



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Ein multidimensionaler Item Response Theorie
(MIRT) Ansatz zur Untersuchung von
Antworttendenzen im Differentiellen Stressinventar
HR (DSIHR) in Abhängigkeit unterschiedlicher
Fragebogeninhalte

verfasst von

Alina Bugelnig

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl:	A 298
Studienrichtung:	Psychologie
Betreuerin:	Mag. Dr. Lale Khorramdel

Danksagung

Ich möchte mich an dieser Stelle ganz herzlich bei allen Personen bedanken, die mich bei der Erstellung dieser Diplomarbeit unterstützt haben.

Allen voran möchte ich mich bei Mag. Dr. Lale Khorramdel für ihre Geduld, die großartige Betreuung und die vielen Fertigkeiten, die ich von ihr lernen durfte, bedanken.

Vielen Dank an Claudia Krutis, die mit mir gemeinsam Daten erhoben und während der Erstellung der Diplomarbeit so manches motivationales Gespräch mit mir geführt hat.

Danke auch an meine Eltern Stefan und Gislinde Bugelnig, die mich während meiner gesamten Studienzeit unterstützt haben und mir in meinen Entscheidungen zur Seite standen.

Vielen Dank an Bernhard Slanic, der mir motivational immer wieder auf die Sprünge geholfen hat und wochenlang ertragen hat, dass sich in unserem Wohnzimmer die Fachartikel und Bücher stapelten.

Herzlichen Dank auch an die Personen, die mich durch das Korrekturlesen meiner Diplomarbeit unterstützt haben.

Anmerkung

Die Daten für diese Diplomarbeit wurden gemeinsam mit einer anderen Diplomandin (Claudia Krutis) erhoben. Es können dementsprechend Ähnlichkeiten bei der Beschreibung der demographischen Daten auftreten. Des Weiteren beruhen beide Diplomarbeiten auf derselben Basisliteratur, bearbeiten aber verschiedene Fragestellungen.

Abstract

In dieser Diplomarbeit wurde die Tendenz zu extremen Urteilen, sowie die Tendenz zur Mitte in Ratingskalen mittels eines von Böckenholt (2012) postulierten und von Khorramdel und von Davier (in Druck) sowie von Davier und Khorramdel (in Druck) erweiterten multidimensionalen Item Response Theorie-Ansatzes untersucht. Die erhobenen Daten wurden dafür in binäre Pseudoitems, welche die Tendenzen als Antwortsubprozesse darstellen sollten, zerlegt. Es wurde überprüft, ob Antworttendenzen in den Daten vorhanden sind, diese eindimensional erfassbar sind oder zusätzlich die Fragebogenskalen messen und ob bei Vorhandensein dieser Antworttendenzen der Fragebogeninhalt bzw. die persönlichen Einstellungen zu diesem einen Einfluss auf deren Auftreten haben. Zur Erhebung der Daten wurden als Teil eines Online-Self-Assessments das *Differentielle Stressinventar HR* (DSIHR, Lefevre & Kubinger, 2004) sowie zusätzlich Fragen nach der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) vorgegeben. Die Ergebnisse zeigten, dass Antworttendenzen in den Daten vorhanden waren, diese sich allerdings nicht für alle Messungen unidimensional darstellen ließen. Die binären Pseudoitems, die den extremen Antwortstil abbilden sollten, konnten in den Primärskalen des DSIHR *Manifestation*, *Coping* und *Stabilisierung* am besten dadurch erklärt werden, dass sie sowohl die Antworttendenz, als auch die Fragebogenskalen darstellen. Es zeigten sich außerdem Hinweise darauf, dass die Antworttendenzen durch den Fragebogeninhalt beeinflusst sein könnten. Des Weiteren zeigte sich sowohl für die Pseudoitems der Tendenz zur Mitte, als auch für jene der Tendenz zu extremen Antworten, dass die Verhaltensintention (Ajzen, 1985; Ajzen, 1991) bezüglich eines abgefragten Verhaltens einen Einfluss auf das Auftreten von Antworttendenzen haben könnte.

In this thesis the extreme response style and the midpoint response style in rating scales were examined using a multidimensional item response theory approach (MIRT) presented by Böckenholt (2012) and expanded by Khorramdel and von Davier (in press) and von Davier and Khorramdel (in press). The data was decomposed into binary pseudoitems, each representing one response style through a latent response subprocess. It was examined whether the response styles were unidimensionally measurable or in addition measured the scales of the used questionnaire. Furthermore, it was investigated whether or not the content of the questionnaire or the personal attitude towards it

influenced the frequency of the occurring response styles. To gather the data a questionnaire was used, which was part of an Online Self Assessment addressing different dimensions of stress called Differentielles Stressinventar HR (DSIHR, Lefevre & Kubinger, 2004). Additionally, questions based on the theory of planned behavior (Ajzen, 1985; Ajzen, 1991) were used. Results found response styles in the data, though they cannot always be measured unidimensionally. They also showed that the binary pseudo items representing the extreme response style can be explained most accurately by the response style and the intended psychological constructs for three out of four main scales of the DSIHR. Indices were found, that there is an effect of the content of the questionnaire on the frequency of using response styles. Furthermore, the results suggest that behavioral intention (Ajzen, 1985, Ajzen, 1991) of the behaviour in question could have an influence on the occurrence of response styles (extreme and midpoint).

Inhaltsverzeichnis

1 EINLEITUNG	1
2 PERSÖNLICHKEITSFRAGEBÖGEN.....	3
2.1 RATINGFORMATE.....	3
2.1.1 unipolar/bipolar.....	4
2.1.2 Abstufungen.....	4
2.1.3 Gerade/ungerade Anzahl an Abstufungen	4
2.1.4 Benennung.....	5
3 PROBLEME BEI DER BEANTWORTUNG VON FRAGEBÖGEN.....	5
3.1 FORMALE ANWORTTENDENZEN.....	6
3.1.1 Akquieszenz/Disakquieszenz.....	6
3.1.2 Tendenz zur Mitte	7
3.1.3 Tendenz zu extremen Urteilen.....	7
3.1.4 Tendenz zu milden Urteilen.....	8
3.1.5 Response Range.....	8
3.1.6 Positionseffekte.....	8
3.1.7 Reihenfolgeeffekte.....	8
3.2 INHALTLICHE ANWORTTENDENZEN.....	8
3.2.1 Soziale Erwünschtheit	9
3.2.2 Simulation/Dissimulation.....	9
3.3 DEMOGRAPHISCHE UND KULTURELLE UNTERSCHIEDE FORMALER ANWORTTENDENZEN.....	9
3.3.1 Alter	9
3.3.2 Geschlecht.....	10
3.3.3 Bildung.....	10
3.3.4 Kulturelle Unterschiede.....	10
4 FRAGEBOGENINHALT IN BEZUG AUF FORMALE ANWORTTENDENZEN.....	11
4.1 MODELLE DER FRAGEBOGENBEARBEITUNG IN BEZUG AUF DEN FRAGEBOGENINHALT	11
4.2 ARTEN VON FRAGEBOGENITEMS	15
4.2.1 Kategorisierung von Itemformulierungen	15
4.2.2 Kategorisierung von Iteminhalten	16
4.3 STUDIEN ZUM FRAGEBOGENINHALT IN BEZUG AUF ANWORTTENDENZEN.....	18
4.4 DIE THEORIE DES GEPLANTEN HANDELNS.....	19
5 METHODEN ZUR ERFASSUNG VON FORMALEN ANWORTTENDENZEN	21
5.1 ITEM RESPONSE THEORIE-ANSÄTZE ZUR MESSUNG VON ANWORTTENDENZEN.....	22
5.2 AKTUELLSTER IRT-ANSATZ ZUR MESSUNG VON ANWORTTENDENZEN.....	24
6 BESCHREIBUNG DES IN DER DIPLOMARBEIT VERWENDETEN IRT-MODELLS.....	31
6.1 DAS ZWEI-PARAMETER-LOGISTISCHE MODELL (2PL) VON BIRNBAUM (1968).....	31
6.2 ERWEITERUNG DES 2PL-MODELLS.....	32
7 EMPIRISCHER TEIL	33
8 METHODE.....	33

8.1	DATENERHEBUNG.....	33
8.2	STICHPROBE	34
8.3	ERHEBUNGSINSTRUMENT.....	35
8.4	DIFFERENTIELLES STRESSINVENTAR HR.....	35
8.4.1	<i>Kennwerte des Fragebogens</i>	36
8.4.2	<i>Skalen des DSIHR.....</i>	36
8.5	ZUSATZFRAGEN NACH DER THEORIE DES GEPLANTEN HANDELNS.....	37
8.6	DATENAUFBEREITUNG.....	39
8.7	SPEZIELLE VERWENDETE SOFTWARE ZUR AUSWERTUNG	39
9	HYPOTHESEN.....	40
10	ANALYSEN UND ERGEBNISSE.....	41
10.1	ANALYSEN DER GESAMTDATEN.....	41
10.2	ANALYSEN GETRENNT FÜR ALLE PSEUDOITEMS.....	43
10.3	BIFACTOR-MODELL.....	44
10.4	ANALYSEN BEZÜGLICH DES FRAGEBOGENINHALTS	45
10.5	EVALUATION DER MODELLE	47
10.6	IRT-MODELLE MIT MULTIDIMENSIONALEN ANWORTTENDENZ- UND MERKMALSAKTOREN... 47	
10.6.1	<i>Primärskalen als Faktoren</i>	47
10.6.2	<i>Subskalen je Primärskala als Faktoren</i>	49
10.7	IRT-MODELLE MIT EINEM UNIDIMENSIONALEN ANWORTTENDENZFAKTOR UND MULTIDIMENSIONALEN MERKMALSAKTOREN.....	52
10.7.1	<i>Primärskalen und Subskalen als Faktoren</i>	52
10.7.2	<i>Subskalen je Primärskala als Faktoren.....</i>	59
10.8	BIFACTOR-MODELLE	63
10.8.1	<i>Primärskalen und Subskalen als Faktoren.....</i>	64
10.8.2	<i>Subskalen je Primärskala als Faktoren</i>	66
10.9	FRAGEBOGENINHALT.....	70
10.9.1	<i>Pseudoitems des Typs m.....</i>	70
10.9.2	<i>Pseudoitems des Typs e.....</i>	72
10.9.3	<i>Mehrgruppenmodelle für die Primärskala Coping.....</i>	73
11	DISKUSSION.....	78
11.1	MULTIDIMENSIONALE ANWORTTENDENZFAKTOREN.....	79
11.2	UNIDIMENSIONALER ANWORTTENDENZFAKTOREN.....	80
11.2.1	<i>Pseudoitems des Typs m</i>	80
11.2.2	<i>Analysen der Pseudoitems des Typs e.....</i>	80
11.2.3	<i>Analysen der Pseudoitems des Typs d.....</i>	81
11.3	BIFACTOR-MODELLE.....	81
11.4	PSEUDOITEMS DES TYP S M.....	82
11.5	PSEUDOITEMS DES TYP S E.....	83
11.6	FRAGEBOGENINHALT	83
11.6.1	<i>Pseudoitems des Typs m.....</i>	83
11.6.2	<i>Pseudoitems des Typs e.....</i>	85
11.6.3	<i>Pseudoitems des Typs d.....</i>	86
11.7	KRITIK	86
12	ZUSAMMENFASSUNG.....	87
13	LITERATURVERZEICHNIS	90

14 ABBILDUNGSVERZEICHNIS.....	95
15 TABELLENVERZEICHNIS.....	96
16 ANHANG.....	99

1 Einleitung

Fragebögen bieten die Möglichkeit verschiedenste psychologische Konstrukte zu erfassen und werden auf Grund ihrer Ökonomie in vielen Bereichen der psychologischen Diagnostik eingesetzt. Die Ausprägung des gemessenen Konstrukts wird unter Anderem mit sogenannten Rating-Skalen erfasst. Sie bestehen aus mehreren Antwortmöglichkeiten, die als Abstufungen des erfragten Merkmals dienen. Testpersonen sollen bei der Beantwortung jene Möglichkeit wählen, die ihnen am für sie passendsten erscheint.

Diese Art der Messung bringt allerdings verschiedene Probleme mit sich. Unter anderem zählen dazu die Antworttendenzen. Antworttendenzen sind systematische Einflüsse auf das Testergebnis, die nicht durch die Ausprägung des gemessenen Merkmals der Person bedingt sind. Es wird angenommen, dass sie vom Iteminhalt unabhängig sind. Für diese Studie sind insbesondere zwei Tendenzen wichtig, nämlich die Tendenz zur Mitte und die Tendenz zu extremen Urteilen. Die Tendenz zur Mitte beschreibt das Verhalten, wenn Personen beim Beantworten von Fragen häufig den mittleren Bereich der Antwortskala wählen, um entweder ihre Einstellung zur gestellten Frage nicht zu offenbaren oder auch aus Unsicherheit, welche Antwort sie wählen sollen. Im Gegensatz dazu beschreibt die Tendenz zu extremen Urteilen das häufige Wählen der extremen Antwortkategorien einer Rating-Skala (Rost, 2004).

Mit Hilfe eines multidimensionalen Item Response Theorie- Ansatzes wird in der vorliegenden Diplomarbeit zum einen das Auftreten der oben erwähnten Antworttendenzen über alle Skalen eines Fragebogens hinweg untersucht, sowie der mögliche Einfluss des Fragebogeninhaltes auf die gemessenen Antworttendenzen, sollten diese vorhanden sein.

Böckenholt (2012) stellte einen multidimensionalen IRT-Ansatz vor, der es ermöglicht, multiple Antwortprozesse zu erfassen, indem dichotome „Pseudoitems“ aus den beobachteten Daten kodiert werden. Diese stellen einen latenten Antwortsubprozess dar. Khorramdel und von Davier (in Druck) erweiterten diesen Ansatz und verglichen unidimensionale Antworttendenzfaktoren mit multidimensionalen

Persönlichkeitsfaktoren. Durch die unidimensionale Messung von Antworttendenzen entsteht der Vorteil, dass die Ergebnisse detailliertere Informationen über die Dimensionalität der Antworttendenzen bieten und daher eine höhere Validität aufweisen und besser von Persönlichkeits- oder Merkmalsausprägungen unterschieden werden können. Schließlich werden die Daten auch noch mit dem von von Davier und Khorramdel (in Druck) beschriebenen Bifactor-Modell untersucht, da die gemessenen Antworttendenzen teilweise zusätzlich die Merkmalsausprägung messen könnten. Das Bifactor-Modell erlaubt eine gleichzeitige Ladung der Items auf einen Hauptfaktor (Antworttendenz) und einen Nebenfaktor (Merkmalsausprägung) und sollte damit die Ergebnisse noch besser beschreiben können.

Im ersten Teil der Arbeit wird auf Persönlichkeitsfragebögen und deren häufig benutzte Antwortformate eingegangen. Ebenso werden die Probleme mit diesen Antwortformaten erläutert und es wird speziell auf die Antworttendenzen eingegangen. Es werden verschiedene Methoden zur Erfassung von Antwortstilen vorgestellt und ein genauer Überblick über die angewendete Methode gegeben.

Der zweite Teil befasst sich mit den Methoden und der Durchführung der Studie sowie der Darstellung der Ergebnisse und der Diskussion.

2 Persönlichkeitsfragebögen

In der psychologischen Diagnostik verwendete Fragebögen dienen der Erfassung von Persönlichkeitsmerkmalen, Einstellungen, Interessen oder Ähnlichem. Im Vergleich zu Objektiven Persönlichkeitstests und projektiven Verfahren dient die Selbst- oder Fremdbeschreibung der Testpersonen als Grundlage der Interpretation. Da die Inhalte von Fragebögen meist keine „richtige“ oder „falsche“ Antwort haben, wird die Höhe der Ausprägung des interessierenden Merkmals interpretiert (Mossbrugger & Kelava, 2012). Um von einer Itemantwort auf eine Merkmalsausprägung schließen zu können, müssen nach Rost (2004) Anforderungen bezüglich des Antwortprozesses erfüllt sein. Die Testperson muss über genug Selbstkenntnis verfügen, um die abgefragten Inhalte beantworten zu können. Dieses Wissen über die eigene Persönlichkeit wird als Metakognition bezeichnet. Ist diese nicht vorhanden oder entspricht nicht der Realität, wird mit einem Fragebogen mehr das Selbstbild als eine tatsächliche Merkmalsausprägung abgefragt. Zudem muss die Testperson auch bereit sein, diese Merkmalsausprägung im Fragebogen korrekt darzustellen. Neben der Fähigkeit die eigenen Merkmalsausprägungen einzuschätzen, müssen auch die sozialen Vergleichsmaßstäbe gegeben sein, um beurteilen zu können, wie sehr man sich in der Ausprägung eines Merkmals von anderen unterscheidet (Rost, 2004).

2.1 Ratingformate

Fragebögen werden meist mit Ratingformaten bzw. Ratingskalen vorgegeben. Diese Antwortformate gehören nach Rost (2004) zur Gruppe der gebundenen Antwortformate und besitzen mehr als zwei abgestufte Antwortkategorien. Diese sollen eine Rangordnung für die Testpersonen darstellen. Ihre Formulierung ist item-unspezifisch, das heißt, die Antwortkategorien sind für alle Items eines Fragebogens gleich. Im Vergleich zu dichotomen Antwortformaten kann man mit Hilfe von Ratingformaten mehr Informationen sammeln, da die Testpersonen differenziertere Antworten geben können (Rost, 2004).

Ratingformate lassen sich durch vier Gestaltungsmerkmale (Rost, 2004) unterscheiden:

2.1.1 *unipolar/bipolar*

Die Antwortskalen können *unipolar* (z.B. 0-1-2-3-4) oder *bipolar* (z.B. -, -, 0, +, +) dargestellt werden (Rost, 2004). Unipolare Skalen gehen nur in eine Richtung und besitzen einen „Nullpunkt“, der das geringste Maß der Zustimmung oder Ablehnung beschreibt, wohingegen bipolare sowohl einen negativen, als auch einen positiven Pol aufweisen.

2.1.2 *Abstufungen*

Rating-Skalen können unterschiedlich viele *Abstufungen* aufweisen. Je mehr Stufen, desto differenzierter kann ein Merkmal erfasst werden (Rost, 2004).

- *mehrkategoriell*: Im Vergleich zu dichotomen Antwortalternativen erhält man durch mehrkategoriale Skalen genauere Informationen über die Ausprägung eines Merkmals, wobei die Ökonomie der Vorgabe sich nicht maßgeblich zum Schlechteren verändert (Rost, 2004). Zusätzlich sollte bei mehrkategorialen Antwortmöglichkeiten besonders auf die Exhaustivität (Vollständigkeit) Wert gelegt werden, da Testpersonen sich sonst eventuell für die am ehesten auf sie zutreffende Antwortmöglichkeit entscheiden, obwohl keine Antwortalternative zur Gänze zutrifft (Moosbrugger & Kelava, 2012). Wie viele Abstufungen letztendlich gewählt werden, ist abhängig davon, wie viel Differenzierung zwischen den Antwortalternativen man den Testpersonen zutraut (Rost, 2004).
- *analog*: Nach Bühner (2011) bietet dieses Antwortformat den Testpersonen die Möglichkeit auf einer kontinuierlichen Skala zwischen zwei Extremen jenen Bereich auszuwählen, der am ehesten zutrifft. Diese Art der Vorgabe hat sich durch die Verwendung von Computern in der psychologischen Diagnostik entwickelt.

2.1.3 *Gerade/ungerade Anzahl an Abstufungen*

Die Abstufungen können eine *gerade oder ungerade Anzahl* haben. Dies ist insofern von Bedeutung, als bei einer ungeraden Anzahl eine „neutrale“ mittlere Kategorie entsteht, die oft eher im Sinne einer „Ablehnung“ des Items als durch die Ausprägung des Merkmals ausgewählt wird (Rost, 2004).

2.1.4 Benennung

Die *Benennung* der Antwortmöglichkeiten spielt bei Rating-Skalen ebenfalls eine Rolle. Die Möglichkeiten reichen von Zahlen, über verbale/figurale Antwortstufen bis hin zu Häufigkeiten oder auch visuellen Analogskalen (Bühner, 2011).

- *Numerische Ratingskalen*: Die Antwortalternativen sind durch Zahlen markiert und vermitteln so den Eindruck einer präzisen Messung. Allerdings stellt die numerische Skala nicht sicher, dass die Abstände zwischen den Alternativen für die Testpersonen dieselbe Bedeutung haben (Mossbrugger & Kelava, 2011).
- *Verbale Ratingskala*: Sie zeichnet sich nach Mossbrugger & Kelava (2011) dadurch aus, dass alle Skalenpunkte mit Worten beschriftet sind und sich die Testpersonen keine subjektive Vorstellung davon machen müssen, was einzelne Skalenpunkte bedeuten, allerdings gibt es auch hier Schwankungen hinsichtlich der Bedeutung der sprachlichen Bezeichnung. Für die Beschreibung werden hauptsächlich Bewertungen, Häufigkeits- und Intensitätsangaben herangezogen.
- *optische Ratingskala*: Die Skalenpunkte werden durch Symbole beschrieben (z.B. Smileys). Dadurch sollen die Schwankungen in der Bedeutung im Vergleich zu sprachlichen Bezeichnungen verringert werden und zugleich nicht der Eindruck entstehen, dass die Skala übertrieben genau misst, wie es bei den numerischen Ratingskalen der Fall sein kann (Rost, 2004).

3 Probleme bei der Beantwortung von Fragebögen

Bei der Vorgabe von Fragebögen mit Rating-Skalen können verschiedene Probleme entstehen, die die Ergebnisse systematisch beeinflussen und damit die Interpretation erschweren. Böckenholt (2012) beschreibt, dass zum Umgehen dieses Problems Rating-Skalen oftmals binär ausgewertet werden (das heißt, alle zustimmenden Antworten und alle ablehnenden Antworten werden jeweils zu „Zustimmung“ und „Ablehnung“ zusammengefasst). Allerdings kann dadurch wertvolle Information verloren gehen. Rating-Skalen werden, trotz der Probleme, nach wie vor häufig verwendet.

Quellen für das Auftreten von Antworttendenzen können das Format der Rating-Skala, die Art und Weise, wie die Daten gesammelt werden (z.B. über Telefon oder persönlich), Interviewereffekte, die Höhe der kognitiven Belastung, die durch die Fragen verursacht werden, in welcher Sprache der Fragebogen vorgelegt wird oder auch die Wichtigkeit des abgefragten Inhaltes sein (Van Vaerenbergh & Thomas, 2012).

Die bei Rating-Skalen auftretenden Antworttendenzen können in zwei Formen unterschieden werden (Arnold, Eysenck & Meili, 1980, zit. Nach Seiwald, 2003): *Formale Antworttendenzen* und *inhaltliche Antworttendenzen*.

3.1 Formale Antworttendenzen

Formale Antworttendenzen sind durch die Gestaltung des Antwortformates bedingt und treten hauptsächlich bei Rating-Skalen auf (Arnold, Eysenck & Meili, 1980, zit. nach Seiwald, 2003). Wie von Baumgartner & Steenkamp (2001) sowie Paulhus (1991) beschrieben, wird bei der Verwendung von Rating-Skalen nämlich nicht nur das fragliche Konstrukt gemessen, sondern auch latente Variablen, die durch die Verwendung von Rating-Skalen hervorgerufen werden. Viele Studien weisen darauf hin, dass formale Antworttendenzen als stabile Antwortcharakteristik auftreten (Bachman & o'Malley, 1984; Weijters, Geuens & Schillewaert, 2010a; Weijters, Geuens & Schillewaert, 2010b) und vom Fragebogeninhalt unabhängig sind (Paulhus, 1991).

In der Literatur (vgl. Seiwald, 2003; Van Vaerenbergh & Thomas, 2012) werden einige Arten von Antworttendenzen unterschieden, nachfolgend werden die wichtigsten aufgelistet:

3.1.1 Akquieszenz/Disakquieszenz

Die Akquieszenz bezeichnet die Tendenz, allen Items unabhängig von ihrem Inhalt zuzustimmen, wohingegen die Disakquieszenz jenen Antwortstil beschreibt, bei dem Items unabhängig vom Inhalt abgelehnt werden (Bühner, 2011). Auf einer fünfstufigen Antwortskala würden diese Tendenzen folgendermaßen aussehen:



Durch diese Antworttendenzen werden die beobachteten Mittelwerte erhöht bzw. verringert und es können – unabhängig vom tatsächlich gemessenen Konstrukt – multivariate Zusammenhänge entstehen (Van Vaerenbergh & Thomas, 2012).

Eine Mischform daraus ist der „directional bias“, der die Tendenz beschreibt, mehr Akquieszenz als Disakquieszenz zu zeigen, also unabhängig vom gemessenen Konstrukt häufiger zuzustimmen als abzulehnen (Baumgartner & Steenkamp, 2001).

3.1.2 *Tendenz zur Mitte*

Unter der „Tendenz zur Mitte“ versteht man, die Tendenz die mittlere Kategorie einer Rating-Skala den anderen vorzuziehen (Rost, 2004).



Rost (2004) beschreibt, dass sich die mittlere/neutrale Kategorie in vielen Untersuchungen eher ungünstig auswirkte (vgl. Rost, Carstensen & von Davier, 1999). Die neutrale Antwortmöglichkeit wird von Testpersonen oft nicht nur als eine mittlere Position zwischen zwei Polen verstanden, sondern wird auch angekreuzt, wenn das Item abgelehnt und dadurch eine Antwort verweigert wird. Des Weiteren beschreibt Rost (2004), dass motivierte Testpersonen die mittlere Kategorie eher meiden.

Durch diese Antworttendenz verringert sich die Varianz in den Daten, da die mittlere Antwortkategorie häufiger verwendet wird. Außerdem können auch hier – unabhängig vom tatsächlich gemessenen Konstrukt – multivariate Zusammenhänge entstehen (Van Vaerenbergh & Thomas, 2012).

3.1.3 *Tendenz zu extremen Urteilen*

Die Tendenz zu extremen Urteilen beschreibt im Vergleich zur Tendenz zur Mitte die Neigung von Testpersonen, extreme Urteile anzukreuzen (Rost, 2004).



Die Konsequenzen für die Auswertung der Daten entsprechen jenender Akquieszenz (Van Vaerenbergh & Thomas, 2012).

3.1.4 *Tendenz zu milden Urteilen*

Im Vergleich zur Tendenz zu extremen Urteilen beschreibt dieser Antwortstil die Tendenz, extreme Antworten zu vermeiden (Kubinger, 2009).



Durch diese Antworttendenz verringert sich die Varianz in den Daten, da die extremen Urteile weniger häufig gewählt werden; außerdem erhöht sich auch hier das Ausmaß der multivariaten Zusammenhänge (Van Vaerenbergh & Thomas, 2012).

3.1.5 *Response Range*

Beschreibt die Tendenz, eine enge oder weite Spanne von Antwortkategorien – ausgehend von der mittleren Antwortkategorie – zu verwenden.

Wenn die Spanne groß ist, erhöht sich die Varianz in den Daten und das Ausmaß der multivariaten Zusammenhänge (Van Vaerenbergh & Thomas, 2012).

3.1.6 *Positionseffekte*

Positionseffekte beschreiben das Phänomen, dass die Itemposition im Fragebogen einen Einfluss auf die Beantwortung dieses Items hat (Seiwald, 2003; Kubinger, 2009).

3.1.7 *Reihenfolgeeffekte*

Der Reihenfolgeeffekt beschreibt die Beeinflussung der Antwort eines Items auf Grund der Items, die zuvor beantwortet wurden (Rost, 2004).

Des Weiteren werden noch das **Raten** und die **zufällige Beantwortung** von Fragen zu den formalen Antworttendenzen gezählt (Seiwald, 2003; Kubinger, 2009).

3.2 **Inhaltliche Antworttendenzen**

Inhaltliche Antworttendenzen beziehen sich auf bewusste oder unbewusste Verfälschungen der Antworten auf Grund des Fragebogeninhaltes (z.B. in

Bewerbungssituationen, Versicherungsfragen etc.) (Arnold, Eysenck & Meili, 1980, zit. nach Seiwald, 2003).

3.2.1 Soziale Erwünschtheit

Diese Tendenz beschreibt die Neigung einer Person, Antworten in einem Fragebogen nach der sozialen Erwünschtheit auszuwählen. Dies passiert häufig, wenn die Konsequenzen der Testung für die Testperson von Bedeutung sind (Seiwald, 2003).

3.2.2 Simulation/Dissimulation

Die Simulation beschreibt das absichtliche Verschlimmern bestehender (Krankheits-) Symptome, während unter Dissimulation das Gegenteil verstanden wird, also die Bagatellisierung von körperlichen oder seelischen Erkrankungen oder sonstigen Symptomen (Seiwald, 2003; Bühner, 2011).

3.3 Demographische und kulturelle Unterschiede formaler Antworttendenzen

Da in der vorliegenden Diplomarbeit formale Antworttendenzen (Tendenz zu extremen Urteilen und die Tendenz zur mittleren Antwortkategorie) untersucht werden, werden nur Ergebnisse, die sich auf diese zwei formalen Antworttendenzen beziehen, präsentiert.

3.3.1 Alter

De Jong, Steenkamp, Fox und Baumgartner (2008) sowie van Rosmalen, van Herk und Groenen (2010) fanden in ihren Studien Hinweise darauf, dass sowohl jüngere, als auch ältere Menschen die Tendenz zu extremen Urteilen stärker zeigen als Personen im mittleren Alter. Meisenberg und Williams (2008) sowie Weijters et al. (2010a) haben Hinweise darauf gefunden, dass Alter und die Tendenz zu extremen Urteilen positiv zusammenhängen, wohingegen die Studie von Austin, Deary und Egan (2006) das gegenteilige Ergebnis erbrachte. Johnson, Kulesa, Cho und Shavitt (2005) fanden in ihren Studien keinen Zusammenhang zwischen dem extremen Antwortstil und dem Alter.

Für die Tendenz zur mittleren Antwortkategorie fanden Weijters et al. (2010a) einen positiven Zusammenhang mit dem Alter.

3.3.2 *Geschlecht*

Die Ergebnisse von De Jong et al. (2008) sowie Weijters et al. (2010a) weisen darauf hin, dass Frauen eher zur Tendenz zu extremen Urteilen tendieren als Männer. Meisenberg und Williams (2008) fanden in ihrer Studie wiederum Hinweise darauf, dass das Gegenteil der Fall ist. Grimm und Church (1999) konnten in ihren Daten keinen Effekt des Geschlechts nachweisen.

Khorramdel und von Davier (in Druck) fanden keine signifikanten Unterschiede zwischen Männern und Frauen bezüglich der Tendenz zur mittleren Antwortkategorie.

3.3.3 *Bildung*

De Jong et al. (2008) sowie Moors (2008) konnten keine Effekte von Bildung auf die Tendenz zu extremen Urteilen nachweisen. In den Ergebnissen der Studien von Weijters et al. (2010a) sowie van Rosmalen, van Herk und Groenen (2010) zeigte sich ein negativer Zusammenhang zwischen Bildung und der extremen Antworttendenz.

Auch für die Tendenz zur mittleren Antwortkategorie fanden Weijters et al. (2010a) einen negativen Zusammenhang mit Bildung.

3.3.4 *Kulturelle Unterschiede*

Die Ergebnisse der Studie von Khorramdel und von Davier (in Druck) weisen darauf hin, dass asiatische Studierende signifikant seltener die Tendenz zu extremen Urteilen zeigen als afro-amerikanische, weiße oder Studierende mit anderer Abstammung. Grund und Dolnicar (2009) fanden Hinweise darauf, dass internationale Studierende dazu neigen, extreme Antworten zu vermeiden. Afrikanische (Bachman & O'Malley, 1984) und spanische Amerikaner und Amerikanerinnen (Hui & Triandis, 1989) zeigten die Tendenz zu extremen Antworten häufiger als Personen aus Nordamerika mit europäischen Wurzeln. Ebenso zeigte sich die Tendenz zu extremen Urteilen häufiger in mediterranen Ländern (Griechenland, Italien und Spanien) als in westeuropäischen Ländern (England, Deutschland und Frankreich) (van Herk, Poortinga & Verhallen,

2004). Van Rosmalen, van Herk und Groenen (2010) fanden in ihrer Studie heraus, dass Personen aus südeuropäischen Ländern eine höhere Wahrscheinlichkeit zeigen, extreme Antwortkategorien zu verwenden als Personen aus dem United Kingdom und Deutschland.

Khorramdel und von Davier (in Druck) fanden keine signifikanten kulturellen Unterschiede bezüglich der Tendenz zur mittleren Antwortkategorie. Chen, Lee und Stevenson (1995) sowie Hamamura, Heine und Paulhus (2008) konnten in ihren Studien Hinweise darauf finden, dass japanische und chinesische Studierende öfter die mittlere Antwortkategorie wählen als amerikanische Studenten mit europäischen Wurzeln.

Auf das Thema von Antworttendenzen im Zusammenhang mit dem Fragebogeninhalt wird gesondert eingegangen, da dies auch den empirischen Teil dieser Diplomarbeit betrifft.

4 Fragebogeninhalt in Bezug auf formale Antworttendenzen

Es wird angenommen, dass formale Antworttendenzen unabhängig vom Inhalt des Fragebogens auftreten. Da im empirischen Teil dieser Diplomarbeit der Frage nachgegangen wird, ob unter der Voraussetzung, dass Antworttendenzen in den Daten auffindbar sind, doch ein Einfluss bestehen könnte, wird in diesem Kapitel näher auf den Fragebogeninhalt eingegangen. Um eine theoretische Basis für die später dargestellten Hypothesen zu schaffen, werden Einflüsse des Fragebogeninhalts auf verschiedene Stufen des Antwortprozesses, sowie Arten von Itemformulierungen und Inhalten dargestellt und auf diverse Studien betreffend formale Antworttendenzen eingegangen.

4.1 Modelle der Fragebogenbearbeitung in Bezug auf den Fragebogeninhalt

Mossbrugger und Kelava (2012) postulieren, dass dem kognitiven Prozess, der für die Beantwortung eines Items nötig ist, mehr Beachtung geschenkt werden sollte, um ein besseres Verständnis für Fehlerquellen (z.B. Antworttendenzen) zu erhalten.

Ein Modell zur Erklärung der benötigten kognitiven Antwortprozesse zur Beantwortung eines Items stellen Podsakoff, MacKenzie, Lee und Podsakoff (2003) dar:

- *Verstehen*: Die Aufmerksamkeit muss in einem ersten Schritt auf das Item gerichtet werden, um im Anschluss den Iteminhalt zu verstehen, aufzunehmen und eine mentale Repräsentation der Bedeutung zu generieren.

Fehlerquelle: Nach Podsakoff et al. (2003) führt hier Itemmehrdeutigkeit dazu, dass Testpersonen eigene Heuristiken zur Beantwortung der Items verwenden oder die Antworten zufällig auswählen.

- *Abruf*: Diese Stufe beschreibt, wie für die Itembeantwortung relevante Informationen aus dem Gedächtnis abgerufen werden.

Fehlerquellen: Der Kontext der Messung (z.B. Ort, Art der Vorgabe), soziale Erwünschtheit, Priming-Effekte durch die Items sowie die eigenen Stimmungslage (transient mood state) können einen verfälschenden Einfluss auf den Abruf von Informationen haben. Außerdem können Items, die neutral formuliert sind, im Kontext mit positiv und negativ formulierten Items diese wertenden Eigenschaften subjektiv annehmen (item embeddedness). Im Falle, dass Items von verschiedenen Konstrukten nacheinander vorgegeben werden, können die Intrakonstrukt Korrelationen sinken und die Interkonstrukt Korrelationen ansteigen (item intermixing). Die Skalenlänge nimmt insofern einen Einfluss auf den Abruf, als dass bei kürzeren Skalen die vorhergegangenen Items noch leichter erinnert werden und deshalb einen Einfluss auf das Abrufen von Informationen bezüglich des vorliegenden Items nehmen.

- *Urteil*: Die abgerufenen Informationen werden auf ihre Richtigkeit hin beurteilt und im Zusammenhang mit dem Iteminhalt wird daraus ein Urteil gebildet.

Fehlerquellen: Testpersonen können dazu tendieren ihre Antwort möglichst konsistent zu geben (consistency motif). Des Weiteren kann die Formulierung der Items oder deren Inhalt versteckte Hinweise darauf enthalten, wie geantwortet werden soll (item demand characteristics). Beim Urteilen können auch Stimmungseffekte entstehen, die durch die Bearbeitung eines Items oder eines Sets von Items ausgelöst wurden, und Auswirkungen auf die Beantwortung des gesamten Fragebogens haben (context-induced mood).

- *Wahl der Antwortkategorie:* Danach kommt es zur Auswahl einer Antwortkategorie, die optimal zum Urteil passt.

Fehlerquellen: Hier stellen vor allem das Antwortformat, die Benennung der Antwortmöglichkeiten und die Itemposition Fehlerquellen dar.

- *Antwortabgabe:* Die Entscheidung wird noch einmal auf Konsistenz überprüft und die Antwort wird abgegeben.

Fehlerquellen: Auch hier kommen die bereits oben beschriebenen Fehlerquellen consistency motif und item demand characteristics zum Tragen. Zusätzlich beschreiben Podsakoff et al. (2003), dass an diesem Punkt im Antwortprozess Akquieszenz bzw. Disakquieszenz, sowie sozial erwünschtes Antwortverhalten auftreten.

Der Iteminhalt nimmt nach dieser Theorie also in verschiedenen Formen als Fehlerquelle Einfluss auf den kognitiven Bewertungsprozess bei der Beantwortung von Fragebogenitems: unter anderem durch das Hervorrufen von Heuristiken, Priming-Effekte, Beeinflussung der Stimmungslage, dem Wunsch nach konsistenter Beantwortung der Items, soziale Erwünschtheit etc. Formale Antworttendenzen werden als Fehlerquellen in die Theorie im letzten Unterpunkt (Antwortabgabe) beschrieben, auch Podsakoff et al. (2003) sehen die (genannten) Antworttendenzen als inhaltsunabhängig.

Krosnick (1999) beschreibt einen ähnlichen Antwortprozess wie Podsakoff et al. (2003):

First, respondents must interpret the question and deduce its intent. Next, they must search their memories for relevant information and then integrate that information into a single judgment (if more than one consideration is recalled). Finally, they must translate the judgment into a response by selecting one of the alternatives offered. (S. 546)

Zusätzlich postuliert er mit seinem Optimizing-Satisficing-Modell sehr sparsam zwei Zugänge zur Beantwortung von Fragebögen, bei denen zusätzlich die Motivation eine maßgebliche Rolle für die Qualität der Beantwortung spielt.

- *Optimizing*: Viele Items zu verschiedensten Themengebieten sorgfältig beantworten zu können, erfordert einen entsprechend hohen kognitiven Aufwand. Sind Motive (z.B. Altruismus, Wunsch nach Selbstdarstellung, etc.) zur Beantwortung des Fragebogens vorhanden, bewirkt dies eine genaue und sorgfältige Beantwortung der Items, da der kognitive Prozess für jedes Item vollständig durchlaufen wird.
- *Satisfizing*: Sind keine Motive zur Beantwortung des Fragebogens vorhanden, kann das dazu führen, dass die Stufen des Antwortprozesses nur oberflächlich bearbeitet werden und die gewählten Antworten nicht der Merkmalsausprägung entsprechen. Im schlechtesten Fall beantworten die Testpersonen die Items nur mehr entsprechend der vermuteten Fragebogenintention, um den kognitiven Aufwand gering zu halten.

Krosnick (1999) beschreibt, dass die Wahrscheinlichkeit, dass eine Testperson Items nach dem Satisfizing-Prinzip bearbeitet steigt, je höher die Aufgabenschwierigkeit ist und je niedriger die Personenfähigkeit und die Motivation der Person zur sorgfältigen Bearbeitung ist.

Krosnick (1999) erklärt diese drei Einflussfaktoren folgendermaßen: Aufgabenschwierigkeit beschreibt die Probleme, den Iteminhalt sowie die Antwortmöglichkeiten zu interpretieren, Erinnerungen abzurufen und zu manipulieren, externe Störfaktoren und Ähnliches. Die Personenfähigkeit bezieht sich darauf, wie sehr sich eine Person schon mit dem abgefragten Thema gedanklich beschäftigt hat und wie viel Übung sie mit der Bearbeitung von Items hat. Die Motivation kann durch die persönliche Wichtigkeit des Fragebogeninhalts, Annahmen über die Sinnhaftigkeit des Verfahrens oder das Verhalten der testleitenden Person beeinflusst werden.

Der Fragebogeninhalt nimmt laut diesen Annahmen also Einfluss auf die Motivation und demnach darauf, ob ein Fragebogen mit viel oder wenig kognitivem Aufwand bearbeitet wird. Die Vermutung, dass sich bei Personen, die den kognitiven Antwortprozess nur oberflächlich ausführen, möglicherweise angewendete formale Antworttendenzen häufiger zeigen, liegt nahe.

4.2 Arten von Fragebogenitems

Fragebogenitems unterscheiden sich in der Art, welche Inhalte abgefragt werden bzw. in welcher Form dies geschieht. Auf den nächsten Seiten sollen gängige Kategorisierungssysteme diesbezüglich dargestellt werden.

4.2.1 Kategorisierung von Itemformulierungen

4.2.1.1 direkte vs. indirekte Fragen

Es besteht die Möglichkeit ein Item direkt (Sind Sie gesellig?) oder indirekt (Lachen Sie viel, wenn sie unter Leuten sind?) zu formulieren (Moosbrugger & Kelava, 2012). Im Falle der direkten Formulierung muss beachtet werden, dass es kein interindividuelle Übereinstimmung der angesprochenen Inhalte geben muss, die Items von verschiedenen Personen also unterschiedlich interpretiert werden könnten (Moosbrugger & Kelava, 2012).

4.2.1.2 hypothetische vs. biographiebezogene Itemformulierung

Moosbrugger und Kelava (2012) beschreiben, dass Sachverhalte durch hypothetische Situationen (Stellen Sie sich vor, sie müssen sehr viel Stoff für eine wichtige Prüfung lernen. Wie gehen Sie vor?) oder in Anlehnung an ein biographisches Ereignis abgefragt werden können (Wie haben Sie sich auf die letzte Prüfung vorbereitet?).

Des Weiteren beschreiben sie, dass die hypothetische Itemformulierung zu Fehleinschätzungen bezüglich des eigenen Verhaltens führt, deshalb sind biographische Formulierungen vorzuziehen. Allerdings kann es auch dabei zu Problemen kommen, da das abgefragte Merkmal mit einer Situationskomponente konfundiert sein kann oder passende biographische Formulierungen beschränkt bzw. nicht sinnvoll sind.

4.2.1.3 konkrete vs. abstrakte Items

Konkret formulierte Items (Wie verhalten Sie sich, wenn Sie einen Streit in der Familie schlichten sollen?) geben mehr Aufschluss über situationale Bedingungen und sind abstrakten Items (Wie belastend schätzen Sie ein konfliktreiches Familienumfeld ein?) vorzuziehen, da diese größere Interpretationsfreiräume zulassen (Moosbrugger & Kelava, 2012).

4.2.1.4 personalisierte vs. depersonalisierte Items

Moosbrugger und Kelava (2012) beschreiben des Weiteren personalisierte Fragen (Nehmen Sie die Antibabypille?), die zuverlässige Informationen liefern, sofern sie wahrheitsgemäß beantwortet und nicht als Verletzung der Privatsphäre verstanden werden. Ihnen gegenüber stehen depersonalisierte Formulierungen (Sollte die Antibabypille verwendet werden?). Diese können zu nichtssagenden Antworten führen bzw. nicht dieselbe Information liefern wie die personalisierte Form.

4.2.1.5 Stimulusqualität/-intensität

Letztendlich lassen sich Items nach Moosbrugger und Kelava (2012) noch nach ihrer Stimulusqualität einteilen. Das bedeutet, dass sich die emotionale Intensität, mit der Reaktionen bei Testpersonen hervorgerufen werden, unterscheiden kann. Ein Beispiel für ein emotionsneutrales Item wäre: „Sind Sie ängstlich?“. Durch die Frage „Gehen Sie schneller, wenn Ihnen in der Nacht jemand folgt?“ wird ein emotionaler Zustand abgefragt.

4.2.1.6 Schwierigkeit des Items

Rost (2004) beschreibt noch eine weitere Kategorisierungsmöglichkeit für Items: Die Itemformulierung kann sich darin unterscheiden, wie schwer es einer durchschnittlichen Person fällt, einem Item zuzustimmen. „Schwere“ Items (Die Polizei sollte Personen, die auf der Autobahn zu schnell fahren sofort einsperren und strafrechtlich verfolgen) manipulieren im Gegensatz zu „leichten“ Items (Die Polizei sollte Personen, die auf der Autobahn zu schnell fahren, mit den zur Verfügung stehenden Rechtsmitteln bestrafen) die Wahrscheinlichkeit der Zustimmung.

4.2.2 Kategorisierung von Iteminhalten

Angleitner, John und Löhr (1986, S. 69) beschreiben ein Kategoriesystem für Iteminhalte:

4.2.2.1 Fragen zur Selbstbeschreibung

Fragen zur Selbstbeschreibung können in drei Subkategorien unterteilt werden:

Beobachtbare Reaktionen: Beziehen sich auf Reaktionen, die prinzipiell beobachtbar sind. Ein mögliches Item wäre: „Ich bin häufig im Theater.“

Nicht beobachtbare bzw. internale Reaktionen: Beziehen sich auf Reaktionen die privat und von außen nicht beobachtbar sind. Zum Beispiel: „Ich fühle mich oft einsam.“

Symptome: Die dritte Kategorie sind Symptome im Sinne physikalischer Reaktionen, wie zum Beispiel: „Ich habe häufig Kopfweg.“

4.2.2.2 Eigenschaftszuschreibungen

Items, die mit dieser Kategorie beschrieben werden können, beruhen auf der Frage nach den Eigenschaften einer Testperson. Die Eigenschaften werden häufig durch Adjektive ausgedrückt. Zum Beispiel: „Ich bin ein freundlicher Mensch.“ Items dieser Kategorie können zusätzlich dadurch beschrieben werden, ob sie modifiziert oder unmodifiziert sind. Häufige Modifikationen sind Spezifizierungen der Häufigkeit, der Dauer oder des situationalen Kontextes.

4.2.2.3 Wünsche und Interessen

Entsprechende Items beziehen sich auf die Intention, ein bestimmtes Verhalten zu zeigen oder dem Verlangen nach etwas Bestimmten. Zum Beispiel: „Ich möchte gerne ElektrikerIn werden.“

4.2.2.4 Biographische Fakten

Items, die in diese Kategorie fallen, beziehen sich auf die Vergangenheit bzw. vergangene Geschehnisse. Zum Beispiel: „In meiner Jugend hatte ich Probleme mit dem Gesetz.“

4.2.2.5 Einstellungen und Überzeugungen

Diese Kategorie beinhaltet Items bezüglich Überzeugungen, Meinungen und Einstellungen über verschiedene Themengebiete. Zum Beispiel: „Das Gesetz sollte streng durchgesetzt werden.“

4.2.2.6 Reaktionen anderer gegenüber einer Person

Diese Items beschreiben das Verhalten und die Einstellungen von anderen Personen über die Testperson. Zum Beispiel: „Meine Familie ist mit meinen Schulnoten nicht zufrieden.“

4.2.2.7 Bizarre Items

Die meisten der Items, die in diese Kategorie fallen, beschreiben unübliche, sonderbare oder abnormale Verhaltensweisen oder Erlebnisse („Jemand will mich erschießen.“).

Tränkle (1983, S. 243) beschreibt noch eine zusätzliche Kategorie von Frageninhalten:

4.2.2.8 Motive

In diese Kategorie fallen Items, die sich mit den Motiven einer Testperson beschäftigen. Zum Beispiel: „Warum sind Sie dieser Meinung?“

Diese Kategoriesysteme werden im empirischen Teil zur Einteilung der Subskalen des DSIHR verwendet. Im nächsten Unterkapitel sollen Studien zum Thema formale Antworttendenzen und ihrer (Un-) Abhängigkeit vom Fragebogeninhalt dargestellt werden.

4.3 Studien zum Fragebogeninhalt in Bezug auf Antworttendenzen

Paulhus (1991) beschreibt, dass Antworttendenzen unabhängig vom Iteminhalt auftreten und wie oben bereits erwähnt angenommen wird, dass sie stabile Antwortcharakteristiken darstellen. Diese Annahme findet sich in vielen Studien (vgl. Bachman & o'Malley, 1984; Khorramdel & von Davier, in Druck; von Davier & Khorramdel, in Druck; Weijters, Geuens & Schillewaert, 2010a; Weijters, Geuens & Schillewaert, 2010b). Das von Greenleaf (1992) vorgeschlagene Vorgehen heterogene Items in ein bestehendes Verfahren einzubinden, um die Wahrscheinlichkeit zu verringern, dass neben Antworttendenzen auch die Merkmalsausprägung gemessen wird, spricht ebenfalls dafür, dass der Iteminhalt für die Antworttendenzen irrelevant ist.

In der Studie von Weijters et al. (2010) wurden den Testpersonen zwei Fragebögen mit heterogenen Items vorgelegt. Dieselben Fragebögen bearbeiteten die Personen nach einem Jahr nochmals. Die Ergebnisse weisen darauf hin, dass Antworttendenzen (in diesem Fall Akquieszenz und Tendenz zu extremen Urteilen) über längere Zeit hinweg stabil sind und implizieren in Folge dessen, dass sie vom Iteminhalt unabhängig sind (da sie über verschiedene Skalen hinweg auftraten). Die Studie von Gibbons, Zelner und Reduk (1999) wiederum beschäftigte sich unter anderem mit der Frage, welchen Einfluss die wahrgenommene Sinnhaftigkeit des Fragebogenthemas bzw. einzelner Items auf Antworttendenzen hat. Die Ergebnisse zeigten, dass sich die Tendenz zu extremen Urteilen stärker zeigt, wenn der Inhalt des Fragebogens relevant für die Testpersonen ist.

Da der IRT-Ansatz von Khorramdel und von Davier (in Druck) die Möglichkeit bietet, Antworttendenzen in den Daten nachzuweisen, kann bei Vorliegen dieser auch beobachtet werden, ob sie durch den Fragebogeninhalt beeinflusst sind. Dazu wurden in dieser Diplomarbeit zum einen die Skalen des vorgegebenen Fragebogens verwendet, und zum Anderen wurden zusätzliche Fragen auf Grundlagen der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) erstellt, die Aufschluss über die Auswirkung der persönlichen Einstellungen auf Antworttendenzen geben sollten. Die Theorie des geplanten Handelns wird an dieser Stelle näher erklärt .

4.4 Die Theorie des geplanten Handelns

Die Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) ist eine Erweiterung der Theorie des überlegten Handelns (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975), da das ursprüngliche Modell Personen mit fehlender willentlicher Kontrolle zur Ausführung eines Verhaltens nicht miteinbezog.

Die Annahme der Theorie des geplanten Handelns ist, dass die Verhaltensabsicht (Intention) der beste Prädiktor für überlegtes Handeln darstellt. Es wird angenommen, dass Intentionen die motivationalen Faktoren erfassen, die Verhalten beeinflussen (z.B. wie sehr Personen sich anstrengen, um ein bestimmtes Verhalten zu zeigen).

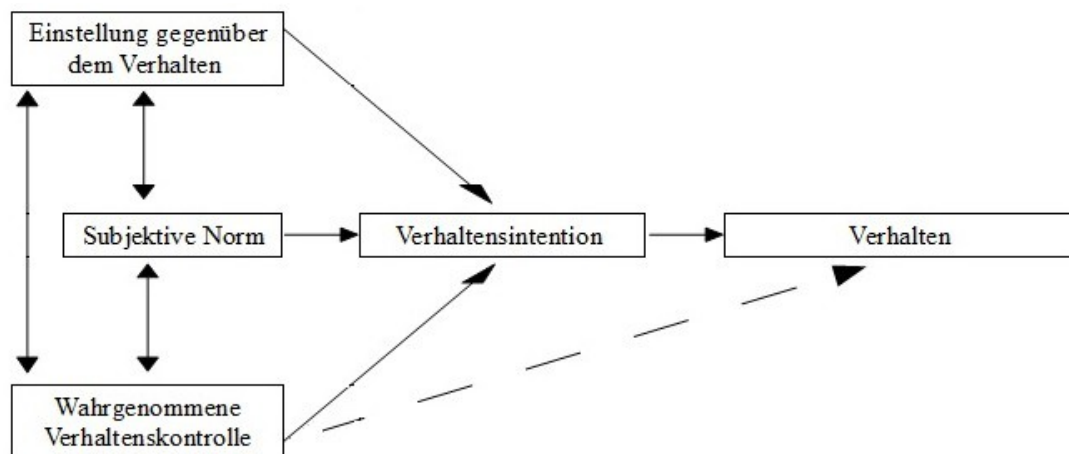


Abbildung 1: Theorie des geplanten Handelns nach Ajzen, 1991, S. 182

Grundsätzlich gilt, je stärker die Intention ein Verhalten zu zeigen ist, um so wahrscheinlicher ist es, dass es auch tatsächlich ausgeführt wird. Die Verhaltensintention kann sich natürlich nur dann zeigen, wenn das fragliche Verhalten unter willentlicher Kontrolle steht (Ajzen, 1991). Die Intention wiederum wird durch drei Faktoren bedingt, die in Abbildung 1 dargestellt sind.

Spezifische Einstellung gegenüber dem Verhalten (attitude toward the behavior): Bezieht sich darauf wie günstig oder ungünstig die subjektive Bewertung des fraglichen Verhaltens ist.

Die Ergebnisse der Studie von Davidson und Jaccard (1979) weisen darauf hin, dass die spezifische Einstellung zu einem Verhalten möglichst konkret abgefragt werden muss, um das Verhalten vorherzusagen. Es zeigte sich in ihrer Studie, dass die Einnahme der Antibabypille bei Frauen durch eine konkrete Frage zur Einnahme der Antibabypille wesentlich besser vorhergesagt werden konnte, als durch eine Frage, die die Geburtenkontrolle im Allgemeinen betrifft. Ajzen (1991) hebt diesen Punkt ebenfalls hervor.

Subjektive Norm (subjective norm): Beschreibt einen sozialen Faktor, nämlich den sozialen Druck ein Verhalten auszuführen bzw. nicht auszuführen und in wie fern das Verhalten von Mitmenschen akzeptiert oder nicht akzeptiert wird (Ajzen, 1991).

Um dies anhand eines Beispiels zu veranschaulichen, stelle man sich eine Person vor, die nicht gerne in nördlichen Ländern Urlaub macht. Es besteht trotzdem die Möglichkeit, dass sie das Verhalten zeigt, da zum Beispiel der Lebenspartner oder die Lebenspartnerin einen Einfluss auf die Wahl des Urlaubsziels hat.

Wahrgenommene Verhaltenskontrolle (perceived behavioral control): Bezieht sich auf die wahrgenommene Schwierigkeit oder Leichtigkeit, mit der ein Verhalten ausgeführt werden kann. Diese Kontrolle kann von Situation zu Situation variieren. Wahrgenommene Verhaltenskontrolle und die Verhaltensabsicht können direkt zur Ausführung des Verhaltens führen. Das Verhalten kann durch diese beiden Faktoren aber nur dann auch vorhergesagt werden, wenn sie im Zusammenhang mit einer konkreten Handlung stehen, der situationale Kontext derselbe ist, die beiden Faktoren bis zur Ausführung des Verhaltens stabil bleiben und die wahrgenommene Verhaltenskontrolle auch der tatsächlichen Verhaltenskontrolle entspricht (Ajzen, 1991).

Die genaue Formulierung der zusätzlichen Items sowie eine Erklärung der Subskala in Bezug auf die vorliegende Diplomarbeit folgen im empirischen Teil. Im nächsten Kapitel wird auf verschiedenen Studien und Ansätze zur Erfassung von Antworttendenz eingegangen.

5 Methoden zur Erfassung von formalen Antworttendenzen

Im Laufe der Zeit haben sich verschiedene Ansätze zur Identifizierung von Antworttendenzen entwickelt. Die einfachsten Methoden sind Summenscores der einzelnen Antwortmöglichkeiten (z.B. die Summe aller extremen Antworten) oder auch die Standardabweichung von Itemscores einer einzelnen Testperson, bei Items mit heterogenem oder negativ formuliertem Inhalt (Bachman & o'Malley, 1984; Baumgartner & Steenkamp, 2001; Buckley, 2009, Johnson et al. 2005).

Von mehreren AutorInnen wurden (Baumgartner & Steenkamp, 2001, Greenleaf, 1992, Johnson et al., 2005) zur Identifizierung von formalen Antworttendenzen Items mit heterogenem Inhalt von unterschiedlichen psychologischen Konstrukten, die möglichst wenig mit den anderen Items korrelieren zu einem bestehenden Instrument hinzugefügt. Die zusätzlichen Items sollen Sicherheit darüber geben, dass Antworttendenzen und

nicht (zusätzlich) das intendierte Konstrukt gemessen werden, was dieses Problem aber nicht zur Gänze löst. Außerdem ergeben sich bei der Suche nach passenden Items zur Ergänzung eines bestehenden Instruments nach Bolt und Newton (2011) zwei Schwierigkeiten: Solche Items sind schwer zu finden und können durch die Einbeziehung von neuen Items in bestehende Verfahren einen negativen Einfluss auf die Validität einer Messung haben.

Bei dieser Art der Erfassung kann nur vermutet werden, dass eine Antworttendenz gemessen wird, da nicht davon ausgegangen werden kann, dass diese Häufigkeiten ausschließlich auf Grund einer Antworttendenz zustande kommen. Es wird keine Rücksicht darauf genommen, dass gegebene Antworten nicht nur Antworttendenzen, sondern auch das abgefragte Konstrukt messen können (vgl. Khorramdel & von Davier, in Druck; von Davier & Khorramdel, in Druck).

Weitere häufig angewendete Ansätze zur Beobachtung von Antworttendenzen basieren auf statistischen Modellen wie Latent-Class-Analysen (Van Rosmalen, Van Herk & Groenen, 2010; Moors, 2010), konfirmatorischen Faktoranalysen (Billiet & McCleod, 2000) oder IRT-Modellen (Bolt & Johnson, 2009; De Jong et al. 2008; Khorramdel & von Davier, in Druck; von Davier & Khorramdel, in Druck). Im nächsten Kapitel werden die IRT-Modelle genauer dargestellt.

5.1 Item Response Theorie-Ansätze zur Messung von Antworttendenzen

Um die oben beschriebenen Probleme zu umgehen, haben sich in den letzten Jahren einige Studien mit IRT-Ansätzen zur Messung von Antworttendenzen beschäftigt (De Jong, Steenkamp, Fox & Baumgartner, 2008; Bolt & Johnson, 2009; Bolt & Newton, 2011; 2009; Rost, Carstensen & von Davier, 1999; Austin, Deary & Egan 2006; Johnson & Bolt, 2010), die den Vorteil haben mehrere Items auf ihre Eindimensionalität zu untersuchen. Sie bieten dadurch die Möglichkeit zu überprüfen, ob tatsächlich der intendierte Aspekt gemessen wird (Khorramdel & von Davier, in Druck) bzw. Personenparameter (Merkmalsausprägungen einer Person) und Itemparameter getrennt zu erfassen, die beide mit Antworttendenzen interagieren können (Bolt & Johnson, 2009; De Jong et al., 2008).

Das von De Jong et al. (2008) verwendete binäre IRT-Modell erlaubt, dass verschiedene Items unterschiedliche Nützlichkeit bei der Messung von (extremen) Antworttendenzen haben und diese Nützlichkeit sich in verschiedenen Gruppen (z.B. verschiedenen Ländern) unterscheiden kann. Das heißt, Personen unterscheiden sich in ihrer Neigung Antworttendenzen zu verwenden und Items unterscheiden sich durch das Ausmaß, in welchem sie Antworttendenzen hervorrufen. Des Weiteren bietet dieses IRT-Modell die Möglichkeit, Berechnungen trotz starker Korrelationen zwischen den Items einer Skala zu ermöglichen und alle zu messenden Parameter gleichzeitig zu erfassen, um Fehler bei der Parameterschätzung zu vermeiden (De Jong et al., 2008).

Rost, Carstensen und von Davier (1999) sowie Austin, Deary und Egan (2005) haben IRT-Modelle (Mixed-Rasch-Modelle) dafür benutzt, um herauszufinden, ob latente Subgruppen von Personen (Klassen) Antworttendenzen in unterschiedlicher Weise zeigen bzw. Antwortskalen unterschiedlich nutzen und welchen Einfluss diese Klassen auf die Eindimensionalität einer Messung haben.

Bolt und Johnson (2009) sowie Bolt und Newton (2011) verwendeten in ihren Studien eine multidimensionale Erweiterung des Nominal-Response-Modells (Bock, 1972, zit. nach Bolt & Johnson, 2009) in dem Antworttendenzen als Eigenschaftsdimensionen dargestellt werden, die einen Einfluss auf die Wahl der Antwortkategorie haben. Insofern unterscheiden sie sich von den oben beschriebenen Untersuchungen (Rost et al., 1999; Austin et al., 2005), als dass sie von einem kontinuierlichen Faktor bei der Berechnung von Antworttendenzen ausgehen und damit eine Kontrolle der Effekte von Antworttendenzen beim Gesamtscore, sowie die Analyse Differential Item Functions (DIF) ermöglichen. DIF-Analysen beschäftigen sich mit der Frage, ob ein Item in verschiedenen Subgruppen/Populationen unterschiedliche Funktionsweisen hat (Böckenholt, 2012).

Johnson und Bolt (2010) verwenden in ihrer Untersuchung ebenfalls ein multidimensionales IRT-Modell, um individuelle Differenzen bei Antworttendenzen zu bestimmen.

Im folgenden Abschnitt wird genauer auf einen neuen IRT-Ansatz (Böckenholt, 2012; Khorramdel & von Davier, in Druck, von Davier & Khorramdeln, in Druck) zur

Identifizierung von Antworttendenzen eingegangen, der auch in dieser Diplomarbeit verwendet wurde.

5.2 Aktuellster IRT-Ansatz zur Messung von Antworttendenzen

Böckenholt (2012) stellte den aktuellsten Ansatz zur Messung von multiplen Antwortprozessen mit Hilfe von Items Response Theorie vor. Das vorgestellte *tree-structure framework for modeling multiple response processes* ermöglicht die vereinfachte Darstellung von beobachteten Antwortprozessen durch binäre Pseudoitems. In einem ersten Schritt werden die latenten Antwortprozesse durch eine multinominale Entscheidungsstruktur dargestellt (ähnlich einem Entscheidungsbaum), wobei jede einzelne Ebene einen Subprozess repräsentiert.

Böckenholt (2012) postuliert, dass zur Beantwortung von nominal oder ordinal skalierten Items multiple latente Antwortprozesse nötig sind, die zu den beobachteten Antworten führen. Durch die Zerlegung von Antwortprozessen in ein Set aus binären Pseudoitems können alle möglichen Antwortsubprozesse dargestellt werden. Ein möglicher multipler Antwortprozess könnte zum Beispiel so aussehen, dass sich eine Testperson bei einem ordinalen Item mit vier Abstufungen zuerst entscheidet, ob sie der Frage zustimmt oder sie ablehnt. Der zweite Subprozess könnte dann beschreiben, ob eine moderate oder eine extreme Zustimmung oder Ablehnung gewählt wird, usw. (vgl. Khorramdell & von Davier, in Druck).

Die Ergebnisse dieser Prozesse können jeweils mit Pseudoitems abgebildet und dann mit *simple structure* multidimensional IRT-Modellen (ein Item wird nur einem Faktor zugeordnet) untersucht werden (Böckenholt, 2012). Diese Modelle gehen davon aus, dass Antwortprozesse durch latente Variablen kontrolliert und durch die Differenz zwischen Personenfähigkeit und Itemschwierigkeit beschrieben werden (vgl. Khorramdel & von Davier, in Druck).

Böckholt (2012) verwendet das graded-response Modell von Samejima (1969, 1997, zit. nach Böckenholt, 2012) zur Analyse der der Ratingskalen.

Dem graded-response Modell liegt die Annahme zu Grunde, dass gegebene Antworten in Fragebögen mit Rating-Skalen einer normalverteilten, latenten Variable unterliegen, die wiederum durch eine Antwortskala abgebildet werden sollen. Eine Antwortkategorie

wird dann ausgewählt wenn der latente Wert zwischen die untere und obere Schwelle, die eine Antwortkategorie definiert, fällt:

$$\Pr(y_{ij} = k) = \Phi(\tau_{k+1} + \theta_i + \gamma_j) - \Phi(\tau_k - \theta_i - \gamma_j)$$

Dementsprechend kann die Wahrscheinlichkeit, mit der eine zufällig gewählte Person i eine Antwortkategorie k ($k = 1, \dots, K$) des Items j mit dem Schwellenwert τ_k auswählt, durch die Differenz zwischen zwei normal kumulativen Verteilungsfunktion dargestellt werden. y_{ij} bezeichnet die Lage des Items j auf dem latenten Kontinuum, beschrieben durch den Personenparameter θ (Böckenholt, 2012).

Zur Untersuchung der Tendenz zur Mitte stellt Böckenholt (2012) ein two-process graded-response-Modell vor, das den Grad untersucht, in dem unterschiedliche Items eine Testperson dazu anleitet, die mittlere Antwortkategorie auszuwählen. Dazu wird eine fünfstufige Rating-Skala (A, B, C, D, E) in zwei Pseudoitems zerlegt. Der erste untersuchte Subprozess beschreibt, ob die Testperson die mittlere Antwortkategorie auswählt oder nicht. Dementsprechend wird aus der ursprünglich fünfstufigen Antwortskala ein binäres Pseudoitem generiert. Der zweite Subprozess beschreibt, ob sich eine Testperson für eine zustimmende oder ablehnende Antwortkategorie entschieden hat. Die fünfstufige Antwortskala wird in ein vierstufiges ordinales Item ohne die Mittelkategorie umgewandelt, welches die Richtung und die Intensität der abgegebenen Selbstbeschreibung anzeigt.

Die Wahrscheinlichkeit, die mittlere Antwortkategorie auszuwählen, kann folgend beschrieben werden:

$$\Pr(y_{ij} = C) = \Phi(\theta_i^{(I)} - \gamma_j^{(I)})$$

Die verbleibenden vier Antwortkategorien werden in dieser Weise dargestellt:

$$\Pr(y_{ij} = A) = [1 - \Phi(\theta_i^{(I)} - \gamma_j^{(I)})]\Phi(\tau_B + \theta_i^{(II)} - \gamma_j^{(II)})$$

$$\Pr(y_{ij} = B) = [1 - \Phi(\theta_i^{(I)} - \gamma_j^{(I)})][\Phi(\tau_D + \theta_i^{(II)} - \gamma_j^{(II)}) - \Phi(\tau_B + \theta_i^{(II)} - \gamma_j^{(II)})]$$

$$\Pr(y_{ij} = D) = [1 - \Phi(\theta_i^{(I)} - \gamma_j^{(I)})][\Phi(\tau_E + \theta_i^{(II)} - \gamma_j^{(II)}) - \Phi(\tau_D + \theta_i^{(II)} - \gamma_j^{(II)})]$$

$$\Pr(y_{ij} = E) = [1 - \Phi(\theta_i^{(I)} - \gamma_j^{(I)})][1 - \Phi(\tau_E + \theta_i^{(II)} - \gamma_j^{(II)})]$$

Die Itemparameter der zwei Prozesse werden durch $\gamma_j^{(I)}$ und $\gamma_j^{(II)}$ beschrieben. $\theta_i^{(I)}$ und $\theta_i^{(II)}$ sind die jeweiligen Personenparameter.

Durch die Zerlegung von polytomen Items in Pseudoitems entsteht nach Böckenholt (2012) eine leicht interpretierbare Datenstruktur, die Aufschluss über latente Antwortprozesse gibt. Des Weiteren beschreibt er, dass Kovariaten (z.B.: Alter) in die Überprüfung von multiplen Antwortprozessen miteinbezogen werden können, die einen möglichen Effekt auf spezifische Subprozesse haben.

Khorramdel und von Davier (in Druck) haben den Ansatz von Böckenholt (2012) erweitert und untersuchten in ihrer Studie das Problem von Antworttendenzen in einem Fragebogen mit mehreren Persönlichkeitsdimensionen. Im Vergleich zu Böckenholt (2012), der die Antworten auf einen Fragebogen mit nur einer Persönlichkeitsskala und multidimensionalen Antworttendenzfaktoren untersuchte, haben Khorramdel und von Davier (in Druck) unidimensionale Antworttendenzfaktoren mit multidimensionalen Persönlichkeitsfaktoren verglichen. Die unidimensionale Messung von Antworttendenzen hat den Vorteil, dass die Ergebnisse detailliertere Informationen über die Dimensionalität der Antworttendenzen bieten und daher eine höhere Validität aufweisen und besser von Persönlichkeits- oder Merkmalsausprägungen unterschieden werden können. Des Weiteren wurden von Khorramdel und von Davier (in Druck) kulturelle Unterschiede und Möglichkeiten, Daten von Antworttendenzen zu korrigieren, untersucht. Die beobachteten Daten wurden in multiple binäre Antwortprozesse zerlegt, um danach mit multidimensionalen IRT-Modellen untersucht zu werden.

Für jedes beobachtete Item wurden von Khorramdel und von Davier (in Druck) drei Pseudoitems kodiert (Bild 1 stellt den zugehörigen multinominalen Prozess-Baum dar):

1. Pseudoitem e für die Tendenz zu extremen Antworten: Alle extrem positiven und extrem negativen Antworten wurden mit 1 kodiert. Bei der verwendeten fünfstufigen Antwortskala bedeutete das, dass jeweils 1 und 5 als extreme Antworten gezählt wurden. Die mittlere Antwortkategorie wurde mit 9 (fehlender Wert) kodiert und die übrigen mit 0.

2. Pseudoitem m für die Tendenz zur Mitte: Immer, wenn die mittlere Antwortkategorie (3) angekreuzt wurde, wurde das Pseudoitem mit 1 kodiert. Alle anderen Antworten wurden mit 0 kodiert.
3. Pseudoitem d für Antworten, die in Richtung der gemessenen Skala gegeben wurden: Immer wenn die angekreuzte Antwort eine Zustimmung zur gemessenen Skala beschrieb, wurde das Item mit 1 kodiert. Die mittlere Antwortkategorie wurde mit 9 (fehlender Wert) kodiert und die übrigen mit 0.

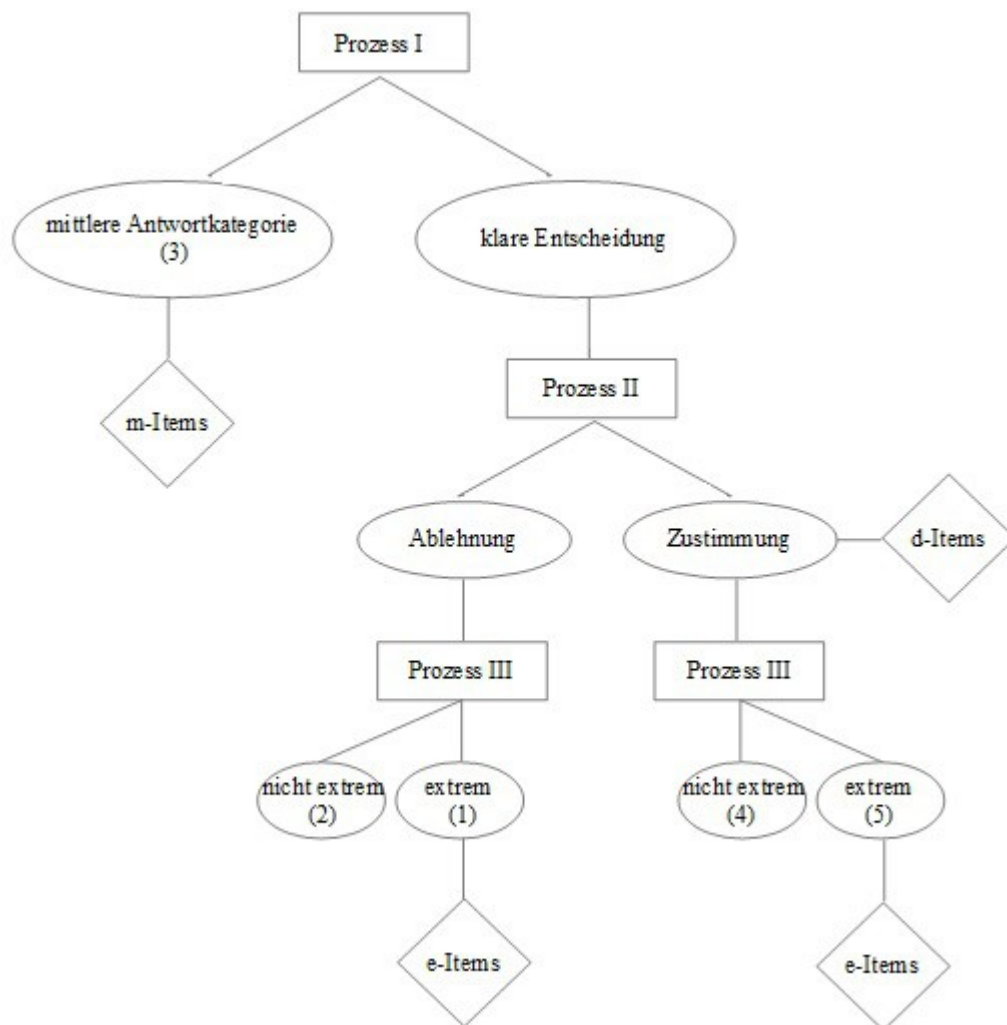


Abbildung 2: Multinomialer Prozess-Baum (Khorramdel & von Davier, in Druck, S. 17)

Die ursprünglich mittlere Antwortkategorie bei den Pseudoitems e und d wurde von Khorramdel und von Davier (in Druck) mit einem fehlenden Wert kodiert, um eine unvollständige Kontingenztafel zu erzeugen, in der eine Quasi-Unabhängigkeit

(Goodman, 1968; Fienberg, 1970) gehalten werden kann. Dieses Vorgehen hat zum Ziel, drei verschiedene latente Variablen mittels der Pseudoitems e , d und m zu untersuchen, ohne dabei die Abhängigkeiten zwischen den drei Pseudoitems (da es keine gibt) gesondert beachten zu müssen.

Khorramdel und von Davier (in Druck) veranschaulichen diese Quasi-Unabhängigkeit mit folgender Tabelle:

Tabelle 1: Veranschaulichung einer unvollständigen Kontingenztafel für die Wahrscheinlichkeiten je Kategorie einer fünfstufigen Antwortskala (Khorramdel & von Davier, in Druck, S. 17)

		Moderate Antwort	Extreme Antwort	Mittlere Antwort	
		0	1		
Negativ	0	P(1)	P(0)	0	P(p=0)
Positiv	1	P(3)	P(4)	0	P(p=1)
Mitte		0	0	P(2)	P(m=1)
		P(e=0)	P(e=1)	P(m=1)	

Die obige Tabelle stellt eine 3x3 Kontingenztafel für die Pseudoitems e und d dar, wobei die kodierten fehlenden Werte (also jeweils die ursprünglich mittleren Antwortkategorien) der e - und d -Items durch eine dritte Antwortkategorie m beschrieben werden. Betrachtet man nur die grau hinterlegten Zellen zeigt sich, dass sich hier ausschließlich *nonzero*-Wahrscheinlichkeiten ergeben. Das heißt, innerhalb dieser vier Zellen dürfte nach Khorramdel und von Davier (in Druck) für die e - und d -Items Unabhängigkeit, perfekte Abhängigkeit bzw. alles dazwischen gelten. Wichtig hierbei ist, dass diese Unabhängigkeit oder Abhängigkeit nicht von $P(2)$ (der Wahrscheinlichkeit, die mittlere Antwortkategorie zu wählen) beeinflusst wird, die einzige Beschränkung kann folgend dargestellt werden: $P(0)+P(1)+P(3)+P(4) = 1-P(2)$ (Khorramdel & von Davier, in Druck; von Davier & Khorramdel, in Druck).

Khorramdel und von Davier (in Druck) konnten mit ihrer Studie zeigen, dass dieser IRT-Ansatz verwendet werden kann, um zu bestätigen, dass Antworttendenzen (Tendenz zur Mitte, Tendenz zu extremen Antworten) in Daten vorhanden sind. Die Messung der extremen Antworttendenz konnte nach Ausscheiden eines einzelnen Items

Eindimensionalität erreichen, bei der Messung der Tendenz zu Mitte mussten zuvor 10 Items ausgeschieden werden. Des Weiteren zeigten sich kulturelle Unterschiede bei der extremen Antworttendenz. Außerdem konnte gezeigt werden, dass durch eine bestimmte Score-Verrechnung die Daten vom Einfluss der Antworttendenzen korrigiert werden können.

Aufbauend auf dieser Studie beschäftigte sich eine weitere Untersuchung von von Davier und Khorramdel (in Druck) mit dem Thema, dass Antworttendenzen mit dem oben erklärten erweiterten IRT-Ansatz (Böckenholt, 2012; Khorramdel & von Davier, in Druck) nicht unidimensional erfasst wurden, sondern die Annahme besteht, dass zusätzlich die Merkmalsausprägungen gemessen werden. Um dieses Problem zu beobachten, wurden zusätzlich zu dem oben beschriebenen Vorgehen Bifactor und second order IRT-Modelle verwendet. Im Vergleich zu *simple structure* MIRT-Modellen erlauben diese Modelle beobachteten Daten von mehr als einem Faktor abhängig zu sein.

Das Bifactor-Modell ist ein hierarchisches IRT-Modell für binäre Daten (Gibbons & Hedeker, 1992, zit. nach von Davier & Khorramdel, in Druck). Es wird postuliert, dass jedes Item eine Hauptdimension und eine von K spezifischen Dimensionen misst, wobei die Hauptdimension jene latente Variable ist, die von höchstem Interesse ist und zur Erfassung der Kovarianz zwischen allen Items herangezogen wird. Im Vergleich dazu sollen die spezifischen Dimensionen zusätzliche Abhängigkeiten unter ausgewählten Gruppen von Items erfassen. Das second order Modell (äquivalent mit den Testlet-Modell postuliert von Bradlow, Wainer & Wang, 1999; Wainer, Bradlow & Wang, 2007, zit. nach von Davier & Khorramdel, in Druck), misst zwar ebenfalls Hauptdimensionen und spezifische Dimensionen, allerdings geht man in diesem Modell davon aus, dass Items nicht direkt von der Hauptdimension abhängig sind, sondern von der jeweiligen spezifischen Dimension (von Davier & Khorramdel, in Druck). Für diese Studie ist vor allem das Bifactor-Modell relevant, daher folgt eine nähere Beschreibung.

Die Modellgleichung des Bifactor-Modells sieht wie folgt aus (von Davier & Khorramdel, in Druck):

$$P(y|\theta) = \prod_{i=1}^I P(y_{i(k)}|\theta_g, +\theta_k)$$

y beschreibt den Vektor aller binär verrechneten Antworten. $y_{i(k)}$ ist die Antwort auf Item i ($i=1, \dots, I$) in der Dimension k ($k=1, \dots, K$). θ_g beschreibt die latente Variable, die alle Items betrifft, wohingegen θ_k die dimensionsspezifischen Variablen beschreibt.

Angewendet auf die Pseudoitems e und m (also der beiden Antworttendenzen „Tendenz zu extremen Urteilen“ und „Tendenz zur Mitte“) bedeutet das, dass die Antworttendenz als Hauptdimension behandelt wird und die Skalen des entsprechenden Fragebogens die spezifischen Dimensionen darstellen, die zusätzliche Abhängigkeiten zwischen Itemgruppen erklären sollen.

Für das Modell wird angenommen, dass die latenten Variablen normalverteilt sind und die statistischen Unabhängigkeit zwischen allen Antworten die in einer bedingten Abhängigkeit zur Hauptdimension und den spezifischen Dimensionen stehen, gilt (von Davier & Khorramdel, in Druck).

Von Davier und Khorramdel (in Druck) konnten zeigen, dass durch das Bifactor-Modell eine bessere Modellpassung für die Messung der extremen Antworttendenz und die Messung der Tendenz zur Mitte erreicht werden kann, als durch *simple structure* IRT-Modelle. Die Ergebnisse der Studie zeigten, dass bei der Untersuchung der Tendenz zu extremen Urteilen nur die Antworttendenz gemessen wird, während die Messung der Tendenz zur Mitte sowohl Antworttendenz als auch Merkmalsausprägung zu messen scheint.

Die Artikel von Khorramdel und von Davier (in Druck), von Davier und Khorramdel (in Druck) sowie Böckenholt (2012) sind die Grundlage für die vorliegende Diplomarbeit. Im nächsten Kapitel wird auf die IRT-Modelle, die zur Auswertung des empirischen Teils dieser Diplomarbeit verwendet wurden, eingegangen.

6 Beschreibung des in der Diplomarbeit verwendeten IRT-Modells

Wie auch bei Khorramdel und von Davier (in Druck) werden zur Messung verschiedener, durch Pseudoitems dargestellter Antwortprozesse (Antworttendenzen, Antworten auf Grund von Merkmalsausprägungen) in dieser Diplomarbeit unidimensionale und multidimensionale IRT-Modelle geschätzt. Bei unidimensionalen IRT-Modellen wird angenommen, dass eine latente Variable ausreicht, um die Daten zu beschreiben, wohingegen man bei multidimensionalen IRT-Modellen davon ausgeht, dass mehrere latente Variablen zu Grunde liegen.

6.1 Das Zwei-Parameter-logistische Modell (2PL) von Birnbaum (1968)

Das 2PL-Modell (Birnbaum, 1968) ist eine Verallgemeinerung des Rasch-Modells. Es baut (wie das Rasch-Modell) auf einer logistischen Funktion auf, hat aber einen Itemparameter mehr. Der zusätzliche Itemparameter wird auch als Diskriminationsparameter bezeichnet, da Items zusätzlich zur Schwierigkeit auch hinsichtlich ihrer Trennschärfe (dem Anstieg der Itemfunktion) untersucht werden sollen. Die Konsequenz daraus besteht darin, dass der Score, den eine Testperson in einem Verfahren erreicht, nicht mehr ausreicht, um den individuellen Grad der Merkmalsausprägung zu beschreiben, sondern jede Itemantwort zuvor mit dem jeweiligen Trennschärfeparameter gewichtet werden müsste (Rost, 2004).

Die Modellgleichung des 2PL-Modells sieht folgendermaßen aus:

$$P(x|\theta_v, \beta_i, \alpha_i) = \frac{\exp(\alpha_i(\theta_v - \beta_i))}{1 + \exp(\alpha_i(\theta_v - \beta_i))}$$

Der Itemdiskriminationsparameter α_i beschreibt den Anstieg der Itemfunktion. Unterschiedliche Itemtrennschärfen führen dazu, dass sich Itemfunktionen überschneiden können. Das bedeutet, zwei Items können für verschiedene Personen eine unterschiedliche Reihenfolge ihrer Lösungswahrscheinlichkeiten aufweisen (Rost, 2004).

6.2 Erweiterung des 2PL-Modells

Das 2PL-Modell kann auch für Items, die multiplen Skalen angehören, berechnet werden. Es wird angenommen (von Davier, Rost & Carstensen, 2007), dass multidimensionaler IRT-Modelle unter der zusätzlichen Bedingung gelten, dass für jedes Set von voneinander unterscheidbaren Itemsubsets (Skalen) verschiedene Personenparameter hinzugefügt werden. Hierbei ist zu unterscheiden, ob zur Lösung eines Items mehr als eine Fähigkeitskomponente (Fertigkeit) benötigt wird (*within-item multidimensionality*) oder ob eine Fähigkeitskomponente ausreicht (*between-item multidimensionality*) (von Davier, Rost & Carstensen, 2007).

Für diese Diplomarbeit ist wie bei Khorramdel und von Davier (in Druck) die *between-item multidimensionality* wichtig, das heißt, die Lösung von Items wird nur durch eine bestimmte Fertigkeit bestimmt; gehören diese Items allerdings verschiedenen Item-Subsets (z. B. verschiedene Skalen) an, kann diese Fertigkeit für jedes Subset eine andere sein. Jedes Item lädt also nur auf eine einzige Skala. Im 2PL-Modell kann die Wahrscheinlichkeit, dass die Testperson v die Antwort x auf Item i ($x=1, \dots, m_i$) der Skala k wählt, folgendermaßen beschrieben werden (Khorramdel & von Davier, in Druck):

$$P(x|\theta_v, \beta_i, \sigma_i) = \frac{\exp(\sigma_i(x\theta_{vk} - \beta_{ix}))}{1 + \sum_{y=1}^{m_i} \exp(\sigma_i(y\theta_{vk} + \beta_{iy}))}$$

Des Weiteren können auch multiple Populationen simultan in die Berechnung des 2PL-Modells inkludiert werden (*multigroup model*). Die IRT-Modelle gelten in diesem Fall unter der Bedingung, dass für unterschiedliche Populationen unterschiedliche Sets von Parametern verwendet werden ($c=1, \dots, C$) (von Davier, Rost & Carstensen, 2007). Die Gleichung würde sich für das 2PL-Modell nach Khorramdel und von Davier (in Druck) folgend verändern:

$$P(x|\theta_v, \beta_i, \sigma_i, c) = \frac{\exp(\sigma_{ixc}(x\theta_{vk} - \beta_{ixc}))}{1 + \sum_{y=1}^{m_i} \exp(\sigma_{iyc}(y\theta_{vk} + \beta_{iyc}))}$$

Der nächste Abschnitt befasst sich mit den Methoden und der Durchführung der Studie sowie der Darstellung der Ergebnisse.

7 Empirischer Teil

Die Ziele dieser Diplomarbeit sind eine weitere Untersuchung von Antworttendenzen mittels des von Böckenholt (2012) und Khorramdel und von Davier (in Druck) vorgestellten IRT-Ansatzes, sowie mittels des Bifactor-Modells (vgl. von Davier & Khorramdel, in Druck). Im Falle, dass Antworttendenzen in den Daten beobachtbar sind, soll der Einfluss des Fragebogeninhalts auf das Auftreten von Antworttendenzen untersucht werden.

Im Vergleich zu Khorramdel und von Davier (in Druck) wurde zur Untersuchung des Einflusses des Fragebogeninhalts in diesem Fall ein bestehender Fragebogen ausgewählt, der nicht fünf sondern neun faktorenanalytisch extrahierte Skalen aufweist, die wiederum zu vier Primärskalen zusammengefasst werden können. Es soll also zum einen beobachtet werden, ob sich Unterschiede im Auftreten von Antworttendenzen in den verschiedenen Skalen zeigen und zum anderen wurden für eine Primärskala, die sich mit dem Thema „*Coping* in Stresssituationen“ beschäftigt, zusätzliche Fragen auf Grundlage der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) gestellt. Die Personen wurden anhand der erreichten Scores in Gruppen eingeteilt, welche wiederum auf Unterschiede untersucht wurden.

8 Methode

8.1 Datenerhebung

Die Daten wurden im Zeitraum September – November 2012 erhoben. Um eine möglichst große Menge an Personen zu erreichen und gleichzeitig etwas anzubieten, dass die Motivation zur Durchführung steigert, wurden die Daten in Form einer Online-Self-Assessments (OSA) erhoben. Auf Grund der Möglichkeit einer Rückmeldung für die Testpersonen anzuzeigen, wurde folgende kostenpflichtige Plattform gewählt, um den Fragebogen online anzubieten: <https://www.onlineumfragen.com/> Nachdem die

Testpersonen die Fragebögen vollständig ausgefüllt hatten, erhielten sie eine Rückmeldung über die Ergebnisse und den von ihnen erreichten Gesamtscore (siehe Anhang). Bei einem vorzeitigen Abbruch wurde keine Rückmeldung für die bisherigen Ergebnisse angezeigt. Da einer der beiden vorgegebenen Fragebögen das Thema Stress behandelt, wurde bei der Bewerbung des OSAs (siehe Anhang) sowie der Rückmeldung auf das Thema Stress, Stressreduktion und Umgang mit Stress eingegangen.

Das OSA wurde hauptsächlich über Facebook verbreitet. Zum einen wurden Personen gebeten, sofern sie die Rückmeldung als hilfreich erachteten, den Link auf ihrer Facebookseite zu teilen, und somit unter all ihren Facebook-Freunden Werbung dafür zu machen. Außerdem wurde der Link auf verschiedene Hochschuleiten auf Facebook bekannt gemacht. Des Weiteren wurde das OSA in Lehrveranstaltungen des Studiengangs Psychologie beworben. Personen konnten hier bei Interesse ihre E-Mail-Adresse in eine Liste eintragen und erhielten dann den Link zum OSA via E-Mail.

8.2 Stichprobe

Die Daten stammen von 635 Personen, die online an einem Self-Assessment zum Thema „Stress“ teilgenommen haben. Insgesamt haben 900 Personen das Online-Self-Assessment (OSA) zumindest aufgerufen, 636 haben den ersten Fragebogen des OSA vollständig bearbeitet. Eine Person musste zusätzlich ausgeschlossen werden, da das Mindestalter auf 16 Jahre festgelegt wurde und die Person 14 war. Die Testpersonen beantworteten zuerst das Differentielle Stressinventar HR (Lefevre & Kubinger, 2004), die Zusatzfragen nach der Theorie des geplanten Handelns und danach das NEO Fünf-Faktoren Inventar (NEO-FFI, Brokenau & Ostendorf, 1993).

An dem OSA nahmen wesentlich mehr Frauen (458, 72%) als Männer (177, 28%) teil. 530 (83%) Personen stammten aus Österreich, 84 (13%) aus Deutschland, 2 (0,3%) aus der Schweiz und 19 (3%) gaben an, eine andere Staatsangehörigkeit zu haben. Eine Person (0,16%) gab an, keinen Schulabschluss zu haben, 19 (3%) hatten die Pflichtschule abgeschlossen, 64 (10%) absolvierten eine Lehre, 351 (55%) gaben an, die Matura zu haben und 200 (31%) einen Hochschulabschluss. Davon gaben 369 (58%) an, derzeit zu studieren. 31 (5%) von den 635 Personen gaben an, eine andere

Muttersprache als Deutsch zu sprechen. Der Altersmittelwert lag bei 29,6 Jahren, wobei die älteste Person 65 und die jüngste 16 Jahre war.

8.3 Erhebungsinstrument

Die Testpersonen bearbeiteten die Items des Differentiellen Stressinventars HR (DSIHR; Lefevre & Kubinger, 2004), sowie jene des NEO-FFI (Brokenau & Ostendorf, 1993) und zusätzlich gestellte Fragen nach der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) zu einer Subskala des DSIHR. Alle drei Verfahren wurden im selben OSA bearbeitet. Beim DSIHR wurden die Fragebogenskalen mit einem verbalen, fünfstufigen Antwortformat in einer Matrix vorgegeben: *trifft nicht zu* - *trifft eher zu* - *weder noch* - *trifft eher zu* - *trifft zu*. Für die Zusatzfragen wurden die Items der entsprechenden Skala noch einmal in einer Matrix dargestellt, wobei hier die vier unter Punkt 8.5 dargestellten Aussagen als Antwortskala dienten und jeweils durch das Anklicken einer Checkbox bestätigt werden konnten. Das NEO-FFI wurde mit einer figuralen, fünfstufigen Antwortskala ebenfalls in einer Matrix vorgegeben. Für diese empirische Arbeit sind nur die Daten des DSIHR sowie die Zusatzfragen von Bedeutung, deshalb wird auf diese näher eingegangen.

8.4 Differentielles Stressinventar HR

Das DSIHR dient zur Erfassung des individuellen Umgangs mit Stress und behandelt folgende Themen: *StressAuslösung*, *StressManifestation*, *Copingstrategien* sowie Risiken der *StressStabilisierung*. Es beruht auf dem Konzept der Leistungsangstdiagnostik von Rost und Schermer (1987).

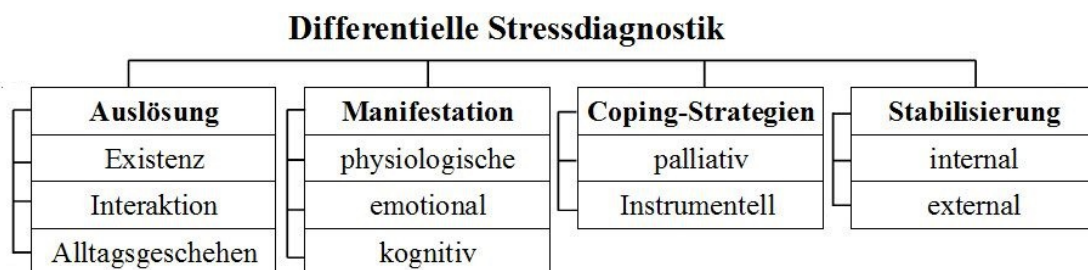


Abbildung 3: Schematische Darstellung der Struktur des DSIHR nach Lefevre und Kubinger, 2004

Das DSIHR besteht aus 9 faktorenanalytisch gewonnenen Dimensionen, die wiederum zu den vier Primärskalen (*StressAuslösung*, *StressManifestation*, *Copingstrategien* und

Stress*Stabilisierung*) zusammengefasst werden. Abbildung 3 zeigt die theoretische Faktorenstruktur. Sie konnte weitgehend faktorenanalytisch belegt werden, nur die beiden theoretischen Subskalen „emotional“ und „kognitiv“ konnten nicht als einzelne Faktoren belegt werden, sondern stellten sich als ein Faktor da. Alle Items des Fragebogens sind in die selbe Richtung formuliert.

Das DSIHR unterscheidet sich vom ursprünglichen Differentiellen Stressinventar (DSI) dadurch, dass vier Items, die sich mit dem Thema Sex im Zusammenhang mit Stress beschäftigen, ausgeschlossen wurden.

8.4.1 Kennwerte des Fragebogens

Auch wenn die Kennwerte aus dem Manual auf die hier verwendete Online-Version, die neben der anderen Darstellung auch fünf statt vier Antwortmöglichkeiten bietet, nicht für das hier angewandte Erhebungsinstrument gelten können, sollen sie kurz dargestellt werden, um die Verwendung des Verfahrens zu begründen.

Da die 9 Skalen durch eine konfirmatorische Faktorenanalyse belegt sind, kann von einer Konstruktvalidität im Sinne der klassischen Testtheorie ausgegangen werden. Die innere Konsistenz ist mit Werten zwischen $r = ,70$ – $r = ,94$ (Cronbach-Alpha) als hoch anzusehen*. Die Normen für das Verfahren wurden in den Jahren 2003-2004 an 378 ÖsterreicherInnen erhoben.

8.4.2 Skalen des DSIHR

Stressauslösung (49 Items): beschreibt vorhandene Auslöser für Stress.

- *Alltagsgeschehen (23 Items):* Diese Variable beschreibt das Ausmaß an Stressauslösern durch unvermeidliche Angelegenheiten des Alltags.
- *Interaktion (16 Items):* Beschreibt das Ausmaß an Belastung durch Interaktion mit Mitmenschen im privaten und beruflichen Bereich.

* Auch für die im Rahmen der Diplomarbeit erhobenen Daten zeigte sich eine hohe innere Konsistenz (Cronbach-Alpha) für die Primärskalen des Fragebogens (Auslösung: $r = ,90$; Manifestation: $r = ,80$; Coping: $r = ,78$ und Stabilisierung: $r = ,78$).

- *Existenzängste (10 Items)*: Diese Variable beschreibt das Ausmaß an Belastung durch Zukunftssorgen und Existenzängste.

Stressmanifestation (20 Items): Die *Stressmanifestation* beschreibt die Häufigkeit von auftretenden StressManifestationen.

- *Physische Manifestation (10 Items)*: Beschreibt die Stärke der körperlichen Beanspruchung durch Stress.
- *Emotional-kognitive Manifestationen (10 Items)*: Beschreibt die Stärke der emotional-kognitiven Belastung durch Stress.

Coping (31 Items): Die Skala *Coping* misst die tatsächlich eingesetzten *Coping*-Strategien einer Person.

- *palliatives Coping (17 Items)*: Die Skala „*palliatives Coping*“ misst, welche positiven Emotionen und Kognitionen zur Bewältigung der Beanspruchung verfügbar sind.
- *instrumentelles Coping (14 Items)*: Die Skala „*instrumentelles Coping*“ misst aktive Vorgehensweisen zur Bewältigung der Beanspruchung durch Stresssituationen.

Stressstabilisierung (20 Items): Misst die Stabilisierungsfaktoren, die zu einer Chronifizierung des subjektiven Stresserlebens führen können.

- *Externale StressStabilisierung (8 Items)*: Misst externale Verstärker, die zu einer Chronifizierung des Stresserlebens führen können.
- *Internale StressStabilisierung (12 Items)*: Misst internale Verstärker, die zu einer Chronifizierung des Stresserlebens führen können.

Items des DSIHR dürfen in der Diplomarbeit nicht dargestellt werden, deshalb wird hier zur Ansicht der genauen Formulierung der Items auf das Manual des DSIHR (Lefevre & Kubinger, 2004) verwiesen.

8.5 Zusatzfragen nach der Theorie des geplanten Handelns

Jedes Item der Skala *Coping* wurde mit neuen Antwortmöglichkeiten die Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) betreffend vorgegeben. In der Skala

Coping werden Verhaltensweisen zur Stressreduktion abgefragt, wobei die Skala „*instrumentelles Coping*“ beobachtbare Handlungen beinhaltet. Auch wenn nicht alle Items planbare Handlungen beschreiben, wurden die zusätzlichen Fragen für alle gestellt. Die zusätzlichen Antwortmöglichkeiten beziehen sich vereinfacht auf die drei Faktoren der Handlungsintention, wie im Theorieteil dargestellt. Sie wurden folgendermaßen formuliert:

Einstellung gegenüber dem Verhalten: Ich habe diese Handlung in der letzten Stresssituation ausgeführt.

Diese Frage soll die persönliche Einstellung zu dem im DSIHR angesprochen Verhalten erfassen. Sowohl Davidson und Jacquard (1979) als auch Ajzen (1991) beschrieben, dass das Verhalten möglichst spezifisch abgefragt werden muss, um das später tatsächlich ausgeführte Verhalten vorhersagen zu können. Da die Fragen im DSIHR im Sinne der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) nicht sehr spezifisch gestellt sind, wurde danach gefragt, ob das Verhalten schon einmal gezeigt wurde, um eine genaue Beschreibung der abgefragten Situation zu umgehen.

Subjektive Norm: Diese Handlung wird von meinen Mitmenschen akzeptiert.

Wahrgenommene Verhaltenskontrolle: Die Ausübung dieser Handlung fällt mir leicht.

Zusätzlich gab es die Antwortmöglichkeit: Keine dieser Aussagen trifft zu. Abbildung 4 veranschaulicht die Vorgabe.

Es konnte maximal drei Antwortmöglichkeiten ausgewählt werden. Wurde die letzte Antwortmöglichkeit gewählt, konnte sonst keine mehr angekreuzt werden.

	<i>Ich habe dieses Verhalten in der letzten Stresssituation ausgeführt.</i>	<i>Dieses Verhalten wird von meinen Mitmenschen akzeptiert.</i>	<i>Die Ausübung dieses Verhaltens fällt mir leicht.</i>	<i>Keine dieser Aussagen trifft zu.</i>
<i>1. Beispielitem der Skala Coping.</i>	☐	☐	☐	☐

Abbildung 4: Darstellung des Vorgabeformats der zusätzlichen Antwortmöglichkeiten für die Primärskala Coping

8.6 Datenaufbereitung

Die Daten wurden in gleicher Weise wie im Artikel von Khorramdel und von Davier (in Druck) kodiert (siehe Kapitel 5.2). Für jedes Originalitem wurden (wie in Tabelle 2 dargestellt) drei binäre Pseudoitems kodiert:

Tabelle 2: Kodierung der Pseudoitems

Pseudoitem	Verrechnung	1 (trifft nicht zu)	2 (trifft eher nicht zu)	3 (weder noch)	4 (trifft eher zu)	5 (trifft zu)
d (Zustimmung zur Skala)		0	0	9	0	0
e (extreme Antworttendenz)		1	0	9	0	1
m(mittlere Antworttendenz)		0	0	1	0	0

Die mittlere Antwortkategorie bei den *d* und *e* Pseudoitems wurden wegen der im Theorieteil beschriebenen Quasi-Unabhängigkeit mit 9 (fehlender Wert) kodiert. Die Pseudoitems *e* und *m* sollen die formalen Antworttendenzen „Tendenz zu extremen Urteil“ und „Tendenz zur mittleren Antwortkategorie“ darstellen, wohingegen das Pseudoitem *d* die Zustimmung zur Skala beschreibt und wenig durch Antworttendenzen beeinflusst sein soll. Statt den ursprünglichen 120 Originalitems entstanden durch die Umkodierung 360 Pseudoitems.

8.7 Spezielle verwendete Software zur Auswertung

Die drei binäre Pseudoitems (*e*, *m* und *d*) wurden mit uni- und multidimensionalen IRT-Modellen untersucht. Diese wurden mit Hilfe des Mixture General Diagnostic Modeling Framework (MGDM; von Davier, 2010) mit der Software MDLTM (von Davier, 2011) geschätzt. Die Software ist für die Analyse von large-scale Daten geeignet und bietet viele Möglichkeiten latente Strukturmodelle zu untersuchen (Hee Seo, Xu & von Davier, 2007 & 2009). Es können multidimensionale IRT Modelle, basierend auf dem Rasch-Modell sowie dem 2PL-Modell und Modellen, die auf dem generalisierten Partial-Credit-Modell basieren, geschätzt werden (Khorramdel & von Davier, In Druck). Das MGDM Framework erlaubt die Spezifikation von diskreten *mixture*-Modellen mit einer hierarchischen Komponente. Die Familie der inkludierenden Modelle kann folgend definiert werden (Khorramdel & von Davier, in Druck):

$$P(X=x|\beta_i, a, q_i, \gamma_i, c) = \frac{\exp(\beta_{ixc} + \sum_{k=1}^K x \gamma_{ikc} h(q_{ik} a_k))}{1 + \sum_{y=1}^{m_i} \exp(\beta_{iyc} + \sum_{k=1}^K y \gamma_{ikc} h(q_{ik} a_k))}$$

$c = 1, \dots, C$ Klassen, die mehrere beobachtete Populationen darstellen können
(z.B. ethnische Gruppen).

K Anzahl der Merkmale (oder Skills)

$a = (a_1, \dots, a_k)$ latente Merkmalsmuster oder Skillprofil

(q_{ik}) binäre Q-Matrix, die Merkmalsfaktoren und Items darstellt

β_{ixc} klassenspezifische Itemschwierigkeitsparameter

$\gamma_{ikcx} = x \gamma_{ikc}$ Steigungsparameter, die Skills und Items verbinden

$h(q_{ik}, a_k)$ Funktion, die Skills und Q-Matrix für das Item abbildet

9 Hypothesen

Die grundlegende Annahme dieser Diplomarbeit ist, wie auch bei Khorramdel und von Davier (in Druck), dass die Korrelationen zwischen den Fragebogenskalen basierend auf den Pseudoitems e und m hoch sind, da sie eher Antworttendenzen als das intendierte Konstrukt messen. Korrelationen zwischen den Fragebogenskalen basierend auf den Pseudoitems d sollen dagegen eher niedrig sein, weil angenommen wird, dass die Dimensionen des Fragebogens durch diese (fast) unbeeinflusst von Antworttendenzen dargestellt werden können. Das heißt, es wird die Hypothese aufgestellt, dass die Pseudoitems e und m über alle Items hinweg Antworttendenzen messen und die Pseudoitems d die Fragebogenskalen abbilden. Dies wird sowohl mit multidimensionalen als auch mit unidimensionalen IRT-Modellen untersucht.

Sollten keine Antworttendenzen in den Daten vorhanden sein (sollten also die jeweiligen Antworttendenzfaktoren die Daten nicht besser beschreiben als die intendierten Fragebogenskalen), wird wie bei von Davier und Khorramdel (in Druck) überprüft, ob ein Bifactor-Modell besser auf die Daten passt. Das heißt, der Antworttendenzfaktor beschreibt im Bifactor-Modell den Generalfaktor und die Primärskalen werden als spezifische Faktoren angesehen. Damit soll untersucht werden,

ob die gegebenen Antworten eine Mischung aus Antworttendenzen und den eigentlichen Fragebogendimensionen darstellen, da die Möglichkeit besteht, dass einige Testpersonen gemäß dem Fragebogeninhalt geantwortet haben, während andere Antworttendenzen zeigten.

Im Fall, dass basierend auf den *e*- und *m*- Pseudoitems Antworttendenzen in den Daten nachgewiesen werden können, wird angenommen, dass fragebogeninhaltsabhängige Unterschiede zwischen den einzelnen Fragebogenskalen vorhanden sind. Des Weiteren werden Gruppenunterschiede basierend auf den zusätzliche gestellten Fragen bezüglich der Theorie des geplanten Handelns betrachtet. Es wird angenommen, dass wenn die Verhaltensintention (Score der Zusatzfragen) für das abgefragte Verhalten hoch ist, weniger Antworttendenzen in den Daten auftreten.

10 Analysen und Ergebnisse

Um diese Annahmen zu untersuchen, wurden die folgenden Analysen durchgeführt, wobei die Skalenzugehörigkeit der Items als fixer Faktor gilt. Alle Analysen wurden auf Grundlage des 2PL-Modells geschätzt.

10.1 Analysen der Gesamtdaten

Zur Analyse, ob Antworttendenzen in den Daten existieren, wurden wie bei Khorramdel und Davier (in Druck) multidimensionale Antworttendenz- sowie Merkmalsfaktoren modelliert.

- Ein 3-dimensionales Modell (mit den Pseudoitems als Faktoren) wurde mit einem 4-dimensionalen Modell (mit den Primärskalen des DSIHR als Faktoren) sowie einem 6-dimensionalen Modell (*d*-Pseudoitems werden den vier Skalen und die *e*- und *m*- Pseudoitems dem jeweiligen Antworttendenzfaktor zugeordnet), verglichen. Eine exemplarische Darstellung der Itemzuordnung für die 3-, 4- und 6-dimensionalen Modelle findet sich in den Tabellen 3-5.

Tabelle 3: Beispiel einer Datenmatrix für ein 3-dimensionales IRT-Modell im Fall von vier Fragebogenitems, die in 12 Pseudoitems zerlegt wurden

3-dimensionales IRT-Modell			
Item	Faktor für die extreme Antworttendenz	Faktor für die mittlere Antworttendenz	Faktor für die Zustimmung zur Skala
e-1	1	0	0
e-2	1	0	0
e-3	1	0	0
e-4	1	0	0
m-1	0	1	0
m-2	0	1	0
m-3	0	1	0
m-4	0	1	0
d-1	0	0	1
d-2	0	0	1
d-3	0	0	1
d-4	0	0	1

Tabelle 4: Beispiel einer Datenmatrix für ein 4-dimensionales IRT-Modell im Fall von vier Fragebogenitems, die in 12 Pseudoitems zerlegt wurden.

4-dimensionales IRT-Modell				
Item	Auslösung	Manifestation	Coping	Stabilisierung
e-1	1	0	0	0
e-2	0	1	0	0
e-3	0	0	1	0
e-4	0	0	0	1
m-1	1	0	0	0
m-2	0	1	0	0
m-3	0	0	1	0
m-4	0	0	0	1
d-1	1	0	0	0
d-2	0	1	0	0
d-3	0	0	1	0
d-4	0	0	0	1

- Ein 3-dimensionales Modell (mit den Pseudoitems als Faktoren) wurde mit einem 9-dimensionalen Modell (mit den Subskalen des DSIHR als Faktoren) sowie einem 11-dimensionalen Modell (*d*-Pseudoitems werden den 9 Subskalen und *e*- und *m*-Pseudoitems dem jeweiligen Antworttendenzfaktor zugeordnet) verglichen.

Tabelle 5: Beispiel einer Datenmatrix für ein 6-dimensionales IRT-Modell im Fall von vier Fragebogenitems, die in 12 Pseudoitems zerlegt wurden.

6-dimensionales IRT-Modell						
Item	extreme Antwort- t.Faktor	mittlerer Antwort- t.Faktor	Auslösung	Manifestation	Coping	Stabilisierung
e-1	1	0	0	0	0	0
e-2	1	0	0	0	0	0
e-3	1	0	0	0	0	0
e-4	1		0	0	0	0
m-1	0	1	0	0	0	0
m-2	0	1	0	0	0	0
m-3	0	1	0	0	0	0
m-4	0	1	0	0	0	0
d-1	0	0	1	0	0	0
d-2	0	0	0	1	0	0
d-3	0	0	0	0	1	0
d-4	0	0	0	0	0	1

Die Berechnung der 9- bzw. 11-dimensionalen Modelle stellte allerdings sehr hohe Anforderungen an die Rechenleistung der zur Verfügung stehenden PCs. Die Modelle konnten deshalb nicht mit der nötigen Genauigkeit berechnet werden (die Modelleinstellungen in der Software MDLTM mussten so getroffen werden, dass die Berechnung solcher komplexen Modelle möglich war: maxlevel=2). Die Ergebnisse werden angeführt, müssen aber mit Vorsicht interpretiert werden. Um genauere Aussagen machen zu können, wurden deshalb für jede Primärskala 3-dimensionale Modelle (mit den Pseudoitems als Faktoren) berechnet und je nach Anzahl der Subskalen 2- bzw. 3-dimensionale Modelle (mit den Subskalen als Faktoren) und 4- bzw. 5-dimensionale Modelle (*d*-Pseudoitems werden den jeweiligen Subskalen und *e*- und *m*-Pseudoitems dem jeweiligen Antworttendenzfaktor zugeordnet) geschätzt und danach verglichen.

10.2 Analysen getrennt für alle Pseudoitems

Wie bei Khorramdel und von Davier (in Druck) werden beim Vorhandensein von Antworttendenzen in den Daten ebenso unidimensionale Antworttendenzfaktoren und multidimensionale Merkmalsfaktoren untersucht. Es wird angenommen, dass die Pseudoitems *e* und *m* die Antworttendenzen messen und Pseudoitem *d* die Fragebogenskalen.

Ein 1-dimensionales Modell (je nach Pseudoitems: extreme Antworttendenz als Faktor, mittlere Antworttendenz als Faktor oder Zustimmung zur Skala als Faktor) wird mit einem 4-dimensionalen Modell (mit den übergeordneten Skalen des DSIHR als Faktoren) verglichen. Ein 1-dimensionales Modell (je nach Pseudoitems: extreme Antworttendenz als Faktor, mittlere Antworttendenz als Faktor oder Zustimmung zur Skala als Faktor) wurde danach auch noch einem 9-dimensionalen (mit den Subskalen als Faktoren) gegenübergestellt. Wie oben bereits beschrieben, konnten auch hier die 9-dimensionalen Modelle auf Grund ihrer Komplexität nicht mit ausreichender Genauigkeit berechnet werden, weshalb zusätzlich das 1-dimensionale Modell pro Primärskala berechnet und dann den Modellen mit den jeweiligen Subskalen als Faktoren gegenübergestellt wurde.

10.3 Bifactor-Modell

Bifactor-Modelle wurden für die IRT-Modelle mit eindimensionalen Antworttendenzfaktoren (Pseudoitems der Typen m und e) berechnet. Dadurch können Daten sowohl auf einen Generalfaktor (Antworttendenz), als auch auf spezifische Faktoren (Fragebogenskalen) laden. Jedes Item wurde dementsprechend kodiert. Tabelle 6 veranschaulicht die Itemkodierung.

Tabelle 6: Beispiel einer Datenmatrix für ein Bifactor-IRT-Modell im Fall von vier Fragebogenitems mit der ursprünglichen Kodierung, die in 8 Pseudoitems zerlegt wurden.

6-dimensionales IRT-Modell					
Item	Auslösgun	Maifestation	Coping	Stabilisierung	Antworttendenzfaktor
e-1	1	0	0	0	1
e-2	0	1	0	0	1
e-3	0	0	1	0	1
e-4	0	0	0	1	1
m-1	1	0	0	0	1
m-2	0	1	0	0	1
m-3	0	0	1	0	1
m-4	0	0	0	1	1

4-dimensionale Modelle getrennt für die Pseudoitems m und e mit den Primärskalen als Faktoren wurden mit Bifactor-Modellen verglichen, in denen die Items zusätzlich zu dem Antworttendenzfaktor auf die Primärskalen luden. Dasselbe Vorgehen wurde auch mit den Subskalen pro Primärskala durchgeführt. Es wurden also 2 bzw. 3-dimensionale Modelle (je nach Anzahl der Subskalen pro Primärskala) getrennt für die Pseudoitems e

und m berechnet und anschließend mit einem Bifactor-Modell verglichen, in dem jedes Item zusätzlich zum Antworttendenzfaktor auf die Subskalen luden.

10.4 Analysen bezüglich des Fragebogeninhalts

In einem ersten Schritt sollte dargestellt werden, ob in einzelnen Primärfragebogenskalen Antworttendenzen nachgewiesen werden konnten (d. h., dass jene Modelle mit den Antworttendenzen als Faktoren getrennt für die Pseudoitems des Typs e und m besser passten als jene Modelle, mit den Fragebogenskalen als Faktoren) und in anderen nicht, um mögliche Hypothesen über den Fragebogeninhalt in Bezug auf das Auftreten von Antworttendenzen ableiten zu können. Des Weiteren wurden jene Primärskalen untersucht, in denen Antworttendenzen nachgewiesen werden konnten, indem die auf Basis des Modells geschätzten Mittelwerte für die Pseudoitems e und m verglichen wurden. Dadurch sollte festgestellt werden, in welcher Skala die meisten m - bzw. e -Antworten gegeben wurden. In Tabelle 6 wurde eine Kategorisierung der Iteminhalte nach Angleitner, John und Löhr (1986) auf den hier relevanten Fragebogen angewandt, um eine weitere Einteilung neben den fragebogenspezifischen Inhalten zu erhalten.

Bezüglich der Itemformulierung (eine genaue Beschreibung dazu findet sich unter Punkt 4.2.1) kann das DSIHR folgendermaßen eingeteilt werden: Die Items des DSIHR sind biographiebezogen, personalisiert und konkret formuliert. Das intendierte Konstrukt wird indirekt und mit einer eher höheren Stimulusintensität abgefragt.

Auf der Ebene der Primärskalen wurden ebenfalls geschätzte Mittelwerte der e - und m -Pseudoitems für jene Primärskalen, in denen Antworttendenzen nachgewiesen werden konnten, auf Unterschiede in der Häufigkeit der e - und m -Antworten untersucht. Hier wurde ebenfalls zusätzlich zu den spezifischen Skaleninhalten auf Tabelle 6 Bezug genommen.

Tabelle 7: Kategorisierung des Iteminhalts der Subskalen des DSIHR nach den Kategorien von Angleitner, John und Löhr (1986).

	Skala	Kategorisierung nach Art des Iteminhalts
<i>Auslösung</i>	<i>Alltagsgeschehen</i>	Modifizierte Eigenschaftszuschreibungen
	<i>Interaktion</i>	Modifizierte Eigenschaftszuschreibungen
	<i>Existenzängste</i>	Modifizierte Eigenschaftszuschreibungen
<i>Manifestation</i>	<i>Physische Manifestation</i>	Fragen zur Selbstbeschreibung: Symptome
	<i>Emotional-kognitive Manifestation</i>	Fragen zur Selbstbeschreibung: beobachtbare und nicht beobachtbare Reaktionen (drei Items könnten auch als Symptome interpretiert werden)
<i>Coping</i>	<i>palliatives Coping</i>	Fragen zur Selbstbeschreibung: nicht beobachtbare Reaktionen (zwei Items könnten auch als beobachtbare Reaktion interpretiert werden)
	<i>instrumentelles Coping</i>	Fragen zur Selbstbeschreibung: beobachtbare Reaktionen (ein Item kann auch als nicht beobachtbare Reaktion interpretiert werden)
<i>Stabilisierung</i>	<i>Externale StressStabilisierung</i>	Fragen zur Selbstbeschreibung: nicht beobachtbare Reaktionen
	<i>Internale StressStabilisierung</i>	Einstellungen und Überzeugungen, Reaktionen anderer gegenüber einer Person

Des Weiteren wurde ein 1-dimensionales Multigruppen-IRT-Modell, ebenfalls basierend auf dem 2PL-Modell, für die Skala *Coping* sowie separat für die Subskala „*instrumentelles Coping*“ getrennt für die Pseudoitems *e* und *m* geschätzt. Gruppenzugehörigkeit der Testpersonen sowie Skalenzugehörigkeit der Items waren fixe Faktoren. Die Gruppen werden nach hohen bzw. niedrigen Scores (Mittelwert als Trennpunkt) der Testpersonen in den zusätzlich gestellten Fragen nach der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) eingeteilt. Bei den Gruppenunterschieden wurde davon ausgegangen, dass Personen mit höheren Werten bei den Zusatzfragen weniger Antworttendenzen zeigen und dementsprechend einen geringeren geschätzten Mittelwert aufweisen würden.

10.5 Evaluation der Modelle

Für die Evaluation der Modelle wurden das Akaike Informationskriterium (AIC, Akaike, 1974) sowie das Bayesian Informationskriterium (BIC; Schwarz, 1978) verwendet. Die Informationskriterien wurden verwendet, um die Modellpassung verschiedener Modelle zu vergleichen. Hierbei sind Modelle mit dem relativ gesehen geringeren Wert gegenüber anderen Modellen mit höheren Werten vorzuziehen.

$$AIC = -2\ln * L + 2n_p$$

$$BIC = -2\ln * L + (\ln N) + n_p$$

n_p beschreibt die Anzahl der Parameter des entsprechenden Modells, N die Stichprobengröße, $\ln$ den natürlichen Logarithmus und L die Likelihood. Während im AIC die Anzahl an Modellparametern mit 2 gewichtet wird, wird beim BIC der Logarithmus der Stichprobengröße als Gewicht verwendet. Das BIC bestraft die Verwendung zu vieler Parameter mehr als das AIC, sobald $\ln(N) > 2$ ist (vgl. Khorramdel & von Davier, in Druck).

10.6 IRT-Modelle mit multidimensionalen Antworttendenz- und Merkmalsfaktoren

10.6.1 Primärskalen als Faktoren

Zuerst wurden die 3-dimensionalen, 4-dimensionalen und 6-dimensionalen IRT-Modelle (basierend auf dem 2PL-Modell, Skalenzugehörigkeit der Items als fixer Faktor) für alle Primärskalen und alle Pseudoitemgruppen gleichzeitig geschätzt, um festzustellen, ob Antworttendenzen in den Daten vorhanden sind und diese von den Persönlichkeitseigenschaften unterschieden werden können.

3-dimensionales Modell: Die Pseudoitems wurden drei Dimensionen (e , d , m) zugeordnet.

4-dimensionales Modell: Die Pseudoitems wurden den vier Primärskalen des DSIHR zugeordnet.

6-dimensionales Modell: Pseudoitems d wurden den vier Primärskalen zugeordnet und die Pseudoitems e und m jeweils einer Dimension.

Das 3-dimensionale Modell wurde in einem ersten Schritt mit dem 4-dimensionalen Modell verglichen, um herauszufinden, ob die Pseudoitems drei Dimensionen oder die vier fraglichen Persönlichkeitseigenschaften messen. Da davon ausgegangen wurde, dass die Pseudoitems des Typs d , anders wie e und m , die Persönlichkeitseigenschaften messen, sollte das 6-dimensionale Modell die Daten am besten beschreiben.

Die Ergebnisse zeigen, dass das 3-dimensionale Modell (AIC = 206616,75, BIC = 209872,35) besser passt als das 4-dimensionale Modell (AIC = 211896,09, BIC = 215147,24). Im Gegensatz zur Hypothese passte das 6-dimensionale Modell (AIC = 209231,20, BIC = 212526,88) allerdings nicht besser als das 3-dimensionale Modell.

Tabelle 8: Ergebnisse der 3-dimensionalen, 4-dimensionalen und 6-dimensionalen IRT-Modelle, für alle Skalen und alle Pseudoitemtypen (360 Items insgesamt).

<i>Primärskalen, alle Pseudoitemtypen</i>	<i>6-dimensionales Modell</i>	<i>4-dimensionales Modell</i>	<i>3-dimensionales Modell</i>
AIC Index	209231,20	211896,09	206616,75
BIC Index	212526,88	215147,24	209872,35
log-penalty (modellbasierend, pro Item)	0,497	0,504	0,491

10.6.1.1 Subskalen als Faktoren

Die selbe Vorgehensweise wie oben beschrieben wird auch für die Subskalen des DSIHR angewandt.

3-dimensionales Modell: Die Pseudoitems wurden drei Dimensionen (e , d , m) zugeordnet.

9-dimensionales Modell: Die Pseudoitems wurden den neun Subskalen des DSIHR zugeordnet.

11-dimensionales Modell: Pseudoitems d wurden den neun Subskalen zugeordnet und die Pseudoitems e und m jeweils einer Dimension.

Das 3-dimensionale Modell wurde mit dem 9-dimensionalen Modell verglichen, um herauszufinden, ob die Pseudoitems besser durch drei Dimensionen oder die neun Faktoren des Fragebogens beschrieben werden können. Da davon ausgegangen wurde, dass die Pseudoitems des Typs d , anders als e und m , die Persönlichkeitseigenschaften messen, sollte das 11-dimensionale Modell die Daten am besten beschreiben.

Die Ergebnisse zeigen, dass entsprechend der Annahme das 3-dimensionale Modell (AIC = 206616,75, BIC = 209872,35) eine bessere Modellpassung hat als das 9-dimensionale Modell (AIC = 220705,67, BIC = 224072,61). Das 11-dimensionale Modell zeigt entgegen der Hypothese die schlechteste Passung (AIC = 268678,61, BIC = 272125,71). Detaillierte Ergebnisse finden sich in Tabelle 9. Wie oben bereits erwähnt, konnten die 9- und 11-dimensionalen Modelle auf Grund der hohen Rechenleistung nicht sehr genau berechnet werden, was auch der Grund für die sehr schlechte Modellpassung des 11-dimensionalen Modells sein könnte.

Tabelle 9: Ergebnisse der 3-dimensionalen, 9-dimensionalen und 11-dimensionalen IRT-Modelle, für alle Subskalen und alle Pseudoitemtypen (360 Items insgesamt).

<i>Subskalen, alle Pseudoitemtypen</i>	<i>11-dimensionales Modell</i>	<i>9-dimensionales Modell</i>	<i>3-dimensionales Modell</i>
AIC Index	268678,61	220705,67	206616,75
BIC Index	272125,71	224072,61	209872,35
log-penalty (modellbasierend, pro Item)	0,640	0,525	0,491

10.6.2 Subskalen je Primärskala als Faktoren

Um das Vorhandensein der Antworttendenzen auch für die Subskalen zu überprüfen, wurden auf Primärskalenebene (*Auslösung, Manifestation, Coping* und *Stabilisierung*) 3-dimensionale und je nach Anzahl der Subskalen 2 bzw. 3-dimensionale und 4 bzw. 5-dimensionale IRT-Modelle (basierend auf dem 2PL-Modell, Skalenzugehörigkeit der Items als fixer Faktor) für alle Subskalen und alle Pseudoitemgruppen gleichzeitig geschätzt.

3-dimensionales Modell (Antworttendenzfaktoren): Ein 3-dimensionales Modell für jede Primärskala getrennt mit den Pseudoitemtypen als Faktoren wurde berechnet.

2 bzw. 3-dimensionales Modell (Subskalen): Ein 2 bzw. 3-dimensionales Modell, indem alle Pseudoitems den jeweiligen Subskalen je Primärskala zugeordnet sind wurde berechnet.

4 bzw. 5-dimensionales Modell: Ein 4 bzw. 5-dimensionales Modell, in dem die Pseudoitems *e* und *m* jeweils einem Antworttendenzfaktor zugeordnet wurden und die Pseudoitems des Typ *d* den zwei bzw. drei Subskalen der jeweiligen Primärskala.

Das 3-dimensionale Modell (Pseudoitems) wurde zuerst mit dem 2 bzw. 3-dimensionalen Modell (Subskalen) verglichen, um herauszufinden, ob die Pseudoitems drei Dimensionen oder die drei fraglichen Persönlichkeitseigenschaften messen. Wie schon oben wurden davon ausgegangen, dass die Pseudoitems des Typs *d*, anders als *e* und *m*, die Fragebogenskalen messen, deshalb sollte das 4 bzw. 5-dimensionale Modell die Daten am besten beschreiben.

10.6.2.1 Auslösung

Die Ergebnisse zeigen, dass das 3-dimensionale Modell (Antworttendenzfaktoren, AIC = 83547,58, BIC = 84905,94) besser passt als das 3-dimensionale Modell (Subskalen, AIC = 85469,55, BIC = 86827,90). Im Gegensatz zur Hypothese passte das 5-dimensionale Modell (AIC = 84408,22, BIC = 85779,94) auch hier nicht besser als das 3-dimensionale Modell. Tabelle 10 zeigt die Ergebnisse im Detail.

Tabelle 10: Ergebnisse der 3-dimensionalen (Pseudoitems), 3-dimensionalen (Subskalen) und 5-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala StressAuslösung und alle Pseudoitemtypen (147 Items insgesamt).

<i>Subskalen, alle Pseudoitemtypen</i>	<i>5-dimensionales Modell</i>	<i>3-dimensionales Modell (Subskalen)</i>	<i>3-dimensionales Modell (Antworttendenzfaktoren)</i>
AIC Index	84408,22	85469,55	83547,58
BIC Index	85779,94	86827,90	84905,94
log-penalty (modellbasierend, pro Item)	0,488	0,493	0,483

10.6.2.2 Manifestationen

Die Ergebnisse für die Primärskala *Manifestation* zeigen, dass das 3-dimensionale Modell (Antworttendenzfaktoren, AIC = 34587,01, BIC = 35170,43) eine bessere Modellpassung hat als das 2-dimensionale Modell (Subskalen, AIC = 35270,60, BIC = 35831,75) und das 4-dimensionale Modell (AIC = 34560,54, BIC = 35152,88) die Daten noch besser beschreibt als das 3-dimensionale Modell. Das Ergebnis entspricht demnach der Hypothese, dass die Pseudoitems des Typ *d* besser durch die Subskalen als Faktoren dargestellt werden und die Pseudoitems des Typs *e* und *m* Antworttendenzen darstellen. Tabelle 11 zeigt die Ergebnisse im Detail.

Tabelle 11: Ergebnisse der 3-dimensionalen (Pseudoitems), 2-dimensionalen (Subskalen) und 4-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala Manifestation und alle Pseudoitemtypen (60 Items insgesamt).

Subskalen, alle Pseudoitemtypen	4-dimensionales Modell	2-dimensionales Modell (Subskalen)	3-dimensionales Modell (Antworttendenzfaktoren)
AIC Index	34560,54	35270,60	34587,01
BIC Index	35152,88	35831,75	35170,43
log-penalty (modellbasierend, pro Item)	0,480	0,491	0,481

10.6.2.3 Coping

Für die Primärskala *Coping* zeigt sich, dass das 3-dimensionale Modell (Antworttendenzfaktoren, AIC = 53243,47, BIC = 54120,84) eine bessere Modellpassung hat als das 2-dimensionale Modell (Subskalen, AIC = 54899,79, BIC = 55754,89) und das 4-dimensionale Modell (AIC = 53203,53, BIC = 54089,81) die Daten noch besser beschreibt als das 3-dimensionale Modell. Auch hier entsprechen die Ergebnisse der bereits oben beschriebenen Hypothese. Tabelle 12 zeigt detaillierte Ergebnisse.

Tabelle 12: Ergebnisse der 3-dimensionalen (Pseudoitems), 2-dimensionalen (Subskalen) und 4-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala Coping und alle Pseudoitemtypen (93 Items insgesamt).

Subskalen, alle Pseudoitemtypen	4-dimensionales Modell	2-dimensionales Modell (Subskalen)	3-dimensionales Modell (Antworttendenzfaktoren)
AIC Index	53203,53	54899,79	53243,47
BIC Index	54089,81	55754,89	53120,84
log-penalty (modellbasierend, pro Item)	0,494	0,509	0,494

10.6.2.4 Stabilisierung

Für die Primärskala *Stabilisierung* zeigt sich ebenfalls, dass das 3-dimensionale Modell (Antworttendenzfaktoren, AIC = 34684,06, BIC = 35267,49) eine bessere Modellpassung hat als das 2-dimensionale Modell (Subskalen, AIC = 34936,18, BIC = 35497,34). Das 4-dimensionale Modell (AIC = 34417,77, BIC = 35010,10) passt noch besser auf die Daten als das 3-dimensionale Modell und entspricht demnach der Hypothese. Tabelle 13 gibt detaillierte Ergebnisse wieder.

Tabelle 13: Ergebnisse der 3-dimensionalen (Pseudoitems), 2-dimensionalen (Subskalen) und 4-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala Stabilisierung und alle Pseudoitemtypen (60 Items insgesamt).

Subskalen, alle Pseudoitemtypen	4-dimensionales Modell	2-dimensionales Modell (Subskalen)	3-dimensionales Modell (Antworttenden- denzfaktoren)
AIC Index	34417,77	34936,18	34684,06
BIC Index	35010,10	35497,34	35267,49
log-penalty (modellbasierend, pro Item)	0,508	0,516	0,512

10.7 IRT-Modelle mit einem unidimensionalen Antworttendenzfaktor und multidimensionalen Merkmalsfaktoren

10.7.1 Primärskalen und Subskalen als Faktoren

1-dimensionale, 4-dimensionale und 9-dimensionale IRT-Modelle (basierend auf dem 2PL-Modell, Skalenzugehörigkeit der Items als fixer Faktor) wurden geschätzt. Im Vergleich zu den vorherigen Analysen wurden die Analysen für jeden Pseudoitemtyp (*e*-, *m*- und *d*-Items) einzeln berechnet.

1-dimensionales Modell: Ein 1-dimensionales Modell für jeden Pseudoitemtyp wurde berechnet, wobei *e*-Items, *m*-Items und *d*-Items jeweils separat einem Faktor (der Antworttendenz) zugeordnet wurden.

4-dimensionales Modell: Ein 4-dimensionales Modell mit den vier Primärskalen als Faktoren wurde für jeden Pseudoitemtyp berechnet, wobei die jeweiligen Pseudoitems (*e*-Items, *m*-Items oder *d*-Items) den 4 Primärskalen zugeordnet wurden.

9-dimensionales Modell: Ein 9-dimensionales Modell mit den 9 Subskalen als Faktoren wurde für jeden Pseudoitemtyp extra berechnet, wobei die jeweiligen Pseudoitems (*e*-Items, *m*-Items oder *d*-Items) den 9 Subskalen zugeordnet wurden.

Das 1-dimensionale Modell wurden mit den 4- bzw. 9-dimensionalen Modellen verglichen, um herauszufinden, ob die *e* und *m* Pseudoitems unidimensional Antworttendenzen messen, oder ob sie sich besser mit den vier Primärskalen bzw. den Subskalen als Faktoren beschreiben lassen und die *d*-Pseudoitems die Fragebogenskalen abbilden.

Die Ergebnisse der IRT-Modelle zeigen für die *m*-Items der Hypothesen entsprechend niedrigere Werte für das 1-dimensionale Modell (AIC = 53924,42, BIC = 55002,20) als für das 4-dimensionale (AIC = 54129,33, BIC = 55242,75). Bei den *e*-Items verhält es sich allerdings umgekehrt. Das 4-dimensionale Modell (AIC = 79786,24, BIC = 80913,01) beschreibt die Daten besser als das 1-dimensionale Modell (AIC = 80070,31, BIC = 81148,09). Die 9-dimensionalen Modelle (*m*: AIC = 57861,33, BIC = 59086,07; *e*: AIC = 85345,97, BIC = 86575,17) zeigen für die Pseudoitems der Typen *m* und *e* wie angenommen keine bessere Modellpassung als das 1-dimensionale Modell. Die 9-dimensionalen Modelle konnten auf Grund ihrer Komplexität nicht mit der gewünschten Genauigkeit berechnet werden. Deshalb wurden die Ergebnisse zusätzliche auf Primärskalenebene überprüft. Siehe Tabelle 14 für genaue Modellinformationen.

Die marginalen Reliabilitäten für die beiden unidimensionalen Antworttendenzmessungen waren, wie auch bei Khorramdel und von Davier (in Druck), hoch (*m*-Items: ,90; *e*-Items: ,94).

In Tabelle 15 und 16 werden die auf Basis der Modelle geschätzten Interkorrelationen für die 4 Primärskalen bzw. die 9 Subskalen für die Pseudoitems *e*, *m* und *d* sowie jene der originalen Items dargestellt (die empirisch beobachteten Skaleninterkorrelationen finden sich im Anhang.). Die Interkorrelationen der Primärskalen sind für beide Pseudoitemtypen relativ hoch (*m*-Items: zwischen ,0,495-0,645; *e*-Items: zwischen 0,589-0,663). Für die Interkorrelationen der Subskalen sind sie geringer (*m*-Items: zwischen ,0,189-0,509; *e*-Items: zwischen 0,262 – 0,502).

Zur graphischen Überprüfung, ob sich die Itemschwierigkeiten der 1-dimensionalen 2PL-Modelle (*e*-Items, *m*-Items und *d*-Items wurden jeweils einem separaten Faktor – also der Antworttendenz – zugeordnet) unterteilt in zwei verschiedenen Gruppen (Median der Testscores als Teilungskriterium) unterscheiden, wurden Grafische Modelltests (GMT, Abbildung 5 und 6) erstellt.

Tabelle 14: Ergebnisse für die 1-dimensionalen, 4-dimensionalen und 9-dimensionalen IRT-Modelle, für alle Pseudoitemtypen getrennt (120 Items pro Pseudoitemtyp).

<i>Subskalen, alle Pseudoitemtypen</i>	<i>1-dimensionales Modell</i>	<i>4-dimensionales Modell</i>	<i>9-dimensionales Modell</i>
<i>Skalen, m-Items</i>			
AIC Index	53924,42	54129,33	57861,33
BIC Index	55002,20	55242,75	59086,08
log-penalty (modellbasierend, pro Item)	0,350	0,352	0,376
<i>Skalen, e-Items</i>			
AIC Index	80070,31	79786,24	85345,97
BIC Index	81148,09	80913,01	86575,17
log-penalty (modellbasierend, pro Item)	0,601	0,598	0,640
<i>Skalen, d-Items</i>			
AIC Index	72667,92	71670,46	76092,10
BIC Index	73745,70	72797,22	77317,74
log-penalty (modellbasierend, pro Item)	0,544	0,537	0,570

Koller et al. (2012) beschreiben, dass ein GMT üblicherweise zur grafischen Modellkontrolle der Itemschwierigkeiten zweier Personengruppen im Rasch-Modell verwendet wird. Bei der Teilung eines Datensatzes in zwei Gruppen entsteht die Möglichkeit, die Itemparameter beider Gruppen gegeneinander in einem Streudiagramm aufzutragen. Die waagrechten und senkrechten Koordinaten der Itemschwierigkeitsparameter der ersten und zweiten Gruppe stellen einen Punkt im Streudiagramm dar. Ist die Modellpassung für das überprüfte Teilungskriterium gut (weichen die Itemschwierigkeitsparameter in den beiden überprüften Gruppen also nicht systematisch voneinander ab), so liegen alle Punkte annähernd auf einer 45°-Geraden durch den Ursprung (Koller et al., 2012).

Tabelle 15: Geschätzte Interkorrelationen der Score-Verteilung der vier Primärskalen auf Grundlage des 4-dimensionalen IRT-Modells der Pseudoitems *m*, *e* und *d* sowie der Originaldaten.

	<i>Auslösung</i>	<i>Manifestation</i>	<i>Coping</i>	<i>Stabilisierung</i>
<i>m-Items</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,56982	1		
<i>Coping</i>	0,55224	0,64476	1	
<i>Stabilisierung</i>	0,49538	0,54975	0,58351	1
<i>e-Items</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,58908	1		
<i>Coping</i>	0,58695	0,62130	1	
<i>Stabilisierung</i>	0,57245	0,57270	0,66303	1
<i>d-Items</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,52755	1		
<i>Coping</i>	0,09973	0,22227	1	
<i>Stabilisierung</i>	0,52106	0,48421	0,14460	1
<i>Originaldaten</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,49723	1		
<i>Coping</i>	0,32532	0,45769	1	
<i>Stabilisierung</i>	0,45003	0,51877	0,45861	1

Ein GMT bietet die Möglichkeit einer globalen Einschätzung der Qualität der Items (je weiter die Items von der 45°-Geraden entfernt sind, desto schlechter) und es können einzelne Items, die für die zwei überprüften Gruppen starke Unterschiede in den Itemschwierigkeitsparametern aufweisen, identifiziert werden (Koller et al., 2012). Das 2PL-Modell hat im Vergleich zum Rasch-Modell neben dem Itemschwierigkeitsparameter zusätzlich einen Itemdiskriminationsparameter, dementsprechend überprüft der GMT im Fall des 2PL-Modells eigentlich nicht das „Modell“ sondern nur die Itemschwierigkeitsparameter. Es kann also lediglich pro Item überprüfen, ob sich die Itemschwierigkeiten (ohne die Itemdiskriminationsparameter miteinzubeziehen) in zwei verschiedenen Gruppen ähnlich verhalten, um eine Hypothese aufstellen zu können, welche Items vielleicht problematisch sind.

Deshalb wurden auf Grundlage des in Abbildung 6 dargestellten GMTs Items sowohl im 1-dimensionalen als auch im 4-dimensionalen Modell der *e*-Items ausgeschlossen, um zu überprüfen, ob dadurch eine bessere Modellpassung für die 1-dimensionalen Modelle im Vergleich zu den 4-dimensionalen Modellen erreicht werden kann. Die

Tabelle 16: Geschätzte Interkorrelationen der Score-Verteilung der neun Subskalen auf Grundlage des 9-dimensionalen IRT-Modells der Pseudoitems *m*, *e* und *d* sowie der Originaldaten.

	<i>Exist.</i>	<i>Inter.</i>	<i>Alltags.</i>	<i>physiol.</i>	<i>em./ko.</i>	<i>pallia.</i>	<i>instr.</i>	<i>intern.</i>	<i>extern.</i>
<i>m-Items</i>									
Existenz	1								
Interaktion	0,449	1							
Alltagsgeschehen	0,441	0,509	1						
physiologisch	0,297	0,325	0,318	1					
emotional/kognitiv	0,380	0,356	0,348	0,306	1				
palliativ	0,359	0,359	0,371	0,337	0,377	1			
instrumentell	0,356	0,358	0,352	0,333	0,395	0,443	1		
internal	0,211	0,210	0,198	0,189	0,257	0,284	0,293	1	
external	0,310	0,317	0,323	0,247	0,348	0,340	0,308	0,236	1
<i>e-Items</i>									
Existenz	1								
Interaktion	0,345	1							
Alltagsgeschehen	0,441	0,452	1						
physiologisch	0,318	0,316	0,370	1					
emotional/kognitiv	0,361	0,296	0,375	0,395	1				
palliativ	0,323	0,330	0,414	0,385	0,444	1			
instrumentell	0,290	0,338	0,393	0,394	0,405	0,502	1		
internal	0,262	0,290	0,272	0,316	0,285	0,373	0,370	1	
external	0,329	0,298	0,412	0,340	0,350	0,469	0,405	0,367	1
<i>d-Items</i>									
Existenz	1								
Interaktion	0,399	1							
Alltagsgeschehen	0,463	0,359	1						
physiologisch	0,340	0,257	0,291	1					
emotional/kognitiv	0,323	0,228	0,404	0,352	1				
palliativ	0,042	0,121	-0,033	0,126	0,039	1			
instrumentell	0,125	0,120	0,120	0,216	0,264	0,205	1		
internal	0,005	0,014	0,111	0,118	0,077	0,202	0,211	1	
external	0,351	0,193	0,445	0,303	0,383	0,013	0,206	0,135	1
<i>Originaldaten</i>									
Existenz	1								
Interaktion	0,300	1							
Alltagsgeschehen	0,345	0,404	1						
physiologisch	0,170	0,218	0,271	1					
emotional/kognitiv	0,194	0,248	0,358	0,306	1				
palliativ	0,128	0,180	0,146	0,224	0,218	1			
instrumentell	0,092	0,181	0,236	0,242	0,303	0,357	1		
internal	0,008	0,028	0,104	0,184	0,135	0,276	0,275	1	
external	0,154	0,200	0,322	0,262	0,324	0,220	0,212	0,167	1

Ergebnisse in Tabelle 17 zeigen aber, dass sich auch nach dem Ausschluss von 29 Items keine bessere Modellpassung für das 1-dimensionale Modell (AIC = 60941,17, BIC = 61760,64) im Vergleich zum 4-dimensionalen Modell (AIC = 60874,44, BIC = 61742,90) zeigte.

Tabelle 17: Ergebnisse für die 1-dimensionalen und 4-dimensionalen IRT-Modelle, für die Pseudoitems des Typ *e* (120 Items) vor und nach der Ausscheidung von 29 Items.

Skalen, Pseudoitemtyp <i>e</i>	1-dimensionales Modell	4-dimensionales Modell	1-dimensionales Modell	4-dimensionales Modell
			nach Ausscheiden von 29 Items	
AIC Index	80070,31	79786,24	60941,17	60874,44
BIC Index	81148,09	80913,01	61760,64	61742,90
log-penalty (modellbasierend, pro Item)	0,601	0,598	0,56	0,599

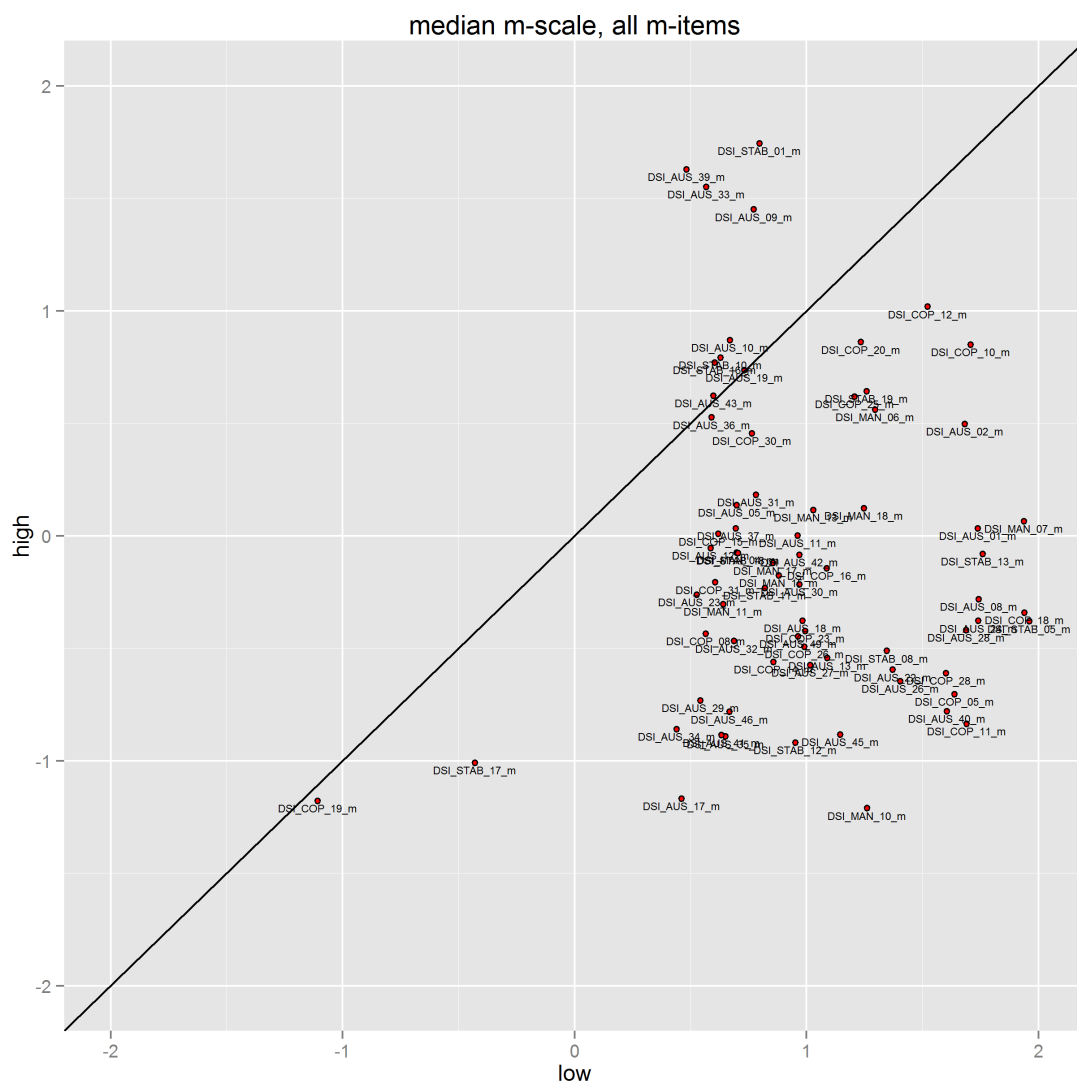


Abbildung 5: GMT auf Basis der Itemschwierigkeiten für die 1-dimensionalen 2PL-Modelle der *m*-Items (alle *m*-Items) mit dem Median der Scores als Teilungskriterium.

Die Ergebnisse der IRT-Modelle der Pseudoitems d entsprachen für die Primärskalen der aufgestellten Hypothese, dass die Pseudoitems dieses Typs die Fragebogenskalen messen. Das 4-dimensionale Modell (AIC = 71670,46, BIC = 72797,22) beschrieb die Daten besser als das 1-dimensionale Modell (AIC = 72667,92, BIC = 73745,70). Das 9-dimensionale Modell (AIC = 76092,10, BIC = 77317,74) beschrieb allerdings eine schlechtere Modellpassung als das 1-dimensionale Modell. Die marginalen Reliabilitäten pro Primärskala fielen hoch aus (*Auslösung*: ,90; *Manifestation*: ,80; *Coping*: ,69 und *Stabilisierung*: ,71) Detaillierte Ergebnisse finden sich in Tabelle 14.

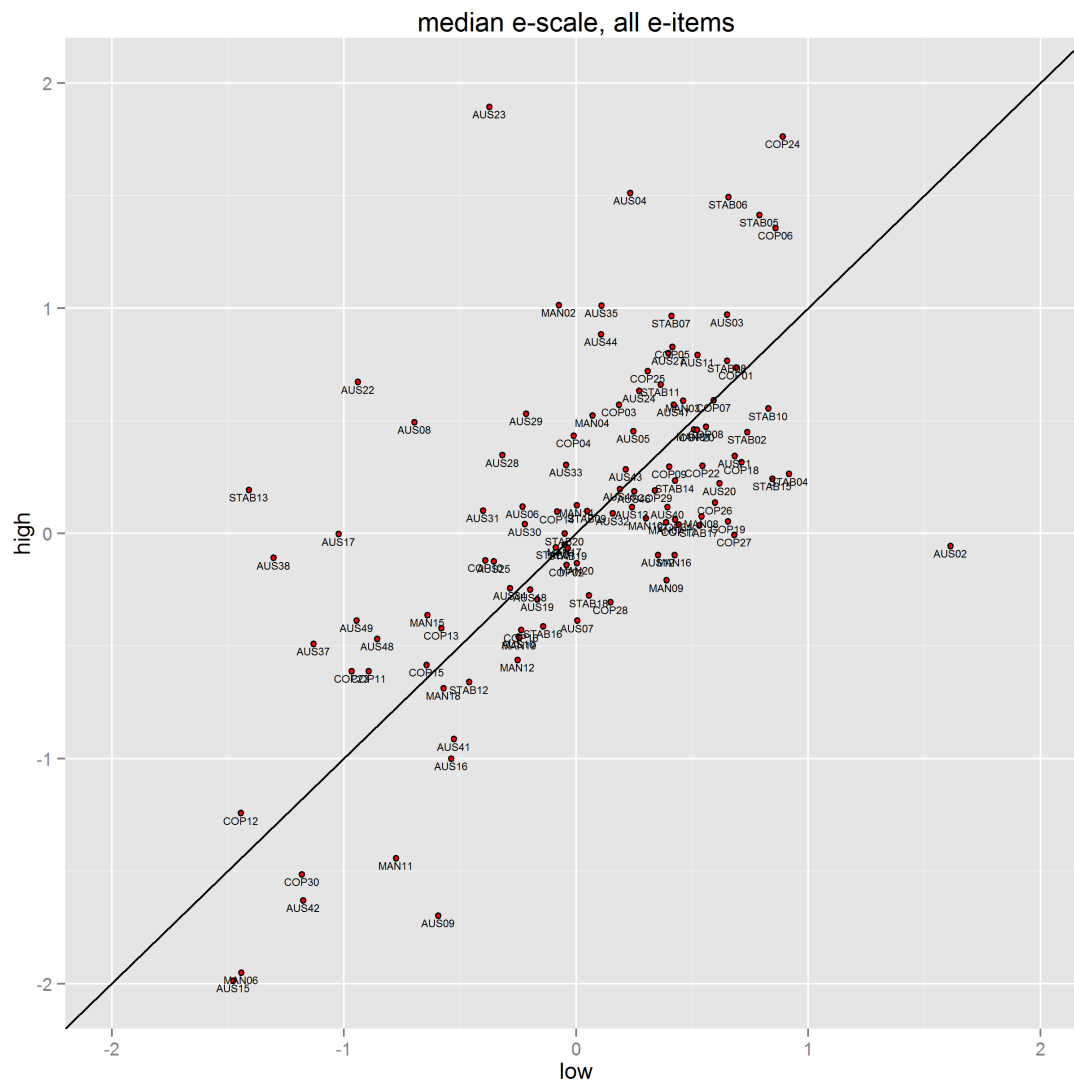


Abbildung 6: GMT auf Basis der Itemschwierigkeiten für das 1-dimensionale 2PL-Modell der e -Items (vor dem Ausschluss der 29 Items) mit dem Median der Scores als Teilungskriterium.

Des Weiteren kann der Tabelle 15 entnommen werden, dass die Skaleninterkorrelationen für die Primärskalen sowie die Subskalen zwar nicht niedrig, aber geringer sind (Primärskalen: zwischen 0,100-0,528; Subskalen: -0,033-0,463), als jene der *e* und *d* Pseudoitems.

10.7.2 Subskalen je Primärskala als Faktoren

Es wurden 1-dimensionale und 2- bzw. 3-dimensionale IRT-Modelle (je nach Anzahl der Subskalen pro Primärskala, basierend auf dem 2PL-Modell, Skalenzugehörigkeit der Items als fixer Faktor) für jeden Pseudoitemtyp und die Subskalen geschätzt.

1-dimensionales Modell (Pseudoitems): Ein 1-dimensionales Modell für jeden Pseudoitemtyp wurde berechnet, wobei e-Items, m-Items und d-Items jeweils separat einem Faktor (der Antworttendenz) zugeordnet wurden.

2- bzw. 3-dimensionales Modell (Subskalen): Je nach Subskalen pro Primärskala wurde ein 2 bzw. 3-dimensionales Modell berechnet, wobei die jeweiligen Pseudoitems (e-Items, m-Items oder d-Items) den 2 bzw. 3 Subskalen zugeordnet wurden.

Hier wurden die 1-dimensionalen Modelle mit den 2 bzw. 3-dimensionalen Modellen verglichen, um herauszufinden, ob die Pseudoitems auf Primärskalenebene unidimensional Antworttendenzen messen, oder ob sie sich besser mit den Subskalen pro Primärskala als Faktoren beschreiben lassen.

10.7.2.1 Auslösung

Die Ergebnisse der IRT-Modelle für die Primärskala *Auslösung* zeigten sich alle den Hypothesen entsprechend. Für die *m*- und *e*-Pseudoitemtypen beschrieben die jeweiligen 1-dimensionalen Modelle (*m*-Items: AIC = 21032,81, BIC = 21478,18; *e*-Items: AIC = 33589,25, BIC = 34034,61) eine bessere Modellpassung als die 3-dimensionalen Modelle (*m*-Items: AIC = 21301,63, BIC = 21787,07; *e*-Items: AIC = 33681,71, BIC = 34167,16) und sprechen dafür, dass Antworttendenzen in den Daten vorhanden sind. Die Pseudoitems des Typs *d* hingegen wurden durch das 3-dimensionale Modell (AIC = 28965,16, BIC = 29450,60) besser beschrieben als durch das 1-dimensionale (AIC = 29214,02, BIC = 29659,38). Tabelle 18 zeigt eine Übersicht.

Tabelle 18: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Auslösung, für alle Pseudoitemtypen getrennt (49 Items pro Pseudoitemtyp).

	3-dimensionales Modell	1-dimensionales Modell
<i>Primärskala Auslösung, m-Items</i>		
AIC Index	21301,63	21032,81
BIC Index	21787,07	21478,18
log-penalty (modellbasierend, pro Item)	0,339	0,335
<i>Primärskala Auslösung, e-Items</i>		
AIC Index	33681,71	29214,02
BIC Index	34167,16	29659,38
log-penalty (modellbasierend, pro Item)	0,611	0,529
<i>Primärskala Auslösung, d-Items</i>		
AIC Index	28965,16	29214,02
BIC Index	29450,60	29659,38
log-penalty (modellbasierend, pro Item)	0,525	0,529

Die auf Basis des 3-dimensionalen Modells geschätzten Skaleninterkorrelationen (genaue Werte in Tabelle 19) zeigen für alle Pseudoitemtypen hohe Zusammenhänge (*m*-Items: zwischen ,607-,672; *e*-Items: zwischen ,585-,655; *d*-Items zwischen ,575-,645).

Tabelle 19: Geschätzte Interkorrelationen der Score-Verteilung der Subskalen der Skala Auslösung auf Grundlage des 3-dimensionalen IRT-Modells der Pseudoitems *m*, *e* und *d*.

	Existenz	Interaktion	Alltagsgeschehen
<i>m-Items</i>			
Existenz	1		
Interaktion	0,61507	1	
Alltagsgeschehen	0,60690	0,67231	1
<i>e-Items</i>			
Existenz	1		
Interaktion	0,58549	1	
Alltagsgeschehen	0,59962	0,65525	1
<i>d-Items</i>			
Existenz	1		
Interaktion	0,57890	1	
Alltagsgeschehen	0,64541	0,57540	1

10.7.2.2 Manifestationen

Die Ergebnisse der IRT-Modelle für die Primärskala *Manifestation* entsprachen für die Pseudoitems *m* und *d* den Hypothesen. Das 1-dimensionale Modell (AIC = 7608,48, BIC = 7795,54) hat eine bessere Passung als das 2-dimensionale Modell (AIC = 7647,40, BIC = 7852,27) für die Pseudoitems des Typs *m*. Bei den Pseudoitems des Typ

d verhält es sich wie angenommen andersherum (1-dimensional: AIC = 12961,90, BIC = 13148,95; 2-dimensional: AIC = 12820,68, BIC = 13025,56). Bei den Pseudoitems des Typs *e* beschreibt das 2-dimensionale Modell (AIC = 14090,88, BIC = 14295,74) eine bessere Modellpassung als das 1-dimensionale (AIC = 14113,83, BIC = 14300,88), was gegen die Hypothese spricht, da die Daten hier durch die Subskalen als Faktoren besser beschrieben werden als durch die Antworttendenz als 1-dimensionaler Faktor. Detaillierte Informationen finden sich in Tabelle 20.

Die geschätzten Skaleninterkorrelationen auf Basis des 2-dimensionalen Modells betrugen für die *m*-Items: ,421, für die *e*-Items: ,601 und für die *d*-Items: ,563. Auch hier zeigen sich entgegen der Hypothese für alle drei Pseudoitemtypen hohe Korrelationen.

Tabelle 20: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Manifestation, für alle Pseudoitemtypen getrennt (20 Items pro Pseudoitemtyp).

	2-dimensionales Modell	1-dimensionales Modell
<i>Primärskala Manifestation, m-Items</i>		
AIC Index	7647,40	7608,48
BIC Index	7852,27	7795,54
log-penalty (modellbasierend, pro Item)	0,297	0,296
<i>Primärskala Manifestation, e-Items</i>		
AIC Index	14090,88	14113,83
BIC Index	14295,74	14300,88
log-penalty (modellbasierend, pro Item)	0,609	0,610
<i>Primärskala Manifestation, d-Items</i>		
AIC Index	12820,68	12961,90
BIC Index	13025,56	13148,95
log-penalty (modellbasierend, pro Item)	0,554	0,560

10.7.2.3 Coping

Die Ergebnisse der Primärskala *Coping* entsprechen für alle drei Pseudoitemtypen den Hypothesen. Für die *m*- und *e*- Pseudoitems passt jeweils das 1-dimensionale Modell besser (*m*-Items: AIC = 14850,61, BIC = 15135,64; *e*-Items: AIC = 20172,71, BIC = 20457,74) als das 2-dimensionale (*m*-Items: AIC = 14896,45, BIC = 15199,29; *e*-Items AIC = 20234,85, BIC = 20537,70). Die Pseudoitems des Typs *d* wurden hingegen durch das 2-dimensionale Modell (AIC = 18106,55, BIC = 18409,40) besser

beschrieben als durch das 1-dimensionale (AIC = 18304,04, BIC = 18589,07). In der Tabelle 21 sind die genauen Ergebnisse aufgelistet.

Die geschätzten Skaleninterkorrelationen auf Basis der 2-dimensionalen Modelle betrugen für die *m*-Items ,634, für die *e*-Items ,685 und für die *d*-Items ,699.

Tabelle 21: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Coping, für alle Pseudoitemtypen getrennt (31 Items pro Pseudoitemtyp).

	2-dimensionales Modell	1-dimensionales Modell
<i>Primärskala Coping, m-Items</i>		
AIC Index	14896,45	14850,61
BIC Index	15199,29	15135,64
log-penalty (modellbasierend, pro Item)	0,375	0,374
<i>Primärskala Coping, e-Items</i>		
AIC Index	20234,85	20172,71
BIC Index	20537,70	20457,74
log-penalty (modellbasierend, pro Item)	0,594	0,593
<i>Primärskala Coping, d-Items</i>		
AIC Index	18106,55	18304,04
BIC Index	18409,40	18589,07
log-penalty (modellbasierend, pro Item)	0,531	0,537

10.7.2.4 Stabilisierung

In der Primärskala *Stabilisierung* entsprachen die Ergebnisse der IRT-Modelle für die Pseudoitems der Typen *m* und *e* nicht der Hypothese. In beiden Fällen hatte das 1-dimensionale Modell (*m*-Item: AIC = 10994,65, BIC = 11181,7; *e*-Items: AIC = 12558,93, BIC = 12745,98) eine schlechtere Modellpassung als das 1-dimensionale (*m*-Items: AIC = 10986,17, BIC = 11191,04, *e*-Items: AIC = 12437,6, BIC = 12642,52). Auch wenn die Differenz zwischen den Modellkennwerten gering ist, deutet das doch darauf hin, dass die Pseudoitems *m* und *e* in der Primärskala *Stabilisierung* nicht durch den jeweiligen Antworttendenzfaktor beschrieben werden können. Tabelle 21 zeigt die Ergebnisse im Detail. Die Pseudoitems des Typs *d* werden der Hypothese entsprechend durch das 2-dimensionale Modell (AIC = 10899,39; BIC = 11104,26) besser beschrieben als durch das 1-dimensionale Modell (AIC = 11281,29, BIC = 11468,35).

Die geschätzten Skaleninterkorrelationen des 2-dimensionalen Modells sind allerdings für die *m*- und *e*-Items (*m*-Items: ,493, *e*-Items: ,551) wesentlich größer als für die *d*-Items (,195) und sprechen für die Hypothese.

Tabelle 22: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Stabilisierung, für alle Pseudoitemtypen getrennt (20 Items pro Pseudoitemtyp)

	2-dimensionales Modell	1-dimensionales Modell
<i>Primärskala Stabilisierung, m-Items</i>		
AIC Index	10986,17	10994,65
BIC Index	11191,04	11181,70
log-penalty (modellbasierend, pro Item)	0,429	0,430
<i>Primärskala Stabilisierung, e-Items</i>		
AIC Index	12437,65	12558,93
BIC Index	12642,52	12745,98
log-penalty (modellbasierend, pro Item)	0,590	0,597
<i>Primärskala Stabilisierung, d-Items</i>		
AIC Index	10899,39	11281,29
BIC Index	11104,26	11468,35
log-penalty (modellbasierend, pro Item)	0,517	0,536

10.8 Bifactor-Modelle

Sowohl für die *e*- als auch für die *m*-Items wurden Bifactor-Modelle berechnet. Die Ergebnisse der Pseudoitems des Typs *e* lassen, wie oben dargestellt, darauf schließen, dass diese sich nicht unidimensional durch einen Antworttendenzfaktor erklären lassen. Die Pseudoitems des Typs *m* ließen sich zwar weitgehend besser durch die Antworttendenz als unidimensionalen Faktor erklären, die Ergebnisse zwischen jenen Modellen mit der Antworttendenz als Faktor, und jenen, in denen die Pseudoitems des Typs *m* den Fragebogenskalen zugeordnet waren, lagen aber teilweise sehr knapp beieinander. Des Weiteren zeigt der GMT, dass sich die Itemschwierigkeiten einzelner Items in den zwei fraglichen Gruppen (geteilt nach Median des Testscores) stark unterschieden. Auf Grund dessen, wurden auch für die *m*-Pseudoitems zusätzlich Bifactor-Modelle berechnet.

10.8.1 Primärskalen und Subskalen als Faktoren

Die Bifactor-Modelle wurden mit den jeweiligen, im Kapitel 10.7 beschriebenen am besten passenden IRT-Modellen für die Pseudoitemtypen m und e verglichen.

Bifactor-Modell: Bifactor-Modelle (mit der Antworttendenz als Generalfaktor und den Primärskalen bzw. den Subskalen als spezifische Faktoren) wurden getrennt für die Pseudoitems m und e berechnet.

Das jeweilige Bifactor-Modell wurde mit dem bisher am besten auf die Daten passenden IRT-Modell verglichen, um herauszufinden, ob die Daten besser beschrieben werden können, wenn sie sowohl auf die Primärskalen, als auch einen Antworttendenzfaktor laden.

Die Werte (siehe Tabelle 23) zeigen, dass für die Pseudoitems des Typs m das AIC und das BIC widersprüchliche Ergebnisse liefern (Bifactor-Modells: AIC = 53753,43, BIC = 55334,47; 1-dimensionale Modell: AIC = 53924,42, BIC = 55002,20). Dementsprechend kann keine konkrete Aussage bezüglich der aufgestellten Hypothese getroffen werden. Für die Pseudoitems des Typs e konnte ebenfalls eine bessere Modellpassung beschrieben durch das AIC des Bifactor-Modell (AIC = 79541,33, BIC = 81140,18) erreicht werden, als durch das 4-dimensionale *simple structure* Modell (e -Items: AIC = 79786,24, BIC = 80913,01). Für das BIC verhält es sich genau umgekehrt.

Tabelle 23: Ergebnisse für die Bifactor-Modelle (Primärskalen als spezifische Faktoren) und die 4-dimensionalen IRT-Modelle für alle Primärskalen für die Pseudoitemtypen m und e getrennt (120 Items pro Pseudoitemtyp).

	Bifactor-Modell (Primärskalen)	1-dimensionales Modell
<i>m</i> -Items		
AIC Index	53753,43	53924,42
BIC Index	55334,47	55002,20
log-penalty	0,348	0,352
(modellbasierend, pro Item)		
	Bifactor-Modell (Primärskalen)	4-dimensionales Modell
<i>e</i> -Items		
AIC Index	79541,33	79786,24
BIC Index	81140,80	80913,01
log-penalty	0,595	0,598
(modellbasierend, pro Item)		

Die Hypothese, dass die Pseudoitems des Typs *e* am besten durch Antworttendenzen und Merkmalsausprägungen beschrieben werden können, kann auf Grund der Ergebnisse nicht bestätigt werden.

Tabelle 24 gibt zusätzlich einen Überblick über die erklärte Varianz (SD^2) durch jeden latenten Faktor des Bifactor-Modells (Antworttendenz als Generalfaktor, Primärskalen als spezifische Faktoren). Bei den Pseudoitems des Typs *m* wird die meiste Varianz (,805) durch den Antworttendenzfaktor erklärt, was darauf hinweist, dass diese hauptsächlich Indikatoren für die mittlere Antworttendenz sind. Bei den Pseudoitems des Typs *e* hingegen wird die meiste Varianz nicht durch den Antworttendenzfaktor beschrieben.

Tabelle 24: Varianzen (SD^2) der geschätzten spezifischen Faktoren (Primärskalen) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen *m* und *e* nach dem Bifactor-Modell.

	<i>Auslösung</i>	<i>Manifestation</i>	<i>Coping</i>	<i>Stabilisierung</i>	<i>Antworttendenz</i>
<i>m</i> -Items, SD^2	0,373	0,209	0,259	0,219	0,805
<i>e</i> -Items, SD^2	0,549	0,912	0,711	1,309	1,055

Das Bifactor-Modell mit der jeweiligen Antworttendenz als Generalfaktor und den neun Subskalen als spezifische Faktoren zeigte für die Pseudoitems des Typs *m* eine schlechtere Modellpassung für das Bifactor-Modell ($AIC = 55852,56$, $BIC = 57411,33$) als für das *simple structure* 1-dimensionale Modell ($AIC = 53924,42$, $BIC = 55002,20$). Detaillierte Ergebnisse finden sich in Tabelle 25. Wie in Tabelle 24 ersichtlich, wird im Fall der *m*-Items auch die meiste Varianz durch den Generalfaktor (Antworttendenz, $SD^2 = 1,980$) erklärt. Für die *e*-Items entsprechen die Ergebnisse ebenfalls nicht der Hypothese. Das Bifactor-Modell ($AIC = 87202,60$, $BIC = 88801,45$) hat zwar eine bessere Modellpassung als das 9-dimensionale *simple structure* Modell, kann die Daten aber auch nicht besser erklären als das 4-dimensionale Modell ($AIC = 79786,24$, $BIC = 80913,01$). Genaue Ergebnisse finden sich in Tabelle 26. Auch kann, wie Tabelle 25 zu entnehmen ist, die meiste Varianz nicht durch den Generalfaktor (Antworttendenz) erklärt werden.

Tabelle 25: Varianzen (SD^2) der geschätzten spezifischen Faktoren (Primärskalen) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen *m* und *e* nach dem Bifactor-Modell.

	<i>Exist.</i>	<i>Inter.</i>	<i>Alltag.</i>	<i>physio.</i>	<i>em./ko.</i>	<i>pallia.</i>	<i>instr.</i>	<i>intern.</i>	<i>extern.</i>	<i>Ant.te.</i>
<i>m</i> -Items, SD^2	1,446	1,405	1,861	1,944	1,229	1,052	1,889	1,044	1,741	1,980
<i>e</i> -Items, SD^2	4,567	2,913	1,091	1,279	2,812	9,241	3,913	15,736	48,158	0,385

Tabelle 26: Ergebnisse für die Bifactor-Modelle (Subskalen als spezifische Faktoren) und die 9-dimensionalen IRT-Modelle für alle Primärskalen für die Pseudoitemtypen *m* und *e* getrennt (120 Items pro Pseudoitemtyp).

	Bifactor-Modell (Subskalen)	1-dimensionales Modell
<i>m</i> -Items		
AIC Index	55852,56	53924,42
BIC Index	57411,33	55002,20
log-penalty (modellbasierend, pro Item)	0,362	0,352
	Bifactor-Modell (Subskalen)	4-dimensionales Modell
<i>e</i> -Items		
AIC Index	87202,60	79786,24
BIC Index	88801,45	80913,01
log-penalty (modellbasierend, pro Item)	0,653	0,598

10.8.2 Subskalen je Primärskala als Faktoren

Die Bifactor-Modelle wurden mit den jeweiligen, im Kapitel 10.7 beschriebenen am besten passenden IRT-Modellen für die Pseudoitemtypen *m* und *e* verglichen.

Bifactor-Modell: Ein Bifactor-Modell (mit der entsprechenden Antworttendenz als Generalfaktor und den Subskalen je Primärskala als spezifische Faktoren) wurde getrennt für die Pseudoitems *m* und *e* berechnet.

Die Bifactor-Modelle wurden mit den bisher die Daten am besten beschreibenden IRT-Modellen verglichen, um herauszufinden, ob die Daten durch Antworttendenzfaktoren und Eigenschaftsfaktoren gemeinsam am besten beschrieben werden können.

10.8.2.1 Auslösung

Für die Subskalen der Primärskala *Auslösung* zeigt sich, für die IRT-Modelle der *m*-Items, dass das Bifactor-Modell (AIC = 21282,40, BIC = 21932,63) keine bessere Modellpassung hat als das 1-dimensionale Modell (AIC = 21032,91, BIC = 21478,18). Detaillierte Informationen finden sich in Tabelle 28. Die meiste Varianz wird im

Bifactor-Modell durch die Antworttendenz erklärt (1,527, siehe Tabelle 27). Für die *e*-Items zeigte sich ebenfalls keine bessere Modellpassung für das Bifactor-Modell (AIC = 34073,74, BIC = 34723,97) im Vergleich zum 1-dimensionalen Modell (AIC = 29214,02, BIC = 29659,38). Die meiste Varianz wird durch die Subskalen als Faktoren beschrieben, wie Tabelle 27 zu entnehmen ist.

Tabelle 27: Varianzen (SD^2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Auslösung) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen *m* und *e* nach dem Bifactor-Modell.

	Existenz	Interaktion	Alltagsgeschehen	Antworttendenz
<i>m</i> -Items, SD^2	1,108	0,436	0,585	1,527
<i>e</i> -Items, SD^2	2,665	1,328	1,216	1,535

Tabelle 28: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Auslösung für die Pseudoitemtypen *m* und *e* getrennt (49 Items pro Pseudoitemtyp).

	Bifactor-Modell	1-dimensionales Modell
<i>m</i> -Items		
AIC Index	21282,40	21032,81
BIC Index	21932,63	21478,18
log-penalty	0,337	0,335
(modellbasierend, pro Item)		
<i>e</i> -Items		
AIC Index	34073,74	29214,02
BIC Index	34723,97	29659,38
log-penalty	0,616	0,529
(modellbasierend, pro Item)		

10.8.2.2 Manifestation

Für die Subskalen der Primärskala *Manifestation* zeigte sich für die *m*-Items keine bessere Modellpassung des Bifactor-Modells (AIC = 7672,83, BIC = 7935,59) im Vergleich zum *simple structure* 1-dimensionalen Modell (AIC = 7608,48, BIC = 7795,54). Der Generalfaktor (Antworttendenz) erklärt die meiste Varianz (,896, genaue Ergebnisse sind in Tabelle 24 dargestellt). Die *e*-Items können durch das Bifactor-Modell (AIC = 13985,59, BIC = 14248,36) besser beschrieben werden als durch das 2-dimensionale Modell (AIC = 14090,88, BIC = 14295,74). Der Antworttendenzfaktor beschreibt allerdings nicht die meiste Varianz (siehe Tabelle 30). Genaue Informationen zu den jeweiligen Modellpassungen finden sich in Tabelle 29.

Tabelle 29: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Manifestation für die Pseudoitemtypen *m* und *e* getrennt (20 Items pro Pseudoitemtyp).

	Bifactor-Modell	1-dimensionales Modell
<i>m</i> -Items		
AIC Index	7672,83	7608,48
BIC Index	7935,59	7795,54
log-penalty	0,297	0,296
(modellbasierend, pro Item)		
	Bifactor-Modell	2-dimensionales Modell
<i>e</i> -Items		
AIC Index	13985,59	14090,88
BIC Index	14248,36	14295,74
log-penalty	0,603	0,609
(modellbasierend, pro Item)		

Tabelle 30: Varianzen (SD^2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Manifestation) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen *m* und *e* nach dem Bifactor-Modell.

	physiologisch	emotional-kognitiv	Antworttendenz
<i>m</i> -Items, SD^2	0,749	0,623	0,896
<i>e</i> -Items, SD^2	1,241	0,286	1,165

10.8.2.3 Coping

Die *m*-Items der Subskalen für die Primärskala *Coping* werden durch das *simple structure* 1-dimensionale Modell (AIC = 14850,61, BIC = 15135,64) besser beschrieben als durch das Bifactor-Modell (AIC = 14922,28, BIC = 15332,01). Die meiste Varianz wird, wie in Tabelle 30 angeführt, durch die Antworttendenz beschrieben. Das Bifactor-Modell beschreibt die Daten gegensätzlich zur Hypothese nicht besser. Für die *e*-Items zeigt sich, dass das Bifactor-Modell (AIC = 20030,14, BIC = 20439,87) eine bessere Passung hat als das 1-dimensionale Modell (AIC = 20172,71, BIC = 20457,74). Auch hier wird die meiste Varianz durch den Antworttendenzfaktor erklärt (siehe Tabelle 31). Eine genaue Beschreibung der Ergebnisse finden sich in Tabelle 32.

Tabelle 31: Varianzen (SD^2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Coping) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen *m* und *e* nach dem Bifactor-Modell.

	palliativ	instrumentell	Antworttendenz
<i>m</i> -Items, SD^2	0,412	0,446	0,912
<i>e</i> -Items, SD^2	0,360	0,180	1,226

Tabelle 32: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Coping für die Pseudoitemtypen *m* und *e* getrennt (31 Items pro Pseudoitemtyp).

	Bifactor-Modell	1-dimensionales Modell
<i>m</i> -Items		
AIC Index	14922,28	14850,61
BIC Index	15332,01	15135,64
log-penalty (modellbasierend, pro Item)	0,374	0,374
<i>e</i> -Items		
AIC Index	20030,14	20172,71
BIC Index	20439,87	20457,74
log-penalty (modellbasierend, pro Item)	0,587	0,593

10.8.2.4 Stabilisierung

Für die *m*-Items zeigten sich für das BIC und das AIC für das Bifactor-Modell (AIC = 10982,18, BIC = 11244,94) und das *simple structure* 2-dimensionale Modell (AIC = 10986,17, BIC = 11191,04) gegensätzlich Werte. Das AIC zeigt eine bessere Modellpassung für das Bifactor-Modell, wohingegen das BIC eine bessere Passung für das *simple structure* Modell beschreibt. Auf Grund der Ergebnisse kann keine klare Aussage bezüglich der Modellpassung vorgenommen werden. Tabelle 32 gibt einen Überblick über die erklärte Varianz der einzelnen Faktoren. Die *e*-Items können durch das Bifactor-Modell (AIC = 12362,69, BIC = 12625,46) besser beschrieben werden als durch das *simple structure* 2-dimensionale Modell (AIC = 12437,65, BIC = 12642,52). Wie Tabelle 33 zeigt, erklärt der Antworttendenzfaktor allerdings nicht die meiste Varianz. Genaue Ergebnisse zu den IRT-Modellen finden sich in Tabelle 34.

Tabelle 33: Varianzen (SD^2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Stabilisierung) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen *m* und *e* nach dem Bifactor-Modell.

	<i>internal</i>	<i>external</i>	Antworttendenz
<i>m</i> -Items, SD^2	0,386	0,891	0,740
<i>e</i> -Items, SD^2	0,239	1,809	1,722

Tabelle 34: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Stabilisierung für die Pseudoitemtypen *m* und *e* getrennt (20 Items pro Pseudoitemtyp).

	Bifactor-Modell	1-dimensionales Modell
<i>m</i> -Items		
AIC Index	10982,18	10986,17
BIC Index	11244,94	11191,04
log-penalty	0,428	0,429
(modellbasierend, pro Item)		
<i>e</i> -Items		
AIC Index	12362,69	12437,65
BIC Index	12625,46	12642,52
log-penalty	0,586	0,590
(modellbasierend, pro Item)		

10.9 Fragebogeninhalt

Wie oben bereits kurz beschrieben, sollten inhaltliche Unterschiede zwischen Primärskalen, in denen Antworttendenzen auftraten, und jenen Primärskalen, in denen keine Antworttendenzen gefunden werden konnten, untersucht werden. Des Weiteren wurden Mittelwerte der geschätzten Verteilungen für jene Primär- und Sekundärskalen der Pseudoitems *e* und *m* verglichen, für die Antworttendenzen festgestellt werden konnten. Je höher der geschätzte Mittelwert ausfällt, desto mehr mittlere bzw. extreme Antworten wurden gegeben.

10.9.1 Pseudoitems des Typs *m*

Für die Pseudoitems des Typs *m* konnten sowohl für den gesamten Fragebogen (alle Primärskalen), als auch für die Primärskalen *Auslösung*, *Manifestation* und *Coping* Antworttendenzen festgestellt werden.

10.9.1.1 Vergleich der Fragebogeninhalte der Primärskalen

Wie im Kapitel 10.7 dargestellt, konnte nur für die Primärskala *Stabilisierung* keine mittlere Antworttendenz nachgewiesen werden. In der Skala *Stabilisierung* werden (nach der Einteilung in Tabelle 7) neben nicht beobachtbaren Reaktionen auch Einstellungen und Überzeugungen sowie Reaktionen anderer gegenüber einer Person abgefragt. Konkret beschäftigen sich die Fragen mit Stabilisierungsfaktoren von Stress, die als Verstärker des Stresserlebens wirken können. Zusätzlich anzumerken ist, dass die Fragen der Primärskala *Stabilisierung* die einzigen sind, die im Fragebogen als vollstän-

dige Aussagen dargestellt wurden und nicht alle mit demselben Satzteil (wie z.B.: Ich erlebe Stress, weil...) beginnen, der nur einmal am Anfang der Skala angeführt ist. In den anderen drei Primärskalen können die Iteminhalte (siehe Tabelle 7) in nicht beobachtbare und beobachtbare Reaktionen, Symptome sowie modifizierte Eigenschaftszuschreibungen eingeteilt werden. Konkret befasst sich die Skala *Auslösung* mit spezifischen Gründen, durch die Stress erlebt werden kann. In der Skala *Manifestation* werden konkrete Auswirkungen von Stress auf den Körper oder das Denken einer Person abgefragt und die Skala *Coping* fragt nach konkreten Verhaltensmustern, die man zeigt bzw. umsetzt, um Stress entgegenzuwirken.

10.9.1.2 Unterschiede der geschätzten Mittelwerte der Primärskalen

Für die Primärskalen wurden die meisten *m*-Antworten in der Skala *Stabilisierung* (Mittelwert: -2,09, SD: 1,39) gegeben. In dieser Skala werden stabilisierende Stressfaktoren abgefragt, die laut der von der Autorin vorgenommen Einteilung in Tabelle 7 Fragen zu nicht beobachtbaren Reaktionen, Einstellungen und Überzeugungen, sowie Reaktionen anderer gegenüber einer Person beinhalten. Die wenigsten *m*-Antworten wurden in der Subskala *Manifestation* (Mittelwert: -3,28, SD: 1,93) gegeben. Die Skala beschäftigt sich mit Formen der Stress*Manifestation* und kann laut Tabelle 7 in Fragen bezüglich beobachtbare und nicht beobachtbare Reaktionen sowie Symptome eingeteilt werden. Genaue Ergebnisse können Tabelle 35 entnommen werden.

Tabelle 35: Geschätzte Mittelwerte pro Primärskala auf Basis des 4-dimensionalen 2PL-Modells für die Pseudoitems des Typs *m*.

	<i>Auslösung</i>	<i>Manifestation</i>	<i>Coping</i>	<i>Stabilisierung</i>
<i>Mittelwert</i>	-2,89	-3,28	-2,54	-2,09
<i>SD</i>	1,54	1,93	1,61	1,39

10.9.1.3 Unterschiede der geschätzten Mittelwerte der Subskalen der Primärskala Auslösung

In allen drei Subskalen werden modifizierte Eigenschaftszuschreibungen abgefragt. Bedeutend weniger *m*-Antworten (Mittelwert: -2,98, SD: 1,86) wurden in der Subskala *Existenz* gegeben, in der das Ausmaß der Belastung durch Zukunftssorgen und Existenzängste abgefragt wird. Die geschätzten Mittelwerte der Subskalen *Interaktion*

(Mittelwert: -1,15, SD: 2,19) und *Alltagsgeschehen* (Mittelwert: -1,14, SD: 1,98) liegen sehr nahe beieinander. In diesen beiden Subskalen geht es um das Ausmaß der Belastung durch die Interaktion mit Mitmenschen bzw. durch unvermeidliche Angelegenheiten des Alltags.

10.9.1.4 Unterschiede der geschätzten Mittelwerte der Subskalen der Primärskala Manifestation

Der geschätzte Mittelwert (-3,40, SD: 1,87) für die Subskala „*physiologische Manifestationen*“ ist bedeutend geringer als jener für die Subskala „*emotional-kognitive Manifestationen*“ (-,98, SD: 1,95). Dies könnte darauf hinweisen, dass in der Subskala „*physiologische Manifestationen*“ in der konkreten Symptome (laut Tabelle 7 Selbstbeschreibung: Symptome) abgefragt werden, tendenziell weniger *m*-Antworten gegeben werden als in der Subskala „*physiologische Manifestationen*“, deren Items ebenfalls Selbstbeschreibungen darstellen, aber beobachtbare und nicht beobachtbare Reaktionen erfragen.

10.9.1.5 Unterschiede der geschätzten Mittelwerte der Subskalen der Primärskala Coping

Auch für die Subskalen der Skala *Coping* wurden Unterschiede in den geschätzten Mittelwerten erwartet, da die Subskala *palliatives Coping* aus Items besteht, die sich mit der Selbstbeschreibung nicht beobachtbarer Reaktionen beschäftigen, wohingegen die Items der Skala *instrumentelles Coping* sich eher auf beobachtbare Reaktionen beziehen. Der geschätzte Mittelwert der Skala *palliatives Coping* (-2,28, SD: 1,45) fiel geringer aus als jener der Skala *instrumentelles Coping* (-1,23, SD: 2,33).

10.9.2 Pseudoitems des Typs e

Eine extreme Antworttendenz konnte bei den *e*-Items nur für die Subskalen der Primärskalen *Auslösung* und *Coping* nachgewiesen werden.

10.9.2.1 Vergleich der Fragebogeninhalte der Primärskalen

Die extreme Antworttendenz konnte für die Primärskalen *Manifestation* und *Stabilisierung* nicht nachgewiesen werden. In diesen beiden Skalen werden (nach Tabelle 7) beobachtbare und nicht beobachtbare Reaktionen, Symptome, Einstellungen und Überzeu-

gungen sowie Reaktionen anderer gegenüber einer Person abgefragt. Die Inhalte der Fragen der Skalen *Auslösung* und *Coping* können als modifizierten Eigenschaftszuschreibungen sowie beobachtbares und nicht beobachtbares Verhalten kategorisiert werden. Die konkreten Inhalte der einzelnen Skalen wurden unter Punkt 10.9.1.1 bereits ausführlicher beschrieben.

10.9.2.2 Unterschiede der geschätzten Mittelwerte der Subskalen der Primärskala Auslösung

Die geschätzten Mittelwertsunterschiede für die drei Subskalen Existenz, Interaktion und Alltagsgeschehen hielten sich eher gering. Die wenigsten *e*-Antworten wurden in der Subskala *Interaktion* (Mittelwert: -,02, SD: 1,19) gegeben. Die geschätzten Mittelwerte für die Subskalen *Existenz* (Mittelwert: ,01, SD: 1,23) und *Alltagsgeschehen* (Mittelwert: ,01, SD: 1,23) lagen sehr nah beieinander.

10.9.2.3 Unterschiede der geschätzten Mittelwerte der Subskalen der Primärskala Coping

Wie auch schon für die *m*-Items wurde die geschätzten Mittelwerte der Subskalen der Primärskala *Coping* verglichen. Auch wenn die Unterschiede hier eher gering sind, fiel der geschätzte Mittelwert der Skala *palliatives Coping* (-,65, SD: 1,38) höher aus als jener für die Subskala *instrumentelles Coping* (-,14, SD: 1,23).

10.9.3 Mehrgruppenmodelle für die Primärskala Coping

Bezüglich der zusätzlichen Antwortmöglichkeiten für die Subskala *Coping* nach der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) wurden Gruppenmodelle berechnet. Die Gruppen wurden anhand der Scores auf die zusätzlichen Antwortmöglichkeiten berechnet. Die Testpersonen wurden hierzu in zwei Gruppen unterteilt: jene mit Scores die geringer oder gleich dem Mittelwert waren und jene, die höhere Scores erzielt hatten. Pro Pseudoitemtyp wurden vier Scores berechnet: drei für die Häufigkeit mit der eine Testperson eine der zusätzlichen Antwortmöglichkeiten gewählt hat (also einen Score für die Frage zur Einstellung gegenüber des Verhaltens (Score 1), einen für die subjektive Norm (Score 2), einen für die wahrgenommene Verhaltenskontrolle (Score 3) und einen der dann mit 1 kodiert wurde, wenn alle drei zusätzlichen Antwortmöglichkeiten die Theorie betreffend angekreuzt wurden (Score 4).

Mehrgruppenmodelle wurden für alle drei Pseudoitemtypen berechnet. Bei den *m*- und *e*-Items wurde die Hypothese aufgestellt, dass Personen, die niedrigere Scores (also $\leq$ Mittelwert) erreichten, und damit eher geringe Verhaltensintentionen für das abgefragte Verhalten zeigen sollten, häufiger Antworttendenzen nutzen und dementsprechend einen höheren, auf Basis des Modells geschätzten Mittelwert zeigen. Für die *d*-Items wurde angenommen, dass Personen mit niedrigen Werten auch geringere, auf Basis des Modells geschätzte Mittelwerte zeigen, da die Pseudoitems Zustimmung zur Skala (und damit zum abgefragten Verhalten) darstellen. Mittelwerte der einzelnen Scores über alle Testpersonen hinweg finden sich für die Primärskala *Coping* sowie die Subskala *instrumentelles Coping* in Tabelle 36.

Tabelle 36: Mittelwerte der Scores 1, 2, 3 und 4 der zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns.

	Score 1	Score 2	Score 3	Score 4
Mittelwert Primärskala (31 Items)	7,63 (SD 5,44)	6,38 (SD 5,46)	5,62 (SD 4,46)	3,41 (SD 3,72)
Mittelwert Subskala (14 Items)	3,94 (SD 2,75)	3,02 (SD 2,48)	2,81 (SD 2,42)	1,18 (SD 1,59)

Des Weiteren wurde, wie auch bei Khorramdel und von Davier (in Druck), eine Jackknife-Methode in die Berechnung der Gruppenmodelle implementiert, um die Unterschiede zwischen den Gruppen auf Signifikanz zu untersuchen. Die Idee der Jackknife-Methode besteht darin, eine oder mehrere Beobachtungen der vorliegenden Stichprobe auszulassen, um die statistische Schätzung systematisch neu zu berechnen. Die Jackknife-Methode ist weniger abhängig von Modellannahmen im Vergleich zu einer einfachen Schätzung der zufälligen Stichprobenvarianz und es wird eine bessere Annäherung an den Standardfehler erreicht (vgl. Shao & Tu, 1995). Das Jackknife wurde auf der Grundlage von 80 Pseudo-Datensätzen geschätzt. Um Konfidenzintervalle (KI) zu berechnen, wurden die Differenzen der durchschnittlichen Gruppenmittelwerte und der zugehörige quadrierte Standardfehler verwendet ($KI = \text{Differenz der durchschnittlichen Gruppenmittelwerte} \pm (2 \cdot \text{quadrierter Standardfehler})$). Das Jackknife-KI weist dann auf ein signifikantes Ergebnis hin, wenn es Null nicht überlappt (Khorramdel & von Davier, in Druck).

10.9.3.1 Mehrgruppenmodelle für die Pseudoitems des Typs m

Alle Gruppenmodelle wurden auf Basis des 2PL-Modells berechnet. Wie in Tabelle 37 dargestellt, entsprechen die Ergebnisse nicht der Hypothese, dass Leute mit höheren Werten und damit einer höheren angenommenen Verhaltensintention weniger Antworttendenzen zeigen als jene, die niedrigere Werte erzielten. Die geschätzten Mittelwerte der Personen mit höheren Scores waren durchgehend größer, als jene der Personen mit niedrigen Scores und sprechen damit dafür, dass sich mehr Antworttendenzen gezeigt haben. Die Unterschiede sind nach der Jackknife-Methode alle signifikant.

Auf Grund der Ergebnisse wurden die Scores auch noch separat für die Subskala *instrumentelles Coping* berechnet, da hier konkrete beobachtbare Handlungen abgefragt wurden, was der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) besser entspricht.

Tabelle 37: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs m für die Primärskala Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.

	<i>Einstellung gegen- über dem Verhalten</i>	<i>Subjektive Norm</i>	<i>wahrgenommene Verhaltenskontrolle</i>	<i>Theorie</i>
Niedriger Score				
<i>n</i>	330	318	354	342
<i>Geschätzter MW</i>	-2,43	-2,52	-2,42	-2,39
<i>SD</i>	1,46	1,61	1,48	1,57
Hoher Score				
<i>n</i>	277	289	253	274
<i>Geschätzter MW</i>	-0,74	-0,59	-0,71	-0,93
<i>SD</i>	1,72	1,46	1,74	1,87
Jackknife				
<i>Standardfehler</i>	0,18	0,18	0,18	0,25
<i>KI</i>	[-2,04; -1,30]	[-2,26; -1,55]	[-1,97; -1,25]	[-1,98; -0,96]

Tabelle 38 gibt die Ergebnisse wieder. Auch hier zeigten sich ähnliche Ergebnisse, nur für die zusätzliche Antwortmöglichkeit zur wahrgenommenen Verhaltenskontrolle konnte kein signifikantes Ergebnis beobachtet werden, aber auch hier war der geschätzte Mittelwert der Personen mit hohen Scores größer als jener der Personen mit niedrigen.

Tabelle 38: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs *m* für die Subskala instrumentelles Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.

	<i>Einstellung gegen- über dem Verhalten</i>	<i>Subjektive Norm</i>	<i>wahrgenommene Verhaltenskontrolle</i>	<i>Theorie</i>
Niedriger Score				
<i>n</i>	342	342	381	406
<i>Geschätzter MW</i>	-2,75	-2,86	-2,80	-2,89
<i>SD</i>	1,68	1,81	1,83	1,66
Hoher Score				
<i>n</i>	267	267	228	209
<i>Geschätzter MW</i>	-1,05	-0,70	-1,34	0,16
<i>SD</i>	2,17	1,60	2,76	1,21
Jackknife				
<i>Standardfehler</i>	0,27	0,96	0,92	0,33
<i>KI</i>	[-2,18; -1,10]	[-4,22; -0,40]	[-2,95; 0,74]	[-3,63; -2,31]

10.9.3.2 Mehrgruppenmodelle für die Pseudoitems des Typs *e*

Für die Pseudoitems des Typs *e* zeigte sich ein ähnliches Bild wie auch bei den Items des Typs *m*. Wie Tabelle 39 zu entnehmen ist, zeigten Personen mit niedrigen Scores – widersprüchlich zur aufgestellten Hypothese – weniger Antworttendenzen. Das Jackknife-KI deutet für den Score *Wahrgenommene Verhaltenskontrolle* und den Score *Theorie* auf ein signifikantes Ergebnis hin.

Tabelle 39: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs *e* für die Primärskala Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.

	<i>Einstellung gegen- über dem Verhalten</i>	<i>Subjektive Norm</i>	<i>wahrgenommene Verhaltenskontrolle</i>	<i>Theorie</i>
Niedriger Score				
<i>n</i>	330	318	354	342
<i>Geschätzter MW</i>	-0,34	-0,33	-0,42	-0,46
<i>SD</i>	1,36	1,31	1,34	1,33
Hoher Score				
<i>n</i>	277	289	253	274
<i>Geschätzter MW</i>	-0,25	-0,26	-0,15	-0,10
<i>SD</i>	0,97	1,02	0,97	0,97
Jackknife				
<i>Standardfehler</i>	0,11	0,11	0,11	0,10
<i>KI</i>	[-0,33; 0,13]	[-0,30; 0,16]	[-0,49; -0,05]	[-0,57; -0,16]

Tabelle 40: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs e für die Subskala instrumentelles Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.

	<i>Einstellung gegen- über dem Verhalten</i>	<i>Subjektive Norm</i>	<i>wahrgenommene Verhaltenskontrolle</i>	<i>Theorie</i>
Niedriger Score				
<i>n</i>	342	342	381	406
<i>Geschätzter MW</i>	0,03	0,06	-0,00	-0,11
<i>SD</i>	1,24	1,18	1,21	1,12
Hoher Score				
<i>n</i>	267	267	228	209
<i>Geschätzter MW</i>	-0,41	-0,43	-0,48	-0,55
<i>SD</i>	0,99	1,00	0,10	1,37
Jackknife				
<i>Standardfehler KI</i>	0,11 [0,22; 0,67]	0,11 [0,27; 0,70]	0,11 [0,27; 0,69]	0,16 [0,13; 0,79]

Für die Subskala *instrumentelles Coping* entsprachen die Ergebnisse der aufgestellten Hypothese. Eine Darstellung findet sich in Tabelle 40. Hier zeigten Personen, die seltener eine der zusätzlichen Antwortmöglichkeiten bezüglich der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) gewählt hatten, mehr Antworttendenzen, als Personen die sie häufiger gewählt hatten. Die Jackknife-KIs weisen bei allen Modellen auf ein signifikantes Ergebnis hin.

10.9.3.3 Mehrgruppenmodelle für die Pseudoitems des Typs d

Die Ergebnisse der Mehrgruppenmodelle für die Pseudoitems des Typs *d* entsprachen, wie man den Tabellen 41 und 42 entnehmen kann, sowohl für die gesamte Skala *Coping* als auch für die Subskala *instrumentelles Coping* den Hypothesen. Personen mit niedrigen Scores hatten einen geringeren Mittelwert. Das heißt, Personen, die die zusätzlichen Antwortmöglichkeiten bezüglich der Theorie des geplanten Handelns seltener auswählten, stimmten den Items der Skala *Coping* weniger häufig zu.

Tabelle 41: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs d für die Primärskala Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.

	<i>Einstellung gegen- über dem Verhalten</i>	<i>Subjektive Norm</i>	<i>wahrgenommene Verhaltenskontrolle</i>	<i>Theorie</i>
Niedriger Score				
<i>n</i>	330	318	354	342
<i>Geschätzter MW</i>	-0,09	-0,00	-0,03	0,05
<i>SD</i>	0,97	0,99	0,99	1,00
Hoher Score				
<i>n</i>	277	289	253	274
<i>Geschätzter MW</i>	0,31	0,30	0,28	0,38
<i>SD</i>	0,47	0,71	0,47	0,77
Jackknife				
<i>Standardfehler KI</i>	0,07 [-0,53; -0,25]	0,09 [-0,48; -0,12]	0,08 [-0,47; -0,15]	0,09 [-0,51; -0,15]

Tabelle 42: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs d für die Subskala instrumentelles Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.

	<i>Einstellung gegen- über dem Verhalten</i>	<i>Subjektive Norm</i>	<i>wahrgenommene Verhaltenskontrolle</i>	<i>Theorie</i>
Niedriger Score				
<i>n</i>	342	342	381	406
<i>Geschätzter MW</i>	-1,36	-1,20	-1,27	-0,79
<i>SD</i>	1,09	1,71	1,08	1,25
Hoher Score				
<i>n</i>	267	267	228	209
<i>Geschätzter MW</i>	0,32	0,27	0,34	-0,26
<i>SD</i>	0,51	0,99	0,59	1,26
Jackknife				
<i>Standardfehler KI</i>	0,12 [-1,92; -1,45]	0,17 [-1,84; -1,14]	0,10 [-1,79; -1,41]	0,16 [-0,85; -0,22]

11 Diskussion

Es wurde unter Anwendung eines von Böckenholt (2012) sowie Khorramdel und von Davier (in Druck) und von Davier & Khorramdel (in Druck) vorgestellten IRT-Ansatzes untersucht, ob Antworttendenzen (Tendenz zur Mitte, Tendenz zu extremen Antworten) in den im Rahmen der Diplomarbeit erhobenen Daten gefunden werden können. Die Daten wurden mittels eines Online-Self-Assessments erfasst. Testpersonen konnten dort

die Fragen des DSIHR (Differentiellen Stressinventar HR) mit einer fünfstufigen Antwortskala bearbeiten und erhielten zusätzlich Fragen bezüglich der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991). Es wurde untersucht, ob Antworttendenzen (Tendenz zur Mitte, Tendenz zu extremen Antworten) eindimensional erfassbar sind oder zusätzlich die Fragebogenskalen messen, und ob bei Vorhandensein dieser Antworttendenzen der Fragebogeninhalt bzw. die persönliche Einstellung zu diesem einen Einfluss haben. Die Daten wurden zu binäre Pseudoitems umkodiert, die jeweils einen latenten Antwortsubprozess darstellen sollen (Böckenholt, 2012). Es wurden IRT-Analysen mit multidimensionalen und unidimensionalen Antworttendenzfaktoren (Böckenholt, 2012; Khorramdel & von Davier, in Druck) sowie Bifactor-Modellen und Mehrgruppenmodellen (von Davier & Khorramdel, in Druck) durchgeführt.

11.1 Multidimensionale Antworttendenzfaktoren

In den Untersuchungen mittels multidimensionalen Antworttendenz- und Merkmalsfaktoren konnte wie bei Khorramdel und von Davier (in Druck) aufgezeigt werden, dass die Daten besser durch jene Modelle beschrieben werden konnten, in denen die Pseudoitems den jeweiligen Antworttendenzfaktoren zugeordnet wurden, als wenn sie den jeweiligen Primär- bzw. Subskalen zugeteilt wurden. Diese Ergebnisse weisen darauf hin, dass Antworttendenzen in den Daten vorhanden sind. Allerdings zeigte sich sowohl über den gesamten Fragebogen hinweg, als auch in der Primärskala *Auslösung*, dass die Modellpassung nicht erhöht werden konnte, wenn die *d*-Items den entsprechenden Primär- bzw. Subskalen zugeordnet wurden (während die *e*-Items und *m*-Items weiterhin den Antworttendenzen zugeordnet wurden). Dies könnte daran liegen, dass die laut DSIHR-Manual intendierte Faktorenlösung auf die erhobenen Daten nicht zutrifft oder die *d*-Items nicht die Faktoren des Persönlichkeitsfragebogens messen. Für die Primärskalen *Manifestation*, *Coping* und *Stabilisierung* konnten allerdings hypothesenkonforme Ergebnisse gefunden werden. In allen drei Primärskalen passte jenes Modell, in welchem die *m*- und *e*-Items jeweils einem Antworttendenzfaktor und die *d*-Items den entsprechenden Subskalen zugeordnet waren, am besten.

11.2 Unidimensionaler Antworttendenzfaktoren

Nachfolgend werden die Ergebnisse der Analysen bezüglich unidimensionaler Antworttendenzfaktoren einzeln für jeden Pseudoitemtyp diskutiert.

11.2.1 *Pseudoitems des Typs m*

Die Modellvergleiche für die *m*-Pseudoitems (also das 1-dimensionale Modell mit der Zuordnung der Pseudoitems des Typs *m* zu einem Antworttendenzfaktor verglichen mit den 4- bzw. 9-dimensionalen Modellen mit der Zuordnung der Pseudoitems des Typs *m* zu den 4- bzw. 9 Primär- bzw. Subskalen als Faktoren) zeigten, dass das unidimensionale IRT-Modell die Daten am besten beschreiben kann und sprechen dafür, dass die Pseudoitems *m* die mittlere Antworttendenz messen. Die Ergebnisse lagen allerdings sehr knapp beieinander, deshalb ist trotzdem nicht auszuschließen, dass die mittlere Antworttendenz nicht eindimensional gemessen wurde, sondern möglicherweise auch durch die abgefragten psychologischen Konstrukte beeinflusst wird. Des Weiteren zeigt der GMT, dass sich die Itemschwierigkeiten einzelner Items in den zwei fraglichen Gruppen (geteilt nach Median des Testscores) stark unterschieden. Die auf dem 1-dimensionalen IRT-Modell basierende Messung der mittleren Antworttendenz zeigen eine hohe marginale Reliabilität. Auf Primärskalenebene konnten die *m*-Items nur im Fall der Primärskala *Stabilisierung* nicht besser durch einen unidimensionalen Antworttendenzfaktor, als durch die entsprechenden Subskalen beschrieben werden, dies könnte darauf hinweisen, dass diese Skala auf Grund des abgefragten Inhaltes weniger anfällig für Antworttendenzen ist.

11.2.2 *Analysen der Pseudoitems des Typs e*

Der unidimensionale Antworttendenzfaktor konnte die *e*-Items (also das 1-dimensionale Modell mit der Zuordnung der Pseudoitems des Typs *e* zu einem Antworttendenzfaktor) nicht besser beschreiben als das 4-dimensionale Modell (Zuordnung der Pseudoitems des Typs *e* zu den 4 Primärskalen als Faktoren). Das 9-dimensionale Modell (Zuordnung der Pseudoitems des Typs *e* zu den 9 Subskalen als Faktoren) hatte zwar eine schlechtere Modellpassung, hat aber auf Grund der ungenauen Schätzung weniger Aussagekraft. Der GMT deutet für die Itemschwierigkeiten des 1-dimensionalen

Modells der *e*-Items darauf hin, dass einige Items sich in den beiden Gruppen (geteilt nach Median des Testscores) in ihrer Schwierigkeit systematisch unterscheiden. Nach Ausschluss von 29 Items, die sich im GMT auffällig darstellten, konnte keine bessere Modellpassung erreicht werden, weshalb hier davon ausgegangen wird, dass entgegen der aufgestellten Hypothese die Pseudoitems des Typs *e* die extreme Antworttendenz nicht eindimensional messen. Die marginale Reliabilität des 1-dimensionalen Modells zeigte sich aber auch hier hoch. Auf Primärskalenebene konnte für die Skalen *Auslösung* und *Coping* eine bessere Modellpassung für den unidimensionalen Antworttendenzfaktor gefunden werden. Die Primärskalen *Manifestation* und *Stabilisierung* wurden durch die Fragebogenskalen als Faktoren besser beschrieben.

11.2.3 Analysen der Pseudoitems des Typs *d*

Die *d*-Items konnten – bis auf den Vergleich mit dem 9-dimensionalen Modell (alle Pseudoitems des Typs *d* wurden den 9 Subskalen zugeordnet) – der Hypothese entsprechend immer besser durch die Primär- bzw. Subskalen als Faktoren beschrieben werden als durch einen eindimensionalen Faktor. Es kann in Anbetracht der Tatsache, dass das 9-dimensionale Modell nicht genau geschätzt werden konnte, davon ausgegangen werden, dass die *d*-Items die intendierten Konstrukte des DSIHR messen. Es zeigten sich hohe, auf dem 4-dimensionalen Modell (Zuordnung der *d*-Items zu den vier Primärskalen als Faktoren) basierenden marginale Reliabilitäten pro Primärskala.

Die Skaleninterkorrelationen der *d*-Items waren allerdings – außer für die Primärskala *Stabilisierung* – nicht wesentlich geringer als jene der *m*- und *e*-Items. Im Falle, dass die *d*-Items die psychologischen Konstrukte des Fragebogens messen, sollten sie nach der aufgestellten Hypothese geringere Skaleninterkorrelationen haben, da sie im Gegensatz zu den Antworttendenzen (von denen angenommen wird, dass sie unabhängig vom Iteminhalt über alle Skalen hinweg auftreten), unterschiedliche Merkmale in den einzelnen Skalen des Fragebogens erfassen.

11.3 Bifactor-Modelle

Auf Grund der oben diskutierten Ergebnisse der unidimensionalen Messungen der Antworttendenzen für die *m*- und *e*- Pseudoitems wurden, wie bei von Davier und Khorram-

del (in Druck), Bifactor-Modelle berechnet, die untersuchen sollten, ob die Pseudoitems m und e sowohl Antworttendenzen als auch die intendierten Fragebogenskalen messen.

11.4 Pseudoitems des Typs m

Für die Analysen des Bifactor-Modells für die Pseudoitems des Typs m (alle Pseudoitems des Typs m wurden einem Antworttendenzfaktor und den Primärskalen zugeordnet) konnte keine konkrete Aussage getroffen werden, da das AIC und das BIC jeweils etwas Gegenteiliges aussagen. Die meiste Varianz wurde durch den Generalfaktor (Antworttendenz) erklärt. Das Bifactor-Modell, in dem die Pseudoitems des Typs m dem Generalfaktor (Antworttendenz) und den neun Subskalen als spezifische Faktoren zugeordnet waren, zeigte keine bessere Modellpassung als das 1-dimensionale Modell (die meiste Varianz wurde durch den Generalfaktor erklärt) und spricht demnach gegen die Hypothese, dass durch die Pseudoitems des Typs m sowohl die Antworttendenz als auch das intendierte Konstrukt des Fragebogens erfasst wird. Hier muss angemerkt werden, dass das Bifactor-Modell mit den Subskalen als spezifische Faktoren sehr rechenintensiv war und dementsprechend mit einer geringeren Genauigkeit geschätzt werden musste. Eine klare Aussage, ob die mittlere Antworttendenz die m -Pseudoitems unidimensional beschreibt oder das Bifactor-Modell besser passt, kann in Anbetracht der nicht eindeutig interpretierbaren Ergebnisse bzw. auf Grund der ungenauen Messung nicht getroffen werden.

Auf Primärskalenebene der Pseudoitems des Typs m konnten für die Bifactor-Modelle mit den Subskalen der Primärskala *Stabilisierung* als spezifischen Faktoren und der Antworttendenz als Generalfaktor ebenfalls keine konkreten Aussagen getroffen werden, da das AIC und das BIC etwas Gegenteiliges aussagen. Für die Bifactor-Modelle der Primärskalen *Auslösung*, *Coping* und *Manifestation* konnte die meiste Varianz jeweils durch den Generalfaktor (Antworttendenz) erklärt werden, die Daten konnten durch das Modell aber nicht besser beschrieben werden als durch die unidimensionale Messung. Für diese drei Primärskalen liegt also die Annahme nahe, dass die mittlere Antworttendenz eindimensional erfasst wird.

11.5 Pseudoitems des Typs e

Für das Bifactor-Modell in dem alle Pseudoitems des Typs *e* jeweils einem Generalfaktor (Antworttendenz) sowie den vier Primärskalen als spezifische Faktoren zugeordnet waren, konnte ebenfalls keine konkrete Aussage getroffen werden, da die Informationskriterien (AIC und BIC) gegenteilige Ergebnisse erbrachten. Das Bifactor-Modell mit den neun Subskalen als spezifische Faktoren (auch hier ergab sich das oben erwähnte Problem der intensiven Rechenleistung auf Kosten der Genauigkeit der Schätzung) und der Antworttendenz als Generalfaktor zeigte keine bessere Modellpassung als das 1-dimensionale Modell. Bei diesen beiden Modellen konnte die meiste Varianz nicht durch den Generalfaktor (Antworttendenz) erklärt werden. Eine konkrete Aussage bezüglich der Hypothese, ob durch die Pseudoitems des Typs *e* neben der Antworttendenz auch noch die intendierten Konstrukte des Fragebogens erfasst werden, kann aber nicht getroffen werden.

Auf Ebene der Primärskalen der Pseudoitems des Typs *e* zeigten sich für *Manifestation*, *Coping* und *Stabilisierung*, dass die Bifactor-Modelle mit den entsprechenden Subskalen als spezifische Faktoren und der Antworttendenz als Generalfaktor eine bessere Modellpassung zeigten als die *simple structure* Modelle. Allerdings konnte nur im Fall der Primärskala *Coping* auch die meiste Varianz durch den Generalfaktor erklärt werden. Das Bifactor-Modell mit den Subskalen der Primärskala *Auslösung* als spezifische Faktoren und der Antworttendenz als Generalfaktor zeigte eine schlechtere Modellpassung, als das entsprechende *simple structure* Modell (die meiste Varianz wurde durch den Generalfaktor erklärt).

11.6 Fragebogeninhalt

11.6.1 Pseudoitems des Typs m

Die Tatsache, dass in der Skala *Stabilisierung* keine mittlere Antworttendenz nachgewiesen werden konnte, während sie aber in den drei anderen Primärskalen sowie dem gesamten Fragebogen auftrat, kann als Hinweis darauf gedeutet werden, dass der Fragebogeninhalt einen Einfluss auf die Auftretenshäufigkeit der fraglichen Antworttendenz haben könnte. Was die Primärskala *Stabilisierung* von den anderen drei

Skalen unterscheidet ist, dass einerseits natürlich stabilisierende Faktoren des Stresserlebens abgefragt werden, andererseits lassen sich die Iteminhalte (neben nicht beobachtbaren Reaktionen, die auch zum Teil in den anderen Primärskalen abgefragt werden) in Einstellungen, Überzeugungen und Reaktionen anderer gegenüber einer Person kategorisieren. Eine weitere Erklärungsmöglichkeit für das Zustandekommen der Ergebnisse wäre aber auch, dass die Fragen der Primärskala *Stabilisierung* als vollständige Aussage dargeboten und nicht – wie in den anderen drei Primärskalen – jeweils mit demselben Satzanfang dargestellt werden, der nur einmal am Anfang der Skala eingeblendet wird.

Die auf Basis des jeweiligen Modells geschätzten Mittelwerte für die Pseudoitems des Typs *m* geben ebenfalls Hinweise darauf, dass beim Vorhandensein von Antworttendenzen in den Daten der Fragebogeninhalt einen Einfluss auf die mittlere Antworttendenz haben kann. Der stärkste Mittelwertsunterschied zeigt sich zwischen den Subskalen der Primärskala *Manifestation*. Die Subskala *physiologische Manifestationen* (Stärke der körperlichen Beanspruchung durch Stress, Selbstbeschreibung: Symptome) scheint weniger anfällig für die mittlere Antworttendenz zu sein als die Subskala *emotional-kognitive Manifestationen* (Stärke der emotional-kognitiven Belastungen durch Stress, Selbstbeschreibung: beobachtbare und nicht beobachtbare Reaktionen). Insgesamt scheint die Primärskala *Manifestation* (Häufigkeit von auftretenden *Stressmanifestationen*) auch am wenigsten anfällig für die mittlere Antworttendenz zu sein, verglichen mit den anderen drei Primärskalen. Auch für die Subskalen der Primärskalen *Auslösung* zeigten sich Mittelwertsunterschiede. In der Subskala *Existenz* (Ausmaß der Belastung durch Zukunftssorgen und Existenzängste, modifizierte Eigenschaftszuschreibung) zeigen sich weniger häufig *m*-Antworten, als in den beiden Subskalen *Interaktion* (Ausmaß der Belastung durch Interaktion mit Mitmenschen, modifizierte Eigenschaftszuschreibung) und *Alltagsgeschehen* (Ausmaß der Belastung durch unvermeidliche Angelegenheiten des Alltags, modifizierte Eigenschaftszuschreibung). Die geringsten, aber trotzdem erwähnenswerten Mittelwertsunterschiede finden sich schließlich für die Subskalen der Primärskalen *Coping*. In der Subskala *palliatives Coping* (Vorhandensein von positiven Emotionen und Kognitionen zur Bewältigung von Beanspruchung, Selbstbeschreibung:

nicht beobachtbare Reaktionen) konnten weniger *m*-Antworten gefunden werden als in der Subskala *instrumentelles Coping* (aktive Vorgehensweisen zur Bewältigung von Beanspruchung, Selbstbeschreibung: beobachtbare Reaktionen).

Bezüglich der Mehrgruppenmodelle zeigten sich zur Hypothese widersprüchliche, signifikante Ergebnisse, sowohl für die gesamte Skala *Coping*, als auch für die Subskala *instrumentelles Coping*. Personen, die öfter angaben das fragliche Verhalten schon einmal ausgeführt zu haben (Score 1), für das Verhalten Akzeptanz durch Mitmenschen zu erhalten (Score 2) und/oder dass ihnen die Ausführung der Handlung leicht fällt (Score 3 bzw. 4, wenn alle drei Antworten angekreuzt wurden), gaben gegensätzlich zur Annahme häufiger *m*-Antworten. Im Anhang finden sich Grafiken der Verteilung der demographischen Daten auf die beiden Gruppen (hoher und niedriger Score, gesamte Skala *Coping*). Da diese relativ ähnlich ausfielen, ist eine Begründung für die Ergebnisse durch demographische Unterschiede in den Gruppen eher unwahrscheinlich. In der Studie von Gibbons et. al. (1999), zeigte sich, wie im Theorieteil beschrieben, dass extreme Antworttendenzen häufiger auftreten, wenn der Fragebogeninhalt als sinnvoll eingestuft wurde, eventuell lassen sich diese Ergebnisse auch auf die mittlere Antworttendenz umlegen.

11.6.2 Pseudoitems des Typs e

In den Subskalen *Manifestation* und *Stabilisierung* konnten keine extremen Antworttendenzen nachgewiesen werden, allerdings zeigte sich hier auch für den gesamten Fragebogen keine Tendenz zu extremen Urteilen. In den Primärskalen *Auslösung* und *Coping* hingegen konnten die Pseudoitems des Typs *m* am besten durch den Antworttendenzfaktor erklärt werden. Auch dies kann als Hinweis eines Einflusses des Fragebogeninhalts auf das Auftreten von Antworttendenzen gewertet werden. Auch hier gilt für die Skala *Stabilisierung*, dass sie sich auf Grund des konkret abgefragten Inhalts (stabilisierende Faktoren von Stress) und durch die Kategorisierung des Iteminhalts in Einstellungen, Überzeugungen und Reaktionen anderer gegenüber einer Person von den anderen Primärskalen unterscheidet. Die Möglichkeit, dass das Ergebnis auf Grund der Tatsache entstanden ist, dass die Fragen der Skala – im Vergleich zu den anderen Primärskalen – als vollständige Aussage dargestellt werden und nicht jeweils

mit demselben Satzanfang beginnen, der nur einmal am Anfang der Skala eingeblendet wird, besteht auch hier. In der Primärskala *Manifestation* wird im Vergleich zu den anderen Skalen abgefragt, welche Manifestationen des Stresserlebens Personen zeigen. Außerdem fallen einige Items der Skala in die Kategorie „Abfragen von Symptomen“ und unterscheiden sich damit ebenfalls von den anderen Skalen.

Für die Pseudoitems des Typ *e* konnten keine nennenswerten Mittelwertsunterschiede für die Subskalen der Primärskalen *Auslösung* und *Coping* gefunden werden. Hier zeigten sich keine Hinweise darauf, dass der Fragebogeninhalt einen Einfluss auf die Häufigkeit des Auftretens der extremen Antworttendenz hatte.

Die Mehrgruppenmodelle der Pseudoitems des Typs *e* zeigten für die Gesamtskala *Coping* ebenfalls Ergebnisse, die nicht der Hypothese entsprachen. Möglicherweise entsprechen die Ergebnisse hier denen von Gibbons et. al. (1999). Dadurch, dass eher eine Verhaltensintention vorhanden ist, könnte auch die wahrgenommene Sinnhaftigkeit der Items erhöht worden sein. Allerdings waren in diesem Fall die Ergebnisse für Score 3 und Score 4 nicht signifikant. Für die Subskala *instrumentelles Coping* konnten der Hypothese entsprechende, signifikante Ergebnisse gefunden werden. Personen mit hohen Scores zeigten weniger häufig extreme Antworttendenzen.

11.6.3 Pseudoitems des Typs *d*

Die Ergebnisse der Mehrgruppenmodelle für die Pseudoitems des Typs *d* entsprachen sowohl für die gesamte Skala *Coping*, als auch für die Subskala *instrumentelles Coping* der Hypothese. Personen mit hohen Scores und dementsprechend einer höheren angenommenen Verhaltensintention gaben mehr *d*-Antworten und zeigten damit einen höhere Zustimmung für die Inhalte der Fragen der Skala *Coping*.

11.7 Kritik

Auf Grund des im Vorfeld nicht bedachten hohen Rechenaufwands für die 9- bzw. 11-dimensionale IRT-Modelle sind wichtige Informationen zur Interpretation der aufgestellten Hypothesen verloren gegangen.

Außerdem muss für die Mehrgruppenmodelle an dieser Stelle angemerkt werden, dass die zusätzlichen Fragen auf denen diese begründet sind, eher unspezifisch gestellt wurden und daher eventuell nicht die Verhaltensintention gemessen haben. Es wären Unter-

suchungen mit auf die Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) passenderen Items sowie konkreter formulierte Antwortmöglichkeiten notwendig, um hier genauere Ergebnisse zu erhalten.

Insgesamt sind weitere Untersuchungen bezüglich des Einflusses des Fragebogeninhaltes auf in den Daten vorhandene Antworttendenzen nötig, um genauere Aussagen machen zu können. Es könnte ein eigenes Erhebungsinstrument für diesen Zweck erstellt werden, in dem klar getrennte Iteminhalte abgefragt werden, um konkretere Ergebnisse zu erzielen.

12 Zusammenfassung

In dieser Diplomarbeit wurde mittels eines neuen IRT-Ansatzes (Böckenholt, 2012; Khorramdel & von Davier, in Druck, von Davier & Khorramdel in Druck) untersucht, ob die Tendenz zur Mitte sowie die extreme Antworttendenzen in den erhobenen Daten gefunden werden können, diese eindimensional erfassbar sind oder zusätzlich die Fragebogenskalen messen, und ob bei Vorhandensein dieser Antworttendenzen der Fragebogeninhalt bzw. die persönliche Einstellung zu diesem einen Einfluss haben. Um die IRT-Analysen durchführen zu können wurden binäre Pseudoitems aus den Daten kodiert, die einen latenten Antwortsubprozess darstellen sollen (Böckenholt, 2012). Es wurden IRT-Analysen mit multidimensionalen und unidimensionalen Antworttendenzfaktoren (Böckenholt, 2012; Khorramdel & von Davier, in Druck) sowie Bifactor-Modellen und Mehrgruppenmodellen (von Davier & Khorramdel, in Druck) durchgeführt.

Dafür wurde mittels eines Online-Fragebogens das DSIHR (Lefevre und Kubinger, 2004) vorgegeben, sowie zusätzliche Fragen bezüglich der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) gestellt. Es wurde angenommen, dass Antworttendenzen sowohl multidimensional als auch unidimensional in den Daten nachgewiesen werden können. Sollte es bei den unidimensionalen Messungen Hinweise darauf geben, dass die Pseudoitems nicht eindimensional messen, wurde die Hypothese aufgestellt, dass die jeweiligen Pseudoitems besser durch Bifactor-Modelle, die es den Daten erlauben, sowohl auf einen Generalfaktor als auch auf spezifischen Faktoren zu laden, erklärt werden können. Sollten sich Antworttendenzen in den Daten zeigen,

wurden Unterschiede in deren Auftreten anhand geschätzter Mittelwerte (auf Basis des jeweiligen Modells) verglichen, sowie die Ergebnisse von Mehrgruppenmodellen (aufgeteilt nach hohen bzw. niedrigen Scores in den Zusatzfragen nach der Theorie des geplanten Handelns) betrachtet.

Wie bei Khorramdel und von Davier (in Druck) konnte auch in dieser Untersuchung mittels *multidimensionalen Antworttendenz- und Merkmalsfaktoren* gezeigt werden, dass Antworttendenzen in den Daten vorhanden sind. Allerdings schienen die *d*-Items über die gesamten Daten hinweg nicht die intendierten DSIHR Skalen zu messen. Auf Primärskalenebene konnten allerdings hypothesenkonforme Ergebnisse gefunden werden. Bei den unidimensionalen Messungen der *d*-Items konnte weitgehend gezeigt werden, dass die *d*-Items durch die Fragebogenskalen als Faktoren besser beschrieben werden konnten, allerdings fielen die Skaleninterkorrelationen nicht wie angenommen niedrig aus. Es kann also vermutet werden, dass die *d*-Items nicht nur die Fragebogenskalen messen oder dass die intendierte Dimensionalität des verwendeten Fragebogens auf die erhobenen Daten nicht zutrifft.

Die Ergebnisse der *unidimensionalen Messungen* der Antworttendenzen lassen vermuten, dass die extreme und mittlere Antworttendenz durch die Pseudoitems nicht eindimensional dargestellt werden können. Die Messungen weisen aber, wie bei Khorramdel und von Davier (in Druck) sowie von Davier und Khorramdel (in Druck), hohe marginale Reliabilitäten auf. Eine Erklärungsmöglichkeit dafür, dass die Pseudoitems *m* und *e* Antworttendenzen nicht eindimensional erfassen ist, dass sie ebenfalls die Fragebogenskalen messen, weshalb wie bei von Davier und Khorramdel (in Druck) Bifactor-Modelle berechnet wurden. Bei den Pseudoitems des Typs *m* kann die Hypothese, dass die *m*-Pseudoitems am besten durch das Bifactor-Modell erklärt werden zwar nicht verworfen werden (da sich teilweise widersprüchliche Ergebnisse für die beiden Informationskriterien AIC und BIC zeigten), aber es gibt Hinweise darauf, dass die unidimensionale Messung die Daten genauer beschreibt, da die 1-dimensionalen Messungen auf Primärskalenebene für drei von vier Primärskalen eine bessere Modellpassung aufwiesen als die Bifactor-Modelle, und für die vierte konnte auf Grund von widersprüchlichen AIC- und BIC-Werten keine konkrete Aussagen getroffen werden. Des Weiteren wurde die meiste Varianz – außer im Fall der Primärskala *Stabilisierung* – immer durch den Generalfaktor

erklärt. Für die *e*-Items konnte die Hypothese für die Gesamtdaten ebenfalls nicht beantwortet werden. Auf Primärskalenebene zeigte sich allerdings bei drei von vier Skalen eine bessere Passung des Bifactor-Modells. Die meiste Varianz konnte nur im Fall der Primärskalen *Auslösung* und *Coping* durch den Generalfaktor erklärt werden.

Die Untersuchungen bezüglich des *Fragebogeninhalts* ergaben, dass sich Antworttendenzen nicht in allen Primärskalen zeigten, dies könnte ein Hinweis darauf sein, dass der Fragebogeninhalt einen Einfluss auf das Auftreten dieser hat. Bei Vorhandensein der mittleren Antworttendenz in den Daten zeigten sich sowohl Unterschiede in der Häufigkeit der mittleren Antworten für verschiedene Fragebogeninhalte als auch Hinweise darauf, dass die persönliche Einstellung zum Fragebogeninhalt in den Daten einen Einfluss auf deren Auftretenshäufigkeit haben könnte (auch wenn die signifikanten Ergebnisse der Mehrgruppenmodelle umgekehrt zu den aufgestellten Hypothesen ausfielen).

Für die *e*-Items konnten in Skalen, in denen eine extreme Antworttendenz nachgewiesen wurde, keine Unterschiede der Häufigkeit der *e*-Antworten in verschiedenen Fragebogenskalen gefunden werden. Es zeigten sich allerdings auch hier signifikante Ergebnisse für die Mehrgruppenmodelle. Für zukünftige Studien bezüglich der Theorie des geplanten Handelns (Ajzen, 1985; Ajzen, 1991) wurde angeraten, für die Theorie passendere Items zu verwenden und die Fragen konkreter zu Stellen. Insgesamt sind mehr Studien zum Einfluss des Fragebogeninhalts auf die Auftretenshäufigkeit von Antworttendenz nötig, um spezifischere Aussagen treffen zu können.

13 Literaturverzeichnis

- Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior*. In J. Kuhl, & J. Beckmann (Hrsg.), Springer series in social psychology (S. 11-39). Berlin: Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Angleitner, A., John, O. P. & Löhr, F. J. (1986). It's What You Ask and How You Ask it. An Itemmetric Analysis of Personality Questionnaires. In A. Angleitner & J. Wiggins (Hrsg.) *Personality Assessment via Questionnaires* (S. 61–108). Berlin:Springer.
- Arnold, W., Eysenck, H. J. & Meili, R. (Hrsg.). (1980). *Lexikon der Psychologie*. Freiburg: Herder.
- Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, 40, 1235-1245.
- Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, 48, 491-509.
- Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: a cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Billet, J. B. & McCledon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M., & Novick, M. R. (Hrsg.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Bock, R. D. (1972). Estimating itemparameters and latent ability when responses are scored two or more nominal categories. *Psychometrika*, 37, 29-51.

- Böckenholt, U. (2012). Modeling Multiple Response Processes in Judgment and Choice. *Psychological Methods*. Advance online publication.
- Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, 33, 335-352.
- Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, 71, 814-833.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae*. Handanweisung. Göttingen: Hogrefe.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153-168.
- Buckley, J. (2009, June). *Cross-national response styles in international educational assessments: Evidence from PISA 2006*. NCES Conference on the Program for International Student Assessment: What we can learn from PISA, Washington DC. Heruntergeladen von: <http://edsurveys.rti.org/PISA/>
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Education Deutschland GmbH.
- Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6, 170-75.
- Davidson, A. R. & Jaccard, J. J. (1979). Variables that moderate the attitude-behavior relation: Results of a longitudinal survey. *Pres. Soc. Psychol*, 37, 1364-1376.
- De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P. & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, 45, 104-115.
- Dolnicar, S. & Grun, B. (2009). Response Style Contamination of Student Evaluation Data. *Journal of Marketing Education*, 31, 160-172.
- Fienberg, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. *Journal of the American Statistical Association*, 65, 1610-1616.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

- Gibbons, J. L., Zelner, J. A. & Reduk, D. J. (1999). Effects of language and meaningfulness on the use of extreme response styles by Spanish-English bilinguals. *Cross-Cultural Research*, 33, 369-381.
- Gibbons, R. D., & Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423-436.
- Goodman, L. A. (1994). On quasi-independence and quasi-dependence in contingency tables, with special reference to ordinal triangular contingency tables. *Journal of the American Statistical Association*, 89, 1059-1063.
- Greenleaf, Eric A. (1992), Measuring Extreme Response Styles. *Journal of Marketing Research*, 29, 176-188.
- Grimm, D. D., & Church, A. T. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, 33, 415-441.
- Hamamura, T., Heine, S. J., & Paulhus, D. L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences*, 44, 932-942.
- Hee Seo, M., Xu, X. & von Davier M. (2007, 2009). *Software for Multidimensional Discrete Latent Trait Models. Manual*.
- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 296-09.
- Johnson, T. R., & Bolt, D. M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style. *Journal of Educational and Behavioral Statistics*, 35, 92-114.
- Johnson, T., Kulesa, P., Cho, Y. L. & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36, 264-277.
- Khorramdel, L. & von Davier, M. (in Druck). Measuring Response Styles across the Big Five: A Multiscale Extension of an Approach using Multinomial Processing Trees. *Multivariate Behavioral Research*.
- Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: Eine Einführung in eRm*. Wien: Facultas Verlags- und Buchhandels AG.
- Krosnick, J. A. (1999). Survey Research. *Annual Reviews Psychology*. 50, 537-567.
- Kubinger, K. D. (2009). *Psychologische Diagnostik*. Göttingen: Hogrefe Verlag GmbH & Co.KG.

- Lefevre, S. & Kubinger, K. D. (2004). *Differentielles Stressinventar HR. Manual*. Mödling: Schuhfried GmbH.
- Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education?. *Personality and Individual Differences*, 44, 1539-1550.
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, 37, 277-302.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Testkonstruktion*. Berlin, Heidelberg: Springer-Verlag.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Hrsg.), *Measures of personality and social psychological attitudes* (S. 17-59). San Diego, CA: Academic Press.
- Podsakoff, P. M., MacKenzie S. B., Lee, J. & Podsakoff, N. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88, 879-903.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche (Expanded Edition, Chicago, University of Chicago Press, 1980).
- Rost, D. H. & Schermer, F. J. (1987). Auf dem Weg zu einer differentiellen Diagnostik der Leistungsangst. *Psychologische Rundschau*, 38, 14-36.
- Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion*. Bern: Verlag Hans-Huber.
- Rost, J., Carstensen, C. H. & von Davier, M. (1999). Sind die Big Five Raschskalierbar? Eine Reanalyse der NEO-FFI-Normierungsdaten. *Diagnostica*, 45 (3), 119-127.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34(4, Teil 2).
- Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Hrsg.), *Handbook of modern item response theory* (S.85-100). New York, NY: Springer.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of statistics*, 6, 461-464.
- Seiwald, B. (2003). Antworttendenzen. In K. Kubinger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 29-32). Weinheim, Basel, Berlin: Beltz Verlag.

- Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap*. New York, NY: Springer-Verlag, Inc.
- Tränkle, U. (1983). Fragebogenkonstruktion. In: H. Feger & J. Bredekamp (Hrsg.). *Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band 2: Datenerhebung* (S. 222-301). Göttingen: Hogrefe.
- Van Herk, H., Poortinga, Y. H. & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, 35, 346-360.
- Van Rosmalen, J., Van Herk, H. & Groenen, P. J. F. (2010). Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model. *Journal of Marketing Research*, XLVII, 157-172.
- Van Vaerenbergh, Y. & Thomas, D. T. (2012). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*, 1-23.
- von Davier, M. (2010). Hierarchical mixtures of diagnostic models. *Psychological Test and Assessment Modeling*, 52, 8-28.
- von Davier, M. & Khorramdel, L. (in press). The problem of differentiating between response styles and construct related responses: A new IRT approach using bifactor and second-order IRT models. In: *New Developments in Quantitative Psychology: Presentations from the 77th Annual Psychometric Society Meeting*, New York: Springer.
- von Davier, M., Rost, R., & Carstensen, C. H. (2007). Introduction: Extending the Rasch Model. In M. von Davier & C. H. Carstensen (Hrsg.), *Multivariate and Mixture Distribution Rasch Models: Extensions and applications* (S. 1-12). New York: Springer.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. New York, NY: Cambridge University Press.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The stability of individual response styles. *Psychological Methods*, 15, 96-110.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The Individual Consistency of Acquiescence and Extreme Response Style in Self-Report Questionnaires. *Applied Psychological Measurement*, 34, 105-121.

14 Abbildungsverzeichnis

Abbildung 1: Theorie des geplanten Handelns (Ajzen, 1991).....	20
Abbildung 2: Multinominaler Prozess-Baum (Khorramdel & von Davier, in Druck)...	27
Abbildung 3: Schematische Darstellung DSIHR (Lefevre & Kubinger, 2004).....	35
Abbildung 4: Vorgabeformat der zusätzlichen Antwortmöglichkeiten.....	38
Abbildung 5: GMT auf Basis der Itemschwierigkeit (1-dimensionales 2PL-Modell; m-Items).....	57
Abbildung 6: GMT auf Basis der Itemschwierigkeit (1-dimensionales 2PL-Modell; e-Items).....	58

Tabellenverzeichnis

Tabelle 1: Veranschaulichung einer unvollständigen Kontingenztafel für die Wahrscheinlichkeiten je Kategorie einer fünfstufigen Antwortskala (Khorramdel & von Davier, in Druck, S. 17).....	28
Tabelle 2: Kodierung der Pseudoitems.....	39
Tabelle 3: Beispiel einer Datenmatrix für ein 3-dimensionales IRT-Modell im Fall von vier Fragebogenitems, die in 12 Pseudoitems zerlegt wurden.....	42
Tabelle 4: Beispiel einer Datenmatrix für ein 4-dimensionales IRT-Modell im Fall von vier Fragebogenitems, die in 12 Pseudoitems zerlegt wurden.....	42
Tabelle 5: Beispiel einer Datenmatrix für ein 6-dimensionales IRT-Modell im Fall von vier Fragebogenitems, die in 12 Pseudoitems zerlegt wurden.....	43
Tabelle 6: Beispiel einer Datenmatrix für ein Bifactor-IRT-Modell im Fall von vier Fragebogenitems mit der ursprünglichen Kodierung, die in 8 Pseudoitems zerlegt wurden.....	44
Tabelle 7: Kategorisierung des Iteminhalts der Subskalen des DSIHR nach den Kategorien von Angleitner, John und Löhr (1986).....	46
Tabelle 8: Ergebnisse der 3-dimensionalen, 4-dimensionalen und 6-dimensionalen IRT-Modelle, für alle Skalen und alle Pseudoitemtypen (360 Items insgesamt).....	48
Tabelle 9: Ergebnisse der 3-dimensionalen, 9-dimensionalen und 11-dimensionalen IRT-Modelle, für alle Subskalen und alle Pseudoitemtypen (360 Items insgesamt).....	49
Tabelle 10: Ergebnisse der 3-dimensionalen (Pseudoitems), 3-dimensionalen (Subskalen) und 5-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala StressAuslösung und alle Pseudoitemtypen (147 Items insgesamt).....	50
Tabelle 11: Ergebnisse der 3-dimensionalen (Pseudoitems), 2-dimensionalen (Subskalen) und 4-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala Manifestation und alle Pseudoitemtypen (60 Items insgesamt).....	51
Tabelle 12: Ergebnisse der 3-dimensionalen (Pseudoitems), 2-dimensionalen (Subskalen) und 4-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala Coping und alle Pseudoitemtypen (93 Items insgesamt).....	51
Tabelle 13: Ergebnisse der 3-dimensionalen (Pseudoitems), 2-dimensionalen (Subskalen) und 4-dimensionalen IRT-Modelle, für alle Subskalen der Primärskala Stabilisierung und alle Pseudoitemtypen (60 Items insgesamt).....	52
Tabelle 14: Ergebnisse für die 1-dimensionalen, 4-dimensionalen und 9-dimensionalen IRT-Modelle, für alle Pseudoitemtypen getrennt (120 Items pro Pseudoitemtyp).....	54
Tabelle 15: Geschätzte Interkorrelationen der Score-Verteilung der vier Primärskalen auf Grundlage des 4-dimensionalen IRT-Modells der Pseudoitems m, e und d sowie der Originaldaten.....	55
Tabelle 16: Geschätzte Interkorrelationen der Score-Verteilung der neun Subskalen auf Grundlage des 9-dimensionalen IRT-Modells der Pseudoitems m, e und d sowie der Originaldaten.....	56
Tabelle 17: Ergebnisse für die 1-dimensionalen und 4-dimensionalen IRT-Modelle, für die Pseudoitems des Typ e (120 Items) vor und nach der Ausscheidung von 29 Items...57	57
Tabelle 18: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Auslösung, für alle Pseudoitemtypen getrennt (49 Items pro Pseudoitemtyp).....	60

Tabelle 19: Geschätzte Interkorrelationen der Score-Verteilung der Subskalen der Skala Auslösung auf Grundlage des 3-dimensionalen IRT-Modells der Pseudoitems m, e und d.....	60
Tabelle 20: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Manifestation, für alle Pseudoitemtypen getrennt (20 Items pro Pseudoitemtyp).....	61
Tabelle 21: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Coping, für alle Pseudoitemtypen getrennt (31 Items pro Pseudoitemtyp).....	62
Tabelle 22: Ergebnisse für die 1-dimensionalen und 3-dimensionalen IRT-Modelle für die Primärskala Stabilisierung, für alle Pseudoitemtypen getrennt (20 Items pro Pseudoitemtyp).....	63
Tabelle 23: Ergebnisse für die Bifactor-Modelle (Primärskalen als spezifische Faktoren) und die 4-dimensionalen IRT-Modelle für alle Primärskalen für die Pseudoitemtypen m und e getrennt (120 Items pro Pseudoitemtyp).....	64
Tabelle 24: Varianzen (SD2) der geschätzten spezifischen Faktoren (Primärskalen) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen m und e nach dem Bifactor-Modell.....	65
Tabelle 25: Varianzen (SD2) der geschätzten spezifischen Faktoren (Primärskalen) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen m und e nach dem Bifactor-Modell.....	66
Tabelle 26: Ergebnisse für die Bifactor-Modelle (Subskalen als spezifische Faktoren) und die 9-dimensionalen IRT-Modelle für alle Primärskalen für die Pseudoitemtypen m und e getrennt (120 Items pro Pseudoitemtyp).....	66
Tabelle 27: Varianzen (SD2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Auslösung) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen m und e nach dem Bifactor-Modell.....	67
Tabelle 28: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Auslösung für die Pseudoitemtypen m und e getrennt (49 Items pro Pseudoitemtyp).....	67
Tabelle 29: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Manifestation für die Pseudoitemtypen m und e getrennt (20 Items pro Pseudoitemtyp).....	68
Tabelle 30: Varianzen (SD2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Manifestation) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen m und e nach dem Bifactor-Modell.....	68
Tabelle 31: Varianzen (SD2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Coping) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen m und e nach dem Bifactor-Modell.....	68
Tabelle 32: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Coping für die Pseudoitemtypen m und e getrennt (31 Items pro Pseudoitemtyp).....	69
Tabelle 33: Varianzen (SD2) der geschätzten spezifischen Faktoren (Subskalen der Primärskala Stabilisierung) und dem Generalfaktor (Antworttendenzfaktor) für die Pseudoitemtypen m und e nach dem Bifactor-Modell.....	69

Tabelle 34: Ergebnisse für die Bifactor-Modelle und die 3-dimensionalen IRT-Modelle für die Subskalen der Primärskala Stabilisierung für die Pseudoitemtypen m und e getrennt (20 Items pro Pseudoitemtyp).....	70
Tabelle 35: Geschätzte Mittelwerte pro Primärskala auf Basis des 4-dimensionalen 2PL-Modells für die Pseudoitems des Typs m.....	71
Tabelle 36: Mittelwerte der Scores 1, 2, 3 und 4 der zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns.....	74
Tabelle 37: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs m für die Primärskala Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.....	75
Tabelle 38: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs m für die Subskala instrumentelles Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.....	76
Tabelle 39: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs e für die Primärskala Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.....	76
Tabelle 40: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs e für die Subskala instrumentelles Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.....	77
Tabelle 41: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs d für die Primärskala Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.....	78
Tabelle 42: Ergebnisse der Mehrgruppenmodelle der Pseudoitems des Typs d für die Subskala instrumentelles Coping für Personen mit niedrigen bzw. hohen Scores bei den zusätzlichen Antwortmöglichkeiten nach der Theorie des geplanten Handelns, sowie die Ergebnisse der Jackknife-Methode.....	78

15 Anhang

Anhang I: Begrüßungstext für den Online Fragebogen

Hallo!

Hast du Stress in Studium oder Arbeit? Nutzt du deine Ressourcen optimal? Weißt du, wo deine Stärken und Schwächen liegen? Im Rahmen unserer Diplomarbeiten* kannst du ein Self-Assessment zum Thema Stress und Persönlichkeit ausfüllen, bei dem du eine individuelle Rückmeldung erhältst.

Mit dem folgenden Fragebogen wird dein Umgang mit Stress im Alltag erhoben sowie Persönlichkeitseigenschaften, die für ein erfolgreiches Studieren oder im Arbeitsleben relevant sein können. Mit Stress gut umzugehen, ist beispielsweise bei Prüfungen oder Projektabgaben wichtig. Auch Persönlichkeitseigenschaften, wie gute soziale Kompetenzen, ein positiver Umgang mit den eigenen Emotionen oder ein gewissenhafter Arbeitsstil können hilfreich sein. Probleme in diesen Bereichen können unter anderem ein Vorankommen im Studium bzw. die Zusammenarbeit mit Kollegen und Kolleginnen erschweren.

Die Bearbeitungsdauer liegt ungefähr zwischen 20 und 30 Minuten. Wir würden dich ersuchen, den Fragebogen vollständig zu bearbeiten. Dies ist auch in deinem Sinne, da wir dir nur in diesem Fall eine individuelle Rückmeldung deiner Ergebnisse mit hilfreichen Links und Tipps anbieten können.

Selbstverständlich werden deine Daten vertraulich behandelt.

Falls Fragen oder Unklarheiten auftauchen, kannst du dich gerne an uns wenden.

Wir bedanken uns für deine Mitarbeit!

Unsere Kontaktdaten:

Alina Bugelnig
(alina.bugelnig@univie.ac.at)

Claudia Krutis
(claudia.krutis@univie.ac.at)

* Es handelt sich um Diplomarbeiten bezüglich der Qualitätssicherung von Verfahren am Fachbereich psychologische Diagnostik der Fakultät für Psychologie der Universität Wien.

Anhang II: Rückmeldetexte des Onlinefragebogens

Ihre persönliche Auswertung

Fragebogen 4728950 (System-ID) vom 05.10.2012 - 11:15

Dimensionen	Punkte	Grafik	Max.	Optimal	Erreicht %
Stressauslösung	3.55		10	... - ...	35 %
Stressmanifestation	4.9		10	... - ...	49 %
Coping	5.87		10	... - ...	59 %
Stressstabilisierung	4.3		10	... - ...	43 %
Neurotizismus	6		10	... - ...	60 %
Extraversion	6.83		10	... - ...	68 %
Offenheit	7		10	... - ...	70 %
Verträglichkeit	9.5		10	... - ...	95 %
Gewissenhaftigkeit	7.16		10	... - ...	72 %

Ihre persönliche Auswertung:

In der oben angeführten Grafik sehen Sie Ihre erreichte Punktzahl als farbigen Strich auf dem grauen Balken der jeweiligen Dimension. Zusätzlich können Sie Ihre genaue Punktzahl (Spalte "Punkte"), die maximale Punktzahl (Spalte "Max.") sowie die erreichten Punkte als Prozentwert (Spalte "Erreicht %") ablesen.

Darstellung der Scores für jede Skala aus dem Onlinefragebogen.

--- Stressauslösung

Sie haben **2 - 4 von 10 Punkten**.

Die Dimension Stress*Auslösung* beschäftigt sich damit, durch welche Bereiche Stress ausgelöst wird. Diese Stressauslöser können sich auf drei unterschiedliche Bereiche beziehen. Diese sind das Alltagsgeschehen (unvermeidliche Angelegenheiten des täglichen Lebens), Interaktion (berufliche und private Kontakte mit anderen Menschen) sowie Existenzängste und Zukunftssorgen. Ein **hoher Punktwert** gibt an, dass die Gesamtanzahl an vorhandenen Auslösern von Stress hoch ist. Personen mit einem **niedrigen Punktwert** geben an, durch wenige Dinge gestresst zu sein.

Tipps

Bei einem hohen Punktwert gibt es unterschiedliche Möglichkeiten den Stressauslösern entgegenzuwirken. Das Ziel besteht darin, äußere Belastungen zu verändern. Der Alltag sollte möglichst stressfrei sein. Dies wiederum kann auf unterschiedlichen Wegen passieren: Zum einen ist es wichtig, egal ob beruflich, im Studium oder im privaten Bereich, die (Arbeits-)Organisation zu optimieren. Dazu gehört, dass Prioritäten gesetzt werden, eine realistische Zeiteinteilung bereits im Vorfeld geplant wird und auch das Delegieren von Aufgaben, die nicht unbedingt selbst erledigt werden müssen. Des Weiteren können Kommunikations-, Sozial und Problemlösekompetenzen verbessert bzw. kritisch reflektiert werden, um eine Atmosphäre mit Mitmenschen zu schaffen, in der man sowohl um Hilfe bitten, als auch Grenzen setzen kann (z.B. „nein“ sagen oder „ohne mich“), um zusätzlichen Stress zu vermeiden.

--- Stress*Manifestation*

Sie haben **4 - 6 von 10 Punkten**.

Die Dimension Stress*Manifestation* beschreibt die Anzahl der auftretenden Ausprägung von Stress. Diese Ausprägungen können sich sowohl auf der körperlichen (z.B. Schmerzen, Infektionsanfälligkeit, Magenbeschwerden) als auch auf der emotional-kognitiven Ebene (z.B. Unlustgefühl, Konzentrationsschwierigkeiten,

Stimmungsschwankungen) zeigen. Ein **hoher Punktwert** gibt an, dass die Gesamtanzahl an auftretenden Ausprägungen von Stress hoch ist und viele unterschiedliche Formen von Stress verspürt werden. Personen mit einem **niedrigen Punktwert** hingegen beschreiben, wenige unterschiedliche Formen von Stress zu erleben.

Tipps

Bei einem hohen Punktwert ist das Ziel, die körperlichen sowie die emotional-kognitiven *Manifestationen* von Stress zu verringern. Wichtig ist dabei, regelmäßig für Entspannungs- und Erholungsmöglichkeiten zu sorgen, um langfristig die eigene Belastbarkeit zu stärken. Eine gesunde Ernährung, regelmäßiger Ausgleich durch Hobbys, Freizeitaktivitäten und Sport sowie außerberufliche soziale Kontakte können die Stress*Manifestationen* verringern. Des Weiteren helfen kleine Pausen, die im Tagesablauf eingeplant und auch ausgeführt werden sowie genügend Schlaf. Ein weiterer Tipp sind Entspannungstechniken, die bei regelmäßiger Anwendung eine Verbesserung des Stresserlebens bewirken können. (z.B. Progressive Relaxation nach Jacobson http://de.wikipedia.org/wiki/Progressive_Muskelentspannung)

--- Bewältigungsmöglichkeiten

Sie haben **4 - 6 von 10 Punkten**.

Die Dimension Bewältigungsmöglichkeiten beschreibt, wie viele unterschiedliche Arten und Strategien mit Stress umzugehen und ihn zu verringern man tatsächlich verwendet. Diese Möglichkeiten teilt man in jene auf, die durch positive Emotionen und Kognitionen eine Bewältigung ermöglichen (z.B. sich selber Mut zusprechen, an etwas Schönes denken) und jene, die dies durch ein aktives Vorgehen gegen die Stressursache tun (z.B. Prioritätensetzung, um Hilfe bitten, Delegieren von Aufgaben). Personen mit einem **hohen Punktwert** geben an, dass sie viele verschiedene Bewältigungsmöglichkeiten anwenden, um ihren Stress zu verringern. Ein **niedriger Punktwert** hingegen sagt aus, dass wenige unterschiedliche Stressbewältigungsstrategien verfolgt werden.

Tipps

Bei einem niedrigen Punktwert ist es hilfreich, sich mit sich selbst und dem erlebten Stress auseinanderzusetzen. Unter dem unten angeführten Link findet man kurz zusammengefasst die hier bereits unter den Punkten Stress*Manifestation*, Stress*Auslösung* und Stress*Stabilisierung* angeführten Möglichkeiten zur Stressbewältigung.

<http://www.studentenberatung.at/studentenberatung/de/stressbewaeltigung-iv.htm#stressbewaeltigung>

Für eine detaillierte Beschreibung empfehlen wir das Buch: Gert Kaluza, Stressbewältigung, ISBN 978-3-642-13719-8

--- Stress*Stabilisierung*

Sie haben **4 - 6 von 10 Punkten**.

Die Dimension Stress*Stabilisierung* beschreibt, welche Faktoren dazu führen, dass es zu einer chronischen Belastung durch Stress kommen kann. Diese Faktoren können von außen (z.B. Zeitdruck, Gefühl von Machtlosigkeit) oder von innen (z.B. gedankliche Beschäftigung mit Stress, die Annahme, Stress sei wichtig um Erfolg zu haben) eine Stressbelastung verstärken. Personen mit einem **hohen Punktwert** geben an, dass sie durch viele persönliche Einstellungen ein stärkeres Beanspruchungsgefühl erleben als andere und sich dadurch eine chronische Belastung ergeben könnte. Ein **niedriger Punktwert** hingegen sagt aus, dass die persönlichen Einstellungen das Belastungsgefühl nicht erhöhen.

Tipps

Bei einem hohen Punktwert besteht das Ziel darin, sich kritisch mit den eigenen stresserzeugenden und verstärkenden Einstellungen auseinanderzusetzen und diese im Lauf der Zeit zu verändern, um förderliche Denkweisen zu entwickeln. Einige Beispiele für Einstellungsveränderung: Leistungsansprüche sollen überprüft werden, um sie den eigenen Leistungsgrenzen anpassen zu können. Schwierigkeiten sollen positiv, als Herausforderung gesehen, anstatt als Bedrohung verstanden zu werden. Persönliche Distanzierung von alltäglichen Aufgaben solle gefördert werden, anstatt sich mit den Problemen zu identifizieren.

--- Neurotizismus

Sie haben **4 - 6 von 10 Punkten**.

Die Dimension Neurotizismus erfasst die emotionale Stabilität eines Menschen. Personen mit einem **hohen Punktwert** geben an, dass sie öfters negative Gefühle haben und diese zum Teil als überwältigend empfinden. Sie geben an beispielsweise traurig, ängstlich, nervös, erschüttert, betroffen, beschämt, unsicher und verlegen zu reagieren. Sie beschreiben, eher leicht aus dem seelischen Gleichgewicht zu geraten. Sie geben an, sich viele Sorgen zu machen und eine Neigung zu unrealistischen Ideen zu haben. Personen mit einem **niedrigen Punktwert** beschreiben sich als eher emotional stabil. Sie schreiben sich selbst Attribute wie sorgenfrei, ruhig und ausgeglichen zu und geben an, dass sie nur schwer aus der Ruhe zu bringen sind und in Stresssituationen nicht so schnell aus der Fassung geraten.

In Bezug auf einen positiven Umgang mit Stress (bzw. dessen Reduktion oder Vermeidung) ist hier ein niedriger Punktwert vorteilhaft. Dieser spricht für eine hohe Frustrationstoleranz bzw. Belastbarkeit, die für die positive Bewältigung von unvorhergesehenen Hindernissen und Rückschlägen in Studium oder Beruf wichtig ist. Geringer Neurotizismus deutet weiters auf eine gute Distanzierungsfähigkeit hin – dies bedeutet, dass man nicht ständig über die Arbeit nachgrübelt, sondern sich gedanklich auch mit anderen Dingen, die nicht mit Stress verbunden sind, beschäftigen kann.

Tipps

Bei einem hohen Punktwert kann es hilfreich sein, den „Blick auf das Positive“ in Stresssituationen zu verstärken und die positiven Aspekte bewusst hervorzuheben („Finde ich etwas Gutes in dieser Situation?“). Ebenso solle man sich seine eigenen Stärken und Ressourcen („Hat es in meinem Leben schon schwierige Situationen gegeben? Und wie habe ich diese gemeistert?“) bewusst

machen, um dadurch ein „Nachgrübeln“ über einen negativen Ausgang der Situation zu verringern. Zusätzlich kann es hilfreich sein, den Blickwinkel zu verändern und zu versuchen, das stressauslösende Problem in einem anderen Licht zu betrachten („Wie sehen andere Menschen die Situation? Warum sind sie dadurch weniger belastet?“). Zu guter Letzt ist es förderlich, sich mit den positiven Konsequenzen eines Problems auseinanderzusetzen und sich darüber Gedanken zu machen, wie es sein wird, wenn man die Situation erfolgreich bewältigt hat.

--- Extraversion

Sie haben **6 - 8 von 10 Punkten**.

Die Dimension Extraversion erfasst die Geselligkeit und Aktivität eines Menschen. Personen mit einem **hohen Punktwert** in dieser Dimension erleben sich selbst als optimistisch, heiter, energisch, gesprächig, aktiv und selbstsicher. Sie geben an gerne unter Menschen, in Gruppen und auf gesellschaftlichen Versammlungen zu sein. Sie beschreiben außerdem eine Neigung zu einem heiteren Gemüt. Personen mit einem **niedrigen Punktwert** in dieser Dimension beschreiben sich als eher zurückhaltend, unabhängig und weniger aktiv bzw. ausgeglichen. Sie geben an gerne allein Zeit zu verbringen, jedoch ohne dabei pessimistisch oder unglücklich zu sein. In Bezug auf einen positiven Umgang mit Stress (bzw. dessen Reduktion oder Vermeidung) kann Extraversion wichtig sein, weil es eher extravertierten Personen oft leichter fällt, Kontakte zu knüpfen und sich ein soziales Netzwerk aufzubauen. Ein soziales Netzwerk kann wiederum in bestimmten Lebenssituationen und vor allem bei Stress wichtige und hilfreiche Unterstützung bieten. Hohe Extraversion wird oft von hoher sozialer Kompetenz begleitet. Mit einer eher hohen sozialen Kompetenz fällt es auch leicht, gut mit Kollegen und Kolleginnen in Studium oder Beruf auszukommen und zusammenzuarbeiten, sowie diesen die eigenen Meinungen und Ideen mitzuteilen. Ein gutes Auskommen mit anderen Personen kann wiederum dazu beitragen, sozialen Stress zu reduzieren.

Tipps

Bei niedrigen Punktwerten kann ein aktiveres Zugehen auf anderen Menschen in Stresssituationen insofern förderlich sein, als dass diese Aufgaben übernehmen oder emotionale Unterstützung anbieten können. Auch wenn Personen mit niedrigen Werten oft lieber alleine arbeiten, sollte man sich darüber bewusst sein, dass in Stresssituationen soziale Kontakte eine Entlastung bringen können. Diesbezüglich kann eventuell auch ein soziales Kompetenztraining helfen (http://de.wikipedia.org/wiki/Training_sozialer_Kompetenzen).

--- Offenheit für Erfahrungen

Sie haben **6 - 8 von 10 Punkten**.

Die Dimension Offenheit für Erfahrungen erfasst den Umgang mit neuen Erfahrungen, Eindrücken und Erlebnissen. Personen mit einem **hohen Punktwert** beschreiben sich als fantasievoll, wissbegierig, künstlerisch interessiert, intellektuell und experimentierfreudig. Sie geben an, häufig die eigenen positiven und negativen Gefühle

sehr deutlich wahrzunehmen und sich für viele öffentliche und persönliche Begebenheiten zu interessieren. Sie beschreiben, dass sie eine höhere Bereitschaft zur kritischen Hinterfragung bestehender Normen und zum Eingehen auf neuartige Wertvorstellungen im ethischen, sozialen und politischen Bereich haben. Sie geben an, dass ihre Urteile eher unabhängig sind, sie Abwechslung bevorzugen und gerne neue Handlungsweisen ausprobieren. Ihr Verhalten beschreiben sie als unkonventionell. Personen mit einem **niedrigen Punktwert** beschreiben dagegen eher eine Neigung zu konservativen Einstellungen und konventionellem Verhalten. Sie geben an, häufig Bewährtes und Bekanntes lieber zu mögen als Neues und tragen ihre Emotionen weniger stark nach außen. In Bezug auf den Umgang mit Stress kann eine niedrige Offenheit zu Frustration und Stress führen, da man oft in alten Lösungsansätzen und Ideen verhaftet bleibt, selbst wenn diese sich nicht umsetzen lassen.

Tipps

Bei einem niedrigen Punktwert sollte man versuchen, andere Meinungen, Erfahrungen und Ansichten zu berücksichtigen, um zum Beispiel neue Lösungsmöglichkeiten für Probleme zu finden. So können auch eigene neue Ideen angeregt werden. Außerdem erweitert das Interesse an neuen Inhalten das eigene Wissen und somit kann die Leistung in Studium oder Beruf verbessert werden. Dies kann zu Reduktion oder Vermeidung von Stress führen. Dabei ist jedoch zu beachten, dass nicht zu viele Veränderungen auf einmal vorgenommen werden sollten, um für eine gewisse Stabilität in Lern- bzw. Berufsumfeld zu sorgen.

--- Verträglichkeit

Sie haben **8 - 10 von 10 Punkten**.

Die Dimension Verträglichkeit erfasst den eigenen Umgang mit anderen Menschen. Personen mit einem **hohen Punktwert** beschreiben sich als verständnisvoll, wohlwollend und mitfühlend gegenüber anderen Menschen und versuchen, diesen zu helfen. Dabei gehen sie davon aus, dass diese ebenso hilfsbereit sind. Ein hoher Punktwert steht für ein starkes Harmoniebedürfnis und die Neigung zu Nachgiebigkeit, Kooperation und zwischenmenschlichen Vertrauen. Personen mit einem **niedrigen Punktwert** beschreiben sich als eher misstrauisch gegenüber den Absichten anderer Menschen und als egozentrisch und antagonistisch. Sie geben an, dass sie lieber mit anderen wetteifern, als mit diesen zusammenzuarbeiten. Sie beschreiben, dass es ihnen oft leichter fällt für ihre eigenen Interessen zu kämpfen und neigen meist zu Skepsis und Misstrauen. In Bezug auf einen positiven Umgang mit Stress (bzw. dessen Reduktion oder Vermeidung) ist Verträglichkeit – ebenso wie Extraversion – für den sozialen Umgang wichtig. Wie oben bereits erwähnt (siehe Dimension Extraversion), geht es hier um eine gute Zusammenarbeit mit Kollegen und Kolleginnen, das Knüpfen von neuen Kontakten und eine gute Kommunikation. Bei zu großem Misstrauen oder Skepsis gegenüber anderen kann es zu Konflikten kommen, die in weitere Folge wieder den Stress erhöhen.

Tipps

Bei einem niedrigen Punktwert ist (ähnlich wie in der Dimension Extraversion) das Zugehen auf andere Personen wichtig. So kann man zum Beispiel Kollegen und Kolleginnen um Feedback bitten, wie man auf Andere wirkt beziehungsweise wie das Gesagte von Anderen aufgefasst/aufgenommen wird. So kann man Missverständnissen vorbeugen oder diese klären und zu einer Konfliktlösung beitragen. Auch kann man

üben, anderen zuzuhören oder sie ausreden zu lassen und Konfliktgespräche auf eine eher sachliche Ebene zu verlagern. Hier kann ebenfalls ein soziales Kompetenztraining helfen (http://de.wikipedia.org/wiki/Training_sozialer_Kompetenzen).

--- Gewissenhaftigkeit

Sie haben **6 - 8 von 10 Punkten**.

Die Dimension Gewissenhaftigkeit erfasst eine Art der Selbstkontrolle, nämlich die Planung, Organisation und Durchführung von Aufgaben. Personen mit einem **hohen Punktwert** beschreiben sich als ausdauernd, systematisch, zielstrebig, ordentlich, ehrgeizig, fleißig, penibel, genau, willensstark, pünktlich, zuverlässig und diszipliniert. Extrem hohe Ausprägungen können sich eventuell in zwanghafter Ordentlichkeit, einem übertrieben hohen Anspruchsniveau oder in Formen von Arbeitssucht äußern. Personen mit einem **niedrigen Punktwert** beschreiben eher eine Neigung zur Nachlässigkeit, Gleichgültigkeit und Unbeständigkeit. Sie geben an, meist ein geringeres Engagement im Verfolgen der eigenen Ziele zu zeigen. In Bezug auf einen positiven Umgang mit Stress (bzw. dessen Reduktion oder Vermeidung) erweist sich eine hohe Gewissenhaftigkeit oft als vorteilhaft, da Personen mit einem genauen Arbeitsstil Fehler vermeiden, die zu Stress führen können. Auch ein gutes Zeitmanagement sowie eine gute Planungs- und Organisationsfähigkeit können Stress im Studium oder im Arbeitsalltag vermeiden oder verringern.

Tipps

Bei einem niedrigen Punktwert kann es sehr hilfreich sein, ein gezieltes Zeitmanagement und eine gute Prioritätensetzung in den Alltag zu integrieren. Hier sind hilfreiche Links zu den angesprochenen Themen für Studium und Beruf:

<http://www.metacom.com/zeitmanagement-tipps.php>

<http://www.studentenberatung.at/studentenberatung/de/zeitmanagement.htm#Bedeutung>
www.studentenberatung.at/studentenberatung/de/lernen-mit-erfolg.htm#Lern-und-Arbeitsplanung

Im Internet gibt es viele hilfreiche Tipps und Tricks zur Stressbewältigung sowie Erklärungen, was Stress ist bzw. wie er zustande kommt. Um Ihnen eine mühsame Suche zu ersparen, haben wir unten eine Sammlung von passenden Links zusammengestellt und hoffen, dass diese in Kombination mit den obigen Ergebnissen der Fragebögen für Sie hilfreich sind:

Weiterführende Links:

<http://de.wikipedia.org/wiki/Stress>

<http://arbeitsblaetter.stangl-taller.at/TEST/STRESS/Test.shtml>

<http://arbeitsblaetter.stangl-taller.at/STRESS/Stressbewaeltigung-Uebungen.shtml>

<https://www.gesundheit.gv.at/Portal.Node/ghp/public/content/gesund-leben-stress-arbeit.html>

<http://www.arbeiterkammer.com/online/stress-19509.html>

http://www.pfiiffprojekt.de/workshop/fundus/stress_stressbewaeltigung/uebungen%20Stress%20und%20Coping%20komplett.pdf

<http://www.zeitbluten.com/news/2010/entspannungsuebungen/>

<http://www.hauptsache-bildung.de/2012/burnout-praevention-10-uebungen-zur-stressbewaeltigung-am-arbeitsplatz>

<http://www.stresscoach.at/tipps/index.html>

<http://studium.lerntipp.at/stress/stressuebungen.shtml>

Anhang III: empirische Skaleninterkorrelationen

Empirisch beobachtete Interkorrelationen der Score-Verteilung der vier Primärskalen auf Grundlage des 4-dimensionalen IRT-Modells der Pseudoitems m , e und d sowie der Originaldaten.

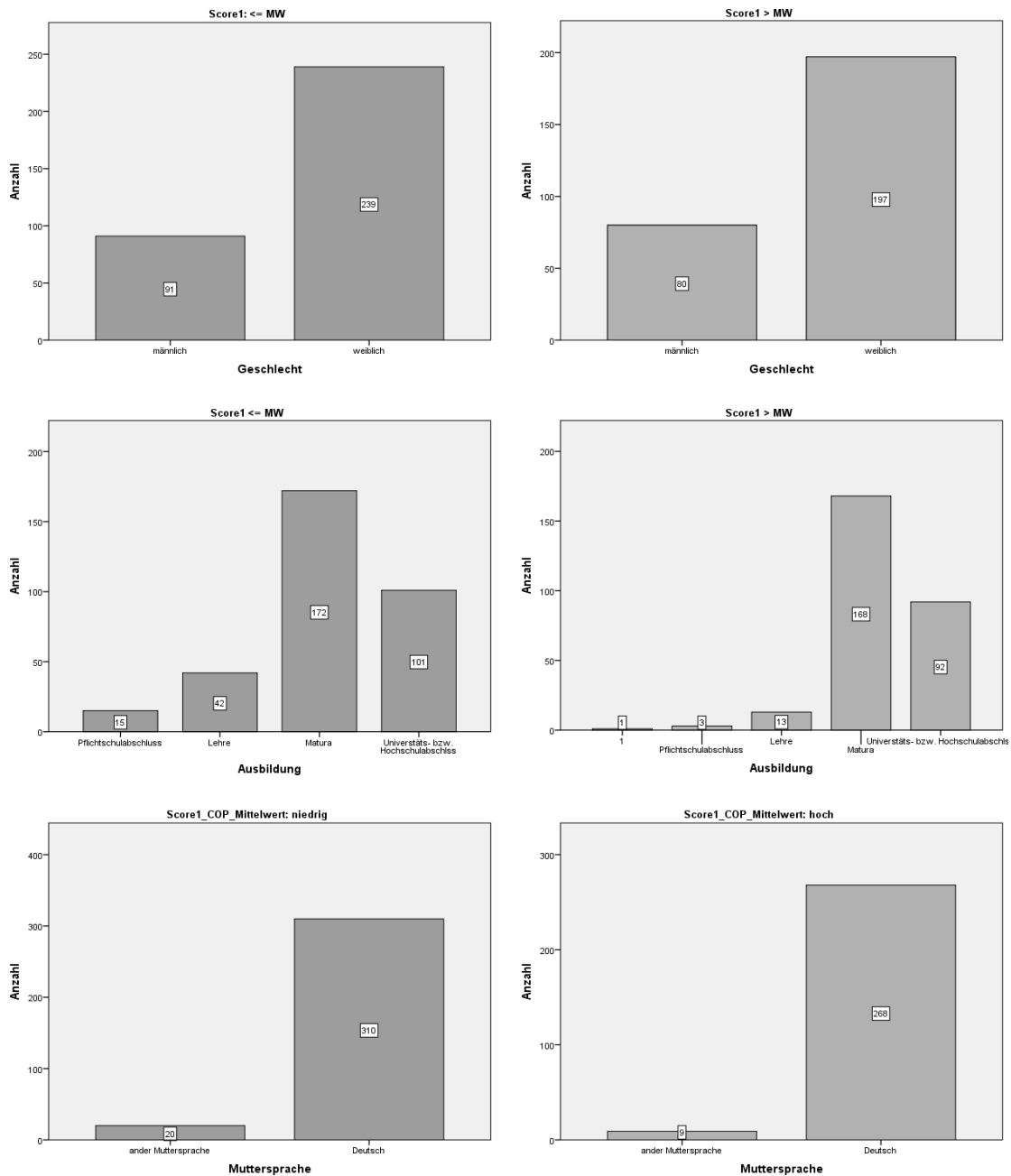
	<i>Auslösung</i>	<i>Manifestation</i>	<i>Coping</i>	<i>Stabilisierung</i>
<i>m-Items</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,59281	1		
<i>Coping</i>	0,56610	0,67708	1	
<i>Stabilisierung</i>	0,50824	0,57540	0,60392	1
<i>e-Items</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,59520	1		
<i>Coping</i>	0,59125	0,62920	1	
<i>Stabilisierung</i>	0,57767	0,58121	0,67239	1
<i>d-Items</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,53322	1		
<i>Coping</i>	0,09959	0,22820	1	
<i>Stabilisierung</i>	0,52823	0,49537	0,14665	1
<i>Originaldaten</i>				
<i>Auslösung</i>	1			
<i>Manifestation</i>	0,49987	1		
<i>Coping</i>	0,32699	0,46188	1	
<i>Stabilisierung</i>	0,45179	0,52492	0,46287	1

Empirische Interkorrelationen der Score-Verteilung der neun Subskalen auf Grundlage des 4-dimensionalen IRT-Modells der Pseudoitems *m*, *e* und *d* sowie der Originaldaten.

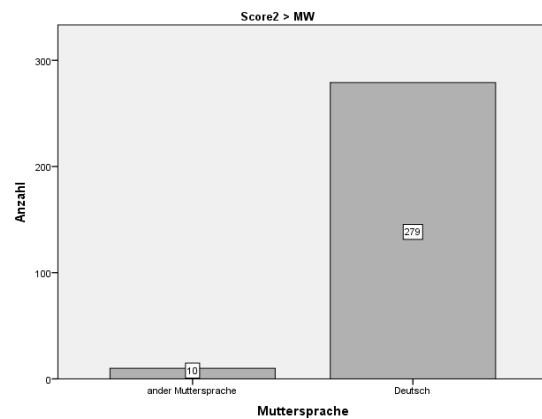
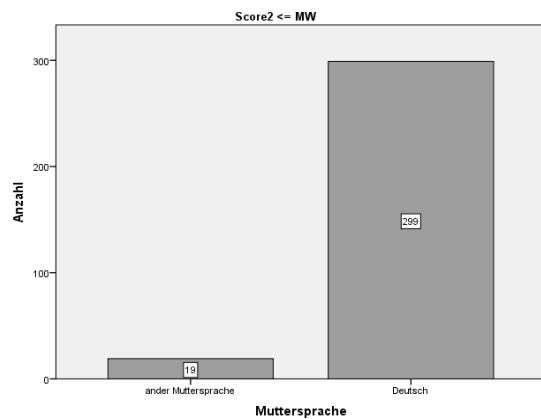
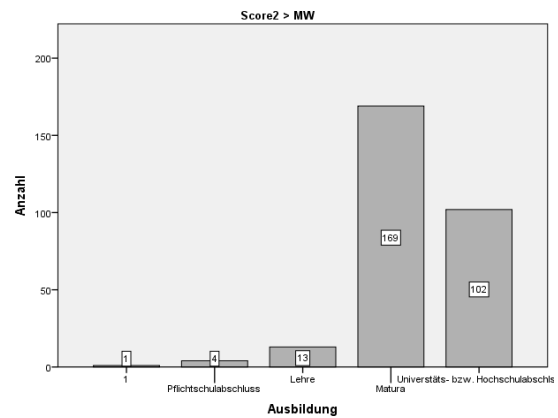
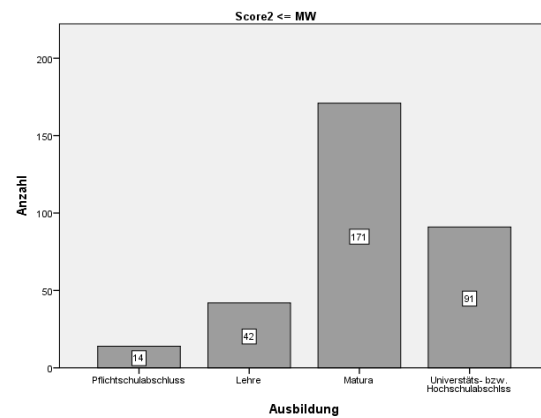
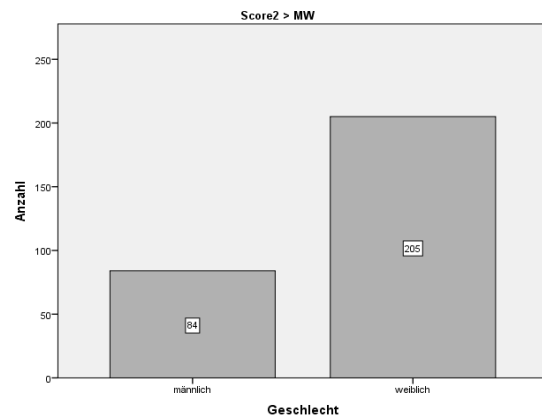
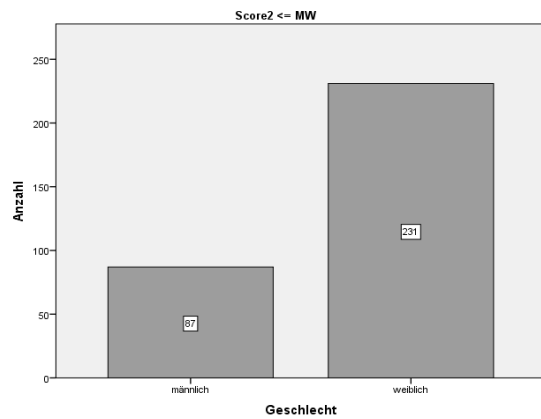
	<i>Exist.</i>	<i>Inter.</i>	<i>Alltags.</i>	<i>physiol.</i>	<i>em./ko.</i>	<i>pallia.</i>	<i>instr.</i>	<i>intern.</i>	<i>extern.</i>
<i>m-Items</i>									
Existenz	1								
Interaktion	0,457	1							
Alltagsgeschehen	0,448	0,516	1						
physiologisch	0,302	0,330	0,323	1					
emotional/kognitiv	0,388	0,363	0,345	0,312	1				
palliativ	0,363	0,363	0,375	0,341	0,382	1			
instrumentell	0,363	0,365	0,357	0,339	0,403	0,449	1		
internal	0,212	0,212	0,199	0,190	0,258	0,285	0,294	1	
external	0,316	0,324	0,328	0,251	0,356	0,345	0,316	0,237	1
<i>e-Items</i>									
Existenz	1								
Interaktion	0,347	1							
Alltagsgeschehen	0,441	0,452	1						
physiologisch	0,319	0,316	0,372	1					
emotional/kognitiv	0,363	0,297	0,377	0,397	1				
palliativ	0,323	0,331	0,416	0,386	0,447	1			
instrumentell	0,292	0,340	0,396	0,396	0,406	0,504	1		
internal	0,263	0,291	0,273	0,318	0,288	0,375	0,372	1	
external	0,330	0,299	0,414	0,340	0,351	0,470	0,407	0,365	1
<i>d-Items</i>									
Existenz	1								
Interaktion	0,400	1							
Alltagsgeschehen	0,464	0,361	1						
physiologisch	0,339	0,259	0,291	1					
emotional/kognitiv	0,324	0,230	0,405	0,352	1				
palliativ	0,043	0,122	-0,033	0,127	0,040	1			
instrumentell	0,125	0,120	0,198	0,216	0,265	0,206	1		
internal	0,005	0,013	0,107	0,119	0,077	0,203	0,211	1	
external	0,352	0,195	0,337	0,303	0,384	0,013	0,206	0,133	1
<i>Originaldaten</i>									
Existenz	1								
Interaktion	0,299	1							
Alltagsgeschehen	0,345	0,404	1						
physiologischemo-	0,170	0,218	0,271	1					
tional/kognitiv	0,194	0,248	0,358	0,307	1				
palliativ	0,128	0,180	0,146	0,224	0,218	1			
instrumentell	0,093	0,181	0,236	0,242	0,304	0,357	1		
internal	0,008	0,028	0,104	0,184	0,135	0,276	0,275	1	
external	0,154	0,120	0,322	0,262	0,324	0,220	0,212	0,167	1

Anhang IV: Histogramme der demographischen Daten separat für jede Gruppe der Mehrgruppenmodelle für die Skala *Coping*

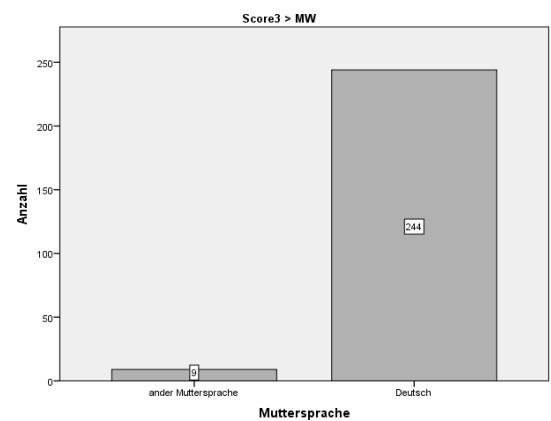
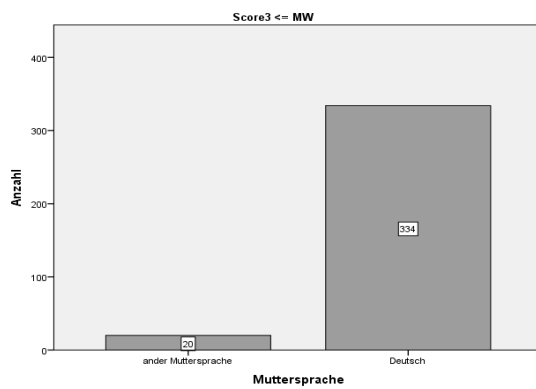
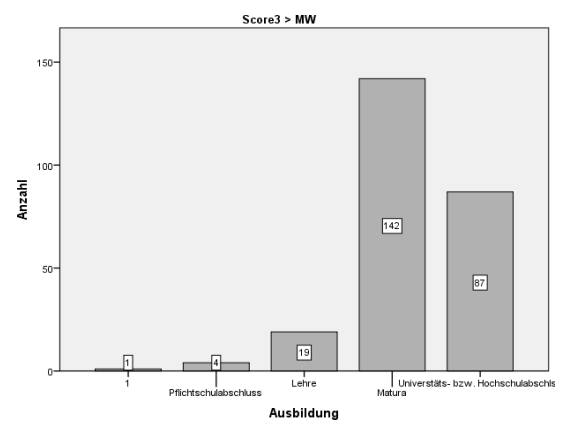
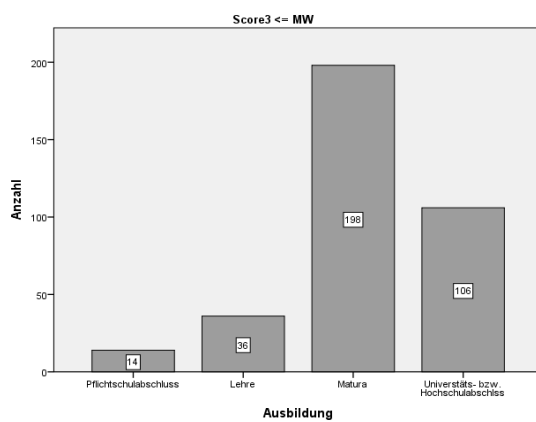
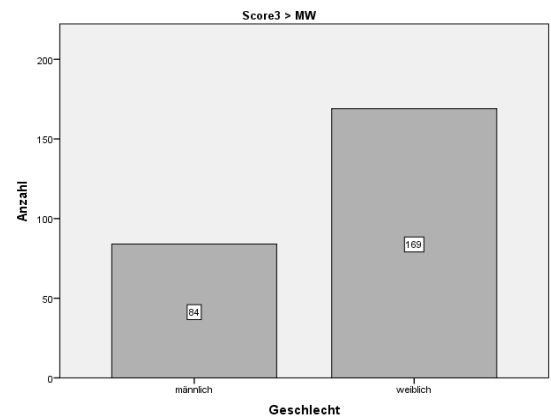
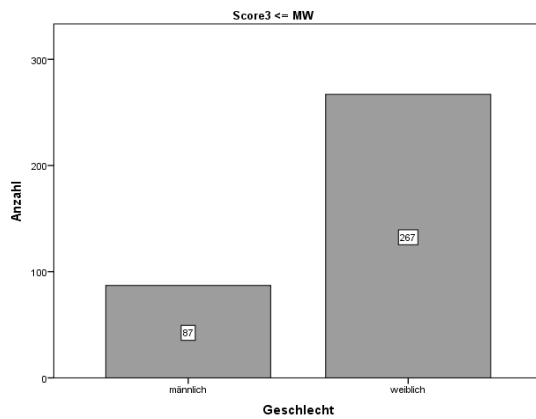
Verteilung Score 1



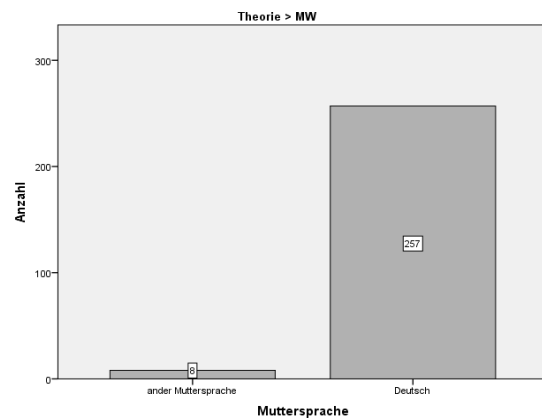
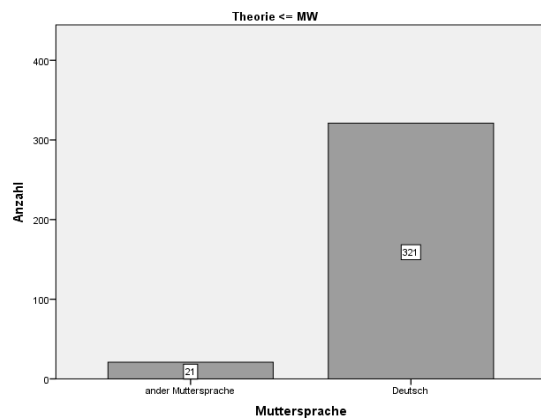
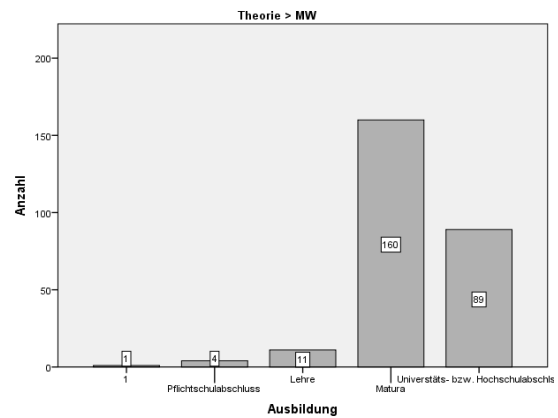
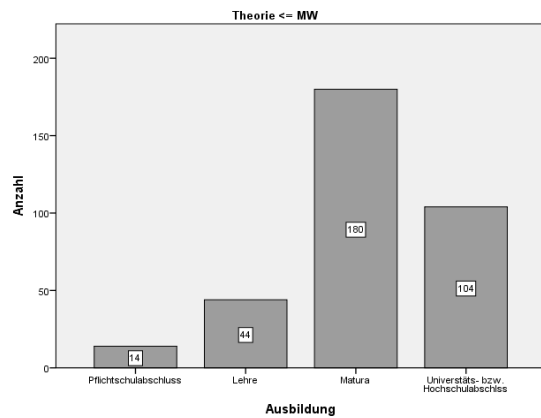
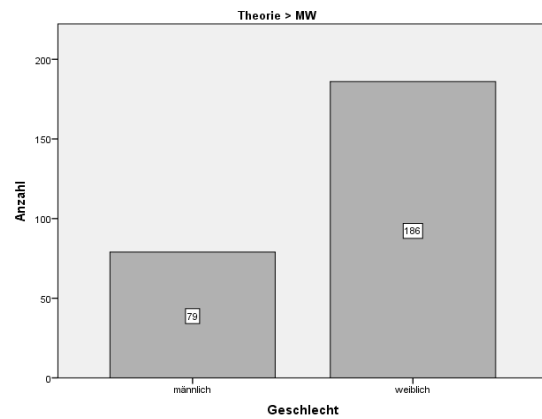
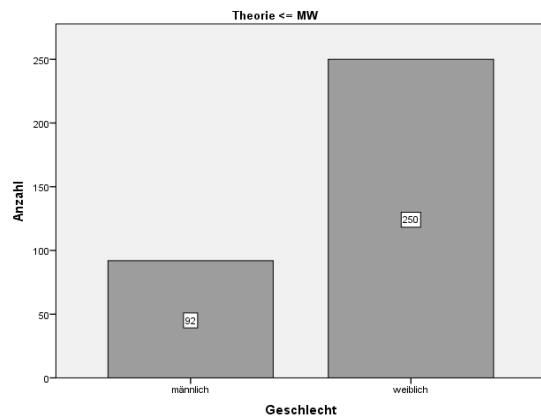
Verteilung Score 2



Verteilung Score 3



Verteilung Score 4



Alina Bugelnig

Curriculum Vitae

Ausbildung	
seit 2007	Universität Wien, Diplomstudium Psychologie
2002 – 2007	Handelsakademie Spittal/Drau; <i>Schwerpunkt Wirtschaftsinformatik</i>
1998 – 2002	Musikhauptschule Seeboden
1994 – 1998	Volksschule Obermillstatt
Berufliche Erfahrung	
11/2011 – 01/2014	Studienassistentin am <i>Arbeitsbereich psychologische Diagnostik</i> der Universität Wien
07/2013 – 09/2013	Projektmitarbeit im <i>Center for Teaching and Learning</i> der Universität Wien
02/2012 – 04/2012	6-Wochen-Praktikum in der <i>Test- und Beratungsstelle</i> des Arbeitsbereichs psychologische Diagnostik der Universität Wien
11/2010 – 02/2011	Testleiterin für Normierungstestungen am <i>Arbeitsbereich psychologische Diagnostik</i> der Universität Wien

Wien, Jänner 2014