleben mitzunehmen.

Die Persönlichkeitsdimension Selbstbewusstsein, das Selbstbild und die Sicherheit im Auftreten wurden durch die Skala „Selbstbild bzw. Selbstbewusstsein“ erfragt. Nach Khorramdel und Maurer (ebd.: S. 110) geben Personen mit hohen Werten in dieser Skala an selbstbewusst und sicher im Auftreten zu sein, zu ihren eigenen Ideen und Meinungen stehen zu können und sich nicht leicht durch andere verunsichern zu lassen.

Die Items des WSP wurden als situative Fragen konzipiert. Diese bestehen aus hypothetischen Aufgaben- oder Problemsituationen aus dem Alltag von Studierenden sowie möglichen Reaktion darauf. Zum Beispiel lautet ein Item der Subskala „Kontaktfähigkeit“, welche der Skala „Soziale Kompetenz“ angehört, folgendermaßen: „Wenn nach dem Seminar noch Fragen offen sind, suche ich das Gespräch mit dem Lehrveranstaltungsleiter.“ Anhand der Antwortmöglichkeiten sollen die Beantwortenden kundtun, in welchem Ausmaß die in der Frage formulierte Reaktion auf sie persönlich zutrifft. Die Antwortmöglichkeiten sind für alle Items gleich. In dem folgenden Unterkapitel wird näher auf sie eingegangen.

3.3.2 Das Antwortformat des WSP

Die Beantwortung des Fragebogens WSP durch die TeilnehmerInnen erfolgt durch eine Selbstbeurteilung in Form einer abgestuften Bejahung oder Verneinung, welche nachfolgend dargestellt ist:

ja	eher ja	eher nein	nein
----	---------	-----------	------

Abb. 2: Vierkategorielles Antwortformat des WSP

Die Beantwortung dieser Beurteilungsaufgaben kann nach Rost (2004: S. 61) als bipolare Ratingskala mit vier verbal ausformulierten Antwortmöglichkeiten beschrieben werden. Zudem handelt es sich um ein gebundenes Antwortformat, welches durch die bereits vorgegebenen Antwortkategorien im Vergleich zu freien Antwortformaten den Vorteil einer

höheren Auswertungsökonomie sowie Auswertungsobjektivität aufweist. Diese erhöhte Auswertungsobjektivität¹⁴ kann Einbußen der Testvalidität nach sich ziehen, wenn die Antwortkategorien der Items nicht alle möglichen Antwortmöglichkeiten ausschöpfen, also exhaustiv sind (ebd.: S. 61). Bipolare Ratingformate sind Antwortformate mit mehr als zwei Kategorien, die von einem negativen Pol zu einem positiven Pol gehen und von den befragten Personen in eine Rangordnung gebracht werden (ebd.: S. 63f). Die Antwortkategorien des WSP sind itemunspezifisch formuliert, was bedeutet, dass jedes Item von den befragten Personen mit Hilfe der gleichen Kategorien zu beantworten ist. Die aus der Beantwortung des Fragebogens resultierenden Daten weisen Ordinalskalenniveau auf.

In die Entscheidung für das Antwortformat der Items des WSP gingen Überlegungen zur Anzahl der Antwortkategorien sowie deren Formulierung und den *response styles* ein, letztere werden von Rost (ebd.: S. 66f) als „die von der zu messenden Personeneigenschaft unabhängige Neigung einer Person, die Ratingskala in einer bestimmten Art und Weise zu gebrauchen“ beschrieben. Es lassen sich drei *response styles* unterscheiden: Die Tendenz zum mittleren Urteil, die Tendenz zum extremen Urteil sowie die Tendenz der Zustimmung unabhängig von den Iteminhalten, welche auch Aquieszenz genannt wird.

Die Tendenz zum mittleren Urteil kann zum Beispiel darauf hindeuten, dass die befragte Person die Antwort verweigert oder die zu beantwortende Frage für unzumutbar hält, wodurch der Test nicht mehr eindimensional messen würde. Um diese Tendenz zu vermeiden, wurde eine gerade Anzahl an Antwortmöglichkeiten, die keine neutrale Mitte aufweisen, verwendet. Um die Zustimmungstendenz, welche unabhängig von den Iteminhalten auftritt, zu vermeiden wurde der Itempool des WSP so konstruiert, dass er eine unterschiedliche Itempolung aufweist. Das bedeutet, dass bei etwa einer Hälfte der Items die Itembejahung bzw. die Itemverneinung für das Vorhandensein des geprüften Merkmals spricht (Maurer, 2009: S. 64). Um Reaktanzphänomene seitens der Befragten (wie zum Beispiel willkürliches Antwortverhalten) zu vermeiden wurden exhaustive Antwortkategorien angestrebt, welche alle möglichen Antwortmöglichkeiten abdecken sollten und den antwortenden Personen überdies mehr als zwei Antwortkategorien zur Verfügung stellen. Allerdings wurde auch ein noch zumutbarer Differenzierungsgrad angestrebt, da zu viele Antwortkategorien die Urteilsfähigkeit der Befragten anhand der Kategorien einschränken kann. Rost (2004: S. 66) empfiehlt diesbezüglich eine

¹⁴ Im Sinne einer intersubjektiv eindeutigen Auswertung.

Kategorienanzahl zwischen drei und zehn. Eine geringe Anzahl an Kategorien kann auch einer extremen Urteilstendenz, bei der vor allem mittels Randkategorien geantwortet wird, Vorschub leisten. Aus den oben genannten Gründen wählte Maurer (2009) eine vierkategoriale Ratingskala zur Beantwortung des WSP.

Die Antwortkategorien wurden verbal formuliert, wodurch nach Rost (ebd.: S. 67) auch eine bessere intersubjektiv ähnliche Bedeutungsauffassung erreicht werden kann. Dabei wurde neben der möglichst kurzen und klaren Formulierung auch darauf geachtet, dass die Rangordnung der Kategorien durch die verbale Gestaltung nicht verletzt wird.

Im nächsten Kapitel wird auf den Entwicklungsprozess des WSP eingegangen, im Zuge dessen die Anforderungsanalyse und Itementwicklung erfolgte.

3.3.3 Die Entwicklung des Wiener Studieneignungs-Persönlichkeitsinventars (WSP)

Das WSP wurde als Persönlichkeitsstrukturfragebogen, welcher studienrelevante Persönlichkeitsdimensionen erfassen soll (Maurer, 2009: S. 71) entwickelt, was in zwei Phasen stattfand: Der Erstellungsprozess des WSP gliederte sich in die Anforderungsanalyse, aus welcher ein Anforderungsprofil resultierte und die Itementwicklung, beide werden nachfolgend kurz skizziert.

Der Fragebogen basiert auf einer umfassenden Anforderungsanalyse¹⁵ an Wiener Universitäten bezüglich der Studienfächer Soziologie, Psychologie, Architektur, Publizistik- und Kommunikationswissenschaften, Maschinenbau, Betriebswirtschaft und Internationale Betriebswirtschaft, Volkswirtschaft sowie Rechtswissenschaften wobei nach der *Critical Incident Technique* von Flanagan (1954, zit. nach Khorramdel und Maurer, 2012: S. 103) vorgegangen wurde. Diese neun Studienrichtungen wurden unter Berücksichtigung der Studiennachfrage ausgewählt. Die Anforderungsanalyse beinhaltete die Befragungen von Studierenden mittels Fragebögen sowie Experteninterviews mit in der Lehre tätigen Personen an vier Wiener Universitäten¹⁶. Die Anforderungen wurden mittels qualitativer Inhaltsanalyse nach Mayring (2008) aus den angegebenen Verhaltensweisen gewonnen (ebd.: S. 105-106). Die Ergebnisse wurden in ein gemeinsames Anforderungsprofil für ein

¹⁵ Aus dieser Anforderungsanalyse resultierten acht Anforderungsprofile (Khorramdel und Maurer, 2012: S. 107).

¹⁶ Es wurden hierfür 339 StudentInnen mittels standardisierter Fragebögen und 78 Lehrpersonen mittels halbstandardisierter Interviews befragt (Khorramdel und Maurer, 2012: S. 105).

erfolgreiches Studieren zusammengefasst – aus dem später die sieben Persönlichkeitsdimensionen (Skalen) des Wiener Studieneignungs-Persönlichkeitsinventars abgeleitet wurden. Die Items des Fragebogens wurden unter der Berücksichtigung der genannten Verhaltensweisen in erfolgskritischen Situationen aus der Anforderungsanalyse mittels eines eigens dafür generierten Regelkatalog zur Itemgenerierung erstellt (ebd. S. 108). Der Itempool des WSP basiert also inhaltlich auf dem Anforderungsprofil für die universitäre Hochschulausbildung und wurde anhand des Regelkatalogs konstruiert. Im Anschluss wurden die Items ExpertInnen (MitarbeiterInnen und DiplomandInnen des Arbeitsbereichs Psychologische Diagnostik der Fakultät Psychologie an der Universität Wien) zur Prüfung vorgelegt und wurden unter Einbezug derer Kritik überarbeitet.

Der resultierende Itempool wurde mit Hilfe einer Online-Plattform an einer Stichprobe von Wiener Studierenden unterschiedlichster Studienrichtungen von Dörner (2009) getestet. Dabei wurden die gesammelten Daten einer Analyse nach dem Rasch-Modell unterzogen, um die Eindimensionalität der sieben Skalen prüfen. Die Eindimensionalität nach dem Rasch-Modell konnte nach Ausschluss von insgesamt 22 Items erreicht werden (Khorramdel & Maurer, 2012: S. 112).

Der Fragebogen wurde in Self-Assessments verschiedener¹⁷ Wiener Studienrichtungen verwendet, unter anderem ist der WSP Bestandteil des *Wiener Self-Assessment Psychologie* – eines webbasierten Self-Assessment für die Studienrichtung Psychologie¹⁸. Den Studierenden wird dabei die Möglichkeit einer individuellen und anonymen Ergebnismeldung in Form eines Stärken- und Schwächenprofils angeboten, wobei die *cut-off-Werte* pro Skala durch Fachpsychologen und Lehrende der jeweiligen Studienrichtungen auf Basis des jeweiligen studienspezifischen Anforderungsprofils festgelegt wurden (ebd.: S. 113). Die Ergebnisse des WSP sollen individuell Informationen für die Studienwahl bieten, womit die Test- und Beratungsstelle der Universität Wien hiermit einen ambitionierten, förderdiagnostischen Ansatz zur Studienberatung verfolgt.

In einer der drei bis jetzt durchgeführten Validierungsstudien untersuchten Mitschek und Müller die prognostische Validität des WSP – Mitschek (2011) führte eine Studie zur Voraussagekraft des WSPs im Self-Assessment Maschinenbau für den Studienerfolg durch,

¹⁷ Für die Wiener Studienrichtungen Architektur, Maschinenbau, Elektrotechnik und Informatik wurden jeweils nach den studienfachspezifischen Anforderungsprofilen adaptierte Versionen des WSP in Form von Self-Assessments realisiert (Khorramdel & Maurer 2012: S. 113).

¹⁸ www.studienberatungtest.at; besucht am 14.08.2012

wobei die Skala „Organisation“ zwischen erfolgreich und weniger erfolgreich Studierenden diskriminieren konnte. Müller (2011) untersuchte dieselbe Aufgabenstellung bei ArchitekturstudentInnen, eine Diskriminanzanalyse zeigte, dass die Skala „Selbstbewusstsein“ zwischen erfolgreichen und weniger erfolgreichen StudentInnen unterschied. Grass (2011) ermittelte prototypische Testkennwerte *exzellenter*¹⁹ PsychologiestudentInnen im Wiener Self-Assessment Psychologie, wobei diese StudentInnengruppe in den Skalen „Organisation“, „Motivation“, und „Selbstbewusstsein“ jeweils hohe Werte erzielten und in der Dimension „Frustrationstoleranz“ eher von geringen Ausprägungen berichteten.

Dörrer (2009) untersuchte die Frage, welche Auswirkung zwei verschiedene Varianten der Reihung der Subskalen auf die testtheoretische Qualität des WSP haben. Dazu ließ er 520 StudentInnen verschiedener Studienrichtungen zwei Varianten des WSP bearbeiten. Die erste Bedingung stellte eine Version mit inhaltshomogenen Itemblöcken dar. Unter der Alternativbedingung wiederum wurden die Items thematisch durchmischt in einer gleichverteilten Itemreihenfolge vorgegeben. Diese beiden Varianten wurden mittels Modelltests (grafischen Modellkontrollen und Likelihood-Ratio-Test nach Andersen) auf Konformität nach dem Rasch-Modell auf Personenebene überprüft. Für beide Varianten konnte – nach Ausschluss einiger Items – Eindimensionalität im Sinne des Rasch-Modells angenommen werden (Dörrer, 2009: S. 42-49). Daraus kann gefolgert werden, dass der Fragebogen in Populationen, für welche die erhobene Stichprobe repräsentativ ist, eindimensionale Persönlichkeitseigenschaften erhebt. Die Ergebnisse dieser Arbeit deuteten darauf hin, dass der manipulierte Itemkontext einen Einfluss auf die Beantwortung des WSP ausübt, da in der Stichprobe, in der das WSP mit *gleichverteilten*²⁰ Items vorgegeben wurde, für das Erreichen von Konformität mit dem Rasch-Modell weniger Items ausgeschlossen werden mussten, als dies in der Stichprobe mit geblockter Itemvorgabe der Fall war (ebd.: S. 61). In der Gesamtstichprobe zeigten Likelihood-Ratio-Tests mit dem Teilungskriterium Vorgabevariante (geblockt oder durchmischt), dass der Itemkontext die Itemschwierigkeiten beeinflusste (Dörrer, 2009: S. 51-58). Diese Befunde liefern Hinweise dafür, dass eine

¹⁹ Als *exzellente* Studierende wurden nach Khorramdel und Maurer (2012: S. 116) jene verstanden, welche von LektorInnen von *Übungen zur Statistik und Psychologischen Diagnostik* als solche eingeschätzt wurden.

²⁰ Mit *Gleichverteilung der Items* ist die Verteilung der Items einer Skala und Subskala über den Fragebogen hinweg gemeint, wobei die Items derselben Skala mit der maximalen Inhaltlichen Distanz zueinander dargeboten werden.

durchmischte Itemvorgabe vorteilhafter für die Vorgabe des Persönlichkeitsfragebogens sein könnte.

Um diese Annahme nochmals gezielt zu untersuchen, soll mit der vorliegenden Studie eine durchmischte Itemreihenfolge mit zwei unterschiedlich geblockten Itemreihenfolgen²¹ verglichen werden. Deren Auswertung erfolgte mit dem Linear Logistischen Testmodell. Hierfür wurden die 153 trennschärfsten Items des Itempools des WSP ausgewählt. Die angestrebte Stichprobe sollte aufgrund der bereits durchgeführten Erhebungen mit teilnehmenden StudentInnen Wiener Universitäten nun zusätzlich HochschülerInnen österreichischer Fachhochschulen umfassen. Das folgende Kapitel beschreibt die Rekrutierung der TeilnehmerInnen sowie die Zusammensetzung der erhobenen Stichprobe.

3.4 UntersuchungsteilnehmerInnen

3.4.1 Rekrutierung der UntersuchungsteilnehmerInnen

Die Stichprobe umfasst StudentInnen österreichischer Fachhochschulen sowie Studierende an Wiener Universitäten. Die Studierenden konnten den Fragebogen online über die Homepage www.hochschultest.at²² im Zeitraum vom 01.09.2012 bis 15.11.2012 beantworten. Die Zielgruppe sollte sich aus StudienanfängerInnen verschiedener Fachhochschulrichtungen, StudentInnen, die vorhaben ihre Fachhochschulrichtung zu wechseln, sowie StudentInnen, die an einer Auseinandersetzung mit ihren studienrelevanten Persönlichkeitseigenschaften interessiert sind, zusammensetzen. Auch MaturantInnen beziehungsweise SchülerInnen wurde es ermöglicht den Fragebogen online auszufüllen, es wurde allerdings im Vergleich zur Rekrutierung von StudentInnen keine Werbung für diese Studie im schulischen Kontext durchgeführt.

Um StudentInnen für die Teilnahme an der Studie zu motivieren, wurden verschiedene Werbemöglichkeiten kombiniert: Hierfür wurde Informationsmaterial in Form von Informationsfoldern an die FH-LeiterInnen sowie die VertreterInnen der jeweiligen Fachhochschulrichtungen der österreichischen Hochschülerschaft geschickt sowie Plakate und Flyer für die StudentInnen aller Fachhochschulen Österreichs zur Verfügung gestellt. Im Zeitraum vom 1. September bis 22. Oktober 2012 verteilte ich in Wien und Niederösterreich

²¹ Davon eine nach Skalen geblockte und innerhalb der Skalen auf Subskalenebene randomisierte Itemreihenfolge sowie eine nach Skalen und Subskalen geblockte Itemreihenfolge.

²² Weitere Informationen diesbezüglich liefert das Kapitel 3.5.1 *Die Homepage www.hochschultest.at* (S. 45f).

Flyer und Poster in Seminarräumen, Computerhörsälen, Mensen, Aufenthaltsräumen sowie vor den Inskriptionsschaltern und Informationsveranstaltungen bei verschiedenen Fachhochschulen. Den Fachhochschulen der anderen Bundesländer Österreichs ließ ich per Post Informationsmaterial über die Studie zukommen und bat darum die Plakate gut sichtbar aufzuhängen sowie mehrere Flyer aufliegen zu lassen²³. Damit sollten Studierende alle Fachhochschulrichtungen gleichermaßen an der Studie teilnehmen können.

StudentInnen an Wiener Universitäten (mit Ausnahme der Sigmund-Freud-Universität) sollten ebenfalls auf die Studie aufmerksam gemacht werden. Für die Studienteilnahme wurde von 23. Oktober bis 15. November 2012 mittels Flyern und Plakaten an Wiener Universitäten geworben. Auch auf studentischen Foren²⁴ der verschiedenen Studienrichtungen machte ich meine KollegInnen auf die Möglichkeit zur Studienteilnahme aufmerksam. Die Fachhochschul – bzw. Studienrichtungen der TeilnehmerInnen waren zwar für die Beantwortung meiner Forschungsfragen nicht relevant, wurden aber trotzdem mit anderen demographischen Fragen erhoben um diese Daten eventuell weiter verwenden zu können.

3.4.2 Deskriptive Beschreibung der Stichprobe

Der Datensatz umfasste 465 Personen, welche den Fragebogen beantwortet hatten. Davon waren 139 Fragebögen lediglich unvollständig bearbeitet und wurden somit für das weitere Vorgehen ausgeschlossen. Der für statistische Analysen verwendete Datensatz umfasste daher 326 vollständig bearbeitete Fragebögen. Davon hatten 99 Studierende die Fragebogenversion A mit der durchmischten Reihenfolge, 122 die skalenweise geblockte Version B und 105 StudentInnen den Fragebogen mit skalenweise und subskalenweise geblockten Items beantwortet. Tabelle 4 gibt für jede der drei Versionen an, wie viele Personen die gestellten Fragen vollständig bzw. unvollständig ausgefüllt hatten.

²³ Poster und Flyer wurden per Post an die Fachhochschulen gesendet. Eine Auflistung der Fachhochschulen befindet sich im Anhang (Kapitel 10.5 *Liste der teilnehmenden Fachhochschulen Österreichs*) auf Seite 115f.

²⁴ Alle Foren, auf denen die Studie beworben wurde, sind im Anhang (Kapitel 10.6 *Liste der studentischen, universitätsnahen Foren, auf denen die Teilnahme an der Studie beworben wurde*) auf Seite 116 angeführt.

Tabelle 4: Anzahl der vollständig bzw. unvollständig bearbeiteten Fragebögen der drei Bedingungen A (durchmischte Itemreihenfolge), B (skalenweise geblockte Itemreihenfolge mit randomisierter subskalenweiser Itemzuweisung innerhalb) und C (mit nach skalen- und subskalenweise geblockter Itemreihenfolge).

Version	Vollständig bearbeitete	Unvollständig bearbeitete	%
	Fragebögen	Fragebögen	
A:	99	51	34,0
B:	122	47	27,8
C:	105	41	28,1

Zur demografischen Beschreibung der Stichprobe wurden die Variablen Geschlecht, Alter Muttersprache, Staatsbürgerschaft, Berufstätigkeit und derzeit besuchte Ausbildung erhoben. 91,4 Prozent aller TeilnehmerInnen gaben an Deutsch als Muttersprache zu sprechen, 7,1 Prozent haben eine andere Muttersprache und 1,5 Prozent wuchsen zweisprachig auf. Von den 326 Personen, welche den Fragebogen vollständig ausgefüllt hatten waren 89,0 Prozent österreichische StaatsbürgerInnen, 7,1 Prozent gaben an die deutsche Staatsbürgerschaft zu besitzen, 2,5 Prozent gaben an einem anderen Staat zugehörig zu sein und 1,5 Prozent waren Angehörige zweier Staaten (nicht Österreich). Der folgenden Tabelle kann man die Verteilung der Variable Berufstätigkeit in den Teilstichproben der Fachhochschul- bzw. UniversitätsstudentInnen entnehmen²⁵.

Tabelle 5: Häufigkeiten und Prozentangaben von berufstätigen bzw. nicht berufstätigen StudentInnen von Fachhochschulen und Universitäten in der Stichprobe.

Derzeitige Ausbildung	Berufstätigkeit	n	%
Fachhochschulstudium	Ja	95	59,38
	Nein	65	40,62
Universitätsstudium	Ja	77	51,33
	Nein	73	48,67

²⁵ Von jenen zwei Personen, welche angegeben haben noch in die Schule zu gehen und jenen fünf, welche vor kurzem maturiert hatten hat niemand angegeben berufstätig zu sein. Alle Beiden, welche zur Zeit der Befragung keine Ausbildung machten sowie fünf von denjenigen sieben Personen, welche die Kategorie *Sonstiges* wählten gaben an berufstätig zu sein.

Die Frage nach der derzeitigen Ausbildung wurden von 160 Personen oder 49,1 % aller TeilnehmerInnen mit Fachhochschule und von 150 Studenten bzw. 46 Prozent aller Personen der Stichprobe mit Universitätsstudium beantwortet. Die Kategorie *Sonstiges* wurde von sieben oder 2,2 Prozent aller TeilnehmerInnen ausgewählt, 1,5 % aller Personen gaben an vor Kurzem maturiert zu haben und jeweils zwei oder 0,61 Prozent antworteten auf diese Frage mit Schule oder keine Ausbildung.

In der nachfolgenden Tabelle wurde aufgelistet, wie sich die Häufigkeiten dieser sechs eben genannten Kategorien auf die Versionen verteilten.

Tabelle 6: Häufigkeiten und Prozentangaben jener Personen, welche in den drei Fragebogenversionen die Frage nach der derzeitigen Ausbildung mit den entsprechenden Kategorien beantworteten.

Version	Ausbildung	n	%
A	Fachhochschule	49	49,5
	Universitätsstudium	45	45,5
	Sonstiges	2	2,0
	Vor Kurzem maturiert	2	2,0
	Schule	1	1,0
	Keine Ausbildung	0	0
B	Fachhochschule	56	45,9
	Universitätsstudium	58	47,5
	Sonstiges	3	2,5
	Vor Kurzem maturiert	3	2,5
	Schule	1	0,8
	Keine Ausbildung	1	0,8
C	Fachhochschule	55	52,4
	Universitätsstudium	47	44,8
	Sonstiges	2	1,9
	Vor Kurzem maturiert	0	0
	Schule	0	0
	Keine Ausbildung	1	0,95

Die Verteilung von Geschlecht und Alter der TeilnehmerInnen in den verschiedenen Versionen wird durch die nachfolgenden Tabellen verdeutlicht.

Tabelle 7: Deskriptivstatistiken zur Verteilung des Geschlechts und des Alters in den drei Bedingungen A, B und C.

Version	Geschlecht	n	%
A	männlich	32	32,3
	weiblich	67	67,7
B	männlich	36	29,5
	weiblich	86	70,5
C	männlich	34	32,4
	weiblich	71	67,6

Version	MW	Median	Min.	Max.	SD
A	26,3	24	16	60	7,65
B	25,4	25	18	51	5,40
C	26,7	24	18	50	6,77

In der Teilstichprobe der FachhochschulstudentInnen lag das Alter der Personen im Mittel bei 27,7 Jahren und der Median bei 25 Jahren. Das durchschnittliche Alter der Studierenden an Universitäten lag bei 24,7 Jahren, der Median betrug 24. Alle drei Fragebogenversionen wurden von mehr Frauen als Männern beantwortet.

Die nachfolgenden Tabellen listen die angegebenen Studienrichtungen der Fachhochschul- bzw. UniversitätsstudentInnen der Stichprobe auf.

Tabelle 8: Studienrichtungen jener 148 von 160 StudentInnen österreichischer Fachhochschulen, welche ihre Studienrichtungen angaben.

Fachhochschulrichtungen	n
Gesundheitsmanagement und Gesundheitsförderung	30
Energie- und Umweltmanagement	18
Technik	13
Mechatronik, Robotik	12
Nachhaltige Energiesysteme	11
Sports Equipment Technology	9
Gebäudetechnik und Gebäudemanagement	8
Unternehmensführung, Executive Management	7
Informatik	5
Sozialarbeit, Sozialwirtschaft, Soziale Arbeit, Sozialpädagogik	4
Verkehr und Umwelt, Intelligente Verkehrssysteme	4
Technisches Umweltmanagement und Ökotoxikologie	4
Immobilienwirtschaft, Immobilienmanagement	3
Biomedical Sciences, Biomedical Engineering	3
Telekommunikation und Internettechnologien	2
Tourismusmanagement	2
Ergotherapie	1
Diätologie	1
Freizeitpädagogik	1
Webdesign und Webentwicklung	1
Hebammen	1
Radiologietechnologie	1
Verwaltung	1
HR Management und Arbeitsrecht	1
Public Management	1
High Tech Manufacturing	1
Baumanagement	1
Wirtschaft	1
Risk Management	1

Tabelle 9: Studienrichtungen jener 147 von 150 UniversitätsstudentInnen, welche ihre Studienrichtungen angaben.

Studienrichtungen	n
Ernährungswissenschaften	37
Biologie	22
Pharmazie	21
Pädagogik, Lehramt bzw. Bildungswissenschaften	11
Kultur- und Sozialanthropologie	9
Germanistik	9
Soziologie	7
Lebensmittel- und Biotechnologie	6
Psychologie	5
Informatik	4
Wirtschaft	3
Physik	2
Chemie	1
Individuelles Studium Landschaftsplanung und Landschaftsarchitektur	1
Kommunikationswissenschaft	1
Maschinenbau	1
Pflegewissenschaft	1
Politikwissenschaft	1
Theater-, Film- und Medienwissenschaft	1
Theologie	1
Umwelt- und Bioressourcenmanagement	1
Wildtiermanagement und Wildtierökologie	1
Wissenschaftsforschung	1

3.5 Durchführung der Untersuchung

3.5.1 Die Homepage www.hochschultest.at

Die Umsetzung der Online-Vorgabe der drei Fragebogenversionen erfolgte mittels einer eigens dafür angelegten Homepage. Diese Homepage, www.hochschultest.at, umfasste neben Online-Fragebogen auch Informationen zum Wiener Studieneignungs-Persönlichkeitsinventar und zur durchgeführten Studie sowie Links zur Test- und Beratungsstelle der Universität Wien (www.testzentrum.at) und der Psychologischen Studentenberatung Österreich (www.studentenberatung.at) und Kontaktdaten der Autorin. Diese Webseite wurde nach den nachfolgenden Kriterien gestaltet:

- Es wurden relevante Informationen über den WSP, die Fragebogenbeantwortung, die Studie, sowie die Vorteile einer Teilnahme vermittelt. Darüber hinaus sollte eine niedrigschwellige Möglichkeit bestehen, die Autorin zu kontaktieren und betont werden, dass die Beantwortung anonym und kostenlos sei. Die Texte der Homepage können im Anhang nachgelesen werden (siehe Kapitel 10.2 *Der Text der Homepage www.hochschultest.at* auf Seite 111f).
- Die angebotenen Inhalte sollten auch das Interesse an einer Studienteilnahme steigern.
- Ebenfalls sollte auf weiterführende Literatur über den WSP verwiesen werden.

Die StudentInnen wurden auf der Homepage randomisiert einer der drei Testversionen zugewiesen. Der Beantwortung der Fragen war die Instruktion²⁶ sowie die Erhebung der demografischen Angaben (siehe Kapitel 3.5.2 *Demografische Variablen*, S. 48f) vorangestellt. Dann folgten in jeder Fragebogenversion 153 Items des WSP, wobei die Reihenfolge der ersten 124 zwischen den drei Versionen variierte, während die restlichen 29 dem Linkingteil angehörten und somit in jeder Version identisch waren. Der Linkingteil bestand in allen drei Testversionen aus Items in derselben vollständig randomisierten Reihenfolge. Abschließend wurden den TeilnehmerInnen ihre Ergebnisse automatisiert rückgemeldet (siehe Kapitel 3.5.4 *Ergebnisrückmeldung an die TeilnehmerInnen* auf S. 50 sowie Kapitel 10.4 *Rückmeldung der Ergebnisse an die TeilnehmerInnen* auf S. 113f). Die

²⁶ Der genaue Wortlaut der Instruktion kann im Anhang, Kapitel 10.3 *Instruktion des webbasierten Wiener Studieneignungs-Persönlichkeitsinventares* auf Seite 113 nachgelesen werden.

einzelnen Items wurden nacheinander vorgegeben, wobei man jeweils das derzeitige Item beantworten musste um zum nächsten zu gelangen. Es wurde nicht die Möglichkeit geboten bereits beantwortete Items erneut zu bearbeiten. Die Itembearbeitung erfolgte ohne Zeitbegrenzung und die Bearbeitungszeit pro Item sowie jene der gesamten Beantwortung des Fragebogens wurde nicht erfasst.

In den folgenden Abschnitten werden zuerst die erfassten demografischen Variablen näher beschrieben, dann werde ich auf die Darbietung der Items des Fragebogens WSP eingehen und anschließend die Ergebnissrückmeldung an die TeilnehmerInnen erläutern.

3.5.2 Demografische Variablen

Neben dem Geschlecht, Alter, Muttersprache, Staatsbürgerschaft und der Berufstätigkeit wurde auch nach der derzeit betriebenen Ausbildung gefragt. Dabei mussten die demografischen Fragen – mit Ausnahme der Frage nach der Berufstätigkeit – beantwortet werden, um mit dem Hauptteil fortfahren zu können. Je nachdem, ob eine Filterfrage mit Fachhochschule, Universitätsstudium, Schule, keiner oder einer anderen Ausbildung beantwortet wurde, wurden andere passende Fragen vorgegeben, die allerdings nicht obligatorisch zu beantworten waren. Alle demografischen Fragen und deren Beantwortungsmöglichkeiten sind nachfolgend in Tabelle 10 angeführt, wobei die mit * gekennzeichnete Fragen beantwortet werden mussten um zur nächsten Frage fortschreiten zu können.

Falls angegeben wurde, dass derzeit eine Fachhochschule besucht werde, wurde anschließend nach dem Fachhochschultyp und dem Semester gefragt. Ähnlich folgten, wenn die Antwortmöglichkeit *Besuch eines Universitätsstudiums* ausgewählt wurde, Fragen nach der Studienrichtung und dem Semester. Wenn angegeben wurde, dass erst vor Kurzem maturiert worden war, folgten keine weiteren Fragen. Bei Auswahl der Antwortmöglichkeit Schule wurde nach dem Schultyp und der besuchten Klasse gefragt. Die Beantwortung mittels der Antwortmöglichkeit *keine Ausbildung* zog die Frage nach der höchsten abgeschlossenen Ausbildung nach sich.

Die nachstehende Tabelle dient der Erläuterung des eben Beschriebenen.

Tabelle 10: Demografische Fragen und deren Antwortmöglichkeiten

Item	Demografische Frage:	Antwortmöglichkeiten
1	Geschlecht*:	männlich
		weiblich
2	Alter*:	Numerische Eingabe
3	Muttersprache*:	Textfeld
4	Staatsbürgerschaft*:	Textfeld
5	Ich besuche derzeit*:	Schule
		Fachhochschule
		Universitätsstudium
		ich habe erst vor Kurzem maturiert
		keine Ausbildung
		andere: (Textfeld)
5.1	Schultyp:	Neue Mittelschule
		Gymnasium - Unterstufe
		Gymnasium - Oberstufe
		Bundeslehranstalt (HBLA, HTBL, TGLA, LFLA)
		Sonstige: (Textfeld)
5.1.1	Klasse:	4. Klasse Unterstufe oder neue Mittelschule
		5. Klasse Oberstufe
		6. Klasse Oberstufe
		7. Klasse Oberstufe
		8. Klasse Oberstufe
		Sonstige: (Textfeld)
5.2	Fachhochschultyp:	Textfeld
5.3	Studienrichtung:	Textfeld
5.2.1 / 5.3.1	Semester:	Numerische Eingabe
5.5	Meine höchste abgeschlossene Ausbildung ist*:	Hauptschule
		Matura
		Universitätsstudium
		Fachhochschule
6	Ich bin berufstätig:	ja
		nein

Die mit * versehenen Fragen stellen obligatorische Fragen dar.

3.5.3 Darbietung der Items des WSP

Über die Homepage erfolgte die randomisierte Zuteilung zu einem der drei Fragebogenversionen. Dabei wurden zuerst die demografischen Fragen und anschließend daran die Fragen des WSP vorgegeben, nach der Beantwortung aller Fragen erfolgte die individuelle Rückmeldung an die TeilnehmerInnen. Sowohl die demografischen Fragen, die Items der drei Fragebogenversionen als auch die Rückmeldung erfolgt mittels LimeSurvey²⁷ (The LimeSurvey project team, 2012).

Die 153 Fragen des WSP bestanden aus den 124 Fragen der jeweiligen Version und den anschließend vorgegebenen 29 Items des Linkingteils. Die Fragen enthielten jeweils eine

²⁷ LimeSurvey ist ein Programm, mit dem online Umfragen erstellt werden können.

studienspezifische Situation und eine mögliche Reaktion darauf. Die Antwortkategorien waren für alle Items dieselben. Zur Illustration wird nachfolgend ein exemplarisches Item inklusive der Antwortmöglichkeiten dargestellt.

Ich führe einen Kalender, um mir wichtige Termine zu notieren.				
trifft zu:	ja	eher ja	eher nein	nein

Abb. 3: Beispielimitem des Wiener Studieneignungs-Persönlichkeitsinventars mit den vier Antwortmöglichkeiten

Alle Fragen wurden in der gleichen Weise nacheinander präsentiert, wobei jeweils nur eine Frage gleichzeitig vorgegeben wurde. Um darauf aufmerksam zu machen, dass ein neues Item eingeblendet wurde, unterlegte ich alle ungeradzahligen Items mit einem grünen und alle geradzahligen Items mit einem blauen Hintergrund. Das nächste Item wurde den TeilnehmerInnen nur nach Beantwortung des vorhergehenden eingeblendet. Es gab weder die Möglichkeit die vorhergehenden Fragen noch einmal zu sehen oder zu verändern, noch konnten die TeilnehmerInnen Fragen überspringen. Bei allen Fragen musste eine der vier Antwortmöglichkeiten eingegeben werden, Mehrfachantworten waren nicht möglich. Die Beantwortenden hatten die Möglichkeit den Fragebogen abubrechen und ihre Antworten zu löschen. Es wurde ihnen nicht die Möglichkeit geboten, den Fragebogen zu einem späteren Zeitpunkt fortzusetzen, da diese Pause zwischen der Beantwortung der Fragebogenteile das Antwortverhalten eventuell verändern hätte können. Während der ganzen Beantwortungszeit wurde eine Fortschrittsanzeige eingeblendet, die den Beantwortenden schematisch den prozentuell bereits absolvierten sowie den noch verbleibenden Teil des Fragebogens anzeigte.

3.5.4 Ergebnissrückmeldung an die TeilnehmerInnen

Da anzunehmen ist, dass zwischen den Subskalen einer Skala eine hohe Korrelation besteht, erfolgte die Rückmeldung über das individuelle Abschneiden im Fragebogen basierend auf den übergeordneten Skalen, nicht aber auf den einzelnen Subskalen. Zur besseren

Verständlichkeit enthielt die Ergebnismrückmeldung neben den persönlich erreichten Werten auch die Beschreibungen der sieben Persönlichkeitsdimensionen²⁸ oder -skalen. Da das Wiener Studieneignungs-Persönlichkeitsinventar, zur Zeit in der die Studie durchgeführt worden war, noch nicht über ausreichende Normierungsstudien verfügte, wurden den TeilnehmerInnen lediglich ihre erreichten Punkte pro Persönlichkeitsdimension rückgemeldet. Das maximal erreichbare Ergebnis lag für jede Skala bei zehn Punkten - null Punkte standen für die niedrigste Ausprägung einer Skala.

4 Ergebnisse

In diesem Kapitel werden die Ergebnisse der durchgeführten statistischen Verfahren dargestellt. Die Analysen wurden mittels der Software *R* (The R foundation for statistical computing, 2009) mit dem Plugin *eRm* (Mair & Hatzinger, 2007) vorgenommen.

Da sich die Abbruchraten in den drei erhobenen Stichproben auffällig unterschieden wurde post hoc ein Chi-Quadrat-Test durchgeführt um zu prüfen, ob die vorgefundenen Abbruchraten der drei Fragebogenversionen sich auch signifikant unterschieden, da dies die Datenqualität beeinträchtigen würde (siehe Kapitel 4.1 *Signifikanztest zur Prüfung der Abbruchraten der Versionen* auf S. 52).

Um Analysen mit dem LLTM durchführen zu können, muss die Passung des Rasch-Modells mit den erhobenen Daten angenommen werden können. Hierfür wurde zuerst mittels Martin-Löf-Tests geprüft, ob die einzelnen Subskalen innerhalb einer Skala ein homogenes Konstrukt messen. Wenn einzelne Subskalen der jeweiligen Skala nicht eindimensional waren wurde die Struktur des WSP entsprechend modifiziert. Anschließend wurden die Items jeder Skala einer Analyse mit Likelihood-Ratio-Tests nach Andersen zur Prüfung der Rasch-Modell-Konformität der Items unterzogen. War diese Homogenität nicht a priori gegeben, wurden mittels Itemfit-Charakteristiken und grafischer Modellkontrollen die für die Modellpassung ungünstigsten Items ermittelt und in Folge ausgeschlossen. Waren diese Voraussetzungen erfüllt, so wurde für jedes der drei theoriebasiert aufgestellten Modelle die Passung mit den erhobenen Daten mittels des LLTM analysiert. Schlussendlich wurden die

²⁸ Der genaue Wortlaut der Ergebnismrückmeldung kann im Anhang im Kapitel 10.4 *Rückmeldung der Ergebnisse an die TeilnehmerInnen* auf Seite 113f nachgelesen werden.

Modellpassungen der Modelle mit den erhobenen Daten zwischen den Modellen verglichen, um jenes zu identifizieren, welches die erhobenen Daten am adäquatesten beschreibt.

Für die Likelihood-Ratio-Tests zur Prüfung der Raschhomogenität mussten die Daten im Vorfeld dichotomisiert werden, wobei die vier Antwortkategorien „ja“, „eher ja“, „eher nein“ und „nein“ dabei wie folgt zusammengefasst wurden: Die Kategorien „ja“ und „eher ja“ wurden mit dem Wert Eins und die Kategorien „eher nein“ sowie „nein“ wurden mit dem Wert Null versehen.

In der oben angeführten Reihenfolge sind in den folgenden Abschnitten die Ergebnisse der durchgeführten Analysen aufgelistet, falls nicht anders erwähnt wurde ein Signifikanzniveau von $\alpha = 0,01$ angenommen, um der Akkumulation des Alphafehlers aufgrund der Durchführung mehrerer Tests in derselben Stichprobe entgegenzuwirken.

4.1 Prüfung der Abbruchraten der Versionen

Die Abbruchraten waren in der Version A 34 %, in der Version B 27,8 % und in der Version C 28,1 %. Um festzustellen, ob sich die Abbruchraten der drei Versionen signifikant voneinander unterscheiden, wurde diesbezüglich post hoc eine entsprechende explorative Hypothese aufgestellt. Zur Beantwortung dieser Fragestellung wurde ein Chi-Quadrat-Test durchgeführt. Der errechnete Wert betrug 0,41 und erwies sich, verglichen mit dem kritischen χ^2 -Wert bei $df = 2$ und einem Signifikanzniveau von $\alpha = 0,05$, als nicht signifikant. Die Abbruchraten der Stichproben der drei Versionen unterscheiden sich demnach nicht signifikant voneinander.

4.2 Analyse der Homogenität der Subskalen innerhalb der Skalen des WSP mittels Martin-Löf-Tests

Um die Hypothese ob die Items innerhalb einer übergeordneten Skala verschiedene messen, zu testen wurden Martin-Löf-Tests durchgeführt. Hierfür wurden die einzelnen Items einer Skala den jeweiligen inhaltlich definierten Subskalen zugeordnet. Für diese Itemgruppen kann mittels Martin-Löf-Test festgestellt werden, ob sie innerhalb einer Skala homogen sind oder nicht. Da die Skalen Belastbarkeit (Frustrationstoleranz) und Selbstbild jeweils nur aus einer Subskala bestehen, wurden für sie dementsprechend keine dahingehenden Analysen durchgeführt. Nachfolgend sind die Ergebnisse der Analysen mittels der Martin-Löf-Tests für jede der drei Versionen A, B und C dargestellt. Signifikante Ergebnisse wurden durch das Symbol * gekennzeichnet, nicht signifikanten Ergebnissen wurde die Abkürzung (n. s.) nachgestellt.

Tabelle 11: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die fünf Subskalen der Skala „Organisation“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Organisation	Informiertheit	χ_{emp}^2 :	127,65 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Prioritätensetzung	χ_{emp}^2 :	97,70 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Planen	χ_{emp}^2 :	75,29 (n. s.)
		Degrees of freedom:	119
		$\chi^2_{\text{krit}} \alpha (1\%)$	156,65
	Organisieren	χ_{emp}^2 :	91,08 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Zeitmanagement	χ_{emp}^2 :	91,31 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28

Tabelle 12: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, für die fünf Subskalen des WSP, welche der Skala „Organisation“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Organisation	Informiertheit	χ^2_{emp} :	99,14 (n. s.)
		Degrees of freedom:	119
		$\chi^2_{\text{krit}} \alpha (1\%)$	156,65
	Prioritätensetzung	χ^2_{emp} :	72,71 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Planen	χ^2_{emp} :	83,84 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Organisieren	χ^2_{emp} :	70,80 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Zeitmanagement	χ^2_{emp} :	76,58 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28

Tabelle 13: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die fünf Subskalen der Skala „Organisation“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Organisation	Informiertheit	χ^2_{emp} :	132,65 (n. s.)
		Degrees of freedom:	119
		$\chi^2_{\text{krit}} \alpha (1\%)$	156,65
	Prioritätensetzung	χ^2_{emp} :	87,16 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Planen	χ^2_{emp} :	96,05 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Organisieren	χ^2_{emp} :	78,08 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28
	Zeitmanagement	χ^2_{emp} :	75,97 (n. s.)
		Degrees of freedom:	137
		$\chi^2_{\text{krit}} \alpha (1\%)$	177,28

Tabelle 14: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die fünf Subskalen der Skala „Arbeitsstil und Arbeitsverhalten“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Arbeitsstil und Arbeitsverhalten	Genauer Arbeitsstil	χ^2_{emp} :	84,84 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73
	Eigeninitiative	χ^2_{emp} :	123,28 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73
	Gewissenhaftigkeit	χ^2_{emp} :	77,67 (n. s.)
		Degrees of freedom:	125
		$\chi^2_{\text{krit}} \alpha (1\%)$	163,55
	Selbstständiger Arbeitsstil	χ^2_{emp} :	166,38*
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73

Tabelle 15: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und randomisierter subskalenweiser Itemzuweisung innerhalb, für die vier Subskalen des WSP, welche der Skala „Arbeitsstil und Arbeitsverhalten“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Arbeitsstil und Arbeitsverhalten	Genauer Arbeitsstil	$\chi^2_{\text{emp}}:$	91,30 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73
	Eigeninitiative	$\chi^2_{\text{emp}}:$	145,87 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73
	Gewissenhaftigkeit	$\chi^2_{\text{emp}}:$	106,00 (n. s.)
		Degrees of freedom:	125
		$\chi^2_{\text{krit}} \alpha (1\%)$	163,55
	Selbstständiger Arbeitsstil	$\chi^2_{\text{emp}}:$	133,66 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73

Tabelle 16: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die vier Subskalen der Skala „Arbeitsstil und Arbeitsverhalten“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Arbeitsstil und Arbeitsverhalten	Genauer Arbeitsstil	χ^2_{emp} :	120,31 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73
	Eigeninitiative	χ^2_{emp} :	123,34 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73
	Gewissenhaftigkeit	χ^2_{emp} :	102,65 (n. s.)
		Degrees of freedom:	125
		$\chi^2_{\text{krit}} \alpha (1\%)$	163,55
	Selbstständiger Arbeitsstil	χ^2_{emp} :	135,43 (n. s.)
		Degrees of freedom:	113
		$\chi^2_{\text{krit}} \alpha (1\%)$	149,73

Tabelle 17: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die sechs Subskalen der Skala "Motivation".

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Motivation	Erfolgszuversicht	χ^2_{emp} :	126,14 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Durchhaltevermögen	χ^2_{emp} :	94,28 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Lernbereitschaft	χ^2_{emp} :	137,00 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Zielorientierung	χ^2_{emp} :	180,27 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Engagement	χ^2_{emp} :	173,60 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Zielsetzung	χ^2_{emp} :	170,54 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81

Tabelle 18: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und randomisierter subskalenweiser Itemzuweisung innerhalb, für die sechs Subskalen des WSP, welche der Skala „Motivation“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Motivation	Erfolgszuversicht	$\chi^2_{\text{emp}}:$	135,87 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Durchhaltevermögen	$\chi^2_{\text{emp}}:$	75,23 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Lernbereitschaft	$\chi^2_{\text{emp}}:$	142,15 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Zielorientierung	$\chi^2_{\text{emp}}:$	134,66 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Engagement	$\chi^2_{\text{emp}}:$	139,61 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Zielsetzung	$\chi^2_{\text{emp}}:$	206,11 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81

Tabelle 19: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die sechs Subskalen, welche der Skala „Motivation“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Motivation	Erfolgszuversicht	χ^2_{emp} :	158,39 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Durchhaltevermögen	χ^2_{emp} :	83,46 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Lernbereitschaft	χ^2_{emp} :	122,61 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Zielorientierung	χ^2_{emp} :	138,90 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Engagement	χ^2_{emp} :	134,15 (n. s.)
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81
	Zielsetzung	χ^2_{emp} :	242,45*
		Degrees of freedom:	179
		$\chi^2_{\text{krit}} \alpha (1\%)$	224,81

Tabelle 20: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die vier Subskalen der Skala „Emotionale Kompetenz“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Emotionale Kompetenz	Distanzierungsfähigkeit	χ^2_{emp} :	90,72 (n. s.)
		Degrees of freedom:	63
		$\chi^2_{\text{krit}} \alpha (1\%)$	90,80
	Offenheit	χ^2_{emp} :	90,72 (n. s.)
		Degrees of freedom:	63
		$\chi^2_{\text{krit}} \alpha (1\%)$	90,80

Tabelle 21: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und subskalenweiser randomisierter Itemzuweisung innerhalb, für die beiden Subskalen des WSP, welche der Skala „Emotionale Kompetenz“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Emotionale Kompetenz	Distanzierungsfähigkeit	$\chi^2_{\text{emp}}:$	117,95*
		Degrees of freedom:	63
		$\chi^2_{\text{krit}} \alpha (1\%)$	90,80
	Offenheit	$\chi^2_{\text{emp}}:$	97,47*
		Degrees of freedom:	63
		$\chi^2_{\text{krit}} \alpha (1\%)$	90,80

Tabelle 22: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die zwei Subskalen, welche der Skala „Emotionale Kompetenz“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Emotionale Kompetenz	Distanzierungsfähigkeit	$\chi^2_{\text{emp}}:$	103.51*
		Degrees of freedom:	63
		$\chi^2_{\text{krit}} \alpha (1\%)$	90,80
	Offenheit	$\chi^2_{\text{emp}}:$	103.51*
		Degrees of freedom:	63
		$\chi^2_{\text{krit}} \alpha (1\%)$	90,80

Tabelle 23: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die drei Subskalen der Skala „Soziale Kompetenz“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Soziale Kompetenz	Kontaktfähigkeit	$\chi^2_{\text{emp}}:$	154,49*
		Degrees of freedom:	104
		$\chi^2_{\text{krit}} \alpha (1\%)$	139,30
	Teamfähigkeit	$\chi^2_{\text{emp}}:$	114,37 (n. s.)
		Degrees of freedom:	111
		$\chi^2_{\text{krit}} \alpha (1\%)$	147,41
	Kommunikationsfähigkeit	$\chi^2_{\text{emp}}:$	138,31 (n. s.)
		Degrees of freedom:	104
		$\chi^2_{\text{krit}} \alpha (1\%)$	139,30

Tabelle 24: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und subskalenweise randomisierter Itemzuweisung innerhalb, für die drei Subskalen des WSP, welche der Skala „Soziale Kompetenz“ zugeordnet sind.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Soziale Kompetenz	Kontaktfähigkeit	$\chi^2_{\text{emp}}:$	117,95 (n. s.)
		Degrees of freedom:	104
		$\chi^2_{\text{krit}} \alpha (1\%)$	139,30
	Teamfähigkeit	$\chi^2_{\text{emp}}:$	94,83 (n. s.)
		Degrees of freedom:	111
		$\chi^2_{\text{krit}} \alpha (1\%)$	147,41
	Kommunikationsfähigkeit	$\chi^2_{\text{emp}}:$	122,67 (n. s.)
		Degrees of freedom:	104
		$\chi^2_{\text{krit}} \alpha (1\%)$	139,30

Tabelle 25: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die drei Subskalen der Skala „Soziale Kompetenz“.

Skala	Subskala	Ergebnisse der Likelihood-Ratio-Tests	
Soziale Kompetenz	Kontaktfähigkeit	χ^2_{emp} :	146,49*
		Degrees of freedom:	104
		$\chi^2_{\text{krit}} \alpha (1\%)$	139,30
	Teamfähigkeit	χ^2_{emp} :	121,63 (n. s.)
		Degrees of freedom:	111
		$\chi^2_{\text{krit}} \alpha (1\%)$	147,41
	Kommunikationsfähigkeit	χ^2_{emp} :	95,09 (n. s.)
		Degrees of freedom:	104
		$\chi^2_{\text{krit}} \alpha (1\%)$	139,30

In der Version A mit durchmischter Itemreihenfolge waren die Subskalen „Selbstständiger Arbeitsstil“ der Skala „Arbeitsstil und Arbeitsverhalten“ und die Subskala „Kontaktfähigkeit“ der Skala „Soziale Kompetenz“ signifikant. Für beide Subskalen der Skala „Emotionale Kompetenz“ in der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, konnte die Eindimensionalität nach Rasch nicht angenommen werden. Die vollständig geblockte Itemreihenfolge der Version C enthielt neben den beiden signifikanten Subtests „Distanzierungsfähigkeit“ und „Offenheit“ der Skala „Emotionale Kompetenz“ auch eine auffällige Subskala der Skala „Soziale Kompetenz“ („Kontaktfähigkeit“) sowie die nicht eindimensional messende Itemgruppe der Subskala „Zielsetzung“ der Skala „Motivation“.

Als am auffälligsten erwiesen sich die Subskalen „Distanzierungsfähigkeit“ und „Offenheit“ der Skala „Emotionale Kompetenz“, da für die Subskalen der Skala in zwei Versionen (B und C) nicht angenommen werden kann, dass sie ein homogenes Konstrukt im Sinne des Rasch-Modells messen. Weiters zeigten Martin-Löf-Tests, dass für die Subskala „Kontaktfähigkeit“ der Skala „Soziale Kompetenz“ nicht angenommen werden kann, in der getesteten Stichprobe eindimensional zu messen, da diese in den Versionen A und C signifikant ausgefallen sind. Für die restlichen Subskalen der Skala „Soziale Kompetenz“ „Teamfähigkeit (Kompromissbereitschaft)“, „Teamfähigkeit (Kollegialität)“, „Teamfähigkeit

(Teamorientierung)“ und „Kommunikationsfähigkeit“ kann hingegen Eindimensionalität innerhalb des Konstruktes der Skala angenommen werden.

Bei der Skala „Motivation“ fiel nur die Subskala „Zielsetzung“ in der Version C signifikant aus. Durch Ausschluss der Items der Subskala „Zielsetzung“ wurde für die entstandenen Skalen „Motivation ohne Zielsetzung“ und die Skala „Zielsetzung“ in allen drei Versionen Itemhomogenität erreicht. Allerdings mussten in den anschließenden Tests für die Rasch-Modell-Konformität in zwei von drei Versionen mehr Items als bei der vorhergehenden Skala „Motivation“ ausgeschlossen werden, um Rasch-Modell-Homogenität zu erreichen und die Aufspaltung der Skalen stellte auch in den nachfolgenden LLTM-Analysen keinen Erklärungsgewinn dar. Aufgrund dieser Ergebnisse wurde entschieden die Skala Motivation nicht in die zwei Skalen „Motivation ohne Zielsetzung“ und „Zielsetzung“ aufzuteilen.

Die Subskala „selbstständiger Arbeitsstil“ der Skala „Arbeitsstil und Arbeitsverhalten“ erzielte in der Version A signifikante Ergebnisse. Aufgrund der inhaltlichen Ähnlichkeiten der Items der Subskalen „selbstständiger Arbeitsstil“ und „Eigeninitiative“ sowie „Genauer Arbeitsstil“ und „Gewissenhaftigkeit“ wurden diese in die Itemgruppen „selbstständiger Arbeitsstil und Eigeninitiative“ sowie „genauer Arbeitsstil und Gewissenhaftigkeit“ unterteilt. Für die so entstandenen Itemgruppen wurden anschließend Martin-Löf-Tests durchgeführt um diese Teilung empirisch zu prüfen. Die Skalen erreichten in Version A einen Andersen χ^2 -Wert von 147,74, in der Version B von 122,78 und in der Version mit der Itemreihenfolge der Version C von 180,93. Bei jeweils 155 Freiheitsgraden und einem entsprechenden kritischen χ^2 -Wert (bei $\alpha = 1\%$) von 198,87 fielen alle drei Martin-Löf-Tests nicht signifikant aus, somit kann für die beiden Subskalen „genauer Arbeitsstil und Gewissenhaftigkeit“ und „selbstständiger Arbeitsstil und Eigeninitiative“ der Skala „Arbeitsstil und Arbeitsverhalten“ in allen drei Versionen Eindimensionalität angenommen werden.

Die anschließende Tabelle gibt nochmal die Skalenstruktur des WSP sowie die durch die Ergebnisse der Martin-Löf-Tests daran vorgenommenen Anpassungen wieder.

Tabelle 26: Struktur des WSP vor sowie nach den durchgeführten Martin-Löf-Tests.

Struktur des WSP		Struktur nach den Martin-Löf-Tests	
Skalen	Zugehörige Subskalen	Skalen	Zugehörige Subskalen
Organisation	Zeitmanagement	Organisation	Zeitmanagement
	Prioritätensetzung		Prioritätensetzung
	Planen		Planen
	Organisieren		Organisieren
	Informiertheit		Informiertheit
Arbeitsstil und Arbeitsverhalten	Gewissenhaftigkeit	Arbeitsstil und Arbeitsverhalten	Genauer Arbeitsstil und Gewissenhaftigkeit
	Genauer Arbeitsstil		Selbstständiger Arbeitsstil und Eigeninitiative
	Eigeninitiative		
	Selbstständiger Arbeitsstil		
Motivation	Durchhaltevermögen	Motivation	Durchhaltevermögen
	Lernbereitschaft		Lernbereitschaft
	Engagement		Engagement
	Zielorientierung		Zielorientierung
	Zielsetzung		Zielsetzung
	Erfolgszuversicht		Erfolgszuversicht
Belastbarkeit	Frustrationstoleranz	Belastbarkeit	Frustrationstoleranz
Soziale Kompetenz	Kommunikationsfähigkeit	Kommunikationsfähigkeit	
	Teamfähigkeit	Teamfähigkeit	
	Kontaktfähigkeit	Kontaktfähigkeit	
Emotionale Kompetenz	Distanzierungsfähigkeit	Distanzierungsfähigkeit	
	Offenheit	Offenheit	
Selbstbild	Selbstbewusstsein	Selbstbild	Selbstbewusstsein

Entsprechend dieser geänderten Struktur wurden die Items im Anschluss skalenweise für jede Version auf Rasch-Modell-Konformität mittels Likelihood-Ratio-Tests geprüft.

4.3 Analyse auf Rasch-Modell-Konformität

Aufgrund der modifizierten Struktur, die aus den Ergebnissen der Martin-Löf-Tests resultierte, wurden in allen drei Versionen A, B und C die Skalen „Organisation“, „Motivation“, „Arbeitsstil und Arbeitsverhalten“, „Belastbarkeit“, „Kontaktfähigkeit“, „Teamfähigkeit“, „Kommunikationsfähigkeit“, „Distanzierungsfähigkeit“, „Offenheit“ und „Selbstbild“ auf Skalenebene auf ihre Rasch-Modell-Konformität geprüft.

Der Ausschluss von Items, um Rasch-Modell-Konformität zu erreichen, erfolgte schrittweise. Hierfür wurden nur jene Items aus dem Datensatz entfernt, welche sowohl im grafischen

Modelltest als auch in Bezug auf ihre Itemfit-Parameter am auffälligsten waren. Da die Linking-Items zur Berechnung der LLTMs notwendig sind, wurden so wenige Linking-Items wie möglich ausgeschlossen. Es wurden keine weiteren Items ausgeschlossen, sobald in allen Teilungskriterien ein nicht signifikantes Ergebnis erreicht wurde.

Nachfolgend sind die Ergebnisse der Modelltests jeder Skala separat für jede Version wiedergegeben. Die Likelihood-Ratio-Tests wurden mit den Teilungskriterien Mittelwert, Geschlecht, Ausbildung und Alter durchgeführt; signifikante Ergebnisse sind durch das Symbol * hervorgehoben.

4.3.1 Die Skala „Organisation“

Die nachfolgenden Tabellen geben die Ergebnisse der Modelltests zur Überprüfung der Rasch-Modell-Konformität der Skala „Organisation“ wieder. In den Teilstichproben mit der Itemreihenfolge der Version B und C konnte nach Ausschluss von vier Items in allen Teilungskriterien ein nicht signifikantes Ergebnis erreicht werden. Zur Erreichung der Rasch-Modell-Konformität mussten in der Version A fünf Items entfernt werden. Die genauen Ergebnisse für die Versionen A, B und C können den folgenden Tabellen entnommen werden.

Tabelle 27: Ergebnisse der Likelihood-Ratio-Tests der Skala „Organisation“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	60,95*	41,73	68,50*	23,93
Degrees of Freedom:	28	28	28	28
$\chi^2_{\text{krit } \alpha (1\%)}$	48,28	48,28	48,28	48,28
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	37,96	32,50	35,80	18,90
Degrees of Freedom:	23	23	23	23
$\chi^2_{\text{krit } \alpha (1\%)}$	41,64	41,64	41,64	41,64

Tabelle 28: Ergebnisse der Likelihood-Ratio-Tests der Skala „Organisation“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	61,97*	53,02*	46,10	22,47
Degrees of Freedom:	27	28	28	28
$\chi^2_{\text{krit}} \alpha (1\%)$	46,96	48,28	48,28	48,28
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	28,51	40,91	36,42	17,80
Degrees of Freedom:	24	24	24	24
$\chi^2_{\text{krit}} \alpha (1\%)$	42,98	42,98	42,98	42,98

Tabelle 29: Ergebnisse der Likelihood-Ratio-Tests der Skala „Organisation“ der Version C, mit vollständig geblockter Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	59,23*	46,67	58,20*	27,33
Degrees of Freedom:	27	28	28	28
$\chi^2_{\text{krit}} \alpha (1\%)$	47,00	48,28	48,28	48,28
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	33,55	38,52	37,30	20,26
Degrees of Freedom:	24	24	24	24
$\chi^2_{\text{krit}} \alpha (1\%)$	42,98	42,98	42,98	42,98

4.3.2 Die Skala „Arbeitsstil und Arbeitsverhalten“

Die Ergebnisse der Modelltests zur Überprüfung der Rasch-Modell-Konformität der Skala „Arbeitsstil und Arbeitsverhalten“ wurden in den folgenden Tabellen zusammengefasst. In der Teilstichprobe A mussten drei Items und in der Teichstichprobe B vier Items entfernt werden. Rasch-Modell-Konformität wurde in der Teilstichprobe mit der Itemreihenfolge

Version C mit sieben Itemausschlüssen erreicht. Die Ergebnisse der diesbezüglichen Likelihood-Ratio-Tests können den nachstehenden Tabellen entnommen werden.

Tabelle 30: Ergebnisse der Likelihood-Ratio-Tests der Skala „Arbeitsstil und Arbeitsverhalten“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	40,17	47,81*	54,71*	31,03
Degrees of Freedom:	24	24	24	24
$\chi^2_{\text{krit } \alpha (1\%)}$	42,98	42,98	42,98	42,98
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	34,20	36,30	38,64	30,22
Degrees of Freedom:	21	21	21	21
$\chi^2_{\text{krit } \alpha (1\%)}$	38,93	38,93	38,93	38,93

Tabelle 31: Ergebnisse der Likelihood-Ratio-Tests der Skala „Arbeitsstil und Arbeitsverhalten“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	53,95*	23,85	47,47*	46,60*
Degrees of Freedom:	24	24	24	24
$\chi^2_{\text{krit } \alpha (1\%)}$	42,98	42,98	42,98	42,98
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	28,12	23,10	30,54	28,43
Degrees of Freedom:	20	20	20	20
$\chi^2_{\text{krit } \alpha (1\%)}$	37,60	37,60	37,60	37,60

Tabelle 32: Ergebnisse der Likelihood-Ratio-Tests der Skala „Arbeitsstil und Arbeitsverhalten“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	83,14*	29,40	31,48	24,53
Degrees of Freedom:	22	24	24	24
$\chi^2_{\text{krit}} \alpha (1\%)$	40,29	42,98	42,98	42,98
Likelihood-Ratio-Tests nach Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	28,15	23,90	22,82	11,89
Degrees of Freedom:	17	17	17	17
$\chi^2_{\text{krit}} \alpha (1\%)$	33,41	33,41	33,41	33,41

4.3.3 Die Skala „Belastbarkeit“

In der Teilstichprobe mit der Itemreihenfolge der Version A konnte nach Ausschluss von drei Items und in der Version B nach dem Entfernen von zwei Items mit allen Teilungskriterien für die Skala „Belastbarkeit“ ein nicht signifikantes Ergebnis erreicht werden. In der Teilstichprobe der Itemreihenfolge C musste kein Item entfernt werden um Rasch-Modell-Konformität zu erreichen.

Die Ergebnisse der Likelihood-Ratio-Tests aller Teilstichproben können den nachstehenden Tabellen entnommen werden.

Tabelle 33: Ergebnisse der Likelihood-Ratio-Tests der Skala „Belastbarkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	32,66*	36,21*	30,88*	3,98
Degrees of Freedom:	11	11	11	11
$\chi^2_{\text{krit}} \alpha (1\%)$	24,73	24,73	24,73	24,73
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	7,46	12,10	16,23	3,59
Degrees of Freedom:	8	8	8	8
$\chi^2_{\text{krit}} \alpha (1\%)$	20,09	20,09	20,09	20,09

Tabelle 34: Ergebnisse der Likelihood-Ratio-Tests der Skala „Belastbarkeit“ der Version B, skalenweise geblockte Itemreihenfolge mit subskalenweise randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	46,83*	21,74	5,64	1,54
Degrees of Freedom:	11	11	11	11
$\chi^2_{\text{krit}} \alpha (1\%)$	23,21	23,21	23,21	23,21
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	21,10	7,10	4,00	1,23
Degrees of Freedom:	9	9	9	9
$\chi^2_{\text{krit}} \alpha (1\%)$	21,67	21,67	21,67	21,67

Tabelle 35: Ergebnisse der Likelihood-Ratio-Tests der Skala „Belastbarkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	20,92	17,65	19,10	13,30
Degrees of Freedom:	11	11	11	11
$\chi^2_{\text{krit } \alpha (1\%)}$	24,73	24,73	24,73	24,73

4.3.4 Die Skala „Distanzierungsfähigkeit“

Um Rasch-Modell-Konformität zu erreichen musste bei der Skala „Distanzierungsfähigkeit“ der Teilstichprobe der Itemreihenfolge B ein Item ausgeschlossen werden. In der Version A mussten in der Subskala „Distanzierungsfähigkeit“ zwei und der Version C drei Items entfernt werden um ein nicht signifikantes Ergebnis mit allen Teilungskriterien zu erhalten. Die folgenden Tabellen enthalten diesbezüglich nähere Informationen.

Tabelle 36: Ergebnisse der Likelihood-Ratio-Tests der Skala „Distanzierungsfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	38,24	15,40	7,95	9,33
Degrees of Freedom:	8	8	8	8
$\chi^2_{\text{krit } \alpha (1\%)}$	20,09	20,09	20,09	20,09
Likelihood-Ratio-Tests nach Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	13,46	16,74	7,82	10,12
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit } \alpha (1\%)}$	16,81	16,81	16,81	16,81

Tabelle 37: Ergebnisse der Likelihood-Ratio-Tests der Skala „Distanzierungsfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	20,84*	3,11	12,46	3,11
Degrees of Freedom:	7	7	7	7
$\chi^2_{\text{krit}} \alpha (1\%)$	18,48	18,48	18,48	18,48
Likelihood-Ratio-Tests nach Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	6,66	2,56	11,88	2,96
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81

Tabelle 38: Ergebnisse der Likelihood-Ratio-Tests der Skala „Distanzierungsfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	23,01*	12,93	11,94	12,30
Degrees of Freedom:	7	7	7	7
$\chi^2_{\text{krit}} \alpha (1\%)$	18,48	18,48	18,48	18,48
Likelihood-Ratio-Tests nach Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	4,41	1,25	1,28	8,47
Degrees of Freedom:	4	4	4	4
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	13,28	13,28	13,28

4.3.5 Die Skala „Offenheit“

In der Skala „Offenheit“ musste in allen Teilstichproben ein Item entfernt werden um Rasch-Modell-Konformität zu erreichen.

Die entsprechenden Ergebnisse können den nachfolgenden Tabellen entnommen werden.

Tabelle 39: Ergebnisse der Likelihood-Ratio-Tests der Skala „Offenheit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	12,91	6,78	9,10	3,84
Degrees of Freedom:	6	7	7	7
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	18,48	18,48	18,48
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	12,91	6,49	7,66	3,66
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81

Tabelle 40: Ergebnisse der Likelihood-Ratio-Tests der Skala „Offenheit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	9,85	2,61	4,35	14,40
Degrees of Freedom:	6	6	7	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	18,48	16,81
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	9,85	2,61	4,14	14,40
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	15,09	15,09	15,09	15,09

Tabelle 41: Ergebnisse der Likelihood-Ratio-Tests der Skala „Offenheit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	4,70	3,53	4,70	6,39
Degrees of Freedom:	6	7	6	7
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	18,48	16,81	18,48
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	4,70	3,41	4,70	6,01
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81

4.3.6 Die Skala „Motivation“

In den Teilstichproben der Itemreihenfolgen A und C der Skala „Motivation“ mussten jeweils fünf Items entfernt werden um Rasch-Modell-Konformität zu erreichen. Ein nicht signifikantes Ergebnis wurde in der Teilstichprobe der Itemreihenfolge B nach dem Ausschluss von drei Items erhalten. In den nachstehenden Tabellen wurden die Ergebnisse der diesbezüglichen Likelihood-Ratio-Tests aufgelistet.

Tabelle 42: Ergebnisse der Likelihood-Ratio-Tests der Skala „Motivation“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	63,98*	59,23*	46,23	58,96*
Degrees of Freedom:	33	35	35	35
$\chi^2_{\text{krit}} \alpha (1\%)$	54,78	57,34	57,34	57,34
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	45,65	49,77	39,48	46,74
Degrees of Freedom:	30	30	30	30
$\chi^2_{\text{krit}} \alpha (1\%)$	50,89	50,89	50,89	50,89

Tabelle 43: Ergebnisse der Likelihood-Ratio-Tests der Skala „Motivation“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	46,93	42,83	32,06	30,52
Degrees of Freedom:	32	35	35	35
$\chi^2_{\text{krit}} \alpha (1\%)$	53,49	57,34	57,34	57,34
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	46,93	40,90	29,78	29,95
Degrees of Freedom:	32	32	32	32
$\chi^2_{\text{krit}} \alpha (1\%)$	53,49	52,19	52,19	52,19

Tabelle 44: Ergebnisse der Likelihood-Ratio-Tests der Skala „Motivation“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	57,19*	34,90	40,32	38,97
Degrees of Freedom:	32	35	35	35
$\chi^2_{\text{krit}} \alpha (1\%)$	53,49	57,34	57,34	57,34
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	48,13	33,74	30,65	33,52
Degrees of Freedom:	30	30	30	30
$\chi^2_{\text{krit}} \alpha (1\%)$	50,89	50,89	50,89	50,89

Aufgrund der Ergebnisse der Analysen der Skala „Soziale Kompetenz“ mittels des Martin-Löf-Tests wurde diese in die Skalen „Kontaktfähigkeit“, „Teamfähigkeit“ und „Kommunikationsfähigkeit“ geteilt. Die folgenden drei Kapitel geben die Ergebnisse der Tests auf Rasch-Modell-Konformität dieser Skalen wieder.

4.3.7 Die Skala „Kontaktfähigkeit“

In den Teilstichproben der Itemreihenfolgen A und C der Subskala „Kontaktfähigkeit“ mussten jeweils zwei Items entfernt werden. In der Teilstichprobe der Version B musste ein Item ausgeschlossen werden um ein nicht signifikantes Ergebnis und somit Rasch-Modell-Konformität zu erreichen. Die nachfolgenden Tabellen enthalten die entsprechenden Ergebnisse für die Versionen A, B und C.

Tabelle 45: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kontaktfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	41,22	1,88	20,65	13,01
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	2,91	2,08	8,66	7,28
Degrees of Freedom:	4	4	4	4
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	13,28	13,28	13,28

Tabelle 46: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kontaktfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	26,98*	12,36	7,38	2,48
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	7,41	10,99	5,52	2,34
Degrees of Freedom:	5	5	5	5
$\chi^2_{\text{krit}} \alpha (1\%)$	15,09	15,09	15,09	15,09

Tabelle 47: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kontaktfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	31,61	6,73	10,76	4,03
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	0,68	5,30	10,82	1,70
Degrees of Freedom:	4	4	4	4
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	13,28	13,28	13,28

4.3.8 Die Skala „Teamfähigkeit“

Für die Skala „Teamfähigkeit“ mussten zur Erreichung der Rasch-Modell-Konformität in der Version A ein Item und in den Itemreihenfolgen der Versionen B und C jeweils drei Items entfernt werden.

Tabelle 48: Ergebnisse der Likelihood-Ratio-Tests der Skala „Teamfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	6,90	11,83	9,41	5,80
Degrees of Freedom:	6	7	7	7
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	18,48	18,48	18,48
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	6,90	10,05	8,92	5,06
Degrees of Freedom:	6	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	16,81	16,81	16,81	16,81

Tabelle 49: Ergebnisse der Likelihood-Ratio-Tests der Skala „Teamfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	0,47	8,98	7,46	7,51
Degrees of Freedom:	4	7	7	7
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	18,48	18,48	18,48
Likelihood-Ratio-Tests nach Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	0,47	0,64	6,33	4,389
Degrees of Freedom:	4	4	4	4
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	13,28	13,28	13,28

Tabelle 50: Ergebnisse der Likelihood-Ratio-Tests der Skala „Teamfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	9,90	6.975	3.626	3
Degrees of Freedom:	4	7	7	7
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	18,48	18,48	18,48
Likelihood-Ratio-Tests nach Itemausschluss	Teilungskriterien			
	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	9,90	5,60	3,25	2,20
Degrees of Freedom:	4	4	4	4
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	13,28	13,28	13,28

4.3.9 Die Skala „Kommunikationsfähigkeit“

Rasch-Modell-Konformität wurde für die Skala „Kommunikationsfähigkeit“ der Teilstichprobe der Itemreihenfolge A nach dem Ausschluss von drei Items, in Version B nach dem Entfernen von einem Item und in der Teilstichprobe der Itemreihenfolge C nach der

Exklusion von zwei Items erhalten. Diese Ergebnisse der entsprechenden Likelihood-Ratio-Tests können den nachstehenden Tabellen entnommen werden.

Tabelle 51: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kommunikationsfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	20,32*	3,25	10,95	4,85
Degrees of Freedom:	4	5	5	5
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	15,09	15,09	15,09
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	3,38	2,97	8,25	3,83
Degrees of Freedom:	3	3	3	3
$\chi^2_{\text{krit}} \alpha (1\%)$	11,35	11,35	11,35	11,35

Tabelle 52: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kommunikationsfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	15,08	2,70	5,93	2,74
Degrees of Freedom:	5	6	6	6
$\chi^2_{\text{krit}} \alpha (1\%)$	15,09	16,81	16,81	16,81
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	15,08	1,43	5,48	2,35
Degrees of Freedom:	5	5	5	5
$\chi^2_{\text{krit}} \alpha (1\%)$	15,09	15,09	15,09	15,09

Tabelle 53: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kommunikationsfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	3,94	6,28	5,39	3,27
Degrees of Freedom:	4	5	5	5
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	15,09	15,09	15,09
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	4,27	5,91	2,68	3,03
Degrees of Freedom:	4	4	4	4
$\chi^2_{\text{krit}} \alpha (1\%)$	13,28	13,28	13,28	13,28

4.3.10 Die Skala „Selbstbild“

Für die Teilstichprobe der Itemreihenfolge A der Skala „Selbstbild“ wurde nach dem Ausschluss von drei Items Rasch-Modell-Konformität erzielt. Ein nicht signifikantes Ergebnis und somit Rasch-Modell-Homogenität wurde für die Teilstichproben der Itemreihenfolgen B und C nach dem Entfernen von jeweils zwei Items erreicht. Die Ergebnisse der entsprechenden Likelihood-Ratio-Tests sind in den nachfolgenden Tabellen dargestellt.

Tabelle 54: Ergebnisse der Likelihood-Ratio-Tests der Skala „Selbstbild“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	47,24*	24,08	6,57	23,05
Degrees of Freedom:	12	12	12	12
$\chi^2_{\text{krit}} \alpha (1\%)$	26,22	26,22	26,22	26,22
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	16,40	15,43	2,51	18,30
Degrees of Freedom:	9	9	9	9
$\chi^2_{\text{krit}} \alpha (1\%)$	21,67	21,67	21,67	21,67

Tabelle 55: Ergebnisse der Likelihood-Ratio-Tests der Skala „Selbstbild“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	33,60*	7,96	22,51	19,36
Degrees of Freedom:	12	12	12	12
$\chi^2_{\text{krit}} \alpha (1\%)$	26,22	26,22	26,22	26,22
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	22,70	7,63	14,88	17,09
Degrees of Freedom:	10	10	10	10
$\chi^2_{\text{krit}} \alpha (1\%)$	23,21	23,21	23,21	23,21

Tabelle 56: Ergebnisse der Likelihood-Ratio-Tests der Skala „Selbstbild“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.

Likelihood-Ratio-Tests vor		Teilungskriterien		
Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	26,19*	13,74	14,74	27,28
Degrees of Freedom:	11	12	12	12
$\chi^2_{\text{krit}} \alpha (1\%)$	24,73	26,22	26,22	26,22
Likelihood-Ratio-Tests		Teilungskriterien		
nach Itemausschluss	Mittelwert	Geschlecht	Ausbildung	Alter
Andersen χ^2 :	10,68	10,94	12,76	20,13
Degrees of Freedom:	10	10	10	10
$\chi^2_{\text{krit}} \alpha (1\%)$	23,21	23,21	23,21	23,21

4.4 Analysen mittels Linear Logistischem Testmodell

Durch die im letzten Kapitel vorgenommenen Arbeitsschritte konnte für die Skalen des verwendeten Fragebogens in der beforschten Population Eindimensionalität nach Rasch angenommen werden, somit waren die Voraussetzungen für die weiteren LLTM-Analysen erfüllt. Um die eingangs gestellte Frage, ob die Itemkontexte einen Einfluss auf das Antwortverhalten der TeilnehmerInnen haben, beantworten zu können, müssen überzufällige Unterschiede im Antwortverhalten der TeilnehmerInnen der drei Bedingungen nachgewiesen werden. Hierfür wurden drei Modelle aufgestellt: das Modell Itemkontext, das saturierte Modell sowie das Rasch-Modell (siehe Kapitel 3.2.2 *Postulierte Modelle*, S. 28f). Die folgende Tabelle liefert einen Überblick über die aufgestellten Modelle.

Tabelle 57: Mittels LLTM zu vergleichende, hypothetische Modelle

Quasi-saturiertes Modell:	Inhaltsparameter (Root) * 3 Versionen
Modell Itemkontext:	Inhaltsparameter (Root) + 3 Versionsparameter
Rasch-Modell:	Itemparameter = Inhaltsparameter (Root)

Das saturierte Modell diskriminiert nicht die Itemzugehörigkeit zu den drei Versionen, somit behandelt das Modell die Items der drei Fragereihenfolgen so als würden die Versionen unterschiedliche Items beinhalten – es werden so viele Inhaltsparameter definiert wie Items in allen drei Versionen vorgegeben werden.

Das Itemkontextmodell unterscheidet die Items nach ihrer Zugehörigkeit zu den drei Fragebogenversionen und berücksichtigt somit neben den Inhaltsparametern auch drei Versionsparameter. Dieses Modell vermag einen möglichen Effekt der Darbietung von geblockten gegenüber durchmischten Items sparsam zu beschreiben.

Das Rasch-Modell entspricht der konservativen Hypothese, die postuliert, dass die jeweiligen Schwierigkeitsparameter lediglich durch die diesen zugrundeliegenden Items zustande kommen und somit von Positionseffekten unabhängig sind.

Mittels LLTM wird nun für jedes der drei aufgestellten Modelle die Passung mit den erhobenen Daten analysiert werden. Anschließend werden modellvergleichende Likelihood-Ratio-Tests durchgeführt um beantworten zu können, welches Modell am besten zu den empirischen Daten passt. Um die Modelle miteinander vergleichen zu können, werden

zuerst die Parameter jedes Modells, welche als LLTM, durch die Q-Matrizen (siehe Kapitel 3.2.2 *Postulierte Modelle*, S. 28f mathematisch beschrieben wurden, anhand der Maximum-Likelihood-Methode berechnet. Für jede Kombination eines Modells und einer Skala des WSP wird die Passung des Modells auf die empirischen Daten angegeben. Sowohl mittels weiteren Likelihood-Quotienten-Tests als auch mittels Informationskriterien wie beispielsweise dem *Akaike information criterion* (in Folge kurz: AIC) (siehe hierfür Kapitel 3.1.3 *Das Linear Logistische Testmodell von Fischer* (LLTM), S. 22f), können die Modelle untereinander anschließend verglichen werden. Das Rasch-Modell wird gegen die beiden anderen Testmodelle geprüft, wodurch – unter Berücksichtigung der Parameteranzahl der drei Modelle – das am besten auf die empirischen Daten passende Testmodell ermittelt werden kann.

In den anschließenden drei Unterkapiteln sind nun die Ergebnisse der LLTM-Analysen der drei aufgestellten Modelle skalenweise aufgelistet. Die erhaltenen Prüfgrößen der Likelihoodquotiententests wurden, falls nicht anders angegeben, mittels der Conditional-log-Likelihood-Methode geschätzt. Das Signifikanzniveau wurde auf $\alpha = 0,01$ festgesetzt, um der Kumulierung des Alphafehlers durch multiples Testen in derselben Stichprobe entgegenzuwirken. Signifikante Unterschiede wurden mit dem Symbol * gekennzeichnet.

4.4.1 Ergebnisse des Rasch-Modells

Die folgende Tabelle fasst die Ergebnisse der LLTM-Analysen des Rasch-Modells für jede der Skalen des WSP zusammen.

Tabelle 58: Ergebnisse der Modellgleichungstests des Rasch-Modells für die zehn Skalen des WSP.

Skala	Conditional log-likelihood	n_p
Organisation	-3363,51	28
Arbeitsstil und Arbeitsverhalten	-2647,04	24
Motivation	-4561,14	34
Belastbarkeit	-1169,10	11
Distanzierungsfähigkeit	-598,84	7
Offenheit	-657,82	7
Kontaktfähigkeit	-454,60	6
Kommunikationsfähigkeit	-439,39	5
Teamfähigkeit	-411,72	6
Selbstbild	-1079,69	11

4.4.2 Ergebnisse des quasi-saturierten Modells

Tabelle 59: Ergebnisse der LLTM Analysen für das quasi-saturierte Modell anhand jeder der zehn Skalen des WSP.

Skala	Conditional log-likelihood	n_p
Organisation	-3337,03	67
Arbeitsstil und Arbeitsverhalten	-2626,14	53
Motivation	-4532,22	84
Belastbarkeit	-1157,55	24
Distanzierungsfähigkeit	-591,39	14
Offenheit	-655,11	17
Kontaktfähigkeit	-449,03	13
Kommunikationsfähigkeit	-436,82	12
Teamfähigkeit	-410,46	12
Selbstbild	-1071,48	25

4.4.3 Ergebnisse des Itemkontextmodells

Tabelle 60: Ergebnisse der LLTM Analysen für das Modell Itemkontext anhand jeder der zehn Skalen des WSP.

Skala	Conditional log-likelihood	n_p
Organisation	-3360,19	30
Arbeitsstil und Arbeitsverhalten	-2646,42	26
Motivation	-4559,80	36
Belastbarkeit	-1168,00	13
Distanzierungsfähigkeit	-613,00	9
Offenheit	-657,69	9
Kontaktfähigkeit	-454,32	8
Kommunikationsfähigkeit	-439,17	7
Teamfähigkeit	-411,60	8
Selbstbild	-1078,63	13

Im Anschluss an die Berechnung der Likelihoodfunktionen der Skalen aller Testmodelle werden die drei Testmodelle im nachfolgenden Kapitel miteinander verglichen.

4.5 Modellvergleiche

Als nächstes wurde die Passung der drei Modelle mit dem Datensatz einander gegenübergestellt. Der Modellvergleich erfolgte für jeweils zwei Modelle mittels Likelihood-Ratio-Tests, indem die Likelihoodquotienten der drei verschiedenen Testmodelle für den erhaltenen Datensatz verglichen wurden. Um die Modellpassung mit den empirischen Daten zur Erklärung des Zustandekommens der Daten für die verwendeten Modelle zu testen wurden Likelihoodquotienten und ein informationstheoretisches Maße herangezogen. Für die in Folge dargestellten Ergebnisse wurde hierfür das informationstheoretische Maß *Akaike information criterion*²⁹ (AIC) verwendet, welches die Anzahl der benötigten Parameter n_p und den $-2 \log L$ Wert des jeweiligen Modells gegeneinander abwägt. Aufgrund der Kumulierung des Alphafehlers durch multiples Testen in der erhobenen Stichprobe

²⁹ Siehe Kapitel 3.1.3 *Das Linear Logistische Testmodell von Fischer (LLTM)*, S. 22f.

wurde das Signifikanzniveau auf 1 % festgesetzt. Statistisch signifikante Ergebnisse wurden mit * gekennzeichnet. In den folgenden Abschnitten wird zuerst das quasi-saturierte Modell mit dem Itemkontextmodell verglichen. Jenes Modell, welches besser zu den Daten passt, wird dann im Anschluss noch dem Rasch-Modell gegenübergestellt. Schließlich werden die Passungen der Modelle und ihre jeweiligen informationstheoretischen Maße dargestellt.

4.5.1 Modellvergleich des quasi-saturierten Modells mit dem Itemkontextmodell

Im folgenden Abschnitt werden die Ergebnisse des Vergleichs des quasi-saturierten Modells mit dem Itemkontextmodell dargestellt. Hierfür wurde ein Likelihoodquotiententest durchgeführt. Die resultierenden Ergebnisse sind der folgenden Tabelle zu entnehmen

Tabelle 61: *Ergebnisse der modellvergleichenden LRTs für das quasi-saturierte Modell und das Itemkontextmodell für die zehn Skalen des WSP.*

Skala	LRT	df	$\chi^2_{\text{krit } \alpha (1 \%)}$
Organisation	46,32	37	60,00
Arbeitsstil und Arbeitsverhalten	40,57	27	46,96
Motivation	55,17	48	73,68
Belastbarkeit	20,90	11	24,73
Distanzierungsfähigkeit	43,17*	5	15,09
Offenheit	5,15	8	20,09
Kontaktfähigkeit	10,60	5	15,09
Kommunikationsfähigkeit	4,68	12	26,22
Teamfähigkeit	2,27	4	13,28
Selbstbild	14,31	12	26,22

Die Tabelle enthält die Resultate der skalenweisen Likelihood-Ratio-Tests, welcher die Likelihoodquotienten des quasi-saturierten Modells mit jenem des Itemkontextmodells für den gleichen Datensatz vergleicht. Die resultierenden Werte der Likelihood-Ratio-Tests sind χ^2 -verteilt und werden kritischen χ^2 -Werten gegenübergestellt. Da alle Ergebnisse der Skalen bis auf die Skala „Distanzierungsfähigkeit“ kleiner als die entsprechenden kritischen χ^2 -Werte ausgefallen sind, zeigte sich, dass das Modell Itemkontext im Vergleich zum quasi-saturierten Modell in allen Skalen mit Ausnahme der Skala „Distanzierungsfähigkeit“ eine

bessere Passung mit den empirischen Daten aufweist. Es wird daher das Itemkontextmodell im nächsten Abschnitt mit dem Rasch-Modell verglichen.

4.5.2 Modellvergleich des Itemkontextmodell mit dem Rasch-Modell

In Folge werden die Ergebnisse des Modellvergleichs des Itemkontextmodells mit dem Rasch-Modell wiedergegeben.

Tabelle 62: *Ergebnisse der modellvergleichenden LRTs für das Itemkontextmodell mit dem Rasch-Modell für die zehn Skalen des WSP.*

Skala	LRT	df	$\chi^2_{\text{krit } \alpha (1 \%)}$
Organisation	6,66	2	9,21
Arbeitsstil und Arbeitsverhalten	1,23	2	9,21
Motivation	2,68	2	9,21
Belastbarkeit	2,20	2	9,21
Distanzierungsfähigkeit	28,28*	2	9,21
Offenheit	0,26	2	9,21
Kontaktfähigkeit	0,57	2	9,21
Kommunikationsfähigkeit	0,45	2	9,21
Teamfähigkeit	0,25	2	9,21
Selbstbild	2,11	2	9,21

Da die LRTs aller Skalen bis auf die Skala „Distanzierungsfähigkeit“ kleiner als die zugehörigen kritischen Werte ausgefallen sind kann gefolgert werden, dass das Rasch-Modell mit weniger Parametern im Vergleich zum Itemkontextmodell für alle Skalen des WSP mit Ausnahme der Skala „Distanzierungsfähigkeit“ eine bessere Passung mit den empirischen Daten aufweist.

4.5.3 Gegenüberstellung der drei Modelle

Die drei Modelle quasi-saturiertes Modell, Itemkontextmodell und Rasch-Modell werden nun bezüglich ihrer Passung mit dem erhobenen Datensatz miteinander verglichen. Dafür werden sowohl die Likelihoodquotienten der Modelle für die einzelnen Skalen als auch die *Akaike information criteria* als informationstheoretische Maße in der folgenden Tabelle angeführt. Der kleinste AIC pro Skala gibt die beste Passung mit den Daten an und wurde mit * versehen.

Tabelle 63: Gegenüberstellung der Likelihoodquotienten sowie der AICs für die drei Modelle quasi-saturiertes Modell, Itemkontextmodell und Rasch-Modell für die Skalen des WSP.

Skala	Modell	Conditional log-likelihood	n _p	AIC
Organisation	Quasi-saturiertes Modell	-3337,03	67	6808,05
	Itemkontextmodell	-3360,19	30	6780,37*
	Rasch-Modell	-3363,51	28	6783,03
Arbeitsstil und Arbeitsverhalten	Quasi-saturiertes Modell	-2626,14	53	5358,27
	Itemkontextmodell	-2646,42	26	5344,84
	Rasch-Modell	-2647,04	24	5342,07*
Motivation	Quasi-saturiertes Modell	-4532,22	84	9232,43
	Itemkontextmodell	-4559,80	36	9191,61
	Rasch-Modell	-4561,14	34	9190,29*
Belastbarkeit	Quasi-saturiertes Modell	-1157,55	24	2363,09
	Itemkontextmodell	-1168,00	13	2362,00
	Rasch-Modell	-1169,10	11	2360,19*
Distanzierungsfähigkeit	Quasi-saturiertes Modell	-591,39	14	1210,79*
	Itemkontextmodell	-613,00	9	1243,95
	Rasch-Modell	-598,84	7	1211,68
Offenheit	Quasi-saturiertes Modell	-655,11	17	1344,23
	Itemkontextmodell	-657,70	9	1333,38
	Rasch-Modell	-657,82	7	1329,63*
Kontaktfähigkeit	Quasi-saturiertes Modell	-449,03	13	924,05
	Itemkontextmodell	-454,32	8	924,65
	Rasch-Modell	-454,61	6	921,22*
Kommunikationsfähigkeit	Quasi-saturiertes Modell	-436,82	12	897,65
	Itemkontextmodell	-439,17	7	892,33
	Rasch-Modell	-439,39	5	888,78*
Teamfähigkeit	Quasi-saturiertes Modell	-410,46	12	844,92
	Itemkontextmodell	-411,60	8	839,20
	Rasch-Modell	-411,72	6	835,44*
Selbstbild	Quasi-saturiertes Modell	-1071,48	25	2192,95
	Itemkontextmodell	-1078,63	13	2183,27
	Rasch-Modell	-1079,69	11	2181,38*

Den Ergebnissen des informationstheoretischen Maßes AIC zufolge erzielte das Rasch-Modell die beste Modellpassung mit den Daten bei folgenden Skalen des WSP: „Arbeitsstil und Arbeitsverhalten“, „Motivation“, „Belastbarkeit“, „Offenheit“, „Kontaktfähigkeit“, „Kommunikationsfähigkeit“, „Teamfähigkeit“ und „Selbstbild“. Das saturierte Modell, welches von allen Modellen die höchste Parameteranzahl aufweist, war das am besten zu den Daten passende Modell für die Skala „Distanzierungsfähigkeit“. Die empirischen Daten der Skala „Organisation“ wurden am besten durch das Itemkontextmodell erklärt.

Das Rasch-Modell erklärt im Vergleich zu den beiden anderen Modellen mit den wenigsten Parametern das Zustandekommen der Daten von acht der nun zehn Skalen des WSP am besten – somit ist es für die im Zuge dieser Studie beforschte Population als das am besten geeignete Modell zu betrachten. Die Forschungsfrage dieser Studie lautete, ob der unterschiedliche Itemkontext zu einer Veränderung des Antwortverhaltens der TeilnehmerInnen des Studieneignungs-Persönlichkeitsinventars (WSP) führt. Die Ergebnisse dieser Untersuchung sprechen dafür, dass kein relevanter Effekt des Itemkontextes für eine geblockte oder durchmischte Itemreihenfolge vorliegt.

5 Diskussion (und Ausblick)

Zur Beantwortung der Forschungsfrage, ob die unterschiedlichen Itemkontexte der drei Versionen einen Einfluss auf das Antwortverhalten der StudentInnen haben, wurden Analysen mit dem Linear Logistischen Testmodell (LLTM) vorgenommen werden, welches Eindimensionalität nach Rasch für die erhobene Eigenschaft in der untersuchten Population voraussetzt. Um diese Rasch-Modell-Konformität des Datensatzes zu erreichen wurden einerseits Tests nach Martin Löf, mit welchen überprüft wurde, ob die Subskalen innerhalb einer übergeordneten Skala ein homogenes Konstrukt messen und andererseits Likelihood-Ratio-Tests, zur Analyse der Items einer Skala auf Passung mit dem Rasch-Modell, durchgeführt.

Ausgehend von diesen Ergebnissen wurde die Skalenstruktur des WSP modifiziert. Für die Subskalen der beiden Skalen „Soziale Kompetenz“ und „Emotionale Kompetenz“ zeigte sich, dass diese nicht konform mit dem Skalenkonstrukt messen – sie wurden somit als eigene

Skalen behandelt. Die ursprüngliche Skalenstruktur mit sieben Skalen wurde dementsprechend adaptiert, sodass zehn Skalen resultierten³⁰.

Anschließend wurden Likelihood-Ratio-Tests zur Prüfung der Daten auf Rasch-Modell-Konformität mit den Teilungskriterien Mittelwert, Geschlecht, Ausbildung und Alter durchgeführt. Die anschließenden Analysen mittels LLTM auf Passung der drei Modelle mit den erhobenen Daten ergab, dass für alle Skalen mit Ausnahme der Skala „Distanzierungsfähigkeit“ das Itemkontextmodell im Vergleich zum quasi-saturierten Modell eine bessere Passung mit den empirischen Daten aufwies und das Rasch-Modell mit den wenigsten Parametern im Vergleich zum Itemkontextmodell für alle Skalen des WSP bis auf die Skala „Distanzierungsfähigkeit“ die erhobenen Daten am effizientesten beschrieb. Auch ein durchgeführter Vergleich anhand von Informationskriterien zeigte die überlegene Modellpassung des Rasch-Modells für den Datensatz. Hier erzielte das Rasch-Modell in acht von zehn Skalen³¹ die beste Modellpassung, die Skala „Distanzierungsfähigkeit“ wurde am besten vom quasi-saturierten Modell (mit der höchsten Parameteranzahl) erklärt und das Itemkontextmodell erklärte das Zustandekommen der Daten der Skala „Organisation“ am besten.

Es bestehen verschiedene mögliche Erklärungsansätze dafür, dass beim Vergleich des quasi-saturierten Modells mit dem Itemkontextmodell bzw. der Gegenüberstellung des Itemkontextmodells und des Rasch-Modells die Skala „Distanzierungsfähigkeit“ jeweils eine bessere Passung mit jenem Modell aufwies, welches das Zustandekommen der Daten der restlichen neun Skalen des WSP schlechter erklärte: Einerseits könnte eine suboptimale Passung der Items einzelner Subskalen der Skala „Distanzierungsfähigkeit“ mit dem Rasch-Modell die Auffälligkeiten verursacht haben. Andererseits könnte die Skala „Distanzierungsfähigkeit“ im Vergleich zu den anderen Skalen mehr umgepolte Items als andere Skalen aufweisen, welche das Antwortverhalten der TeilnehmerInnen beeinflusst haben könnten. Da nach Schwarz (Schwarz, 1999: S.93) die Antworten auf Selbstbeurteilungs- und Einstellungsfragebögen stark von den Eigenheiten des Instruments, wie Frageformulierungen, Antwortformat und Kontext beeinflusst sind, könnten auch

³⁰ Der Vergleich der ursprünglichen Skalenstruktur des WSP mit der modifizierten Skalenstruktur kann der Tabelle 26 auf Seite 66 entnommen werden.

³¹ „Arbeitsstil und Arbeitsverhalten“, „Motivation“, „Belastbarkeit“, „Offenheit“, „Kontaktfähigkeit“, „Kommunikationsfähigkeit“, „Teamfähigkeit“ und „Selbstbild“.

mögliche Unterschiede im Text der Items der Skala „Distanzierungsfähigkeit“ zu jenem der übrigen Skalen, zu einer schlechteren Modellpassung geführt haben.

Um zu überprüfen, ob tatsächlich eine suboptimale Passung der Fragen einzelner Subskalen der Skala „Distanzierungsfähigkeit“ mit dem Rasch-Modell vorlag, wurden grafische Modellkontrollen durchgeführt³². Allerdings wurden nicht auffällig viele weit von der Regressionsgeraden entfernte Items bestimmter Subskalen gefunden. Auch der Erklärungsansatz, dass die Skala „Distanzierungsfähigkeit“ über mehr umgepolte Items als die übrigen Skalen verfügte, ließ sich nicht erhärten. Ebenfalls zeigte der Vergleich der Fragen aller Items auf Textebene, dass die Items der Skala „Distanzierungsfähigkeit“ weder auffällig viele Fragen beinhaltete, bei welchen sich die TeilnehmerInnen im Zuge der Beantwortung Mängel im Studienalltag zugestehen müssten noch andere skalenspezifische Auffälligkeiten der Formulierung. Im Zuge dieser Studie wurden dafür keine befriedigenden Erklärungsansätze für das Zustandekommen des geänderten Antwortverhaltens bei den Items der Skala „Distanzierungsfähigkeit“ gefunden.

Gut überlegte Einstellungen mit hoher Zugänglichkeit sind nach Tourangeau, Rips und Rasinski (2000: S. 220-221) weniger von Kontexteffekten beeinflussbar, auch Schwarz (Schwarz, 1999: S. 100) argumentiert, dass chronisch zugängliche Informationen im Gegensatz zu temporär zugänglichen Aspekten eher kontextunabhängig sind. Der für diese Studie verwendete Fragebogen erfasste studienrelevante Persönlichkeitseigenschaften über das Erfragen von Verhaltensweisen und Einstellungen aus Alltagssituationen von StudentInnen. Die erfragten Inhalte wiesen somit wahrscheinlich eine hohe Zugänglichkeit bei geringer Beeinflussung durch Kontexte auf, da durch den Bezug auf spezifische Situationen aus dem Studienalltag davon ausgegangen werden kann, dass chronisch zugängliche Inhalte erfragt wurden. Da eine höhere Zugänglichkeit von Informationen aus dem Gedächtnis mit einer kürzeren Bearbeitungsdauer einhergehen sollte, könnten zukünftige Studien dies untersuchen, indem zusätzlich die Bearbeitungszeit der einzelnen Items erfasst wird. So könnte etwa angenommen werden, dass jene TeilnehmerInnen mit weniger Studienerfahrung längere Bearbeitungszeiten und ausgeprägtere Kontexteffekte aufweisen.

³² Hierfür wurden für jede der drei Bedingungen sowie jede der vier Teilungskriterien (Mittelwert, Geschlecht, Ausbildung und Alter) grafisch die jeweils äußersten Items des Punktschwarms notiert. Im Anschluss daran wurde festgestellt, welchen Subskalen diese inhaltlich zugerechnet werden können.

Knowles (1988: S. 312-318) zeigte, dass die Reliabilität und Konsistenz der Antworten in einem eindimensionalen Persönlichkeitsfragebogen mit der Anzahl der bereits bearbeiteten Fragen ansteigt. Er führt dieses Ergebnis darauf zurück, dass die beantwortenden Personen mit dem Voranschreiten der bearbeiteten Items mehr Einsicht in das dem Fragebogen zugrundeliegende Persönlichkeitskonstrukt erlangen, da die Thematik von später im Fragebogen auftretenden Items im Gedächtnis leichter zugänglich und durch die bereits behandelten Aspekte auch die eigene Position bezüglich des Gegenstandes klarer werde. In jeder der drei Bedingungen der vorliegenden Studie befand sich der Linking-Teil am Ende der jeweiligen Fragereihenfolge und diente dazu die Versionen im LLTM zu verknüpfen. Fraglich ist, ob dieser von Knowles beschriebene *serial position effect* in dieser Untersuchung zu Tragen kam. Da nach diesem Effekt jene Items am Ende von Fragebögen im Vergleich zu den übrigen Fragen eine höhere Konsistenz aufweisen sollten, wären im Falle dieser Studie die Items des Linkingteils davon betroffen. Die Items des Linkingteils beinhalten Fragen aller Skalen in der gleichen durchmischten Itemreihenfolge und gewährleisteten die Vergleichbarkeit der Items der voneinander unabhängig konstruierten Versionen mit unterschiedlichen Itemreihenfolgen. Falls die Items am Ende jeder der drei Versionen, welche in der gleichen Reihenfolge vorgegeben wurden, vom *serial position effect* betroffen wären, könnte sich dieser vermehrt in der Bedingung mit durchmischter Itemreihenfolge ausgewirkt haben, da die durchmischte Darbietung am Ende des Fragebogens in den beiden geblockten Versionen eine Abwechslung von den bis dahin inhaltshomogenen Frageblöcken darstellte. Folgeuntersuchungen könnten über das Postulieren zusätzlicher Positionsparameter³³ Aufschluss darüber geben, ob der von Knowles beschriebene Effekt beim WSP auftritt.

Schwarz (1999: S. 103) stellt im Hinblick auf die in seinem Artikel zusammengefassten Studien zu Kontexteffekten, welche teilweise von gegensätzlichen Effekten berichten sofern praktisch bedeutsame Unterschiede gefunden wurden, fest, dass all unsere Urteile von ihrem jeweiligen Kontext beeinflusst sind, welche bei Bedarf aus den zu diesem Zeitpunkt zugänglichsten Informationen konstruiert werden:

³³ Siehe hierfür: Hahne, J. (2008). Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly*, 50. 379-390.

„...Attitude measurement is subject to pronounced context effects[....]Human judgement is always context dependent, in research situations as in real life, and attitude judgements are no exception to this rule. We construct these judgements on the spot, when needed (Schwarz & Bless, 1992a; Strack, 1994a, 1994b; Wilson & Hodges, 1992), by drawing on the information that is most accessible at that point in time (Higgins, 1996).“.

Trotzdem im Zuge dieser Studie keine relevanten Effekte des Itemkontextes für die unterschiedlichen Darbietungsmodalitäten gefunden werden konnten, soll an dieser Stelle angemerkt werden, dass nur wenige der gängigen Einstellungsfragebögen auf Itemreihenfolgeeffekte geprüft wurden und in wenigen Manualen überhaupt auf eine mögliche Beeinflussung des Antwortverhaltens durch eine veränderte Itemreihenfolge hingewiesen wird.

Weiterführende Studien könnten – speziell im Hinblick auf eine eventuelle computerisierte adaptive Vorgabe der Items des WSP – untersuchen, ob die Itemschwierigkeiten einzelner Items von den davor beantworteten Fragen beeinflusst werden.

6 Zusammenfassung

Das Ziel der vorliegenden Studie bestand darin festzustellen, ob der *Itemkontext*, welcher durch inhaltlich geblockte oder durchmischte Itemreihenfolgen in drei Bedingungen operationalisiert wurde, die Beantwortung eines Persönlichkeitsfragebogens, dem Wiener Studieneignungs-Persönlichkeitsinventar (WSP), beeinflusst. Dies soll Aufschluss darüber geben, ob die in der Praxis übliche Vorgabe einzelner Skalen in Form von inhaltlichen Blöcken zu Verzerrungen des Antwortverhaltens beim WSP führt. Das untersuchte Verfahren, das WSP, ist ein Selbstbeurteilungsfragebogen, dessen Items studienrelevante Verhaltensweisen und Einstellungen behandeln und so Persönlichkeitseigenschaften erfassen, die für erfolgreiches Studieren erforderlich sind. Die Struktur des WSP³⁴ gliedert sich in die sieben Skalen „Organisation“, „Arbeitsstil und Arbeitsverhalten“, „Motivation“,

³⁴Skalen und Subskalen können der Tabelle 26: *Struktur des WSP vor sowie nach den durchgeführten Martin-Löf-Test*, auf Seite 66 entnommen werden.

„Belastbarkeit“, „Soziale Kompetenz“, „Emotionale Kompetenz“ und „Selbstbild“, die wiederum aus 22 Subskalen bestehen.

Kontexteffekte können nach Tourangeau, Rips und Rasinski (2000: S. 228) auftreten, wenn ein konzeptioneller Zusammenhang zwischen den Themengebieten der gerade zu bearbeitenden Frage und jenen schon beantworteter Items besteht. Demnach sollten Kontexteffekte eher bei einer inhaltlich geblockten Vorgabevariante zu Tragen kommen und zu konsistenteren Einstellungsurteilen führen (siehe hierfür z. B. Steinberg, 1994). Eine durchmischte Itemreihenfolge mit thematisch abwechselnder Itemvorgabe kann Augenscheinvalidität und somit auch Antworttendenzen bzw. Urteilsverzerrungen der Beantwortenden reduzieren, *Primingeffekten* entgegenwirken und Monotonie oder Ermüdung vermindern. Im Gegensatz dazu kann eine inhaltsgeblockte Itemreihung durch die gebotene Transparenz eine Steigerung der Kooperation und Motivation der TeilnehmerInnen erreichen und damit die Eindimensionalität ihrer Antworten fördern.

Das experimentelle Design der Untersuchung sah drei Fragebogenversionen mit verschiedenen Abfolgen der Items und Subskalen vor, denen die TeilnehmerInnen randomisiert zugewiesen wurden. Der Itemkontext wurde als *unabhängige Variable* in den drei Bedingungen manipuliert während das Antwortverhalten der Personen als *abhängige Variable* gemessen wurde. Für die Fragebogenversion A (durchmischte Itemreihenfolge), wurden die Items aller sieben Skalen und 22 Subskalen abwechselnd vorgegeben, wobei ein größtmöglicher Abstand von Fragen der gleichen Skala bzw. Subskala angestrebt wurde. Diese Vorgabevariante entspricht der intendierten Fragereihenfolge der meisten mehrspaltigen Einstellungs- und Persönlichkeitsfragebögen. Die anderen Bedingungen präsentierten die Items den StudentInnen in Form von inhaltshomogenen Blöcken: Die Fragen in Version B wurden skalenweise geblockt vorgegeben, die Items der Subskalen wechselten sich dabei innerhalb eines Skalenblocks mit größtmöglichem Abstand ab. Die Fragen der Fragebogenversion C wurden skalen- und subskalenweise geblockt gereiht. Da zur Beantwortung der Forschungsfrage Analysen mit dem Linear Logistischen Testmodell durchgeführt werden sollten, wurde am Ende aller drei Bedingungen ein *Linkingteil* vorgegeben, welcher es erlaubte versionsübergreifende Parameterschätzungen vorzunehmen.

Die TeilnehmerInnen konnten den Fragebogen über eine Homepage online beantworten³⁵, wobei die Items in der Befragungssituation standardisiert in einer der drei fixen Reihenfolgen vorgegeben wurden. Es wurden insgesamt 465 Fragebögen von StudentInnen österreichischer Fachhochschulen und Wiener Universitäten bearbeitet, 139 davon wurden unvollständig beantwortet, weswegen sie von den statistischen Analysen ausgeschlossen wurden. Von den resultierenden 326 vollständig bearbeiteten Fragebögen beantworteten 99 Personen die Fragebogenversion A mit durchmischter Fragereihenfolge, 122 die Bedingung B mit skalenweise geblockten Items und 105 TeilnehmerInnen die Version C mit nach Skalen und Subskalen geblockten Fragen³⁶.

Für die Analysen mittels dem LLTM wurden a priori die drei Modelle Itemkontextmodell, quasi-saturiertes Modell und Rasch-Modell aufgestellt. Das Rasch-Modell postuliert einen Inhaltsparameter für jedes Item. Von diesem grundlegendem Modell ausgehend ergänzt das Itemkontextmodell die drei Versionsparameter, welche Effekte des Itemkontextes der drei unabhängigen Bedingungen erfassen sollen. Zusätzlich wurde ein quasi-saturiertes Modell aufgestellt, in welchem ein Itemschwierigkeitsparameter für jede Kombination aus Item und Fragebogenversion bestimmt wurde.

Da die erhobenen Daten für Analysen mit dem LLTM Rasch-Modell-Konformität aufweisen müssen, wurde zuerst mittels Martin-Löf-Tests geprüft, ob die einzelnen Subskalen einer Skala entsprechend homogen messen, im Anschluss daran wurden die Items skalenweise mittels Likelihood-Ratio-Tests auf Übereinstimmung mit dem Rasch-Modell getestet. Die Martin-Löf-Tests zeigten, dass die Subskalen der Skalen „Arbeitsstil und Arbeitsverhalten“, „Soziale Kompetenz“ und „Emotionale Kompetenz“ keine Eindimensionalität aufwiesen, weswegen diese aufgeteilt wurden und die Struktur des WSP entsprechend modifiziert wurde. Im Anschluss daran wurden die Items für die drei Versionen A, B und C und die resultierenden Skalen „Organisation“, „Motivation“, „Arbeitsstil und Arbeitsverhalten“, „Belastbarkeit“, „Kontaktfähigkeit“, „Teamfähigkeit“, „Kommunikationsfähigkeit“, „Distanzierungsfähigkeit“, „Offenheit“ und „Selbstbild“ mittels Likelihood-Ratio-Tests nach Andersen auf Rasch-Modell-Konformität geprüft. Wenn die Homogenität der Items einer

³⁵ Die Beantwortung der Items des WSP durch die TeilnehmerInnen erfolgte dabei für alle Fragen mittels einer Selbstbeurteilung in Form von vier verbal ausformulierten und abgestuften Antwortmöglichkeiten: ja, eher ja, eher nein und nein.

³⁶ Ein Post hoc durchgeführter Chi-Quadrat-Test zeigte, dass die Abbruchraten in den drei Bedingungen keine signifikanten Unterschiede aufwiesen.

Skala a priori nicht erfüllt war, wurden die für die Modellpassung ungünstigsten Items mittels Itemfit-Charakteristiken und grafischer Modellkontrollen selektiert und danach ausgeschlossen, bis eine Passung mit dem Rasch-Modell erreicht wurde.

Zuerst wurde mittels Analysen des LLTM die Übereinstimmung der theoretisch postulierten drei Modelle Rasch-Modell, Itemkontextmodell und quasi-saturiertes Modell mit den empirischen Daten ermittelt. Anschließend wurden die Modellpassungen der drei Modelle gegeneinander unter Berücksichtigung der Parameteranzahl mittels Likelihoodquotiententests und mittels eines informationstheoretischen Maßes³⁷ verglichen. Den Likelihood-Ratio-Tests zufolge erzielte das Rasch-Modell (mit der niedrigsten Parameteranzahl) für neun Skalen die beste Passung mit dem erhobenen Datensatz³⁸. Der Vergleich anhand des *Akaike information criteriums* ergab, dass das Rasch-Modell das Zustandekommen der Daten von acht der zehn Skalen³⁹ der modifizierten Struktur des WSP am besten erklärte. Das Rasch-Modell kann somit im Vergleich zum Itemkontextmodell und quasi-saturierten Modell als das geeignetste betrachtet werden. Somit konnte diese Untersuchung in Bezug auf die eingangs gestellte Forschungsfrage zeigen, dass kein relevanter Effekt des Itemkontextes durch eine geblockte oder durchmischte Itemreihenfolge für die mit dem WSP untersuchte Population von StudentInnen Fachhochschulen und Wiener Universitäten vorlag.

Anschließend wurden Erklärungsmöglichkeiten überlegt, warum die modellvergleichenden LRTs nur für eine von zehn Skalen - die Skala „Distanzierungsfähigkeit“ – eine bessere Passung mit dem quasi-saturierten Modell statt des Itemkontextmodells zeigten und besser durch das Itemkontextmodell im Vergleich zum Rasch-Modell erklärt wurden. Die aufgestellten Erklärungsansätze sind: Eine suboptimale Passung der Items der Skala mit dem Rasch-Modell, das Vorhandensein von mehr umgepolten Items als bei anderen Skalen und Unterschiede in der Frageformulierung. Da sich diese allerdings nicht erhärten ließen, konnte die vorliegende Studie keine mögliche Erklärung für das von den anderen Skalen abweichende Antwortverhalten bei der Skala „Distanzierungsfähigkeit“ bieten.

³⁷ Das *Akaike information criterion (AIC)*, siehe dafür Kapitel 3.1.3. *Das Linear Logistische Testmodell von Fischer (LLTM)* auf S. 22f.

³⁸ Das Itemkontextmodell stellte sich im Vergleich zum quasi-saturierten Modell in allen Skalen mit Ausnahme der Skala „Distanzierungsfähigkeit“ als besser zu den empirischen Daten passender heraus.

³⁹ Nach den AIC erklärte das quasi-saturierte Modell das Zustandekommen der Daten am besten für die Skala „Distanzierungsfähigkeit“, das Itemkontextmodell erzielte die beste Passung für die Skala „Organisation“.

Diese Studie konnte keine durch die Itemreihenfolge bedingte Kontexteffekte beim Persönlichkeitsfragebogen WSP nachweisen – somit kann angenommen werden, dass einzelne Skalen des WSPs entnommen und deren Items in unterschiedlichen Fragereihenfolgen präsentiert werden können, ohne dadurch Veränderungen im Antwortverhalten zu provozieren.

Generell wird in Manualen von Einstellungsfragebögen selten darauf hingewiesen, dass eine unterschiedliche Fragereihenfolge bzw. die Vorgabe einzelner Skalen zu Itemreihenfolgeeffekten führen könnte. Auch werden überhaupt nur wenige Persönlichkeits- oder Einstellungsfragebögen auf einen eventuellen Einfluss von Itemreihenfolgeeffekten auf das Antwortverhalten untersucht, was notwendig wäre um die gängige Praxis der separaten Vorgabe von Items oder ganzen Skalen bzw. die Itemvorgabe in einer anderen Fragereihenfolge wissenschaftlich zu legitimieren. Die Frage nach der Auswirkung verschiedener Itemreihenfolgen wurde für den WSP mit dieser Arbeit behandelt, da jedoch z. B. Itemformulierungen eine Rolle dabei spielen könnten, ob Positionseffekte auftreten oder nicht und andere Fragebögen Itemformulierungen und Inhalte haben, kann dieses Ergebnis nicht generalisiert werden. Vor diesem Hintergrund erscheint es wünschenswert, dass vermehrt Anstrengungen unternommen werden, auch diesen Aspekt der Fragebogenkonstruktion zu untersuchen, bevor Items oder Skalen entnommen und Testpersonen in unterschiedlichen Reihenfolgen präsentiert werden. Dies ist insbesondere angesichts der zunehmenden Nutzung computerisierter adaptiver Vorgabevariante dringlich, da die unterschiedlichen Kontexte einzelner Fragen zu verschiedenen Itemschwierigkeiten führen und somit das Gütekriterium der Fairness verletzen können.

7 Literaturverzeichnis

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: University Press.
- Borkenau, P. & Ostendorf, F. (2008). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI)*. Manual. Göttingen: Hogrefe.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer Verlag.
- Dörrer, N. (2009). *Über den Einfluss der inhaltshomogenen Itemblockbildung auf die Dimensionalität eines Persönlichkeitsfragebogens*. Unveröffentlichte Diplomarbeit: Universität Wien.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Verlag Hans Huber.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Grass, J. (2011). *Validierung des Wiener Self-Assessment Psychologie anhand prototypischer Studierender*. Unveröffentlichte Diplomarbeit: Universität Wien.
- Hahne, J. (2008). Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly*, 50, 379-390.
- Irtel, H. (1995). *Entscheidungs- und testtheoretische Grundlagen der Psychologischen Diagnostik*. Universität Mannheim.
- Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). *Self-Assessment: Theorie und Konzepte* (S. 103-118). Pabst Science Publishers.
- Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, 55, 321-320.

- Kubinger, K. D. (2003). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jager (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-422). Weinheim: Beltz Verlag.
- Kubinger, K. D. (2006). *Psychologische Diagnostik - Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe Verlag.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (IST 2000 R). Manual*. Göttingen: Hogrefe Verlag.
- Maurer, M. (2009). *Wiener Studieneignungs-Persönlichkeitsinventar (WSP) – Formulierung eines Regelkataloges zur Itemkonstruktion und dessen Anwendung zur Erstellung eines Itempools*. Unveröffentlichte Diplomarbeit: Universität Wien.
- Mayring, P. (2008). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz.
- Mitschek, B. (2011). *Ermittlung prototypischer Testkennwerte im Wiener Self-Assessment Maschinenbau anhand erfolgreich Studierender*. Unveröffentlichte Diplomarbeit: Universität Wien.
- Müller, C. E. (2011). *Ermittlung prototypischer Testkennwerte für das Wiener Self-Assessment für Architektur anhand erfolgreich und wenig(er) erfolgreich Studierender*. Unveröffentlichte Diplomarbeit: Universität Wien.
- Ortner, T. M. (2004). On changing the position of items in personality questionnaires. Analysing effects of item sequence using IRT. *Psychology Science*, 46. 466-476.
- Ortner, T. M. (2008). Effects of changed item order: A cautionary note to practitioners on jumping to computerized adaptive testing for personality assessment. *International Journal of Selection and Assessment*, 16. 249-257.
- Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Software package eRm (extended Rasch modelling)*. Wien: Pabst Science Publishers.
- Rasch, G. (1980). *Probabilistic Models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: 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. 175-182.

Sparfeldt, J. N., Schilling, S. R., Rost, D. H. & Thiel, A. (2006). Blocked versus Randomized Format of questionnaires - A conformatory multigroup analysis. *Educational and Psychological measurement*, 66. 961-974.

Steinberg, L. (1994). Context and serial-order effects in personality measurement: Limits on the generality of measuring changes the measure. *Journal of Personality and Social Psychology*, 66. 341-349.

Schwarz, N. (1999). Self-Reports – How the questions shape the answers. *American Psychologist*, 54. 93-105.

Tourangeau, R. & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103,S. 299-314.

Tourangeau, R., Rips, L. & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: University Press.

The R foundation for statistical computing (2009). *R for Windows (Version 2.10.0)* [Computer software]. Wien: Institute for Statistics and Mathematics. Als Download unter: <http://www.Rproject.org>.

The LimeSurvey project team (2012). LimeSurvey (Version 2.05) [Computer software]. GNU General Public License. Als Download unter: <http://www.limesurvey.org>.

www.studienberatungstest.at; besucht am 14.08.2012.

8 Tabellenverzeichnis

Tabelle 1: Untersuchungsdesign.....	S. 26
Tabelle 2: Mittels LLTM zu vergleichende, hypothetische Modelle.....	S. 29
Tabelle 3: Die Struktur des Wiener Studieneignungs- Persönlichkeitsinventars (WSP).....	S. 32
Tabelle 4: Anzahl der vollständig bzw. unvollständig bearbeiteten Fragebögen der drei Bedingungen A (durchmischte Itemreihenfolge), B (geblockte Itemreihenfolge randomisiert innerhalb) und C (mit vollständig geblockter Itemreihenfolge).....	S. 42
Tabelle 5: Häufigkeiten von berufstätigen bzw. nicht berufstätigen StudentInnen von Fachhochschulen und Universitäten in der Stichprobe.....	S. 42
Tabelle 6: Häufigkeiten und Prozentangaben jener Personen welche in den drei Fragebogenversionen die Frage nach der derzeitigen Ausbildung mit den entsprechenden Kategorien beantworteten.....	S. 43
Tabelle 7: Häufigkeiten und Deskriptivstatistiken zur Verteilung des Geschlechts und des Alters in den drei Bedingungen A, B und C.....	S. 44
Tabelle 8: Studienrichtungen jener 148 von 160 StudentInnen österreichischer Fachhochschulen, welche ihre Studienrichtungen angegeben haben.....	S. 45
Tabelle 9: Studienrichtungen jener 147 von 150 UniversitätsstudentInnen, welche ihre Studienrichtungen angegeben haben.....	S. 46
Tabelle 10: Demografische Fragen und deren Antwortmöglichkeiten.....	S. 49
Tabelle 11: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die fünf Subskalen der Skala „Organisation“.....	S. 53
Tabelle 12: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, für die fünf Subskalen des WSP, welche der Skala „Organisation“ zugeordnet sind.....	S. 54
Tabelle 13: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die fünf Subskalen der Skala „Organisation“.....	S. 55

Tabelle 14: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die fünf Subskalen der Skala „Arbeitsstil und Arbeitsverhalten“.....S. 56

Tabelle 15: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und randomisierter subskalenweiser Itemzuweisung innerhalb, für die vier Subskalen des WSP, welche der Skala „Arbeitsstil und Arbeitsverhalten“ zugeordnet sind.....S. 57

Tabelle 16: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die vier Subskalen der Skala „Arbeitsstil und Arbeitsverhalten“.....S. 58

Tabelle 17: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die sechs Subskalen der Skala „Motivation“.....S. 59

Tabelle 18: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und randomisierter subskalenweiser Itemzuweisung innerhalb, für die sechs Subskalen des WSP, welche der Skala „Motivation“ zugeordnet sind.....S. 60

Tabelle 19: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die sechs Subskalen, welche der Skala „Motivation“ zugeordnet sind.....S. 61

Tabelle 20: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die vier Subskalen der Skala „Emotionale Kompetenz“.....S. 61

Tabelle 21: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und subskalenweiser randomisierter Itemzuweisung innerhalb, für die beiden Subskalen des WSP, welche der Skala „Emotionale Kompetenz“ zugeordnet sind.....S. 62

Tabelle 22: Ergebnisse des Martin Löff-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die zwei Subskalen, welche der Skala „Emotionale Kompetenz“ zugeordnet sind.....S. 62

Tabelle 23: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version A mit durchmischter Itemreihenfolge für die drei Subskalen der Skala „Soziale Kompetenz“	S. 63
Tabelle 24: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version B mit skalenweise geblockter Itemreihenfolge und subskalenweise randomisierter Itemzuweisung innerhalb, für die drei Subskalen des WSP, welche der Skala „Soziale Kompetenz“ zugeordnet sind.....	S. 63
Tabelle 25: Ergebnisse des Martin Löf-Tests der Teilstichprobe der Version C mit vollständig geblockter Itemreihenfolge für die drei Subskalen der Skala „Soziale Kompetenz“	S. 64
Tabelle 26: Struktur des WSP vor sowie nach den durchgeführten Martin-Löf-Tests.....	S. 66
Tabelle 27: Ergebnisse der Likelihood-Ratio-Tests der Skala „Organisation“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 67
Tabelle 28: Ergebnisse der Likelihood-Ratio-Tests der Skala „Organisation“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 68
Tabelle 29: Ergebnisse der Likelihood-Ratio-Tests der Skala „Organisation“ der Version C, mit vollständig geblockter Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 68
Tabelle 30: Ergebnisse der Likelihood-Ratio-Tests der Skala „Arbeitsstil und Arbeitsverhalten“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 69
Tabelle 31: Ergebnisse der Likelihood-Ratio-Tests der Skala „Arbeitsstil und Arbeitsverhalten“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 69
Tabelle 32: Ergebnisse der Likelihood-Ratio-Tests der Skala „Arbeitsstil und Arbeitsverhalten“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 70
Tabelle 33: Ergebnisse der Likelihood-Ratio-Tests der Skala „Belastbarkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 71

Tabelle 34: Ergebnisse der Likelihood-Ratio-Tests der Skala „Belastbarkeit“ der Version B, mit geblockter Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 71
Tabelle 35: Ergebnisse der Likelihood-Ratio-Tests der Skala „Belastbarkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität	S. 72
Tabelle 36: Ergebnisse der Likelihood-Ratio-Tests der Skala „Distanzierungsfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 72
Tabelle 37: Ergebnisse der Likelihood-Ratio-Tests der Skala „Distanzierungsfähigkeit“ der Version B, skalenweise geblockte Itemreihenfolge mit subskalenweise randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 73
Tabelle 38: Ergebnisse der Likelihood-Ratio-Tests der Skala „Distanzierungsfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 73
Tabelle 39: Ergebnisse der Likelihood-Ratio-Tests der Skala „Offenheit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 74
Tabelle 40: Ergebnisse der Likelihood-Ratio-Tests der Skala „Offenheit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 74
Tabelle 41: Ergebnisse der Likelihood-Ratio-Tests der Skala „Offenheit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 75
Tabelle 42: Ergebnisse der Likelihood-Ratio-Tests der Skala „Motivation“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 75
Tabelle 43: Ergebnisse der Likelihood-Ratio-Tests der Skala „Motivation“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 76
Tabelle 44: Ergebnisse der Likelihood-Ratio-Tests der Skala „Motivation“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 76

Tabelle 45: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kontaktfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 77
Tabelle 46: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kontaktfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 77
Tabelle 47: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kontaktfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 78
Tabelle 48: Ergebnisse der Likelihood-Ratio-Tests der Skala „Teamfähigkeit“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 78
Tabelle 49: Ergebnisse der Likelihood-Ratio-Tests der Skala „Teamfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 79
Tabelle 50: Ergebnisse der Likelihood-Ratio-Tests der Skala „Teamfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 79
Tabelle 51: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kommunikationsfähigkeit“ der Version A, durchmischte Itemreihenfolge zur Prüfung der Rasch-Modell-Konformität.....	S. 80
Tabelle 52: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kommunikationsfähigkeit“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 80
Tabelle 53: Ergebnisse der Likelihood-Ratio-Tests der Skala „Kommunikationsfähigkeit“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität	S. 81
Tabelle 54: Ergebnisse der Likelihood-Ratio-Tests der Skala „Selbstbild“ der Version A, durchmischte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 81
Tabelle 55: Ergebnisse der Likelihood-Ratio-Tests der Skala „Selbstbild“ der Version B, geblockte Itemreihenfolge mit randomisierter Itemzuweisung innerhalb, zur Prüfung der Rasch-Modell-Konformität.....	S. 82

Tabelle 56: Ergebnisse der Likelihood-Ratio-Tests der Skala „Selbstbild“ der Version C, vollständig geblockte Itemreihenfolge, zur Prüfung der Rasch-Modell-Konformität.....	S. 82
Tabelle 57: Mittels LLTM zu vergleichende, hypothetische Modelle.....	S. 83
Tabelle 58: Ergebnisse der Modellgleichungstests des Rasch Modells für die zehn Skalen des WSP.....	S. 85
Tabelle 59: Ergebnisse der LLTM Analysen für das quasi-saturierte Modell anhand jeder der zehn Skalen des WSP.....	S. 85
Tabelle 60: Ergebnisse der LLTM Analysen für das Modell Itemkontext anhand jeder der zehn Skalen des WSP.....	S. 86
Tabelle 61: Ergebnisse der modellvergleichenden LRTs für das quasi-saturierte Modell und das Itemkontextmodell für die zehn Skalen des WSP.....	S. 87
Tabelle 62: Ergebnisse der modellvergleichenden LRTs für das Itemkontextmodell mit dem Rasch-Modell für die zehn Skalen des WSP.....	S. 88
Tabelle 63: Gegenüberstellung der Likelihoodquotienten sowie der AICs für die drei Modelle quasi-saturiertes Modell, Itemkontextmodell und Rasch-Modell für die Skalen des WSP.....	S. 90
Tabelle 64: Itemreihenfolgen der drei Fragebogenversionen A, B und C mit den jeweils ersten 54 Items.....	S. 110

9 Abbildungsverzeichnis

Abb. 1: Exemplarische Darstellung der Q-Matrix der Skala „Belastbarkeit“ des Itemkontextmodells.....	S. 31
Abb. 2: Vierkategorielles Antwortformat des WSP.....	S. 39
Abb. 3: Beispielitem des Wiener Studieneignungs-Persönlichkeitsinventars mit den vier Antwortmöglichkeiten.....	S. 50

10 Anhang

10.1 Itemreihenfolgen der drei Fragebogenversionen

Die folgende Tabelle enthält eine Darstellung der Itemreihenfolge der ersten 54 Items der 124 Items des Fragebogenteils für jede der drei Fragebogenversionen. Im Anschluss daran wurde bei jeder der Versionen gleichermaßen ein Linkingteil mit denselben 29 Items in durchmischter Itemreihenfolge vorgegeben.

Tabelle 64: Itemreihenfolgen der drei Fragebogenversionen A, B und C mit den jeweils ersten 54 Items.

	A: durchmischte Itemreihenfolge		B: Itemreihenfolge geblockt – randomisiert - innerhalb		C: geblockt-vollständig	
Item	Subskalen	Skala	Subskalen	Skala	Subskalen	Skala
1	Informiertheit	Organisation	Informiertheit	Organisation	Informiertheit	Organisation
2	Kommunikationsfähigkeit	Soziale Kompetenz	Prioritätensetzung	Organisation	Informiertheit	Organisation
3	Erfolgszuversicht	Motivation	Zeitmanagement	Organisation	Informiertheit	Organisation
4	Gewissenhaftigkeit	Arbeitsstil & -verhalten	Organisieren	Organisation	Informiertheit	Organisation
5	Distanzierungsfähigkeit	Emotionale Kompetenz	Planen	Organisation	Prioritätensetzung	Organisation
6	Frustrationstoleranz	Belastbarkeit	Informiertheit	Organisation	Prioritätensetzung	Organisation
7	Engagement	Motivation	Prioritätensetzung	Organisation	Prioritätensetzung	Organisation
8	Prioritätensetzung	Organisation	Zeitmanagement	Organisation	Prioritätensetzung	Organisation
9	Selbständiger Arbeitsstil	Arbeitsstil & -verhalten	Organisieren	Organisation	Prioritätensetzung	Organisation
10	Kontaktfähigkeit	Soziale Kompetenz	Planen	Organisation	Zeitmanagement	Organisation
11	Zielsetzung	Motivation	Informiertheit	Organisation	Zeitmanagement	Organisation
12	Zeitmanagement	Organisation	Prioritätensetzung	Organisation	Zeitmanagement	Organisation
13	Genauer Arbeitsstil	Arbeitsstil & -verhalten	Zeitmanagement	Organisation	Zeitmanagement	Organisation
14	Selbstbewusstsein	Selbstbild	Organisieren	Organisation	Zeitmanagement	Organisation
15	Durchhaltevermögen	Motivation	Planen	Organisation	Organisieren	Organisation
16	Offenheit	Emotionale Kompetenz	Informiertheit	Organisation	Organisieren	Organisation
17	Organisieren	Organisation	Prioritätensetzung	Organisation	Organisieren	Organisation
18	Teamfähigkeit (Kollegialität)	Soziale Kompetenz	Zeitmanagement	Organisation	Organisieren	Organisation
19	Eigeninitiative	Arbeitsstil & -verhalten	Organisieren	Organisation	Organisieren	Organisation
20	Lernbereitschaft	Motivation	Planen	Organisation	Planen	Organisation
21	Planen	Organisation	Prioritätensetzung	Organisation	Planen	Organisation
22	Distanzierungsfähigkeit	Emotionale Kompetenz	Zeitmanagement	Organisation	Planen	Organisation
23	Frustrationstoleranz	Belastbarkeit	Organisieren	Organisation	Planen	Organisation
24	Kommunikationsfähigkeit	Soziale Kompetenz	Planen	Organisation	Planen	Organisation
25	Informiertheit	Organisation	Kommunikationsfähigkeit	Soziale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz
26	Gewissenhaftigkeit	Arbeitsstil & -verhalten	Kontaktfähigkeit	Soziale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz
27	Zielorientierung	Motivation	Teamfähigkeit (Kollegialität)	Soziale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz
28	Prioritätensetzung	Organisation	Kommunikationsfähigkeit	Soziale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz
29	Selbständiger Arbeitsstil	Arbeitsstil & -verhalten	Kontaktfähigkeit	Soziale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz
30	Erfolgszuversicht	Motivation	Teamfähigkeit (Kompromissbereitschaft)	Soziale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz
31	Zeitmanagement	Organisation	Kommunikationsfähigkeit	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz
32	Kontaktfähigkeit	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz
33	Engagement	Motivation	Teamfähigkeit (Teamorientierung)	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz
34	Offenheit	Emotionale Kompetenz	Kommunikationsfähigkeit	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz
35	Genauer Arbeitsstil	Arbeitsstil & -verhalten	Kontaktfähigkeit	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz
36	Zielsetzung	Motivation	Teamfähigkeit (Kollegialität)	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz
37	Organisieren	Organisation	Kommunikationsfähigkeit	Soziale Kompetenz	Teamfähigkeit (Kollegialität)	Soziale Kompetenz
38	Eigeninitiative	Arbeitsstil & -verhalten	Kontaktfähigkeit	Soziale Kompetenz	Teamfähigkeit (Kollegialität)	Soziale Kompetenz
39	Frustrationstoleranz	Belastbarkeit	Teamfähigkeit (Kompromissbereitschaft)	Soziale Kompetenz	Teamfähigkeit (Kompromissbereitschaft)	Soziale Kompetenz
40	Selbstbewusstsein	Selbstbild	Kommunikationsfähigkeit	Soziale Kompetenz	Teamfähigkeit (Kompromissbereitschaft)	Soziale Kompetenz
41	Teamfähigkeit (Kompromissbereitschaft)	Soziale Kompetenz	Kontaktfähigkeit	Soziale Kompetenz	Teamfähigkeit (Teamorientierung)	Soziale Kompetenz
42	Durchhaltevermögen	Motivation	Teamfähigkeit (Teamorientierung)	Soziale Kompetenz	Teamfähigkeit (Teamorientierung)	Soziale Kompetenz
43	Planen	Organisation	Erfolgszuversicht	Motivation	Erfolgszuversicht	Motivation
44	Gewissenhaftigkeit	Arbeitsstil & -verhalten	Engagement	Motivation	Erfolgszuversicht	Motivation
45	Lernbereitschaft	Motivation	Zielsetzung	Motivation	Erfolgszuversicht	Motivation
46	Distanzierungsfähigkeit	Emotionale Kompetenz	Durchhaltevermögen	Motivation	Erfolgszuversicht	Motivation
47	Selbstbewusstsein	Selbstbild	Lernbereitschaft	Motivation	Erfolgszuversicht	Motivation
48	Zielorientierung	Motivation	Zielorientierung	Motivation	Engagement	Motivation
49	Kommunikationsfähigkeit	Soziale Kompetenz	Erfolgszuversicht	Motivation	Engagement	Motivation
50	Informiertheit	Organisation	Engagement	Motivation	Engagement	Motivation
51	Erfolgszuversicht	Motivation	Zielsetzung	Motivation	Engagement	Motivation
52	Frustrationstoleranz	Belastbarkeit	Durchhaltevermögen	Motivation	Engagement	Motivation
53	Selbständiger Arbeitsstil	Arbeitsstil & -verhalten	Lernbereitschaft	Motivation	Zielsetzung	Motivation
54	Prioritätensetzung	Organisation	Zielorientierung	Motivation	Zielsetzung	Motivation

10.2 Der Text der Homepage www.hochschultest.at

Willkommen!

Auf dieser Seite hast du die Möglichkeit deine studienrelevanten Persönlichkeitseigenschaften kostenlos zu testen und so deine individuellen Stärken und Schwächen für das Studium zu erfahren. Dazu musst du lediglich einige Fragen beantworten, anschließend erhältst du sofort deine persönlichen Ergebnisse. Der Fragebogen, genauer das Wiener Studieneignungs-Persönlichkeitsinventar, ist anonym und deine Daten werden selbstverständlich nicht weitergegeben.

Das Wiener Studieneignungs-Persönlichkeitsinventar (WSP) wurde am [Arbeitsbereich Psychologische Diagnostik](#) an der Fakultät für Psychologie (Universität Wien) entwickelt und dient der Erfassung studienrelevanter Persönlichkeitseigenschaften. Als solche konnten organisatorische Fähigkeiten, Arbeitsstil, Belastbarkeit, (Leistungs-) Motivation, soziale und emotionale Kompetenzen sowie das eigene Selbstbild identifiziert werden.

Der zur Erhebung dieser Eigenschaften konstruierte Fragebogen ermöglicht es somit, dir persönliche Stärken und Schwächen im Bezug auf das Studium aufzuzeigen. Die Beantwortungszeit beträgt in den meisten Fällen zwischen 10 und 15 Minuten. Du erhältst deine individuellen Ergebnisse sobald du den Fragebogen vollständig ausgefüllt hast. Mein Name ist Julia Kober und ich schreibe meine Diplomarbeit am Arbeitsbereich Psychologische Diagnostik der Fakultät für Psychologie (Universität Wien) zum Antwortverhalten beim WSP. Deine Daten werden selbstverständlich vertraulich und anonym behandelt und ausschließlich für statistische Analysen im Zuge meiner Diplomarbeit verwendet.

→ [zum Fragebogen](#)

→ [Infos zum WSP](#)

→ [Skalen des WSP](#)

→ Noch Fragen oder Anregungen

Information zum WSP

Das Wiener Studieneignungs-Persönlichkeitsinventar wurde von Khorramdel und Maurer (2012) zur Erfassung von Persönlichkeitseigenschaften entwickelt, die ein erfolgreiches Studieren wahrscheinlich machen. Dabei wurde in einem ersten Schritt unter Berücksichtigung von Lehrenden und Studierenden ein studienübergreifendes Anforderungsprofil über verschiedene Studienrichtungen an Wiener Universitäten erstellt. Anschließend wurden Fragestellungen entworfen, die typische Eigenschaften und Verhaltensweisen im Studienalltag umfassen und so eine Selbsteinschätzung in Bezug auf die identifizierten Anforderungen ermöglichen.

Wer sich intensiver mit dem Fragebogen auseinandersetzen möchte, findet in diesem Beitrag detaillierte Informationen:

Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). [*Self-Assessment: Theorie und Konzepte*](#) (S.103-118). Pabst Science Publishers.

→ Hier kommst du zum [Fragebogen](#)

→ Hier erfährst du mehr über die unterschiedlichen [Skalen des WSP](#).

→ Hier erfährst du mehr über mich und meine [Diplomarbeit](#).

→ Hier erfährst du mehr zum Thema Studieneignung und Beratung:

- [Arbeitsbereich Psychologische Diagnostik](#) der Fakultät für Psychologie (Universität Wien)
- [Psychologische Studentenberatung Österreich](#)

Skalen des WSP

Das Wiener Studieneignungs-Persönlichkeitsinventar erhebt sieben unterschiedliche Facetten der Persönlichkeit, die auf Basis empirischer Forschung als für den Studienerfolg relevant identifiziert wurden. Er erfasst jedoch keine kognitiven Eigenschaften wie z. B. Konzentrationsfähigkeit oder analytische Fähigkeiten, die ebenfalls wichtig für den Studienerfolg sein können. Mit dem Fragebogen können die folgenden Persönlichkeitsdimensionen beschrieben werden:

- **Organisation**
- **Arbeitsstil und Arbeitsverhalten**
- **Motivation**
- **Belastbarkeit**
- **Soziale Kompetenz**
- **Emotionale Kompetenz**
- **Selbstbild bzw. Selbstbewusstsein**

Eine genauere Beschreibung der sieben Skalen ist ein Bestandteil der Ergebnissrückmeldung, die du erhältst, wenn der Fragebogen fertig ausgefüllt wurde.

→ Hier kommst du zum [Fragebogen](#)

Kontakt

Mein Name ist Julia Kober, ich bin Diplomandin am [Arbeitsbereich Psychologische Diagnostik](#) der Fakultät für Psychologie an der Universität Wien.

Diese Webseite wurde im Rahmen meiner Diplomarbeit erstellt, die erhobenen Ergebnisse des Fragebogens stellen die Grundlage meiner Diplomarbeit dar.

Für etwaige Fragen und Anregungen bin ich jederzeit unter folgender E-Mail-Adresse erreichbar:
a0203009@unet.univie.ac.at

→ Hier kommst du zum [Fragebogen](#)

10.3 Instruktion des webbasierten WSP

Es folgen nun 153 Aussagen über typische Situationen und Verhaltensweisen im Studienalltag. Damit werden Eigenschaften und Verhaltensweisen erfasst, die für die erfolgreiche Bewältigung des Studiums besonders relevant sind.

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

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.

Einige der Aussagen sind sich in der Formulierung ähnlich, bitte lassen Sie sich nicht davon irritieren. Jede der 153 Aussagen unterscheidet sich inhaltlich von den anderen.

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

Bedenken Sie, dass es keine richtigen oder falschen Antworten gibt. Antworten Sie so, wie es für Sie persönlich am ehesten zutrifft.

Wählen Sie „Weiter“, um mit der Befragung zu beginnen.

10.4 Rückmeldung der Ergebnisse an die TeilnehmerInnen

Organisation

Personen mit hohen Werten beschreiben sich so, dass sie bei der Herangehensweise und Bearbeitung von Aufgaben und Tätigkeiten ein günstiges Zeitmanagement aufweisen (indem sie Zeitpläne erstellen, die Lernzeit einteilen, den zeitlichen Arbeitsaufwand gut einschätzen können), Prioritäten setzen können (indem sie die Aufgaben nach ihrer Wichtigkeit einteilen können und mit der wichtigsten beginnen), sich die relevanten Informationen besorgen (indem sie sich über das Studienangebot und bestimmte Termine informieren), Arbeitsschritte planen (indem sie z.B. recherchieren, welche Arbeitsschritte nötig und wichtig sind) und diese organisieren können (indem sie mehrere Aufgaben oder Arbeitsschritte koordinieren und strukturiert vorgehen). Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Arbeitsstil und Arbeitsverhalten

Personen mit hohen Werten beschreiben für sich einen gewissenhaften und genauen Arbeitsstil (indem sie sich an Geplantes halten, sich z.B. ausreichend auf Prüfungen vorbereiten und bei der Arbeit auf Vollständigkeit und Exaktheit Wert legen). Sie beschreiben sich als jemanden, der Eigeninitiative zeigt und selbständig arbeitet (indem sie z.B. eigenständig Lösungsvorschläge und Ideen erarbeiten und versuchen, diese selbständig ohne Hilfe oder Aufforderung von außen umzusetzen). Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Motivation

Personen mit hohen Werten beschreiben sich so, dass sie sich regelmäßig Ziele setzen (z.B. in der Ausbildung), die sie nicht aus den Augen verlieren und von denen sie sich nicht ablenken lassen. Sie geben an, hohes Engagement zu zeigen (indem sie sich z.B. über das vom Studium geforderte hinaus mit Lerninhalten oder fachlichen Inhalten beschäftigen oder mehr Praxiserfahrung sammeln, als vom Studium vorgeschrieben wird) und Zuversicht zu besitzen, Aufgaben erfolgreich bewältigen zu können (indem sie sich von anfänglichen Schwierigkeiten nicht abhalten lassen und zuversichtlich sind, das Studium erfolgreich abzuschließen; neue Aufgaben sehen sie als Herausforderung an). Weiters beschreiben sie für sich die Bereitschaft, Neues zu (er)lernen (indem sie Verbesserungsvorschläge annehmen können und Kritik als Verbesserungspotenzial sehen) sowie Aufgaben und Ziele mit dem entsprechenden Durchhaltevermögen zu verfolgen (indem sie selbst bei großem Aufwand und Anstrengung oder weniger interessanten Inhalten nicht aufgeben). Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Belastbarkeit

Personen mit hohen Werten beschreiben für sich eine hohe Frustrationstoleranz, da sie sich von erlebten Misserfolgen nicht verunsichern oder in ihrer Leistungsbereitschaft beeinträchtigen bzw. demotivieren lassen (z.B. von schlechten Noten oder Kritik auch für Arbeiten, in die viel Zeit investiert wurde). Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Soziale Kompetenz

Personen mit hohen Werten beschreiben für sich die Bereitschaft, mit anderen zu kommunizieren und die Regeln einer angemessenen Kommunikation zu beherrschen (indem sie einen freundlichen und respektvollen Umgang mit anderen pflegen ohne jemanden zu kränken, sowie andere ausreden lassen). Sie beschreiben sich des Weiteren so, dass sie keine Schwierigkeiten damit haben, neue Kontakte zu knüpfen und auf fremde Personen zugehen zu können. Sie geben außerdem an, dass es ihnen leicht fällt, mit anderen in einem Team zu arbeiten, und dass sie sich dabei als kollegial, hilfsbereit und kompromissbereit zeigen. Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Emotionale Kompetenz

Personen mit hohen Werten beschreiben sich so, dass sie neuen Sichtweisen, Erfahrungen und Inhalten offen gegenüber zu stehen und von der eigenen Arbeit distanzieren zu können (indem sie z.B. eigene Emotionen kontrollieren und ein Thema sachlich und objektiv diskutieren können sowie Probleme aus der Arbeit nicht ins Privatleben mitnehmen). Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Selbstbild bzw. Selbstbewusstsein

Personen mit hohen Werten beschreiben für sich ein hohes Selbstbewusstsein und ein sicheres Auftreten vor anderen Personen (da sie zu ihren eigenen Ideen und Meinungen stehen können, sich nicht durch andere verunsichern lassen). Aus: Khorramdel, L., & Maurer, M. (2012). Das Wiener Studieneignungs-Persönlichkeitsinventar. In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder (Hrsg.). Self-Assessment: Theorie und Konzepte (S.103-118). Pabst Science Publishers.)

Du hast von **10** möglichen Punkten **X** Punkte erreicht.

Vielen Dank für deine Mitarbeit!

10.5 Liste der teilnehmenden Fachhochschulen Österreichs

Name	Adresse
FH Campus Wien	1100 Wien, Favoritenstraße 226
FH des BFI Wien	1020 Wien, Wohlmutterstraße 22
FH Technikum Wien	1200 Wien, Höchstädtplatz 5
FH Wien-Studiengänge der WKW	1200 Wien, Währinger Gürtel 97
FH Wiener Neustadt	2700 Wiener Neustadt, Johannes Gutenberg-Straße 3
FH Wiener Neustadt - Standort Wieselburg	3250 Wieselburg, Zeiselgraben 4
Theresianische Militärakademie FH - Studiengang Militärische Führung	2700 Wiener Neustadt, Burgplatz 1
IMC Fachhochschule Krems	3500 Krems, Piaristengasse 1
FH St. Pölten	3100 St. Pölten, Matthias Corvinus-Straße 15
FH Joanneum - Standort Graz	8010 Graz, Körblergasse 126
FH Joanneum - Standort Kapfenberg	8605 Kapfenberg, Werk-VI-Straße 46
FH Joanneum – Standort Bad Gleichenberg	8344 Bad Gleichenberg, Kaiser-Franz-Josef-Straße 24
FH Oberösterreich - Fakultät für Informatik/Kommunikation/Medien	4232 Hagenberg, Softwarepark 11

FH Oberösterreich - Fakultät für Gesundheit/Soziales	4020 Linz, Garnisonstraße 21
FH Oberösterreich - Fakultät für Management	4400 Steyr, Wehrgrabengasse 1–3
FH Oberösterreich - Fakultät für Technik/Umweltwissenschaften	4600 Wels, Stelzhamerstraße 23
FH Kärnten - Standort Spittal	9800 Spittal a. d. Drau, Villacher Straße 1
FH Kärnten - Standort Villach	9524 Villach, Europastraße 4
FH Kärnten - Standort Feldkirchen	9560 Feldkirchen, Hauptplatz 12
FH Kärnten - Standort Klagenfurt	9020 Klagenfurt am Wörthersee, Primoschgasse 10
FH Salzburg GmbH - Campus Urstein	5412 Puch/Salzburg, Urstein Süd 1
FH Salzburg GmbH – Standort Kuchl	5431 Kuchl, Markt 136a
FH Kufstein Tirol	6330 Kufstein, Andreas Hofer-Straße 7
FHG Zentrum für Gesundheitsberufe Tirol	6020 Innsbruck, Innrain 98
FH Vorarlberg	6911 Lochau, Hoferstrasse 26
FH Studiengänge Burgenland- Standort Eisenstadt	7000 Eisenstadt, Campus 1
FH Studiengänge Burgenland - Standort Pinkafeld	7423 Pinkafeld, Steinamangerstraße 21

10.6 Liste der studentischen universitätsnahen Foren, auf denen die Teilnahme an der Studie beworben wurde

Forum	Studienrichtung
http://www.univie.ac.at/stv-biologie/forum/	Biologie
http://www.univie.ac.at/strv-ew/bbsstart.html	Ernährungswissenschaft
http://forum.pharmapoint.at/forums/default.aspx	Pharmazie
http://www.bagru.info/forum/profile.php	Soziologie
www.powi.at	Politikwissenschaft
http://www.univie.ac.at/paedforum/	Pädagogik
http://forum.narrenschiff.org	Germanistik
http://www.univie.ac.at/strv-publizistik/forum/	Publizistik
http://www.univie.ac.at/stv-ksa/forum	Kultur – und Sozialanthropologie
http://www.lbtforum.at/forum	Lebensmittel – und Biotechnologie

11 Abstract

Die Studie untersucht Itemreihenfolgeeffekte beim Wiener Studieneignungs-Persönlichkeitsinventar (WSP; Khorramdel & Maurer, 2012), einem Selbstbeurteilungsfragebogen mit sieben Skalen zur Erfassung studienrelevanter Einstellungen und Verhaltensweisen. Dazu wurde die Reihenfolge der Items in drei Versuchsbedingungen experimentell variiert: eine randomisierte Itemreihenfolge und zwei unterschiedlich geblockte Itemreihenfolgen (skalenweise geblockt mit Randomisierung der Subskalen, sowie nach Skalen und Subskalen geblockt) wurden miteinander verglichen. Die drei Fragebogenversionen wurden StudentInnen österreichischer Fachhochschulen und Wiener Universitäten online vorgegeben. Analysen mittels dem Linear Logistischen Testmodell (LLTM) zeigen, dass sich das Antwortverhalten der StudentInnen im WSP in keiner der drei Itemreihenfolgen signifikant voneinander unterschied.

This study examines the effect of the sequence of items within a personality questionnaire – the Viennese Study-Qualification Personality Inventory (WSP; Khorramdel & Maurer, 2012), a personality questionnaire assessing academic and study relevant attitudes and behaviors with seven scales. The sequence of the items was varied within an experiment resulting in three test conditions: a randomized item sequence and two different blocked item sequences (blocked on the scale level but randomized on the subscale level, blocked on the scale and subscale level) were compared to each other. The three versions of the questionnaire were administered online to students of colleges and universities throughout Austria. The data were analysed using the Linear Logistic Test Modell (LLTM). Results show that there are no significant differences between the response patterns in the three questionnaire versions.

12 Curriculum Vitae

Persönliche Daten

Name: Julia Kober

Geboren am 16.02.1984 in Wien

Staatsbürgerschaft: Österreich

Bildungsweg

2003/04 Studium Anglistik/Amerikanistik an der Universität Wien

2002 Beginn des Studiums der Psychologie an der Universität Wien

2002 Matura am Realgymnasium St. Ursula (Wien)

Beruflicher Werdegang

Seit 2013 *Betriebsrätliche Tätigkeit Komit GmbH.*

Seit 2011 *Trainerin in einer beschäftigungstherapeutischen Einrichtung (Pappel der Komit GmbH, Teilzeit)* Sozialtherapeutische Arbeit mit Menschen mit psychischen Erkrankungen; Pädagogisch-therapeutische Aufgaben zur Kompetenzsteigerung, Zielerreichung, Konfliktbewältigung und Krisenintervention.

2010-11 *Ersatzdienst (Teilzeit) Komit GmbH* Vertretung von TrainerInnen bei Urlauben, Krankenständen etc.

2010-12 *Testsupervisorin bei psychometrischen Testungen für MMag. Daniel Weigl (zeitweise)* .Durchführung und Auswertung von psychometrischen Testungen für Bewerberselektion und Mitarbeiterentwicklung; Testvorgabe mittels Wiener Testsystem (Schuhfried); Ansprechpartnerin für Probleme oder Fragen.

2010 *TQS (Team für Qualitätsentwicklung und Service Management) Mitarbeiterin bei einer Kundenzufriedenheitserhebung (einmalige Projektmitarbeit).* Strukturierte Interviews zur Erhebung der Kundenzufriedenheit.

2008-10 *Allianz- Versicherung Agentur Schnabl, Lulic und Partner GmbH Sekretärin (Teilzeit).* Kundenumgang (Frontoffice, telefonisch, E-Mail etc.); Administrative Tätigkeiten; Korrespondenz und Kalkulationen mit MS Office.

2006-08 *Allianz- Versicherung Landesdirektion Wien Mitarbeiterin im Sekretariat (Teilzeit).* Frontoffice (telefonisch, E-Mail etc.); Administrative Tätigkeiten; Verwaltung von Personalakten; Vorbereitungen von Tagungen, Leitersitzungen, internen Fortbildungen sowie Workshops und Veranstaltungen; Postmanagement; Erstellen von Präsentationen, Korrespondenz und Kalkulationen mit MS Office.

2005 *Institut für empirische Sozialforschung (Ifes GmbH) Interviewerin (Teilzeit).* Telefoninterviews zu politischen Umfragen sowie Umfragen der Markt- und Meinungsforschung.

2003-04 *Copyshop Thaliastraße Mitarbeiterin (zeitweise).* Mitwirken bei Messen, Promotion und im Verkauf.

2002-05 *Diverse Unternehmen* Mitarbeiterin für Promotion und auf Messen (zeitweise).

2002-04 *Englischnachhilfe (zeitweise).* Schularbeitsvorbereitung; Erarbeitung von relevanten Wissensgebieten.