universität wien

DIPLOMARBEIT

Zur Konstruktion eines adaptiven Persönlichkeitsfragebogens auf Basis eines Konstruktionsrationals

Verfasser

Manuel Reif

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, am 11. 11. 2008

Studienkennzahl:	298
Studienrichtung:	Psychologie
Betreuer:	Univ.-Prof. Dr. Mag. Klaus D. Kubinger

Danksagung

Dieses Werk ist über einen langen Zeitraum entstanden, und aus diesem Grund gibt es auch viele Menschen die mich in dieser Zeit begleiteten.

Zuallererst möchte ich mich bei Univ. Prof . Dr. Mag. Klaus D. Kubinger, nicht nur für die intensive persönliche Betreuung meiner Diplomarbeit, sondern auch für das Zutrauen das er als mein Arbeitgeber zu mir hat, und für all die Fertigkeiten die ich im Rahmen vieler Diskurse von ihm lernen durfte, bedanken.

Großer Dank gebührt Prof. Dr. Tuulia M. Ortner, die mich mit diesem Thema begeistert, und mich damit als Diplomand gewonnen hat. Mit vielen zeitintensiven Gesprächen, Diskursen und Korrekturen trug sie maßgeblich zur Itemerstellung und zum Konstruktionsrational bei.

Vielen Dank auch an Dr. Ingrid Preusche, die für meine Fragen immer ein offenes Ohr hatte, und mich mit konstruktiven Vorschlägen und Hilfestellungen unterstützte.

Vielen Dank gilt auch Mag. Harald Schilly der viel Zeit investierte und mittels seiner löblichen Programmierkenntnisse die Online-Version des Fragebogens erschuf.

Vielen Dank auch an alle Freiwilligen, die in einer kreativen Marathonsitzung die grobe Basis für das vorliegende Werk lieferten.

Dank gilt auch Mag. Joachim Fritz Punter der mir Basiskenntnisse in LaTeXvermittelte, und somit eine professionelle Layoutierung meiner Diplomarbeit ermöglichte.

Großer Dank gilt auch meiner Mutter, die es mir ermöglichte zu studieren, und die während meiner gesamten Studienzeit daran glaubte, dass ich das Studium positiv bewältigen kann.

Insbesondere bedanke ich mich ganz herzlich bei Christine, einerseits für den geleisteten Beistand bei allen emotionalen und motivationalen Tiefs während meines Schreibens der Diplomarbeit, und andererseits für die fachliche Expertise, das mehrfache Korrekturlesen und für die Geduld die sie mir jeden Tag entgegenbringt.

Abstract

Die vorliegende Arbeit thematisiert die Konstruktion von Persönlichkeitsfragebögen, und stellt eine regelgeleitete Itemkonstruktion als optimale Variante in den Vordergrund. Ziel der Arbeit war – auf Basis eines Konstruktionsrationals – einen Rasch-Modell-konformen Fragebogen zur Messung von Extraversion zu entwickeln, der damit auch zur adaptiven Vorgabe geeignet ist. Im theoretischen Teil der Arbeit wird ein Überblick der Persönlichkeitsmodelle, die Extraversion beinhalten, gegeben. Außerdem wird auf generelle Konstruktionsprinzipien von Fragebögen und auf die probabilistische Testtheorie zur Fragebogenkonstruktion eingegangen. Auf Basis gängiger Definitionen des Konstrukts Extraversion wurde zunächst ein Konstruktionsrational zur regelgeleiteten Itemgenerierung entwickelt. Anschließend wurde damit ein Fragebogen zur Messung von Extraversion entwickelt, welcher als Computerversion an einer Stichprobe von 560 Personen erprobt wurde. Die Analysen des Fragebogens hinsichtlich Rasch-Modell-Konformität ergaben nach dem sukzessiven Ausscheiden von zirka 15% der Items a posteriori Modellgeltung für den verbleibenden Itempool. Um außerdem die Geltung des Konstruktionsrationals zu prüfen, wurde im nächsten Schritt der Rasch-Modell-konforme Itempool mittels des LLTM analysiert, wobei die Strukturmatrizen entsprechend dem Konstruktionsrational definiert wurden. Aufgrund mangelnder Geltung kamen mehrere, schrittweise modifizierte Strukturmatrizen zum Einsatz. Schließlich hielt keine der unterschiedlichen Versionen dem Vergleich mit dem Rasch-Modell stand, was inhaltlich bedeutet, dass die Itemschwierigkeiten nicht hinreichend genau durch das Konstruktionsrational erklärt werden. Die vorliegende Arbeit stellt einen Rasch-Modell-konformen Fragebogen zur Extraversion zur Verfügung – was auch die adaptive Messung dieses Konstrukts ermöglicht.

The following thesis is concerned with the construction of personality questionnaires and posits rule-based item construction as an optimal construction strategy. The goal of the study was to develop a Rasch model conform „extraversion“ questionnaire suitable for adaptive testing on the basis of this construction principle. The theoretical section of the thesis provides an overview of personality theories encompassing the concept of extraversion. It also introduces general principles of questionnaire construction and explains this construction in light of probabilistic test theory. In the study, a construction principle for rule-based item generation was developed on the basis of common definitions of extraversion. This principle was then used to develop an extraversion questionnaire that was tested on a sample of 560 people. After roughly 15% of the items were successively eliminated, Rasch model analyses of the questionnaire showed an a posteriori model fit of the remaining items. In order to also test the construction principle, the Rasch model

conform item pool was analyzed using the LLTM, with the design matrices defined according to the construction principle. Due to the ill fit of the data, the design matrices were successively modified. None of the modified versions were comparable to the Rasch model, showing that item difficulties were not adequately explained by the construction principle. The following thesis provides readers with a Rasch model conform extraversion questionnaire – which also allows for the adaptive measurement of this construct.

Inhaltsverzeichnis

1. Einleitung

Jeder Psychologe, der zumindest peripher Diagnostik betreibt, hat einen Fragebogen zumindest schon vorgegeben, wenn nicht sogar schon selbst konstruiert. Auch viele Personen die auf dem Gebiet der Psychologischen Diagnostik mit wenig Wissen beschickt sind, konstruieren Fragebögen. Wozu sollte man sich also mit Fragebogenkonstruktion, oder genauer gesagt mit der Itemkonstruktion für Fragebögen auseinandersetzen?

Die Antwort auf diese Frage ist recht schnell gegeben: Weil viele Konstrukteure Items mit mangelnder Sorgfalt erstellen, und die aus ihrem Fragebogen entstandenen Ergebnisse mit fragwürdigen Methoden „validieren“. Das Ergebnis sind dann allzu oft Fragebögen, die sich auf ein bekanntes Persönlichkeitskonstrukt beziehen, dieses aber in unkontrollierter Weise zu messen versuchen. Die gesamte Problematik ergibt sich aber erst bei dem Versuch, auf Basis der gewonnen Daten eine Prognose im Einzelfall zu stellen.

Der hier gewählte Ansatz und dessen Überprüfung ist ein erster Versuch Items systematisch nach einem auf der latenten Dimension „Extraversion“ basierenden Regelwerk, einem so genannten *Konstruktionsrational*, zu erstellen. Der wesentliche Vorteil dieser Vorgehensweise ist der, dass der Inhalt der Items kontrolliert nur im Rahmen der Systematik variiert und man somit Item-Merkmale unterscheiden kann, die das Antwortverhalten der Person entscheidend beeinflussen.

Diese Herangehensweise ist im Bereich der Persönlichkeitsdiagnostik neu und dringend notwendig, um nicht „fiktive“ Konstrukte zu messen die sich durch ein Konglomerat adhoc konstruierter Items definieren. Es stellt sich in der Persönlichkeitsdiagnostik genauso wie in der Leistungsdiagnostik die zentrale Frage, welche kognitiven Operationen bzw. Bewertungen an der Beantwortung der Items beteiligt sind. Denn werden dahingehend keinerlei Vorannahmen getroffen, und daher diese auch nicht geprüft, kann darüber *was* eigentlich gemessen wurde keinerlei Aussage getroffen werden.

2. Theoretischer Teil

2.1. Persönlichkeitskonzepte

Ziel der Differentiellen Psychologie und der Psychologischen Diagnostik ist und war schon immer eine grundlegende Taxonomie der Persönlichkeit zu finden, die zwar hinreichend komplex, aber dennoch sparsam menschliches Verhalten erklärt. In den meisten Fällen wird versucht Persönlichkeit durch die Messung so genannter „traits“ zu erfassen, so wie das Mischel, Shoda und Ayduk (2008) beschreiben:

> Guided by the assumption that stable dispositions exist, trait psychologists try to identify the individual´s position in one or more dimensions (such as neuroticism, extraversion). They do this by comparing people tested under standardized conditions. They believe that positions on these dimensions are relatively stable across testing situations and over long time periods. (S. 56)

Aus dieser Tradition heraus wurden viele Persönlichkeitsmodelle und Ansätze entwickelt, aus denen nun fünf bedeutende kurz dargestellt werden.

2.1.1. Das PEN-System

Dieses Modell zur Beschreibung der Persönlichkeit geht auf Hans-Jürgen Eysenck zurück. Es werden vier verschiedene hierarchische Levels (Type level, Trait level, Habitual response level, Specific response level) unterschieden. Auf dem obersten Level, dem „Type-level“, werden drei große – auch „Giant-three“ genannte – Typen unterschieden, nämlich: *Psychotizismus*, *Extraversion* und *Neurotizismus*. Diese „Typen“ sind methodisch gesehen Faktoren zweiter Ordnung. Das heißt, dass zuerst eine Faktorenanalyse mit den „Responses“ der Items durchgeführt wird (die resultierenden Faktoren sind jene die Eysenck dem Trait level zuordnet) und anschließend eine Weitere mit jenen aus der ersten Faktorenanalyse extrahierten Faktoren, was dann Sekundärfaktoren zum Ergebnis hat, die den „Typen“ nach Eysenck entsprechen. Herausragend an dem Ansatz den Eysenck verfolgte ist vor allem, dass er sich um eine experimentelle Überprüfung der von ihm postulierten Theorie bemühte. Vor allem wurden zahlreiche Studien zu den biologischen Grundlagen der drei postulierten großen „Typen“, insbesondere zur Extraversion, durchgeführt, die zu zahlreichen und unterschiedlichen Ergebnissen führten (Amelang, Bartussek, Stemmler & Hagemann, 2006; Asendorpf, 2007).

2.1.2. Die Big Five

Das derzeit wohl bekannteste Modell, das auf dem trait Ansatz aufbaut, ist das Modell der „Big Five“.

Die Ursprünge des Fünf-Faktoren-Modells der Persönlichkeit liegen in den 1930er Jahren, als Allport und Odbert im Rahmen einer „Psycholexikalischen Studie“ persönlichkeitsrelevante Begriffe aus einem Wörterbuch filterten. Die daraus gewonnene Adjektivliste stellte die Grundlage für zahlreiche Untersuchungen dar, die darauf basierten, dass Testpersonen Selbst– oder Fremdbeschreibungen durchführten. Die Ergebnisse der faktorenanalytischen Auswertung der Daten waren nicht ganz einheitlich, doch ermittelten die meisten Forscher eine Fünf-Faktoren-Struktur. Auch unter der Verwendung unterschiedlicher Adjektivlisten in unterschiedlichen Nationen konnten die fünf Faktoren zumeist repliziert werden. Einzig über die Benennung des fünften Faktors herrschte über die Jahre reger Disput. Dieser wurde mit unterschiedlichen Oberbegriffen wie „Culture“, „Intellect“ oder „Offenheit“ bedacht (Amelang et al., 2006).

Becker (2003a) beschreibt die heutige Bedeutung des Fünf-Faktoren-Modells so:

> Das Fünf-Faktoren-Modell hat viele Anhänger. Es wird von ihnen als gut repliziertes, kulturübergreifendes, umfassendes und orthogonales Bezugssystem für Persönlichkeitseigenschaften dargestellt, das sich auch bei der dimensionalen Erfassung von Persönlichkeitsstörungen bewährt. Ohne Zweifel sind dies attraktive Merkmale des Modells, und aus diesem Grund wurde in zahlreichen Forschungsprojekten darauf Bezug genommen (etwa im Rahmen der verhaltensgenetischen oder der kulturvergleichenden Forschung). (S. 329)

Zur angeblichen Replizierbarkeit der vorgegebenen Struktur merkte Becker (2003a) kritisch an, dass sich in mehreren Untersuchungen der fünfte Faktor eben nicht replizieren lässt, und dass hohe Interkorrelationen zwischen einzelnen der fünf Faktoren gefunden wurden, was deutlich gegen deren Unabhängigkeit spricht.

Noch grundlegendere Kritik als Beckers kommt von Moosbrugger und Hartig (2003, S. 144):„Hypothesen über die Dimensionalität von Variablen lassen sich im Rahmen der Faktorenanalyse nicht prüfen, sie kann lediglich deskriptive Kennwerte zur Kovarianzstruktur beobachteter Variablen liefern.“ Demnach ist eine eindeutige Bestimmung und damit auch eine wirkliche „Replizierbarkeit“ von bestimmten Faktoren oder deren Anzahl mithilfe der Faktorenanalyse gar nicht möglich. Zur Vertiefung der Problematiken der Faktorenanalyse siehe z. B.: Moosbrugger und Hartig (2002).

Die derzeit gebräuchlichste Benennung der Big Five ist wohl jene, die auch im NEO-PI-R (Ostendorf & Angleitner, 2004) zur Anwendung kommt. Die Struktur ist wie nun folgend definiert:

- Neurotizismus
- Extraversion
- Offenheit für Erfahrung
- Verträglichkeit
- Gewissenhaftigkeit

Speziell auf die Definitionen und Facetten der Extraversion wird später in Kapitel 2.2 noch genauer eingegangen.

2.1.3. Das Vier-Plus-X-Faktoren-Modell

Das Vier-Faktoren-Modell entstand quasi aus der Kritik am Fünf-Faktoren-Modell. Postuliert werden vier Persönlichkeitsdimensionen oder mehr (daher: Vier-Plus-X), die allerdings nicht gezwungener Maßen unabhängig voneinander sein müssen. So werden bei diesem Modell jene Punkte berücksichtigt, die beim Fünf-Faktoren-Modell häufig kritisiert wurden.

- Es werden – wenn notwendig – Interkorrelationen zwischen den Faktoren zugelassen, indem im Zuge der Faktorenanalyse eine Oblimin-Rotation durchgeführt wird, weil sich häufig gezeigt hat, dass die extrahierten Faktoren nicht unabhängig voneinander waren.
- Es werden nur vier robuste Persönlichkeitsdimensionen angenommen, da der fünfte Faktor in vielen Untersuchungen nicht replizierbar war. Der gemeinhin als „Offenheit“ bezeichnete Faktor wird hier der Extraversion zugeordnet.
- Es können insgesamt $\geqq 4$ Faktoren angenommen werden, da in zahlreichen Untersuchungen mehr als fünf Faktoren notwendig waren um einen Großteil der Varianz zu erklären.

Die vier postulierten robusten Persönlichkeitsdimensionen sind:

- Neurotizismus/geringe Seelische Gesundheit
- Extraversion/Offenheit
- Unverträglichkeit
- Gewissenhaftigkeit/Kontrolliertheit (Becker, 2003c, 2003a)

Als Beispiel eines Fragebogen-Inventars welches diesem Modell folgt, sei das „Big Five Plus One Persönlichkeitsinventar“ (Holocher-Ertl, Kubinger & Menghin, 2003) erwähnt, welches sechs unabhängige Persönlichkeitsdimensionen misst. Zusätzlich zu den eben genannten vier robusten Dimensionen, wird die in den Big-Five übliche Dimension „Offenheit“ (hier unabhängig von der Dimension Extraversion) sowie eine gänzlich neue, nämlich „Empathie“, postuliert.

2.1.4. Persönlichkeit in Situationen

Die drei bisher dargestellten Persönlichkeitstheorien sind der Eigenschaftstheorie des Personismus zuzuordnen. In diesen ist die betreffende Situation, in der Verhalten auftritt, nachrangig. Die Situation beeinflusst das Verhalten zwar, doch jede Person in der selben Form und Intensität; also quasi als additive Komponente (Amelang et al., 2006). Doch wurden Stimmen gegen diese Sichtweise Ende der 1960er Jahre laut und mündete in einer jahrelang andauernden „Person vs. Situation“-Debatte. Während die „Personismus“-Anhänger das Verhalten einer Person als durch deren Dispositionen verursacht sahen, und die Varianz im Verhalten zweier Personen mit derselben Disposition

quasi als „error-variance“, waren die „Situationismus“-Befürworter der Ansicht, dass das Verhalten einer Person ausschließlich durch die Situation bestimmt sei. Abweichungen im Verhalten zweier Personen in derselben Situation wurden ebenfalls quasi als „error-variance“ betrachtet (Mischel et al., 2008). Aus diesen beiden Extremen entstand ein Modell das beide Ansätze ineinander vereint: das Dispositionismus-Modell. Amelang et al. (2006) beschreiben das Modell folgendermaßen:

> Das Dispositionismus-Modell der Persönlichkeit sagt Verhalten also aus Situation-Verhaltensverknüpfungen, $S \rightarrow V$, gegeben eine Situation S_j, vorher. Dabei wird nur diejenige Gruppe J von $S \rightarrow V$ Verknüpfungen herangezogen, zu er auch S_j funktional äquivalent ist, z. B. verschiedene Beobachtungen von Hilfeverhalten am Arbeitsplatz, nicht aber in der Familie, wenn Hilfsbereitschaft bei der Arbeit vorhergesagt werden soll. (S. 76)

Dispositionen von einzelnen Personen werden in diesem Modell typischerweise als „wenn-dann“-Verknüpfungen dargestellt und werden interindividuell als stabil betrachtet. Eine diesem Modell entsprechende Persönlichkeitstheorie stammt von Mischel. Diese postuliert nun zwei Konsistenzmaße:

- Mittlere, transsituationale Verhaltenstendenz
- Intraindividuell variierende Erlebens- und Verhaltensweisen in Form von „wenn-dann“-Beziehungen

Das Erste stellt sozusagen den situationsübergreifenden „Mittelwert“ der Persönlichkeitseigenschaft dar, und das zweite Konsistenzmaß, die *Kohärenz*, die intraindividuelle, stabile Veknüpfung von Situation und Verhalten (Amelang et al., 2006).

Doch wie sind diese sich scheinbar widersprechenden Konsistenzmaße in Einklang zu bringen. Dazu führten Mischel et al. (2008) aus:

> Is it more useful to try to infer broad traits of situation-specific signatures? The answer always depends on the particular purpose. Inferences about global traits may have limited value for the practical prediction of a person´s specific future behavior in specific situations, or for the design of specific psychological treatment programs to help faciliate constructive change. But broad traits have many other uses. Indeed, they have value for everyday inferences about what other people seem like on the whole, as when you get first impressions about a stranger, or from a quick opinion about a new acquaintance. (S. 81)

2.1.5. Der act-frequency approach (AFA)

Dieser Ansatz unterscheidet sich von den zuvor genannten erheblich, da er sich nicht mit dem Auffinden von traits beschäftigt, sondern bestimmte traits als „kognitive Handlungsklassen“ als gegeben voraussetzt, und diesen traits bestimmte „acts“ zuordnet die mehr oder weniger prototypisch für diese sind (Amelang et al., 2006).

Buss und Craik (1983a) beschreiben diesen Ansatz sehr anschaulich folgendermaßen.

> The frequency analysis of dispositional constructs focuses on specifying the relative incidence of acts within circumscribed categories or domains. From the frequency perspective, the statement „Mary is arrogant“ means that, over a period of observations, she has displayed a high frequency of arrogant acts, relative to a norm for that category of acts. (S. 106)

Ein trait System wir im AFA als nützliche Kategorienbildung für bestimmte Verhaltensweisen gesehen. Buss und Craik (1983a, 1983b) führen ein Beispiel an, in dem sie einen aus ihrer Sicht adäquaten Weg beschreiben, um zu einem trait passende Verhaltensweisen zu generieren. Zuerst wurden Studenten angehalten an Personen zu denken, die eine hohe Ausprägung bezüglich eines bestimmten Adjektivs oder traits aufweisen, um anschließend an in diesem Zusammenhang für sie typische Verhaltensweisen zu denken. Diese Produktionen wurden anschließend zu 100 Handlungsweisen je Dimension zusammengefasst und mehreren Beurteilern vorgelegt, die diese Verhaltensweisen auf einer 7-stufigen Skala auf ihre Prototypizität für die betreffende Disposition beurteilen sollten. Die jeweils 100 Handlungsweisen umfassenden Skalen wurden jeweils in vier Gruppen geteilt, wobei die erste Gruppe die prototypischsten Items beinhaltete und die vierte Gruppe jene die dem Konstrukt nur sehr peripher nah kamen.

Die traits zu denen Verhaltensweisen konstruiert wurden, sind: „Unnahbarkeit“, „Verträglichkeit“, „Dominanz“, „Geselligkeit“, „Streitsucht“ und „Unterwürfigkeit“[1]. Diese wurden anschließend in das „Wiggins circumplex model“ (vgl. Wiggins, 1979), welches 16 Persönlichkeitsdimensionen postuliert, eingebettet. Dieses Modell hat den Vorteil, dass Korrelationen zwischen den einzelnen Dimensionen vorhergesagt werden und diese dann natürlich auch überprüfbar sind.

Interessant ist, dass bei ebendieser Überprüfung überwiegend positive Korrelationen entdeckt wurden, auch zwischen jenen Faktoren die laut Modell eigentlich negativ korreliert sein sollten, was Buss und Craik (1983a) so interpretieren:

> The positive correlations among most multiple-act criteria stem not from methodological artifacts but from a general activity or g factor in the act domain. In this case, individuals may reliably differ in the number of acts performed, regardless of the act category within which they are classified. (S. 114)

Es gibt also offensichtlich Personen, die unabhängig vom Kontext einfach mehr „acts“ zeigen. Der unmittelbare Zusammenhang zur Dimension „Extraversion“ ist auffällig. Eysenck brachte die Extraversion in direkten Zusammenhang mit dem „Aufsteigenden Retikulären Aktivierungssystem“ (ARAS), das bezüglich der Wachheit bzw. des Aktivierungsniveaus einer Person eine entscheidende Rolle spielt. Extravertierte weisen demnach ein unterempfindliches ARAS auf, und suchen daher vermehrt nach Stimulation, was sich in einem „Mehr“ an Verhalten ausdrückt (Amelang et al., 2006).

[1] Übersetzung durch den Autor der vorliegenden Diplomarbeit

2.2. Extraversion

Betrachtet man unterschiedliche Persönlichkeitstheorien und versucht das Gemeinsame, quasi als „kleinster gemeinsamer Nenner“, zu formulieren, fällt auf, dass Extraversion recht durchgängig vorzufinden ist. Interessant in diesem Zusammenhang ist sicherlich die bekannte Entdeckung in den 1940er Jahren, dass sogar Pantoffeltierchen sich in ihrer Verhaltensweise unterscheiden, und zwar was die Tendenz betrifft, sich entweder alleine („free swimmers“) oder in Körperkontakt mit anderen („groupers“) fortzubewegen (Amelang et al., 2006). Diese Verhaltensweisen könnte man als unterschiedliche Ausprägungen der Disposition Extraversion interpretieren.

Wie ist nun Extraversion in unterschiedlichen Persönlichkeitstheorien bzw. in unterschiedlichen Fragebögen operationalisiert?

2.2.1. Myers-Briggs Typenindikator (MBTI)

Gleich zu Beginn sei jenes psychologische Verfahren angeführt, das auf der Typenlehre von C.G. Jung aufbaut, der sich als wohl einer der Ersten mit dem Konstrukt der Extraversion beschäftigte. (Mischel et al., 2008) Insgesamt werden mit dem MBTI vier „Präferenzen“ erfragt:

- Außenorientierung/Extraversion – Innenorientierung/Introversion
- sinnlich Wahrnehmen – intuitiv Wahrnehmen
- analytisch Beurteilen – gefühlsmäßig Beurteilen
- Beurteilung – Wahrnehmung

Diese Präferenzen werden anschließend zu 16 Typen verrechnet, auf die hier nicht näher eingegangen wird. Wesentlich interessanter für diese Arbeit ist die Operationalisierung des Extraversionskonstrukts (Briggs & Myers, 1991):

> Nach außen orientierte (extravertierte) Personen orientieren sich vorwiegend an der Außen– bzw. Umwelt, d.h., sie tendieren dazu, Wahrnehmung und Beurteilung auf Menschen und Gegenständliches zu richten. Nach innen orientierte (introvertierte) Personen nehmen dagegen vor allem die Impulse der eigenen Innenwelt auf, d.h., ihre Wahrnehmung und Beurteilung richten sich vorwiegend auf geistige Vorstellungen und Ideen – sie sehen Objekte aus einer subjektiven Sicht. (S. 6)

Diese Definition von Extraversion unterscheidet sich von den weiter unten angeführten doch deutlich und ist wahrscheinlich am ehesten mit dem heutigen Konzept der „Feldabhängigkeit – Feldunabhängigkeit“ vergleichbar. Kubinger (2006) definiert den kognitiven Stil Feldabhängigkeit so:

> Der kognitive Stil Feldab- *vs.* Feldunabhängigkeit meint, dass bei einer Person des feldabhängigen (Wahrnehmungs-) Stils die Wahrnehmungsumgebung –

> in diesem Zusammenhang (Um-) „Feld“ bezeichnet – einen (zu) starken Einfluss ausübt und vom wahrzunehmenden Zielobjekt ablenkt, wohingegen bei einer Person des feldunabhängigen Stils die Wahrnehmung auf das wahrzunehmende Zielobjekt (leicht) fokussiert werden kann. (S. 248)

2.2.2. Eysenck Personality Profiler (EPP-D)

Im EPP-D der auf Eysencks PEN Modell (siehe: Kapitel 2.1.1) basiert, beschreiben Eysenck, Wilson und Jackson (1998) Extraversion bzw. extravertierte Personen folgendermaßen:

> Der typische Extravertierte sucht sozialen Anschluß, liebt Parties hat viele Freunde, braucht eine Vielzahl an Menschen, mit denen er sprechen kann. Er befasst sich ungern mit der eigenen Person. Extravertierte brauchen andauernd Erregung, suchen Veränderungen oder Risiken. Sie sind im Allgemeinen impulsiv. Extravertierte lieben das Leben (easy going), machen und mögen Witze, haben in jeder Situation den richtigen „Spruch“ bereit und lachen viel. Extravertierte bevorzugen es, in Bewegung zu sein und vielerlei Sachen zu unternehmen. Sie tendieren zu Aggressivität und sind launisch. Sie haben ihre Gefühle nicht immer unter Kontrolle und neigen zur Unzuverlässigkeit. (S. 21)

Die Dimension Extraversion im EPP-D setzt sich durch folgende Subskalen zusammen:

- aktiv – passiv
- kontaktfreudig – kontaktscheu
- selbstbewusst – schüchtern
- ehrgeizig – anspruchslos

Die englische EPP-Originalversion aus dem Jahr 1991 beinhaltete noch 7 Subskalen für die Dimension Extraversion. Diese wurden aber auf vier reduziert, da ein großer Teil der Items im Zuge der deutschen Übersetzung und derer empirischen Überprüfung als ungeeignet befunden, und somit ausgeschieden wurden (Eysenck et al., 1998).

2.2.3. NEO-Persönlichkeitsinventar (NEO-PI-R)

Die geistigen Väter dieses Fragebogens, Costa und McCrae, gingen in ihrer ersten Version von nur drei großen Faktoren aus, zu denen *Neurotizismus*, *Extraversion* und *Offenheit* zählten. In den folgenden Jahren näherte sich der Fragebogen immer mehr dem Fünf-Faktoren-Modell an, und gilt heute als dessen prominentester Vertreter (Ostendorf & Angleitner, 2004).

Im NEO-PI-R beschreiben Ostendorf und Angleitner (2004) Extraversion so:

> Dem alltäglichen Sprachgebrauch entsprechend lassen sich Personen mit hoher Ausprägung als gesellig, gesprächig, freundlich, unternehmensfreudig und aktiv beschreiben. Extravertierte mögen die Gesellschaft anderer, sie fühlen sich wohl in Gruppen, sind aber auch durchsetzungsfähig, selbstbewusst, dominant, lieben aufregende Situationen und Stimulierungen. Sie neigen zu Optimismus, sind eher heiter gestimmt und sind gute Unterhalter. (S. 40)

Entsprechend dieser Definition sind auch die Facetten der Extraversion gewählt, die da wären:

- Herzlichkeit
- Geselligkeit
- Durchsetzungsfähigkeit
- Aktivität
- Erlebnissuche
- Positive Emotionen

Die im Manual dargestellten vergleichenden Untersuchungen von *Neurotizismus* und *Extraversion* zwischen dem NEO- und dem PEN-Modell, kamen weitgehend zu dem Schluss dass zwischen den beiden Modellen einen hohe Korrespondenz bezüglich dieser Dimensionen herrscht (Ostendorf & Angleitner, 2004).

2.2.4. Trierer Integriertes Persönlichkeitsinventar (TIPI)

Das TIPI basiert auf dem in Kapitel 2.1.3 dargestellten Persönlichkeitsmodell. Die Dimension Extraversion setzt sich hauptsächlich aus den nun dargestellten Skalen zusammen.[2]:

- Fröhlichkeit
- Offenheit für Neues
- Geselligkeit
- Streben nach Aufmerksamkeit
- Tatendrang
- Hedonismus
- Risikobereitschaft

[2] Aufgrund der obliquen Rotation sind einige Skalen mehreren der vier Primärfaktoren zugeordnet. Die Skalen stellen die „Hauptladungen“ des Faktors Extraversion dar.

- Selbstbehauptung
- Geldausgeben

Eine Besonderheit am TIPI ist, dass bereits im Zuge der Skalenkonstruktion alle 34 Primärskalen mit dem ordinalen Rasch-Modell erfolgreich überprüft wurden. Kein anderer derzeit am Markt befindlicher Fragebogen wurde auf diese Art konstruiert (Becker, 2002, 2003c). Dies spricht deutlich für eine sehr gewissenhafte und sorgfältige Itemkonstruktion.

Eine weitere Besonderheit ist, wie Becker (2003c) beschreibt, die Konstruktion der Items.

> Unterschiede im Ausprägungsgrad einer Persönlichkeitseigenschaft manifestieren sich in unterschliedlichen Wahrscheinlichkeiten des Auftretens der prototypischen Reaktionen in prototypischen Situationen. Eine Person mit hoher Ausprägung einer Persönlichkeitseigenschaft zeigt das entsprechende Verhalten oder Erleben in den eigenschaftsrelevanten Situationen mit größerer Wahrscheinlichkeit als eine Person mit einer geringeren Ausprägung dieser Persönlichkeitseigenschaft. (S. 12)

Diese Vorgehensweise orientiert sich stark an dem in Kapitel 2.1.5 dargestellten Act-Frequency-Approach.

2.3. Zur Konstruktion von Persönlichkeitsfragebögen

Will man bestimmte Aspekte der Persönlichkeit messen, kommt man in den meisten Fällen nicht umhin Persönlichkeitsfragebögen zu verwenden, vor allem auch weil diese in breit gefächertem Angebot vorliegen, während alternative Verfahren (bspw. Objektive Persönlichkeitstests) kaum vertrieben werden. Der Grund hierfür dürften vor allem wirtschaftliche Aspekte sein, da Fragebögen fast immer Gruppen von Personen vorgegeben werden können, der Testleiter nicht unbedingt ein ausgebildeter Psychologe sein muss und der Zeitaufwand im Vergleich zur erhaltenen Informationsdichte gering ist (Kubinger, 2006).

Der Fragebogen als diagnostisches Instrument erfreut sich also größter Beliebtheit, doch ist zu beachten dass es sich hierbei um Selbstauskünfte der betreffenden Person handelt (Becker, 2003b) und diese somit „besonders stark von Erinnerungsvermögen, Aufmerksamkeit, Selbsterkenntnis etc. abhängig und sowohl für unwillkürliche Fehler und Verzerrungen als auch für absichtliche Verfälschungen viel anfälliger als 'objektive' Testverfahren“ (Bortz & Döring, 2002, S. 190) sind. Der Schluss, den man aus diesen Tatsachen ziehen muss, ist, dass eine besonders sorgfältige Konstruktion vonnöten ist, um nicht noch zusätzliche Unschärfen den eben beschriebenen Problemen hinzuzufügen.

Betrachtet man nun die Struktur eines Fragebogens so kann man laut Rost (2004, S. 55) folgendes feststellen:„Das Item ist die *kleinste Beobachtungseinheit* in einem Test, sozusagen der elementare Baustein, aus dem ein Test aufgebaut ist. An einem Item lassen sich zwei Komponenten unterscheiden, nämlich der sogenannte *Itemstamm* und

das *Antwortformat*." Dies erscheint einleuchtend und trivial. Der Itemstamm präsentiert für gewöhnlich das Reizmaterial (im Falle eines Fragebogen bspw. eine Frage oder eine Aussage) und das Antwortformat legt dann die Art und Weise fest, wie die Testperson ihre Antwort ausdrücken kann.

2.3.1. Iteminhalt

Während bei Intelligenztests die einzelnen Aufgaben zumeist sehr strukturiert sind, und sie auch eine spezielle Fähigkeit zu messen intendieren, haben die Items eines Fragebogens oft nur geringe Gemeinsamkeiten, wie das Janke (1973; zitiert nach Bühner, 2006) pointiert darstellt.

> Die Itemsätze unserer Fragebögen stellen aber ein merkwürdiges Gemisch aus Fragen völlig unterschiedlicher logischer und empirischer Bezüge dar. Fragen nach konkreten biographischen Daten, nach Verhalten und Erleben in spezifischen Situationen, stehen neben solchen, bei denen der Proband über Zeit und Intensität Verhalten und Erleben in spezifischen oder völlig vagen (etwa „aufregenden") Situationen zu integrieren hat oder zu entscheiden hat, ob ein Verhalten bei ihm häufiger oder intensiver als bei einem nicht-spezifizierten anderen ist. Daneben werden Interessen, Präferenzen, Einstellungen, Bewertungen eigener oder fremder Verhaltensweisen erfragt. Zusätzlich wird er als „Konstrukt-Konstrukteur" oder „Konstruktvalidierer" eingesetzt in direkten Fragen wie „Ich bin ein ängstlicher, ein neurotischer Mensch". Items im Sinne von „habits", „traits" und „types" stehen nebeneinander in einem einzigen Fragebogen; Verhaltensweisen und Teilaspekte eben dieser Verhaltensweisen werden bedenkenlos summiert. Einige Items erfragen Häufigkeiten von Verhalten, andere – im gleichen Fragebogen aufgeführte – erfragen Reaktionsintensitäten. Andere Items wiederum stellen Gemische dar aus Häufigkeit und Intensitäten (Beispiel:„Ich reagiere manchmal in bestimmten Situationen stark"). (S. 67)

Diese Kritik hat an Aktualität nichts eingebüßt, und trifft auf heute weit verbreitete Fragebögen immer noch zu. Eine Lösung dieses Dilemmas wäre die strikte Kategorisierung der Fragebogenitems nach deren zugrundeliegender Abfragesystematik wie dies nachfolgend nun dargestellt wird.

Kategoriensysteme

Folgende Kategorisierung von Itemarten ist von Jankisz und Moosbrugger (2007) angeführt[3]:

Direkte vs. indirekte Items Hier entspricht die Kategorie „direkt" jenen Items die das interessierende Merkmal direkt abfragen (z. B.:„Sind sie gesellig?"), während „indirekte" Items für das Merkmal typische Indikatoren abfragen („Sind sie gerne unter

[3] die beispielhaften Verdeutlichungen stammen vom Autor der vorliegenden Diplomarbeit

Menschen?“). Die Abfrage von Verhaltensindikatoren hat den Vorteil, dass Missverständnisse bzw. divergente Interpretationen das Merkmal betreffend, vermieden werden.

Hypothetische vs. biographiebezogene Items Items können sich entweder auf hypothetische („Stellen Sie sich vor Sie hätten im Lotto gewonnen. Wem würden Sie als erstes davon erzählen?“), oder auf tatsächlich schon erlebte Situationen beziehen („In ? von 5 Fällen beginne ich beim Einkaufen mit einer mir unbekannten Person ein Gespräch.“). Problematisch bei hypothetischen Items ist sicherlich die recht hohe Wahrscheinlichkeit einer Fehleinschätzung seitens der Testperson; biographiebezogene Items gelten als zuverlässiger. Andererseits sind biographiebezogene Items oft schwierig zu konstruieren da die vorgegebenen Situationen beschränkt und oft nicht für jede Testperson sinnvoll sind.

Konkrete vs. abstrakte Items Diese Unterscheidung ist der zuvor getroffenen ähnlich. Während konkrete Items, wie der Name schon sagt, sich auf konkrete Sachverhalte beziehen („Wie verhalten Sie sich, wenn Sie ein unbekannter Mann auf der Straße beschimpft?“), lassen abstrakte Items Freiräume für Interpretationen („Wie nervenstark schätzen Sie sich während der Konfrontation mit einem angstauslösenden Reiz ein?“), was natürlich die Gefahr an Fehleinschätzungen mit sich bringt.

Personalisierte vs. depersonalisierte Items Hier wird zwischen Items unterschieden, die konkret die befragte Person ansprechen („Essen Sie gerne Spinat?“), und depersonalisierte Items, die quasi „allgemeingültig“ formuliert werden („Man sollte oft Spinat essen.“). Aus der ehrlichen Beantwortung der ersten Frage bekommt man konkrete Informationen zur Person, während die zweite Art zu fragen eher Einstellungen oder Konzepte erhebt.

Eine weitere, sehr nützliche Systematik für Fragebogen-Items stellen Angleitner, John und Löhr (1986, S. 69) zur Verfügung.[4]

- Description of reactions
  - Overt reactions – Dieser Kategorie werden prinzipiell beobachtbare Verhaltensweisen zugeordnet.
  - Covert reactions – Dieser Kategorie werden internale Reaktionen zugeordnet, die nicht beobachtbar sind. (z. B.:„Ich denke viel über mich nach.“)
  - Symptoms – Diese Kategorie bezieht sich auf physische Reaktionen.
- Trait attributions – Diese Kategorie beinhaltet Items mit Eigenschaftszuschreibungen, die am häufigsten in Form von Adjektiven abgefragt werden.
- Wishes and interests – Diese Kategorie beinhaltet Items die sich auf Wünsche und Interessen beziehen.

[4] Die Ausführungen zu den einzelnen Kategorien stammen vom Autor dieser Diplomarbeit, und sind demnach *keine* wörtlichen Zitate.

- Biographical facts – Diese Kategorie beinhaltet Items die sich auf vergangene Gegebenheiten beziehen.
- Attitudes and beliefs – Hierzu gehören alle Items die Überzeugungen oder Meinungen erheben.
- Others´ reaction to the person – Zu dieser Kategorie gehören jene Items die die Reaktionen bzw. Verhaltensweisen Anderer auf die eigene Person beschreiben.
- Bizarre items – Diese Items beschreiben ein unübliches, abnormales oder seltsames Verhalten.

Diese Systematik ist eng mit der oben dargestellten Kritik von Janke an der inhomogenen Zusammensetzung der meisten Fragebogenskalen verknüpft. Eine Vermischung der eben dargestellten Kategorien innerhalb derselben Fragebogenskala kann zu methodischen Artefakten führen und wird daher nicht empfohlen (Jankisz & Moosbrugger, 2007).

Itemformulierung

Itemkategorien sind also, wie eben erwähnt, insofern sinnvoll, als sie konkret einen bestimmten Rahmen vorgeben, in dem Items konstruiert werden können. Jankisz und Moosbrugger (2007) machen deutlich, dass in einem Fragebogen möglichst nur eine der bei Angleitner et al. (1986) erwähnten Kategorien vorkommen sollte. Nun soll eine Kategorie mit Inhalt gefüllt werden, und auch dazu gibt es zahlreiche Empfehlungen, von denen nun die relevantesten kurz dargestellt werden.

Eine gute und präzise Zusammenfassung, welche Merkmale günstige Fragen haben sollten, präsentieren Westhoff und Kluck (2003, S. 88). Diese beziehen sich zwar auf ein verbal durchzuführendes Interview mit einer Testperson, doch sind die meisten dieser günstigen Merkmale auch als günstig für Items im Rahmen eines Fragebogens anzusehen.

„Günstig sind Fragen und Aufforderungen die

1. inhaltlich:
   - sich immer auf konkretes individuelles Verhalten beziehen,
   - immer in einem eindeutigen Bezugsrahmen stehen,
   - nur einen Aspekt ansprechen
2. im Ausdruck:
   - möglichst kurz und treffend sind,
   - nicht suggestiv sind,
   - neutral sind hinsichtlich der Bewertung des erfragten Verhaltens“

Jankisz und Moosbrugger (2007); Bühner (2006); Bortz und Döring (2002) geben weitere Hinweise wie Items zu formulieren sind, dass bei ihrer Beantwortung möglichst wenige Probleme auftreten.

positive Formulierung Die Items sollten durchwegs positiv formuliert sein, da Verneinungen bzw. noch schlimmer doppelte Verneinungen im Zusammenhang mit einem Zustimmung messenden Antwortformat (z. B.: stimme zu – stimme nicht zu) leicht zu Verwirrungen, und damit auch zu Fehlurteilen führen können.

Kompliziertheit Komplizierte und zu lange Satzkonstruktionen sollten weitgehend vermieden werden. Dennoch sollte das Item präzise und ausführlich formuliert sein, um Interpretationsspielräume möglichst eng zu halten. Zu vermeiden sind auch, abseits der Komplexität der Itemstruktur, all jene Items, die für eine adäquate Reaktion Vorwissen benötigen. Dies wäre der Fall wenn Items z. B. Fremdwörter, Fachbegriffe oder unübliche Abkürzungen beinhalten, von denen man nicht annehmen kann, dass sie die Personen der Zielpopulation kennen.

Quantoren Das Verwenden von Quantoren, insbesondere von Universalausdrücken wie „alle“ oder „nie“ sollte so gut es geht vermieden werden. Vor allem machen Quantoren im Iteminhalt keinen Sinn, wenn im zugehörigen Antwortformat ebenfalls Quantoren zur Beurteilung herangezogen werden. (Zu vermeiden gilt bspw. „Ich gehe oft auf Parties“, kombiniert mit dem Antwortformat: oft – gelegentlich – selten – nie.)

Aussagenanzahl Einem Item darf nie mehr als eine Aussage zugrundegelegt werden, da sonst unklar ist, auf welche der Aussagen die Testperson reagiert hat. (Zu vermeiden gilt bspw. „Ich gehe abends gerne auf Partys und spreche dort mit vielen Menschen.“)

2.3.2. Antwortformat

Das Antwortformat bietet der Testperson den Rahmen, um auf den dargebotenen Iteminhalt zu reagieren. Der Vorteil des gebundenen Antwortformats ist die wesentlich höhere Verrechnungssicherheit und Ökonomie, da die Verrechnung auch mittels eines Computerprogramms erfolgen kann (Seiwald, 2003a). Dem Autor ist auch, abseits von Projektiven Verfahren, kein Persönlichkeitsverfahren bekannt, bei dem freies, schriftliches Antworten der Testperson ermöglicht wird.

Die Wahl, welches Antwortformat zur Anwendung kommt, ist wesentlich, da es ja die Reaktionen der Testpersonen adäquat abbilden soll. Man kann also zwischen vielen verschiedenen Antwortformaten unterscheiden, von denen nun einige dargestellt werden.

Ratingskala

Dieses Antwortformat wird in Persönlichkeitsfragebögen sicherlich am häufigsten verwendet. Gemeinsames Merkmal von Ratingskalen ist, dass der Iteminhalt anhand des dieser Skala zugrundeliegenden Bewertungssystems (bspw. Zustimmung/Ablehnung, nie/immer etc.) von der Testperson beurteilt wird. Doch ist eine Ratingskala nicht wie die Andere. Es gibt relevante Unterschiede bezüglich wichtiger Merkmale, die große Auswirkungen auf das Antwortverhalten der Testpersonen haben können. Also welche Varianten von

Ratingskalen gibt es, bzw. welche Entscheidungen sind zu treffen falls eine Ratingskala verwendet werden soll?

- **Anzahl der Skalenstufen**

  dichotom Das dichotome Antwortformat beinhaltet die geringst mögliche Anzahl an Skalenstufen, nämlich zwei, (bspw. ja – nein) und ist damit das informationsärmste Antwortformat, da pro Item je Person nur ein bit Information erhoben wird (Rost, 2004). Man zwingt die Testperson also, sich entweder für die eine oder für die andere Seite zu entscheiden; daher wird dieses Antwortformat auch häufig „forced choice“ genannt. Diese „Freiheitsbeschränkung“ kann erhebliche Auswirkungen auf das Antwortverhalten haben. Vor allem Reaktanzphänomene treten bei diesem Antwortformat häufig auf, was dazu führt, dass jene Testpersonen “untypisch bzw. willkürlich und unter Umständen ihren wahren Eigenschaften zuwiderlaufend reagieren“ (Kubinger, 2006, S. 133). Problematisch ist jedenfalls auch, dass sich dichotomes Antwortformat zur, bei Persönlichkeitsfragebögen massiv angewandten, faktorenanalytischen Auswertung nicht eignet (vgl. Moosbrugger & Hartig, 2002; Rasch & Kubinger, 2006; Moosbrugger & Hartig, 2003; Amelang et al., 2006).

  analog Das andere Extrem ist die Analogskala, also ein kontinuierliches Antwortformat, die es ermöglicht, sich „stufenlos“ darzustellen. Dieses Format wird vor allem in den letzten Jahren, aufgrund der verstärkten Verwendung von Computern in der psychologischen Diagnostik, häufiger angewandt. Die Testperson muss sich also an irgendeinem beliebigen Punkt auf einem Kontinuum zwischen zwei benannten Endpunkten einordnen. Diese hohe Differenziertheit der Messung ist einerseits wünschenswert, andererseits ist wohl die Differenziertheit des Urteils der Testperson im Regelfall geringer ausgeprägt (Jankisz & Moosbrugger, 2007). Ein Vorteil von Analogskalen ist, wie man bei Kubinger (2006) nachlesen kann, dass es messtheoretisch keinen Unterschied macht, ob nachträglich dichotomisiert wird um mit dem dichotomen Rasch-Modell zu prüfen oder ob das Rasch-Modell für kontinuierliche Antwortskalen (vgl. Müller, 1999) eingesetzt wird.[5] Als Beispiel eines Persönlichkeitsfragebogens mit kontinuierlicher Antwortskala, der auch mittels des Rasch-Modells kalibriert wurde, sei hier der B5PO (Holocher-Ertl et al., 2003) angeführt.

  mehrkategoriell Wie viele Abstufungen sollten nun vorhanden sein für den Fall, dass mehr als zwei Antwortkategorien, aber weniger als „unendlich“ viele zur Anwendung kommen und soll es sich um eine gerade oder ungerade Anzahl von Ausprägungen handeln? Bezüglich der Anzahl der Skalenstufen ist eine pauschale Empfehlung kaum möglich. Die Anzahl sollte zwar nicht zu groß gewählt werden um die Testpersonen nicht zu überfordern, aber dennoch groß

[5] Von Vorteil ist dies vor allem, weil zum kontinuierlichen Rasch-Modell kaum Software zur Verfügung steht, und deswegen in den meisten Fällen eine dichotome Auswertung einfacher zu bewerkstelligen ist.

genug, um diesen die Möglichkeit zu geben, sich differenziert genug darzustellen. Die Frage, ob eine Mittelkategorie Sinn macht oder ob lieber eine gerade Stufenanzahl gewählt werden sollte, mündet unweigerlich im „Ambivalenz-Indifferenz-Problem". Denn wird eine ungerade Stufenanzahl gewählt, steht die Mittelkategorie quasi „zwischen" den Polen, was sie zur neutralen Mitte macht. Wählt nun eine Testperson diese Mittelkategorie, ist unklar, was das in Bezug auf die Frage bedeutet; also ob sie bspw. einfach nicht antworten will, keine Meinung hat, oder vielleicht tatsächlich eine mittlere Ausprägung aufweist. Die Problematik der Mittelkategorie zeigte sich auch bei der Analyse des NEO-PI-R (5-stufiges Antwortformat) durch Rost, Carstensen und Davier (1999), bei der sich herausstellte, dass die Mittelkategorie in den meisten Fällen mit der zweiten Antwortkategorie vertauscht war. Aufgrund dieses Ergebnisses wurden, für die weiteren Analysen die zweite und die dritte Kategorie zusammengelegt.

- **unipolar vs. bipolar**

  Unipolar ist eine Skala dann wenn sie einen „Nullpunkt" hat, also das eine Extrem lediglich das geringste Ausmaß der Wahrscheinlichkeit/Zustimmung etc. darstellt, während das andere Extrem das größtmögliche Vorhandensein ausdrückt. Eine bipolare Skala reicht von einem extremen positiven Pol zu einem negativen Pol. Der eine Pol stellt genau das Gegenteil des Anderen dar (Jankisz & Moosbrugger, 2007). Welches Format nun das Günstigere ist, lässt sich so pauschal nicht beantworten; Seiwald (2003a) beschreibt dies so:

  > Ob ein Item auf einer uni- oder bipolaren Ratingskala zu beantworten ist, hängt zum einen vom jeweiligen Iteminhalt, zum anderen von der zu erfassenden Persönlichkeitseigenschaft ab, welche auch wiederum unipolar (z. B. Ängstlichkeit) oder bipolar (z. B. Introversion vs. Extraversion) definiert sein kann. (S. 27)

- **Bezeichnung der Skalenpunkte**

  Falls eine Ratingskala nun mehrere Ausprägungen zulässt, gibt es unterschiedliche Möglichkeiten, diese zu bezeichnen:

  numerische Skala Eine numerische Skala ist aufgrund ihrer Knappheit sehr attraktiv, jedoch nur dann einzusetzen, wenn Testpersonen mit dieser abstrakten Darstellungsform umzugehen wissen (z. B.: „Ich bin ein geselliger Mensch." 1 – 2 – 3 – 4 – 5).

  verbale Skala Weniger abstrakt, jedoch damit zumeist eindeutiger, sind Abstufungen, die mit bestimmten Bewertungen, Begriffen, Intensitäten oder Häufigkeiten bezeichnet werden. Konkrete Häufigkeitsangaben (bspw. „zweimal täglich") sind aufgrund ihrer Eindeutigkeit besonders zu bevorzugen, da sie nicht an verbale Quantoren (bspw. „eher selten") gekoppelt sind. Ein Beispiel, wie oftmals Wahrscheinlichkeiten zu messen versucht werden, ist: nie – selten –

manchmal – oft – immer. Die besondere Schwierigkeit bei der verbalen Benennung ist Beschreibungen zu finden, die auch einer eindeutigen Rangordnung folgen.

Symbol Skala Diese Art Abstufungen zu kennzeichnen ist besonders anschaulich und wird vor allem bei Fragebögen für Kinder eingesetzt. Häufig kommen Smilies zum Einsatz die bspw. die Zustimmung zu einer bestimmten Aussage ausdrücken. Allerdings werden grafische Abstufungen auch bei einigen Persönlichskeitsfragebogen für Erwachsene verwendet, um die einzelnen Skalenpunkte zu verdeutlichen, wie dies z. B. im AVEM (Schaarschmidt & Fischer, 2003) geschieht.

Forced-choice

Wie schon in Kapitel 2.3.2 erwähnt, wird eine Ratingskala mit nur zwei Ausprägungsstufen (bspw. ja – nein) auch als „forced choice“ bezeichnet. Diese Bezeichnung wird allerdings auch für ein etwas anderes Antwortformat verwendet, das die Testperson quasi zu einer Entscheidung zwischen Alternativen zwingt. Seiwald (2003a) zu dieser Art des Antwortformats:

> Einige Testverfahren verwenden eine spezielle Form des Multiple-choice, bei der nicht einzelne Fragen oder Statements beantwortet bzw. bewertet werden sollen, sondern die bei jedem Item eine Entscheidung zwischen (meist zwei) gleichzeitig dargebotenen Aussagen erfordert. Diese Form des „Erzwingens“ einer Entscheidung trägt die Bezeichnung „Forced-choice“. (S. 28)

Wichtig zu beachten ist wie Bühner (2006, S. 61) schreibt, „dass dieses Antwortformat lediglich Aussagen über die relative Ausprägung einer Eigenschaft im Vergleich zu einer anderen Eigenschaften einer Person zulässt.“ Das ist insofern unpraktikabel, als man doch oft bestrebt ist Personen miteinander zu vergleichen, was voraussetzt, dass die absolute Ausprägung von Personen auf der Eigenschaftsdimension gemessen wird, was bei diesem Antwortformat nicht der Fall ist. Dieses Antwortformat verwendet bspw. der Berufs-Interessen-Test II (Irle & Allehoff, 1984).

Q-Sort

Dies ist ein sehr selten verwendetes Antwortformat. Die Testperson soll Kärtchen mit Statements in Kategorien, die ein mehr oder weniger starkes Zutreffen der Aussage ausdrücken, einordnen. Den Testpersonen werden zumeist Vorgaben bezüglich der Anzahl der Statements pro Kategorie gemacht, um so Antworttendenzen (vgl. Seiwald, 2003b) zu vermeiden (Seiwald, 2003a).

2.3.3. Fehlerquellen bei der Itembeantwortung

Bearbeitet eine Testperson Items eines Fragebogens, ist die implizite Annahme, dass dessen Antworten im direkten Zusammenhang mit dem interessierenden Merkmal stehen.

Treten jedoch systematische Fehler auf (Bias), so kommt es zu einer Verfälschung der Reaktionen auf die Items. Ein typischer, den inhaltlichen Antworttendenzen zuzuordnender Bias im Zusammenhang mit Persönlichkeitsfragebögen ist bspw. die „soziale Erwünschtheit". Es werden auch mannigfache formale Antworttendenzen unterschieden wie bspw. die „Akquieszenz" (ja-sage-Tendenz) (vgl. Seiwald, 2003b; Kubinger, 2006; Bortz & Döring, 2002; Jankisz & Moosbrugger, 2007). Das bloße Aufzählen von unterschiedlichen Fehlerquellen ist notwendig, um diese zu differenzieren. In weiterer Folge ist es nun aber wichtig, Strategien zu entwickeln, die den Einfluss dieser systematischen Fehler minimieren. Um dies zu erreichen muss der typische Antwortprozess einer Testperson betrachtet werden und wie dieser von typischen Fehlern beeinflusst wird. Podsakoff und MacKenzie (2003, S. 886) stellen in ihrer Arbeit den kognitiven Prozess bei der Bearbeitung von Testaufgaben dar.

1. **Comprehension** Der erste Schritt im Beantwortungsprozess ist, die Aufmerksamkeit gezielt auf das Item zu richten und die Aussage, Frage oder ganz allgemein gesagt die Aufgabenstellung zu verstehen. Die wohl häufigste Fehlerquelle in diesem Stadium ist die Mehrdeutigkeit von Items. Sind Items mehrdeutig formuliert, orientieren sich Testpersonen entweder an den bereits bearbeiteten Fragen oder antworten mithilfe einer Heuristik (bspw. konsequent zustimmen oder neutral antworten etc.) (Podsakoff & MacKenzie, 2003). Diese Fehlerquelle beschreibt auch Becker (2003b, S. 335):„Einzelne Wörter oder bestimmte verbale Formulierungen werden von verschiedenen Personen unterschiedlich aufgefasst. Dies gilt beispielsweise für die in vielen Fragebogen verwendeten Quantoren (z. B. 'manchmal' oder 'oft')." Dieses Problem ist bei Fragebögen allgegenwärtig, doch lässt es sich bspw., wie in Kapitel 3.2.1 dargestellt, durch bestimmte Itemkonstruktionsregeln, eindämmen.

2. **Retrieval** Der zweite Schritt beschreibt den Prozess des Abrufens der notwendigen Information aus dem Langzeitgedächtnis. Die hier auftretenden Fehler hängen z. B. mit den situationalen Bedingungen zusammen in der der Fragebogen vorgegeben wird, aber auch mit aktuellen Stimmungslagen der Testperson. Es werden hauptsächlich Aspekte erinnert, die eine zur derzeitigen Stimmungslage kongruente emotionale Färbung aufweisen (Podsakoff & MacKenzie, 2003). Eindämmen oder kontrollieren kann man diese Effekte wohl nur, indem man für die Vorgabe des Fragebogens ein für jede Person gleiches kontrolliertes Setting wählt oder bspw. die derzeitige Stimmungslage mit erhebt.

3. **Judgement** Erinnern alleine, wie in Schritt 2 dargestellt, führt noch nicht zu einem Urteil. Nun werden die abgerufenen Erinnerungen auf ihre Vollständigkeit und Exaktheit geprüft, eventuelle Unstimmigkeiten beseitigt und Schlussfolgerungen gezogen. Es werden alle Informationen zu einem einzigen Urteil verdichtet. Zu Problemen kommt es vor allem dann, wenn eine Testperson unstimmige, ambivalente Erinnerungen hat und unschlüssig ist oder diese Situation/Frage mit keiner Erinnerung verbunden ist. Sie orientieren sich dann häufig an den anderen Fragen und versuchen konsistent im Sinne des von ihnen vermuteten Konstrukts zu antworten oder, im Fall dass keine Erinnerung vorhanden ist, aufgrund von impliziten

Theorien oder Schlussfolgerungen zu einem Urteil zu kommen. Die problemverursachenden Einflussfaktoren aus Stadium 2 haben natürlich ebenfalls noch unmittelbar Auswirkungen auf die Urteilsbildung wie bspw. die aktuelle Stimmungslage der Testperson oder „priming effects“ (bspw. die Beantwortung der ersten Fragen beeinflusst die restlichen Fragen, da bestimmte Informationen in das Kurzzeitgedächtnis gelangen) (Podsakoff & MacKenzie, 2003). Insgesamt kann man diesen Problemen mit exakt formulierten, und idealerweise in der Biographie der Testperson verankerten Items begegnen, sodass dieser Schritt nicht zu viel kognitive Leistung in Anspruch nimmt.

4. Response selection Nachdem das Urteil getroffen wurde, versucht die Testperson dieses mit dem vorgegebenen Antwortformat abzubilden. Demnach sind die hier auftretenden Probleme eng an das verwendete Antwortformat geknüpft. Es handelt sich dabei vor allem um „Formale Antworttendenzen“ (vgl. Seiwald, 2003b), wie bspw. die Tendenz einer Person „extrem“ zu antworten, also die Randkategorien der Antwortskala verstärkt zu wählen (bei Ausprägungsstufen ≥ 3).

5. Response reporting „The final stage in the response process involves the editing of the responses for consistency, acceptability, or other criteria“ (Podsakoff & MacKenzie, 2003, S. 887). Mögliche, in diesem Stadium auftretende Fehlerquellen sind bspw. die Tendenz von Personen „sozial erwünscht“ oder, wie eben auch zuvor erwähnt, „konsistent“ zu antworten. Um dem Problem der „sozialen Erwünschtheit“, im Sinne der absichtlichen Verfälschung eines Fragebogens zu seinen Gunsten („Impression management“), Herr zu werden, wurden schon viele Strategien ersonnen, wie bspw. das Implementieren einer sogenannten „Lügenskala“ (vgl. Seiwald, 2003c), Begrenzung der Bearbeitungszeit (Khorramdel & Kubinger, 2006) oder der Aufforderung zu ehrlichem Antwortverhalten in der Verfahrensinstruktion. Soll eine andere Form, im Sinne der unbewussten Tendenz sozial erwünscht zu antworten („Self-deceptive Enhancement“), verhindert werden, ist es notwendig, den Personen Anonymität zuzusichern und vor allem ist auch darauf zu achten, dass keine wertenden Begriffe in Items oder Antwortalternativen vorzufinden sind.

2.4. Das Rasch-Modell

Die meisten psychologischen Verfahren, seien dies nun Tests oder Fragebögen, werden auf Basis der in der klassischen Testtheorie zur Verfügung stehenden Mitteln und Methoden erstellt. Als die wohl gängigste Methode zur Ermittlung der dimensionalen Struktur eines Datensatzes wird die „Hauptkomponentenanalyse“ oder die „Faktorenanalyse“ verwendet. Diese vermeintliche „Überprüfung“ des Datensatzes ist höchst problematisch, da es sich bei der Faktorenanalyse um ein rein exploratives, und nicht um ein prüfendes statistisches Verfahren handelt. Im Zuge der Anwendung ergeben sich noch weitere Probleme, deren Lösung vor allem an die Willkür des Forschers geknüpft ist; doch darauf wird hier nicht weiter eingegangen (vgl. Moosbrugger & Hartig, 2002; Rasch & Kubinger, 2006; Amelang et al., 2006). Eine wesentlich elegantere Methode ist, sich eines Modells

zu bedienen, das der Familie der Rasch-Modelle zugehörig ist, da es sich bei jenen um hypothesenprüfende Verfahren handelt.

2.4.1. Das Prinzip des dichotom logistischen Rasch-Modells

Kubinger (1989) beschreibt das Gemeinsame der Modelle die der „Item Response Theory"(IRT) zugeordnet werden können:

> Ausgehend von einer Stichprobe S, mit den Testpersonen $\{1,2, \ldots v \ldots n\}$ und einem Itempool I, mit den Items $\{1,2, \ldots i \ldots k\}$, ist ihnen allen eine wesentliche Idee gemeinsam: Für die beobachteten Reaktionen, also die *manifest* werdenden Verhaltensweisen der Testpersonen aus S auf die Items aus I sind weiter nicht näher beobachtbare, sog. *latente* „Eigenschaften" der Person verantwortlich zu machen. Und zwar stehen erstere mit letzteren modellspezifisch in wahrscheinlichkeitsfunktionalem Zusammenhang. (S. 21)

Der einfachste Fall ist jener in dem genau eine latente Dimension angenommen wird, und nur zwei Reaktionen (bspw. richtig/falsch, ja/nein etc.) unterschieden werden. In diesem Fall drückt sich der wahrscheinlichkeitsfunktionale Zusammenhang für die Lösung des Items i in folgender Weise aus:

$$P(+|\xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}} \tag{2.1}$$

Der Fähigkeitsparameter ξ einer Person v und der Schwierigkeitsparameter σ eines Items i sind zwischen $-\infty$ und ∞ lokalisiert. Wie man aus der Formel erkennen kann, hängt die Wahrscheinlichkeit das Item zu lösen bzw. ihm zuzustimmen (im Falle eines Persönlichkeitsfragebogens) nur von der Itemschwierigkeit und von der Lokalisation der Testperson auf der latenten Dimension ab. Je größer ξ und je kleiner σ ist, umso wahrscheinlicher wird die Lösung des Items, und umgekehrt (Kubinger, 1989).
Was sind nun die Besonderheiten dieses Modells?

erschöpfende Statistik Wird ein Test oder ein Fragebogen zur Bearbeitung vorgegeben, so ist der einfachste, und auch gebräuchlichste Verrechnungsmodus zur Ermittlung eines Testkennwerts jener, bei dem die Anzahl der gelösten Items summiert wird, ohne darauf zu achten *welche* Items nun tatsächlich gelöst wurden. Man nimmt also implizit an, dass die *Anzahl der gelösten Aufgaben* eine erschöpfende Statistik für die erzielte Leistung ist. Mit anderen Worten bedeutet dies „wenn man weiß, wieviele der Testaufgaben von einer Person gelöst wurden, hat man bereits die gesamte, im Testprotokoll enthaltene Information über den Leistungsgrad der Vp[6] ausgeschöpft" (Fischer, 1974, S. 195). Fischer (1974) zeigt in diesem Zusammenhang den Beweis durch vollständige Induktion, dass diese spezielle Verrechnungsvorschrift *nur* dann zulässig ist wenn für den betreffenden Itempool die Geltung des Rasch-Modells angenommen werden kann.

[6] Versuchsperson

spezifische Objektivität Das Rasch-Modell ermöglicht spezifisch objektive Vergleiche, was häufig auch als „Stichprobenunabhängigkeit“ bezeichnet wird. Das bedeutet, dass der Unterschied zweier Itemparameter σ_i und σ_j unabhängig von der gewählten Stichprobe bestimmt werden kann. Ebenso kann der Unterschied zweier Personenparameter ξ_v und ξ_w unabhängig von den vorgegebenen Items bestimmt werden. Diese wünschenswerte Eigenschaft ist nicht nur die Grundlage für die besonders elegante *conditional maximum likelihood* Methode, mit der unabhängig von der Verteilung der Parameter ξ_v die Itemparameter σ_i geschätzt werden können, sondern auch für den bedingten Likelihood-Ratio-Test nach Anderson (siehe Formel 2.4). Dieser stellt einen globalen Modellgeltungstest dar, der die Likelihood der Gesamtstichprobe jenen von p beliebigen Teilstichproben gegenüberstellt (Kubinger, 1989). Somit ist diese zentrale Annahme und damit die Geltung des Rasch-Modells prüfbar, was einen weiteren wichtigen Vorteil des Rasch-Modells kennzeichnet.

$$Z = -2 \ln \frac{L(\text{Daten in } S|\hat{\sigma})}{\prod_{l=1}^{p} L(\text{Daten in } S_l|\hat{\sigma}^l)} \tag{2.2}$$

Die Prüfgröße Z ist asymptotisch χ^2 verteilt mit $(p-1) \cdot (k-1)$ Freiheitsgraden.[7]

lokal stochastische Unabhängigkeit Dies bedeutet, dass die Reaktion einer Testperson auf ein Item ausschließlich von ihrer Fähigkeit in der betreffenden Dimension und von der Schwierigkeit des Items abhängt und nicht bspw. davon, welche Items sie schon gelöst hat oder in weiterer Folge noch lösen wird (Kubinger, 2003b). Dies ist quasi eine logische Schlussfolgerung aus der Grundgleichung 2.1 in die ja nur die beiden genannten Parameter eingehen.

2.4.2. Das Rasch-Modell bei Persönlichkeitsfragebögen

Die Anwendung des Rasch-Modells zur Kalibrierung von Persönlichkeitsfragebögen wird unter Experten kontrovers diskutiert. Kubinger (2000) nimmt dazu so Stellung:

> Persönlichkeitsfragebogen sind wohl kaum entsprechend modellkonform zu konstruieren: Die eigentlich interessierende Eigenschaft wird stets von einer zweiten, nämlich der, „wahr zu antworten“, überlagert, so daß alle eindimensional messenden Modelle scheitern müssen – was natürlich nicht gegen das Rasch-Modell, sondern gegen das Konzept von Persönlichkeitsfragebogen spricht. (S. 3)

Hier wird jene Tatsache angesprochen, dass Personen vor allem in Auswahlsituationen, psychologische Verfahren mit hoher Augenscheinvalidität zu ihren Gunsten zu verfälschen versuchen (Kubinger, 2006). Andererseits ist es natürlich so, dass – ebenso wie bei Leistungstests – bei Fragebögen Summenscores gebildet werden, um die Position der Testperson auf der latenten Dimension abzubilden. Dies ist aber, wie in 2.4.1 dargestellt,

[7] df= (Anzahl der Teilstichproben − 1) · (Anzahl der Items − 1)

nur dann zulässig, wenn der Itempool anhand des Rasch-Modells kalibriert wurde. Rost (2002) bringt dies so auf den Punkt:„The second answer is that they must be unidimensional, because we count responses and calculate sum scores from the item scores. This operation of counting and summing up implies that there is something in common in what we are counting or summing up.“ Rost et al. (1999) wenden bspw. das ordinale Rasch-Modell bzw. das Mixed Rasch-Modell (vgl. Rost, 2004) auf die einzelnen Skalen des NEO-FFI an. Dies ermöglicht nicht nur die dimensionale Überprüfung der einzelnen Skalen, sondern auch die Inspektion der Ordnung der Schwellenparameter, das Aufspüren unskalierbarer Personen und das Aufdecken bestimmter Antworttendenzen in unterschiedlichen Subgruppen. Fragebogendaten mittels des Rasch-Modells auszuwerten, bringt also wesentlich mehr und detailliertere Information als die bloße Anwendung korrelationsbasierter Methodik.

Abseits der Überprüfung der Item- und Skalenqualität ermöglicht ein Rasch-Modellkonformer Datensatz die adaptive Vorgabe der kalibrierten Items. Der Vorteil ist, dass durch die Vorgabe ausschließlich informativer Items bei annähernd gleich bleibender Messgenauigkeit ein großer Teil der Items eingespart werden kann. Die Anzahl der publizierten adaptiven Verfahren hält sich allerdings, wohl aufgrund des größeren Konstruktionsaufwands, in Grenzen (Kubinger, 2003a). Der Erforschung adaptiver Vorgabe von Fragebögen widmet sich Ortner (2005). In ihrer Arbeit wurde eine adaptive Version des EPP-D erstellt, diese an einer Realstichprobe erprobt und Analysen mittels simulierter Datensätze durchgeführt. Als problematisch erwies sich hier die insgesamt geringe Itemanzahl und die daraus resultierenden hohen Standardschätzfehler, vor allem bei Personen in extremen Merkmalsbereichen. Auch Embretson und Reise (2000) berichten über die Anwendung adaptiver Persönlichkeitsfragebögen und über noch höheres Einsparungspotential, was die Anzahl der vorzugebenden Items betrifft, wenn es sich dabei um polychotome Items handelt.

2.4.3. Das Linear Logistische Test-Modell (LLTM)

Das LLTM ist im Gegensatz zum Rasch-Modell, das der Kategorie *descriptive item response models* zuzuordnen ist, eines das versucht die Itemschwierigkeiten der einzelnen Items zu erklären, und ist somit zur Gruppe der *explanatory item response models* zu zählen. Konkret beschreibt das LLTM jeden einzelnen Itemparameter σ als Linearkombination von bestimmten *Basisparametern* η. Inhaltlich bedeutet dies, dass die Schwierigkeit eines Items auf bestimmte „Itemkomponenten“ zurückgeführt wird, die gemeinsam die Schwierigkeit der Items ausmachen. Als Gleichung lässt sich der Schwierigkeitsparameter σ nun wie folgt darstellen.

$$\sigma_i = \sum_{j=1}^{m} q_{ij}\eta_j \tag{2.3}$$

Während η_j die einzelnen Basisparameter darstellt, die im LLTM anstelle der Schwierigkeitsparameter σ_i geschätzt werden, beschreibt q_{ij} die Gewichte, die bereits vor der Parameterschätzung in der Strukturmatrix fixiert werden müssen und jeden beliebigen Wert annehmen können (Fischer, 1974).

Die typischste und häufigste Anwendung des LLTM ist sicherlich die Modellierung von, zur Lösung von Items erforderlichen, kognitiven Operationen. Hier wird a-priori je Item festgelegt, welche kognitiven Operation wie häufig durch eine Testperson zur Lösung des Items angewandt werden muss (vgl. bspw. Fischer & Formann, 1982; Hornke & Rettig, 1989; Sonnleitner, 2007). Weiters kann das Modell z. B. in der Veränderungsmessung angewandt werden oder zum Auffinden von anderen Einflussfaktoren wie bspw. Itemreihenfolgeneffekten (Hohensinn et al., 2008). Eine Übersicht verschiedenartiger Anwendungsmöglichkeiten gibt Kubinger (2008).

Ein wesentlicher Vorteil ist, dass – da das LLTM als Spezialfall des Rasch-Modells gesehen werden kann – dieses auch prüfbar ist. Dies geschieht wieder anhand eines Likelihood-Ratio-Test.

$$\mathrm{Z} = -2 \ln \frac{L(\text{Daten in } S|\hat{\eta})}{L(\text{Daten in } S|\hat{\sigma})} \tag{2.4}$$

Hier wird geprüft inwieweit die Likelihood des LLTM sich signifikant von der des Rasch-Modells, das hier quasi als „saturiertes Modell“ fungiert, unterscheidet. Diese Prüfgröße ist wieder asymptotisch χ^2 verteilt, mit $(k-1)-m$ Freiheitsgraden.[8] Wichtig ist hier anzumerken dass für die Durchführung des LLTM die Geltung des Rasch-Modells für den betreffenden Datensatz eine notwendige Voraussetzung darstellt. Unterscheiden sich die beiden Likelihoods nicht signifikant voneinander, erklärt das LLTM die Daten so gut wie das Rasch-Modell (Kubinger, 1989). Der wesentliche Vorzug des LLTM ist dass bei dessen Gültigkeit[9], dem Itempool „Konstruktvalidität“ attestiert werden kann, da man quasi nachgewiesen hat, *was* der Test misst. Ein weiterer Vorzug ist, dass theoretisch bei Geltung des LLTM für einen gewissen Itempool, durch Zusammensetzung der nachgewiesenen schwierigkeitsrelevanten Teiloperationen, quasi unendlich viele Rasch-Modell-konforme Items, mit a-priori bekannter Schwierigkeit, produziert werden können (Kubinger, 2003b).

Häufig ist es von Interesse, zwei (oder mehr) unterschiedliche Modelle, die auf unterschiedlichen Strukturmatrizen beruhen miteinander zu vergleichen, um zu ermitteln, welche den Strukturmatrizen zugrundeliegenden Annahmen den Datensatz besser erklären. Hierzu werden z. B. Informationstheoretische Maße verwendet. Diese werden mithilfe der Log-likelihood, der Anzahl der geschätzten Modellparameter und teilweise auch mit dem Stichprobenumfang berechnet. Der historisch erste Informationsindex ist das „Akaikes information criterion“ (AIC).

$$\mathrm{AIC} = -2 \ln(L) + 2n_p \tag{2.5}$$

Wie man sieht wird n_p, also die Anzahl der geschätzten Parameter, mit 2 gewichtet, und der AIC wird umso höher, je höher die Log-likelihood ist und je mehr Parameter geschätzt werden müssen. Ein weiterer Informationsindex ist das „Bayes Information Criterion“ (BIC).

$$\mathrm{BIC} = -2 \ln(L) + \ln(N) n_p \tag{2.6}$$

[8] df= (Anzahl der Items – 1) – Anzahl der Basisparameter

[9] Vorausgesetzt ist, dass die Basisparameter η des LLTM die zur Lösung bzw. Beantwortung des Items notwendigen kognitiven Operationen darstellen.

Dieses gewichtet die Parameteranzahl n_p, anders als der AIC, mit einem variablen Koeffizienten, nämlich mit dem natürlichen Logarithmus der Stichprobengröße N (Rost, 2004).

Für beide dargestellte Kriterien gilt: je niedriger desto besser. Im Normalfall sinkt die $-2\ln(L)$ mit steigender Parameteranzahl, was bedeutet dass ein Modell mit n Parametern besser passt als eines mit $n-1$ Parametern. Die Informationsindizes addieren allerdings noch die gewichtete Anzahl der geschätzten Parameter hinzu, was praktisch bedeutet, dass eine Steigerung der Anzahl der Parameter nicht zwangsläufig zu einer Verbesserung der Informationsindizes führt. Dies ist insofern wichtig, als man bei der Beschreibung von Daten meist bestrebt ist ein möglichst sparsames Modell zu wählen, also eines mit wenigen Parametern, das die Daten dennoch ausreichend gut erklärt.

Wichtig ist zu betonen, dass es sich bei jenen Indizes nicht um statistische Tests im Sinne einer inferenzstatistischen Überprüfung der Modellgüte handelt, sondern lediglich um einen deskriptiven Vergleich zwischen zwei oder mehreren konkurrierenden Modellen.

3. Empirischer Teil

3.1. Ziel der Untersuchung

Ziel der Untersuchung ist es, einen Fragebogen zur Erfassung von Extraversion zu erstellen, der in weiterer Folge auch adaptiv angewandt werden kann. Hierzu wurde eigens ein Konstruktionsrational entwickelt, auf dessen Basis die Itemerstellung erfolgte. Der erstellte Itempool soll anschließend dahingehend geprüft werden, inwiefern dieser dem dichotom logistischen Modell von Rasch genügt. Weiters soll überprüft werden, inwieweit die inhaltliche Itemstruktur die Schwierigkeit[1] der Items determiniert oder zumindest beeinflusst. Falls sich die Likelihood des auf der Strukturmatrix des Konstruktionsrationals basierenden LLTM als nicht signifikant unterschiedlich zu jener des Rasch-Modell erweist, kann man annehmen, dass die den Items zugrundegelegte Struktur ein passendes Erklärungsmodell darstellt. So hätten Items, die mit jenem Regelwerk konstruiert werden, den Vorteil, dass bei bekannter Struktur a-priori die Itemschwierigkeit bestimmbar ist, und somit mit wenig Aufwand der Itempool beliebig erweitert werden kann.

3.2. Die Verfahrenskonzeption

Die Vorgaben zur Erstellung des Fragebogens waren einerseits die Entwicklung eines Itempools der den trait Extraversion valide misst und andererseits auch genügend viele Items beinhaltet, um eine spätere adaptive Anwendung des Fragebogens zu ermöglichen. Eine mögliche, spätere adaptive Anwendung verlangt einen größeren Itempool (zumindest 60 – 70 Items) als bei einem konventionellen Fragebogen, und noch hinzu kommt, dass die Items in ihrer Schwierigkeit breit über das Kontinuum streuen sollten, um eine möglichst genaue Messung auch in den Extrembereichen zu ermöglichen (vgl. z. B. Kubinger, 2006).
Um diese Vorgaben zu erreichen, wurde ein Konstruktionsrational erstellt, nach welchem einerseits jedes einzelne Item in seiner Struktur erfasst wird, und andererseits das kontrollierte Variieren einzelner Strukturmerkmale zu einem Itempool von akzeptabler Größe führte. Messtechnisch bedeutet das schlicht, dass die Schwierigkeit eines Items durch dessen Strukturmerkmale erklärt werden soll. Man versucht also zu erschließen, aus welchem Grund die eine verbale Reizvorgabe häufiger Zustimmung findet als die andere. Wie in Kapitel 2.4.3 beschrieben, ermöglicht die erfolgreiche Anwendung des LLTM eine Beantwortung einer solchen Frage. Kubinger (1989) empfiehlt im Zusammenhang

[1] Im Falle eines Fragebogens zur Selbstbeschreibung ist die Schwierigkeit eines Items so zu verstehen, dass die Darstellung in Richtung des interessierenden Konstrukts bei ehrlicher Beantwortung unwahrscheinlicher wird, je „schwieriger“ das Item ist.

mit der beabsichtigten Anwendung des LLTM auf einen Itempool, diesen schon a-priori nach bestimmten Regeln (siehe Kapitel: 2.3) zu konstruieren, wie dies in dieser Arbeit auch durchgeführt wurde.

- Die meisten Items beziehen sich auf konkretes, beobachtbares Verhalten
- Die Items behandeln für den fraglichen trait (Extraversion) typische Indikatoren, und sind damit als „indirekte Items“ zu klassifizieren. Es werden also keine mit dem Konstrukt im Zusammenhang stehende Adjektive zur Beurteilung vorgegeben.
- Die Items wurden in der Absicht konstruiert, „biographiebezogen“ und „konkret“ zu sein. Es wurden also gängige Situationen konstruiert von denen man annehmen kann, dass die meisten erwachsenen Personen mit solchen schon konfrontiert wurden.
- Da die Items sich auf konkrete Verhaltensweisen der jeweils betreffenden Person beziehen, sind sie als „personalisierte“ Items zu kategorisieren. Das Antwortformat basiert auf jenem des TIPI (vgl. Becker, 2003c, 2002), da dieses der Testperson eine Wahrscheinlichkeitsabschätzung des dargestellten Verhaltens in der jeweiligen Situation erlaubt. Außerdem lässt dieses Antwortformat sechs Abstufungen zu, sodass sich jede Testperson ausreichend differenziert darstellen kann, und die Problematik einer Mittelkategorie[2] entfällt (siehe Kapitel 2.3.2).

Nach dem Festlegen dieser Kriterien wurden nun für das zu messende Konstrukt (Extraversion) typische Eigenschaftszuschreibungen bzw. Adjektive festgelegt, die in allen der in Kapitel 2.2 dargestellten Verfahren vorkamen. Diese wurden dann für eine Gruppe freiwilliger Studenten verwendet, die zuerst an einen/eine *typische/n Extravertierte/n* denken, und sein/ihr typisches Verhalten beschreiben und nachher zu den genannten Adjektiven und Eigenschaften in Gruppenarbeit prototypische Verhaltensweisen finden sollten. Diese wurden anschließend von mehreren Experten der Psychologischen Diagnostik kontrolliert, unpassende Produktionen ausgeschieden, diese um weitere Verhaltensweisen erweitert und mithilfe des Kategoriensystems variiert, angepasst und klassifiziert (vgl. Buss & Craik, 1983b, 1983a). Das angesprochene Konstruktionsrational sei hier in Folge nun kurz dargestellt.

3.2.1. Iteminhalt

Zweck

Welchem Zweck ist die dargestellte Situation dienlich?

1. Die dargestellte Situation dient der Verrichtung einer Alltagstätigkeit.
2. Die dargestellte Situation dient der Interaktion mit anderen Menschen.

[2] Wobei bei jenem Antwortformat die Anzahl der Ausprägungsstufen wohl nicht so entscheidend ist, da man diese als diskrete Ausprägungen einer von 0 – 100% reichenden Analogskala sehen kann.

3. Die dargestellte Situation dient der Ausbildung oder der Ausübung des Berufs.
4. Die dargestellte Situation dient der Offenbarung von Gefühlen.
5. Die dargestellte Situation ist global formuliert und lässt keine Rückschlüsse auf den eigentlichen Zweck zu.

Itembeispiel für 1:
In ? von 5 Fällen beginne ich beim Einkaufen mit einer mir unbekannten Person ein Gespräch.

Aufforderungscharakter

Wäre das bloße „sich in der Situation befinden“ schon ein Hinweis auf Extraversion?

1. Neutrale Situation: Sich in der dargestellten Situation zu befinden lässt keinen Rückschluss auf das Bedürfnis nach extravertiertem Verhalten zu.
2. Extraversionsgeladene Situation: Die dargestellte Situation wird zur Befriedigung von Extraversionsbedürfnissen aufgesucht. In diesem Sinne wäre anzunehmen, dass Personen mit einer höheren Ausprägung im trait Extraversion häufiger in derartige Situationen kommen oder eben diese aufsuchen.

Itembeispiel für 2:
In ? von 5 Fällen spreche ich in einer geselligen Runde von mir bekannten Personen am meisten.

Bekanntheitsgrad

Wie werden die in der Situation dargestellten Personen bezüglich ihres Bekanntheitsgrades definiert?

1. Bekannte Personen
2. Unbekannte Personen

Itembeispiel für 2:
In ? von 5 Fällen fühle ich mich unwohl, wenn ich in einer Gruppe mir unbekannter Personen im Mittelpunkt stehe.

Aktivität

Wie aktiv wird das Verhalten der Person gegenüber der Situation dargestellt?

1. Die dargestellte Situation wird von der Person aktiv und willkürlich herbeigeführt. Sie wird als deren Urheber dargestellt.
2. Die dargestellte Situation wird entweder passiv „ausgestanden“ oder die Handlung wird als ein „Mitmachen“ der Person beschrieben.

3. Die Situation gestaltet sich so, dass eine Kontrolle kaum möglich ist. Es handelt sich um Situationen die entweder plötzlich oder unvorhergesehen auftreten.

4. Die dargestellte Situation wird von der Person vermieden, was bedeutet, dass sie aktiv Energie aufwendet um aus einer solchen Situation zu entfliehen oder sie antizipiert und meidet.

Itembeispiel für 1:
In ? von 5 Fällen beginne ich beim Einkaufen mit einer mir unbekannten Person ein Gespräch.

Art der Kommunikation

Wie interagieren die dargestellten Personen miteinander?

1. verbal

2. nonverbal

Itembeispiel für 2:
In ? von 5 Fällen setze ich mich in einem öffentlichen Verkehrsmittel zu mir unbekannten Personen, obwohl es noch Einzelplätze gibt.

Kommunikationsstil

Welche Qualität hat die Kommunikation?

1. Diskussion

2. Konfrontation

3. neutrale Interaktion

4. Monolog

5. Führung/Dominanz

6. Persönliches Gespräch

Itembeispiel für 1:
In ? von 5 Fällen beteilige ich mich an ernsthaften Diskussionen mit mir bekannten Personen.

Emotionen

Mit welcher Emotion reagiert die Person auf die dargestellte Situation?

1. Die Person reagiert mit einer positiven Emotion auf die dargestellte Situation, oder zeigt positive Emotionen.

2. Die Person reagiert mit einer negativen Emotion auf die dargestellte Situation, oder zeigt negative Emotionen.

Itembeispiel für 2:
In ? von 5 Fällen ist es mir unangenehm vor einer Gruppe mir bekannter Personen zu sprechen.
Das hier ausführlich dargestellte Konstruktionsrational wird in Abbildung 3.1 grafisch dargestellt.

3.3. Untersuchungsdesign

Der Itempool umfasste 122 zu prüfende Items für die angenommen wird, Extraversion zu messen. Diese wurden online durch eine speziell für diesen Zweck angefertigte Software freiwilligen Teilnehmern zur Beantwortung vorgegeben. Auf die Besonderheit, dass sehr viele Items einer einzigen latenten Dimension vorzugeben waren, wurde im Testdesign Rücksicht genommen.

1. Um Ermüdungseffekte weitestgehend ausschließen zu können, die durch die Wahl eines vollständigen Designs (allen Testpersonen werden alle Items vorgegeben) entstanden wären, wurde ein „unvollständiges Design“ mit sechs unterschiedlichen Testformen gewählt, so dass jede Testperson maximal 62 Items die den trait Extraversion erfassen zur Bearbeitung vorgelegt bekam. Diese Testformen wurden untereinander so verlinkt, dass der Itempool gemeinsam kalibriert werden konnte, wie in Abbildung 3.4 zu sehen ist.

2. Es wurden in die eigentlich interessierenden Items zur Extraversion auch noch *Füll-items* zum Thema Gewissenhaftigkeit, die nicht in die Auswertung mit eingingen, gemischt. Diese dienten dazu, den Fragebogen abwechslungsreich zu halten, und um automatisiertes und stereotypes Antworten der Testpersonen zu vermeiden.

3.3.1. Online-Fragebogen

Zur Datenerhebung wurde aus den zuvor erstellten Items mit der Unterstützung eines Programmierers eine Online-Version entwickelt. Diese musste folgenden Anforderungen genügen.

- Um die Verständlichkeit der Instruktion zu maximieren, wurde der Online Fragebogen nach seiner Fertigstellung 12 Psychologiestudenten und Psychologiestudentinnen zur Bearbeitung vorgegeben. Diese füllten anschließend einen Feedback-Bogen

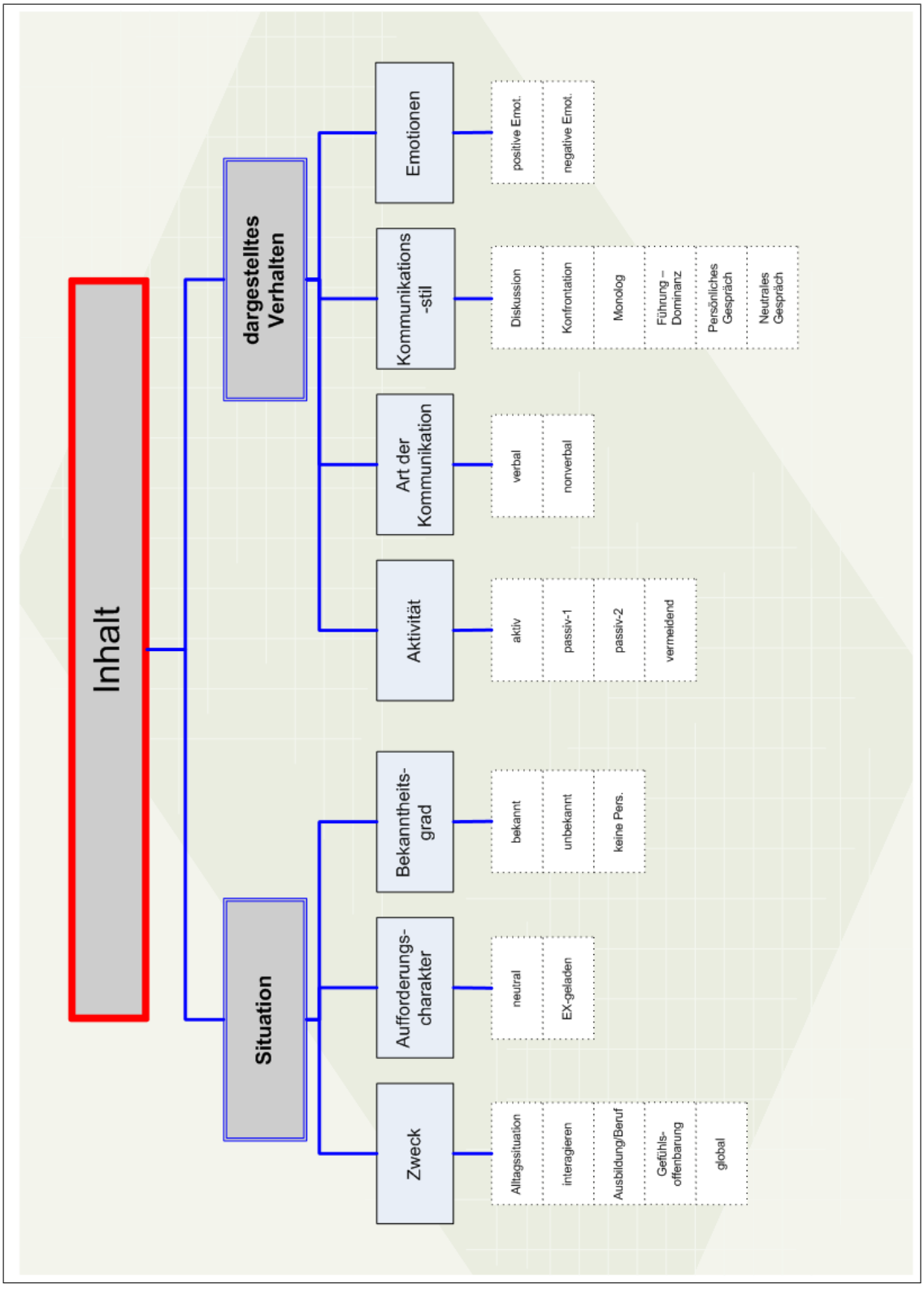


Abbildung 3.1.: Konstruktionsrational zur Entwicklung des Fragebogens

aus, der sich vor allem auf die Verständlichkeit der Instruktion bezog. Genannte und sinnvoll erscheinende Änderungsvorschläge wurden eingearbeitet. Des Weiteren wurde ein „Hilfe“-Button am rechten oberen Bildschirmrand implementiert, nach dessen Anklicken eine kurze, prägnante Zusammenfassung der Instruktion wiedergegeben wird (siehe Abbildung 3.2). Dies dient dazu, das Instruktionsverständnis auch über die gesamte Dauer der Bearbeitung aufrecht zu erhalten.

- Um zu gewährleisten, dass auch Testpersonen mit wenig Computer-Erfahrung den Fragebogen problemlos bearbeiten können, wurde dieser so programmiert, dass der Fragebogen ohne weitere Installation direkt im Browser ausgefüllt wird.
- Um zu verhindern, dass Testpersonen absichtlich oder unabsichtlich Teile der Instruktion zu schnell weiter klicken, erscheint der „Weiter-Button“ erst nach einem bestimmten, von der jeweiligen Textlänge abhängigen, Zeitintervall. Zusätzlich wird bei jedem Item die Bearbeitungszeit gespeichert, um so Personen ausfindig zu machen, die den Fragebogen zu schnell bearbeiteten, sich quasi nur „durchklickten“.
- Um Effekte zwischen den Items zu minimieren, erscheint jedes Item einzeln zur Beurteilung am Bildschirm (siehe Abbildung 3.3).
- Die Testperson wird während der gesamten Bearbeitung über die Anzahl noch zu bearbeitender Items und die geschätzten Restzeit informiert, die aus der Bearbeitungsdauer der ersten zehn Items geschätzt wurde. Diese zusätzliche Information soll die Absprungrate der Teilnehmer vor allem gegen Ende des Fragebogens minimieren.
- Es wurde ein Algorithmus entwickelt, der gewährleistet, dass jede Testform von annähernd gleich vielen Testpersonen bearbeitet wird.
- Um Reihenfolgeneffekte auszumitteln, mussten die Items bei jeder Testperson in randomisierter Reihenfolge vorgegeben werden.

3.3.2. Testdesign

Wie zuvor schon erwähnt, wurde ein unvollständiges Design gewählt, um die zumutbare Menge an Items zum selben Themenkreis nicht zu übersteigen und um so eine hohe Qualität der Daten zu garantieren. Das Design wurde so konstruiert, dass alle Items zusammen kalibriert werden können. Wie aus Abbildung 3.4 hervorgeht, ist jede Testform über die Items der einzelnen Blöcke, mit jeder anderen Testform zumindest indirekt verlinkt. Es wurden je 20 bzw. 21 Items zur Extraversion zu insgesamt 6 Blöcken, und je 6 Items zur Gewissenhaftigkeit zu insgesamt 3 Blöcken, zusammengefasst (siehe Tabelle 3.1). Zwischen den Blöcken wurden die Items so variiert, dass unterschiedliche Themenbereiche sich möglichst auf verschiedene Blöcke aufteilten. Außerdem wurde darauf geachtet, dass das Verhältnis positiv zu negativ gepolten Items in jedem Block

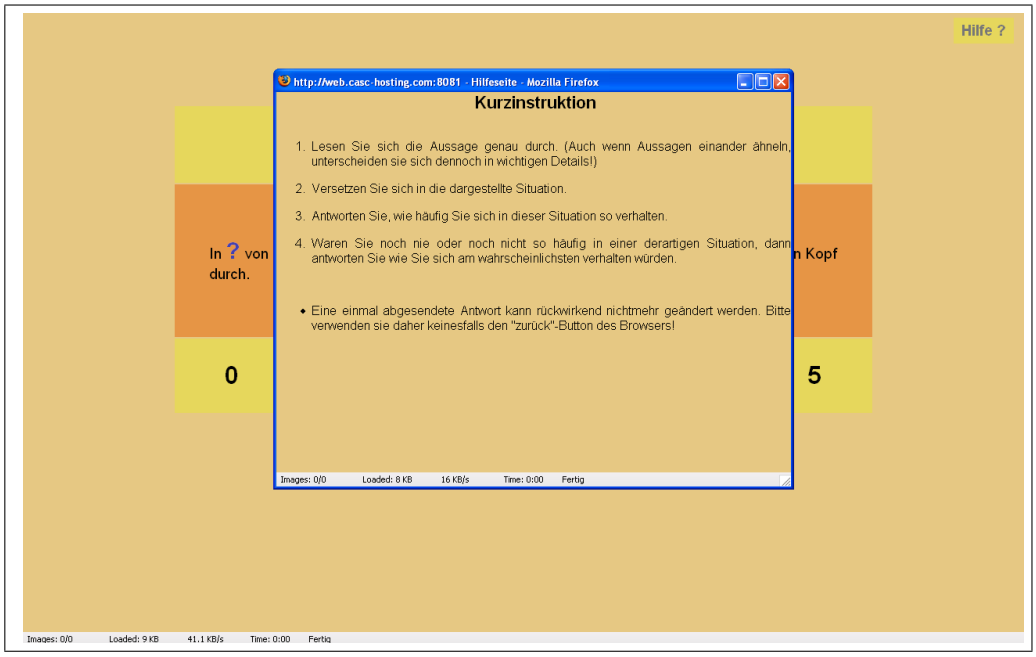


Abbildung 3.2.: Screenshot nach dem Anklicken des Hilfe-Buttons

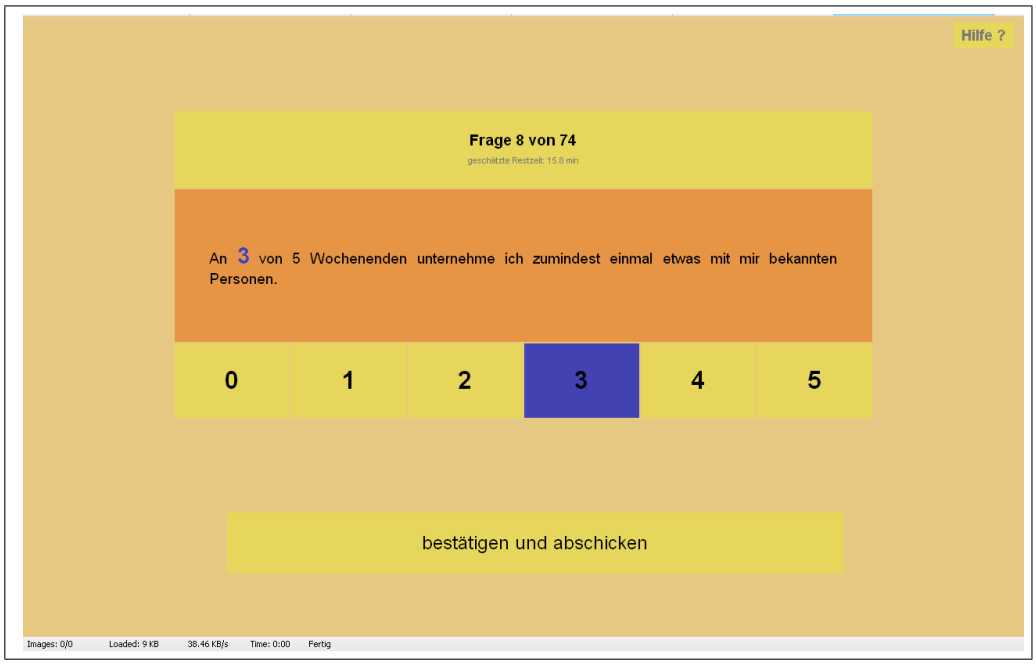


Abbildung 3.3.: Screenshot eines Items des Online Fragebogens

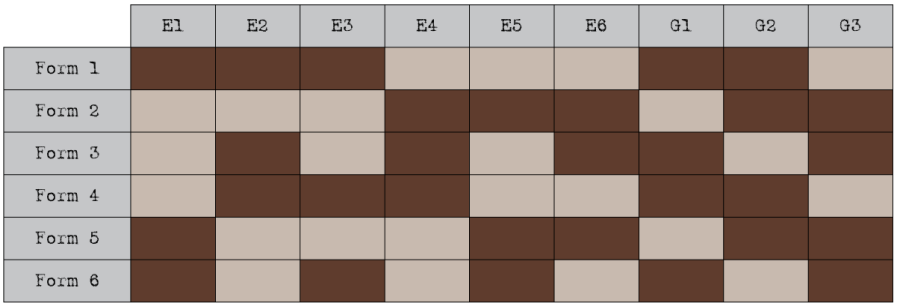


Abbildung 3.4.: Testformen im Überblick

Tabelle 3.1.: Blockgrößen

Itemblöcke	Anzahl der Items
E1	20
E2	20
E3	20
E4	21
E5	20
E6	21
G1	6
G2	6
G3	6

annähernd identisch ist. Wie in Abbildung 3.4 ersichtlich beinhaltet jede Testform drei Blöcke zur Extraversion und zwei zur Gewissenhaftigkeit. Dadurch dass jede Subgruppe von Items in 3 unterschiedlichen Testformen vorgegeben wird, ist erstens garantiert, dass jede Subgruppe mit jeder anderen zumindest einmal in einer Testform kombiniert wird, und zweitens dass jedes Item annähernd gleich oft bearbeitet wird.

3.3.3. Stichprobe

Der Fragebogen wurde zum einen Teil von Studenten und Studentinnen bearbeitet, die im Rahmen der Lehrveranstaltung „Übungen zur Psychologischen Diagnostik“ Stunden zur „Testerfahrung“ absolvieren mussten. Andererseits wurde der Fragebogen in diversen geschlossenen Online-Foren veröffentlicht, um so auch freiwillige Personen mit eventuell niedrigerem Bildungsgrad zu erreichen. Insgesamt bearbeiteten 560 Personen den Fragebogen.

Durch die automatische Miterfassung der jeweiligen Reaktionszeiten pro Item, wurden jene Personen ausgeschlossen, die aufgrund der beobachteten Reaktionszeiten zu wenig Zeit aufwendeten, so dass bei diesen von einer nicht ernsthaften Bearbeitung ausgegangen werden musste. Es wurden zwei Kriterien beachtet.

1. Personen die mehr als 10% der Items jeweils in weniger als drei Sekunden beant-

worteten.

2. Personen deren mittlere Antwortdauer jenen 2,5% zuzuordnen ist, die am schnellsten antworteten (Median der Antwortdauer $\leq 4,7\,sec$). Dies ist in Abbildung 3.5 dargestellt. Der rote Balken beinhaltet jene Personen die aufgrund ihrer niedrigen durchschnittlichen Antwortdauer von den weiteren Analysen ausgeschlossen wurden.

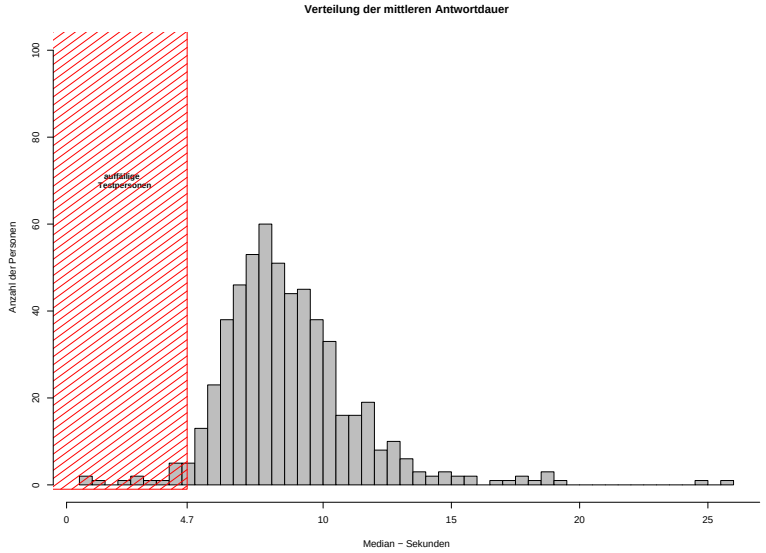


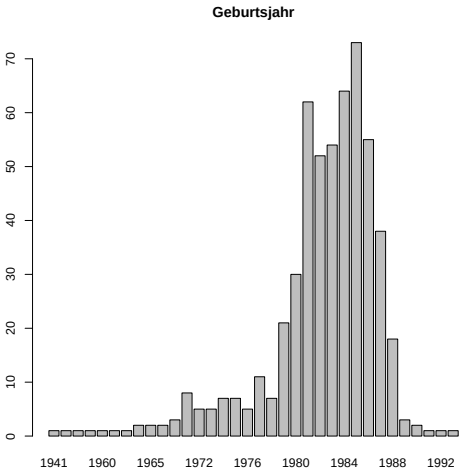
Abbildung 3.5.: Verteilung der Mediane der Gesamtantwortdauer

Alle 11 Personen die bezüglich des ersten Kriteriums auffällig waren, waren dies ebenso nach dem zweiten Kriterium. Insgesamt wurden so 14 Personen von den folgenden Analysen ausgeschlossen. Diese reduzierte Stichprobe setzt sich, wie in Abbildung 3.6 zu sehen ist, zusammen.

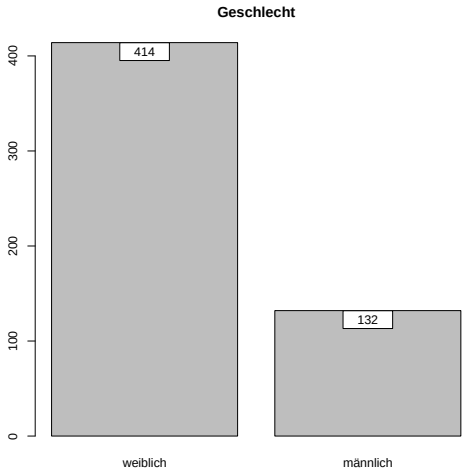
3.4. Ergebnisse

Zur nun folgenden statistischen Analyse des Datensatzes und zur graphischen Datenaufbereitung wurde die statistische Software R (R Development Core Team, 2007) herangezogen. Insbesondere fand das für R programmierte „package“ *eRm* (Mair & Hatzinger, 2007a, 2007b; Poinstingl, Mair & Hatzinger, 2007) Verwendung.

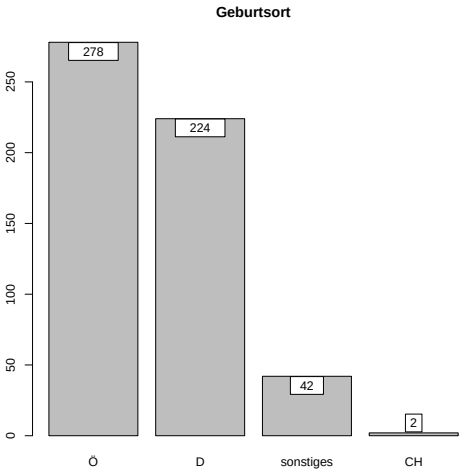
Wie schon in Kapitel 2.4.2 dargestellt, soll der gewonnene Datensatz dem Rasch-Modell genügen, um einerseits späteres adaptives Testen zu ermöglichen, und um andererseits die Architektur der Items mittels LLTM oder einem der mehrkategoriellen Verallgemeinerungen auf deren Geltung überprüfen zu können. Problematisch erwies sich der Versuch den Datensatz mittels des Partial-Credit Models (vgl. Kubinger, 1989;



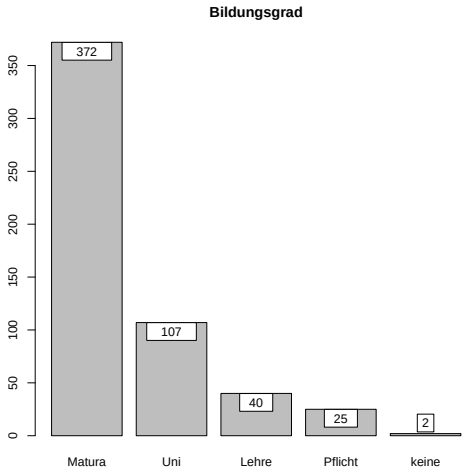
(a) Geburtsjahr



(b) Geschlecht



(c) Geburtsort



(d) Bildungsgrad

Abbildung 3.6.: Demographische Daten

Rost, 2004) auszuwerten, da bei zahlreichen Items eine oder mehrere Antwortkategorien entweder gar nicht, oder nur extrem selten gewählt wurden. Wohl aufgrund dieser geringen Informationsdichte pro Antwortmöglichkeit bei einigen Items, kam es auch beim ersten Versuch einer Parameterschätzung zu einem Abbruch durch das Programm, ohne ein verwertbares Ergebnis zu hinterlassen. Daher wurde der Datensatz nachträglich dichotomisiert, um eine Auswertung nach dem dichotom-logistischen Modell von Rasch zu ermöglichen. Die Antwortkategorien 0 bis 2 wurden zu 0, und 3 bis 5 zu 1 zusammengefasst. Für die inferenzstatistischen Analysen wurde durchgängig ein Signifikanzniveau von $\alpha = 0.01$ gewählt.

3.4.1. Ergebnisse der Analysen hinsichtlich Rasch-Modell-Konformität

Die Stichprobe wurde nach unterschiedlichen Kriterien geteilt, um die Teilstichproben – wie in Kapitel 2.4.1 beschrieben – anhand des Likelihood-Ratio Tests (LRT) nach Anderson und anhand des Grafischen Modelltests mit der Gesamtstichprobe zu vergleichen. Als Teilungskriterien fungierten:

- Score

  Anhand des Medians der erzielten Summenscores jeder Person wurden Personen die weniger extravertiertes Verhalten berichteten, gegen Personen, die mehr extravertiertes Verhalten berichteten, geprüft.

Anhand der demographischen Daten wurde der Datensatz auch nach folgenden Kriterien geteilt, und analysiert:

- Geschlecht („männlich“ vs. „weiblich“)
- Alter (1941–1983 vs. „jünger“)
- Bildungsgrad („kein Schulabschluss“, „Pflichtschule“, „Lehrabschluss“ vs. „Matura“ , „Universitätsabschluss“)
- Geburtsort („Österreich“ vs. „anderes Land“)

Tabelle 3.2.: Ergebnisse LRT – kein Item ausgeschlossen

	Teilungskriterium	Andersen χ^2	df	$\chi^2_{\alpha_{1\%}}$	
1	Score	227.897	121	160.1	sig.
2	Geschlecht	212.503	121	160.1	sig.
3	Alter	148.363	121	160.1	nicht sig.
4	Bildungsgrad	124.416	118	156.65	nicht sig.
5	Geburtsort	201.759	121	160.1	sig.

Nach den ersten Analysen des des Gesamtitempools zur Extraversion war in 3 von 5 Teilungskriterien der LRT nach Andersen signifikant, wie in Tabelle 3.2 dargestellt. Im Teilungskriterium „Bildungsgrad“ waren drei Items „not well-conditioned“, da die Schwierigkeitsparameter dieser Items, aufgrund fehlender Variation der Rohscores innerhalb einer der Teilstichproben, nicht geschätzt werden konnten. Um zumindest eine post-hoc Modellgültigkeit im Sinne des Rasch-Modells zu erzielen, wurden nun sukzessive Items mit niedrigem Item-fit ausgeschieden, solange bis in allen Teilungskriterien ein nicht signifikanter LRT resultierte. Der Item-fit der jeweiligen Items wurde mittels des Grafischen Modelltests und mittels des Wald-Tests ermittelt.

Nach der Reduktion des Itempools um 18 Items, mussten in keinem Teilungskriterium statistisch signifikante Abweichungen, festgestellt werden (siehe Tabelle 3.3). Auch bei der Inspektion der Grafischen Modelltests ist eine leichte Verbesserung, durch das Ausscheiden dieser nicht modellkonformen Items, zu bemerken (vgl. Abbildung 3.7 und 3.8).

Die auszuscheidenden Items wurden einer inhaltlichen Analyse unterzogen, und auf etwaige Gemeinsamkeiten geprüft. Auffällig ist, dass 15 der 18 ausgeschiedenen Items in der Kategorie „Art der Kommunikation“ der Ausprägung „nonverbal“ zuzuordnen sind – vor allem da nur ca. 37% der Items jener Kategorie angehören. Weiters konnten keine gemeinsamen Auffälligkeiten an den nicht modellkonformen Items gefunden werden. Warum der Großteil der auszuscheidenen Items sich auf „nonverbales“ Verhalten beziehen ist unklar. Vermutlich ist gerade das „verbale“ Verhalten jenes, das für die Dimension Extraversion typisch ist.

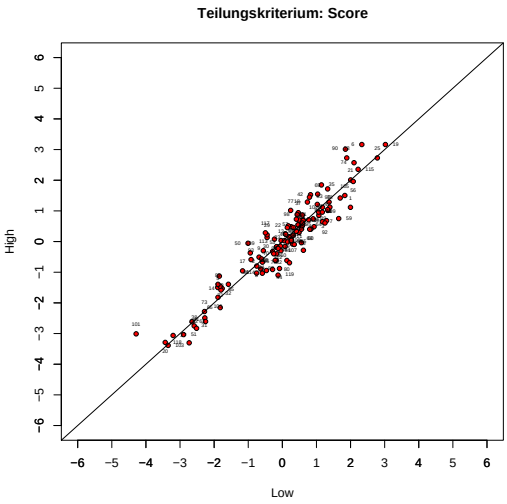
Der auf 104 Items reduzierte Itempool erfüllt nun die Bedingungen die notwendig sind, um weitere Modelltests mittels des LLTM durchzuführen.

Tabelle 3.3.: Ergebnisse LRT – 18 Items ausgeschlossen

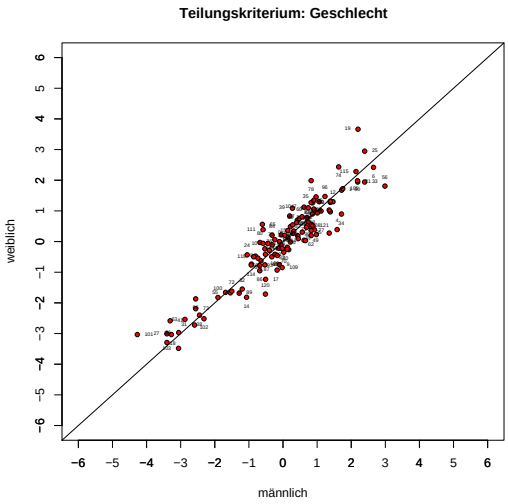
	Teilungskriterium	Andersen χ^2	df	$\chi^2_{\alpha_{1\%}}$	
1	Score	136.471	103	139.3	nicht sig.
2	Geschlecht	119.632	103	139.3	nicht sig.
3	Alter	129.144	103	139.3	nicht sig.
4	Bildungsgrad	100.309	100	135.8	nicht sig.
5	Geburtsort	133.365	103	139.3	nicht sig.

3.4.2. Ergebnisse der Analysen nach dem LLTM

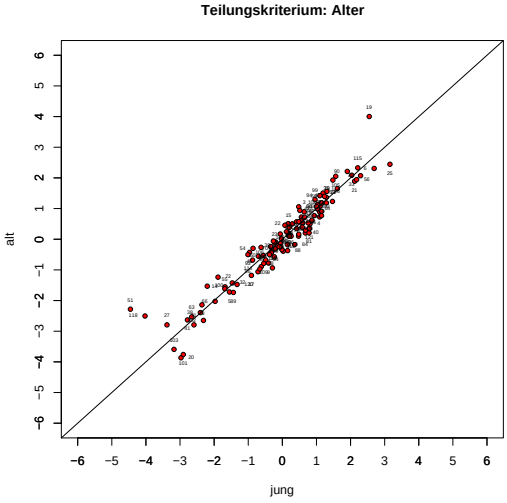
Die Struktur der erstellten Items wurde detailliert in Kapitel 3.2 dargelegt. Nach dem Ausscheiden nicht Rasch-Modell-konformer Items, wird nun das Konstruktionsrational mittels LLTM überprüft. Betrachtet man die Strukturmatrix (siehe Tabelle 3.4) des LLTM, so sieht man, dass die einzelnen Basisparameter (in den Spalten) der Struktur des Konstruktionsrationals entsprechen. Eine eingetragene „1“ bedeutet die Gewichtung des in der Spalte befindlichen Basisparameters mit eins bezüglich des jeweiligen Items.



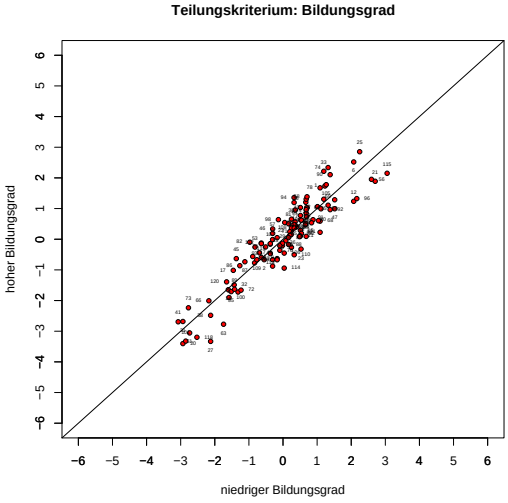
(a) Score



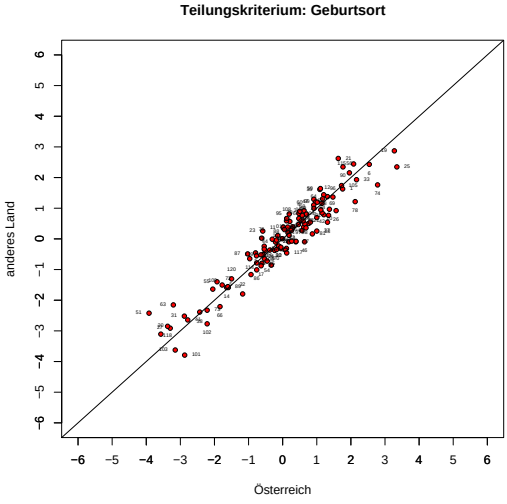
(b) Geschlecht



(c) Alter

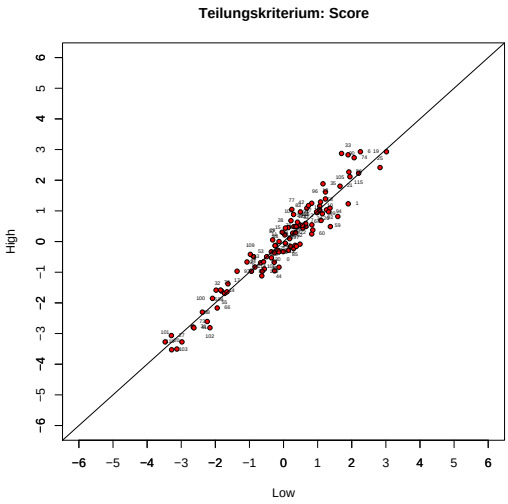


(d) Bildungsgrad

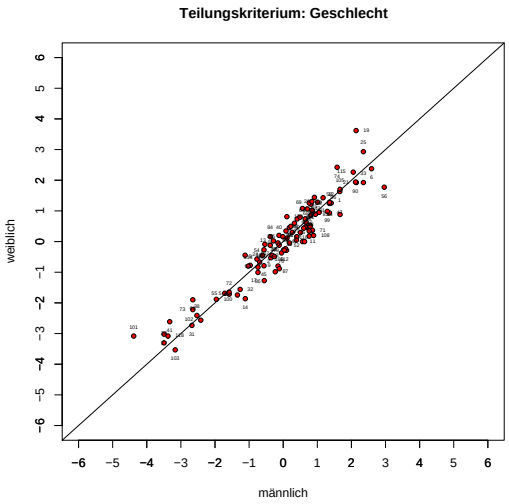


(e) Geburtsort

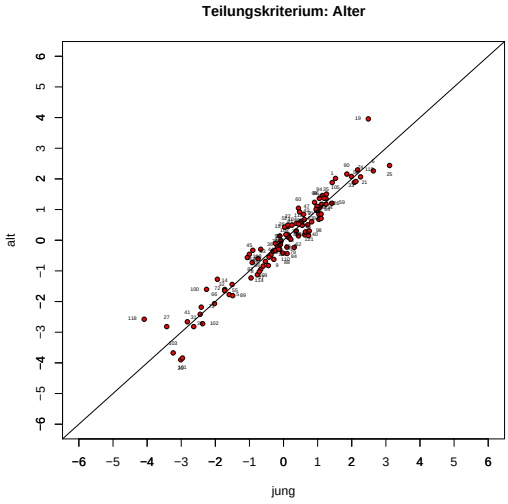
Abbildung 3.7.: Grafische Modelltests – kein Item ausgeschlossen



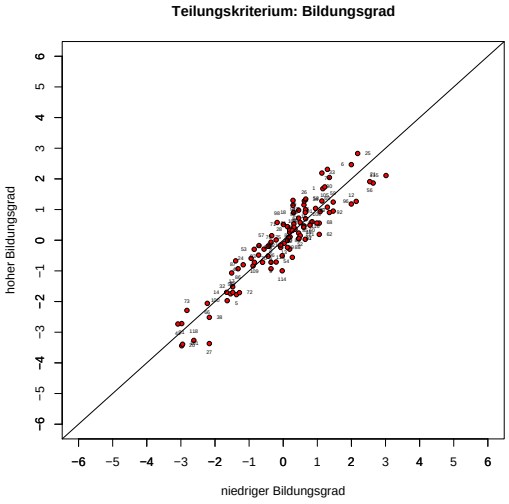
(a) Score



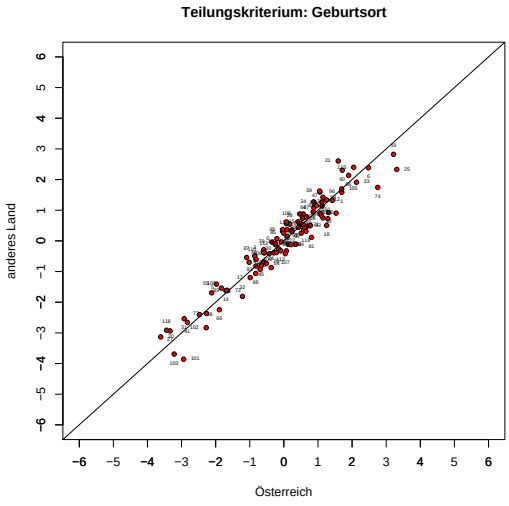
(b) Geschlecht



(c) Alter



(d) Bildungsgrad



(e) Geburtsort

Abbildung 3.8.: Grafische Modelltests – 18 Items ausgeschlossen

„0“ bedeutet die Gewichtung des Basisparameters mit null, was praktisch bedeutet, dass dieser Basisparameter keinen Einfluss auf das jeweilige Item hat. Betrachtet man Item 1 (zu finden in der ersten Zeile), so erkennt man aus Abbildung 3.4 leicht, dass sich dieses aus den Basisparametern „afa“, „ub“ und „ve“ zusammensetzt.

Es soll nun geprüft werden, ob die auf dem Konstruktionsrational aufbauenden Basisparameter des LLTM die Itemschwierigkeiten genauso gut bzw. nur unbeachtlich schlechter erklären wie das Rasch-Modell. Die geschätzten Parameter des Rasch-Modells dienen quasi als Referenz, was in diesem Zusammenhang bedeutet, dass ein nicht signifikanter Unterschied der vom LLTM geschätzten Parameter und der des Rasch-Modells hypothetisiert wird, und dies wiederum bedeutet, dass sich die Itemschwierigkeit additiv durch die, auf der Struktur der Items basierenden, Basisparameter genauso gut erklärt, wie wenn jedes Item einen eigenen Parameter laut Rasch-Modell hätte. Das Rasch-Modell stellt bei jenem Vergleich quasi ein „saturiertes Modell“ dar.

LLTM-1

Tabelle 3.4.: Auszug aus der Strukturmatrix des ersten LLTM

	afi	afb	afg	gg	afa	exg	ub	ak	pt	ve	nv	di	kf	mo	fd	pg	poe	ne
1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0
5	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
8	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0
9	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
11	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
12	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
13	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
14	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0

Es wurden, für die Erstellung der Strukturmatrix, alle in Kapitel 3.2 dargestellten Inhaltskomponenten so berücksichtigt, dass die Strukturmatrix vollen Rang hat (vgl. Fischer, 1974). Insgesamt wurden 18 Basisparameter definiert. Die Berechnungen wurden, wie auch schon in Kapitel 3.4.1, mit dem Software-Package eRm durchgeführt.

Der Likelihood-Quotienten-Test, wie in Tabelle 3.5 zu sehen, als auch der Grafische Modelltest (siehe Abbildung 3.9) zeigen deutlich keine Modellpassung des LLTM-1. Die kritische Grenze der chi^2-Verteilung wurde im Likelihood-Quotienten-Test weit über-

schritten, und der Grafische-Modelltest zeigt starke Abweichungen von der 45°- Geraden. Auch die Korrelation zwischen den von LLTM und Rasch-Modell errechneten Itemparametern ist mit 0,789 nur mäßig. Offensichtlich sind hier weitere, nicht im Modell berücksichtigte Faktoren beteiligt, die wesentlich zur Schwierigkeit der Items beitragen. Aus diesem Grund wurde eine weitere Strukturmatrix erzeugt, welche, ergänzend zu den inhaltlichen Parametern, auch strukturelle Eigenheiten der Items berücksichtigt.

Tabelle 3.5.: Likelihood-Quotienten Test: Rasch-Modell vs. LLTM-1

$logL_{lltm}$	$logL_{rm}$	$-2\log(\frac{L_{lltm}}{L_{rm}})$	$\chi^2_{\alpha_{1\%}}$	df	
-14421.51	-12970.38	2902.26	118.24	85	signifikant

LLTM-2

Bei der Konstruktion dieser Strukturmatrix wurden zu den Basisparametern des LLTM-1 noch weitere angenommen.

Polung der Items Zirka ein Drittel der Items musste aufgrund ihrer negativen Polung vor den eigentlichen Analysen umgepolt werden. Dass die inhaltliche Polung des Items einen Einfluss auf dessen Beantwortung hat, kann nicht ausgeschlossen werden.

Entscheidungsfrage Zirka 10% der Items beinhalten Aussagen, in denen eine Alternative zum dargestellten Verhalten genannt wird. Die Person macht also Angaben zur Häufigkeit des Verhaltens quasi in Relation zum anderen Verhalten. Auch hier kann ein Einfluss dieser strukturellen Unterschiede nicht ausgeschlossen werden.

Häufigkeit der Situation Bestimmte Situationen treten in einem durchschnittlichen Tagesablauf mehr oder weniger häufiger auf. Die Situation „Einkaufen gehen“ wird vermutlich in den meisten Fällen häufiger herbeigeführt als die Situation „zum Frisör gehen“. Die Items wurden in dieser Weise klassifiziert und zwar nach dem dreikategoriellen Schema: „sehr häufig“, „durchschnittlich häufig„ und „selten“. Da selten erlebte Situationen vermutlich schlechter erinnert werden bzw. im Extremfall von der Testperson noch keine fünf Male erlebt wurden, könnte dies einen wichtigen Einfluss auf die Reaktion der Testperson haben.

Mit diesen Erweiterungen beinhaltet die neue Strukturmatrix nun 22 Parameter die es zu schätzen gilt. Aufgrund des analogen Aufbaus zur Matrix wie sie in Tabelle 3.4 zu sehen ist, wird diese hier nicht dargestellt.[3]

Durch das Hinzufügen weiterer Parameter kommt es zwar zu einer Verbesserung gegenüber dem vorherigen Modell wie aus Tabelle 3.6 ersichtlich, dennoch bleiben die

[3]Die vollständigen Strukturmatrizen sind im Anhang abgebildet.

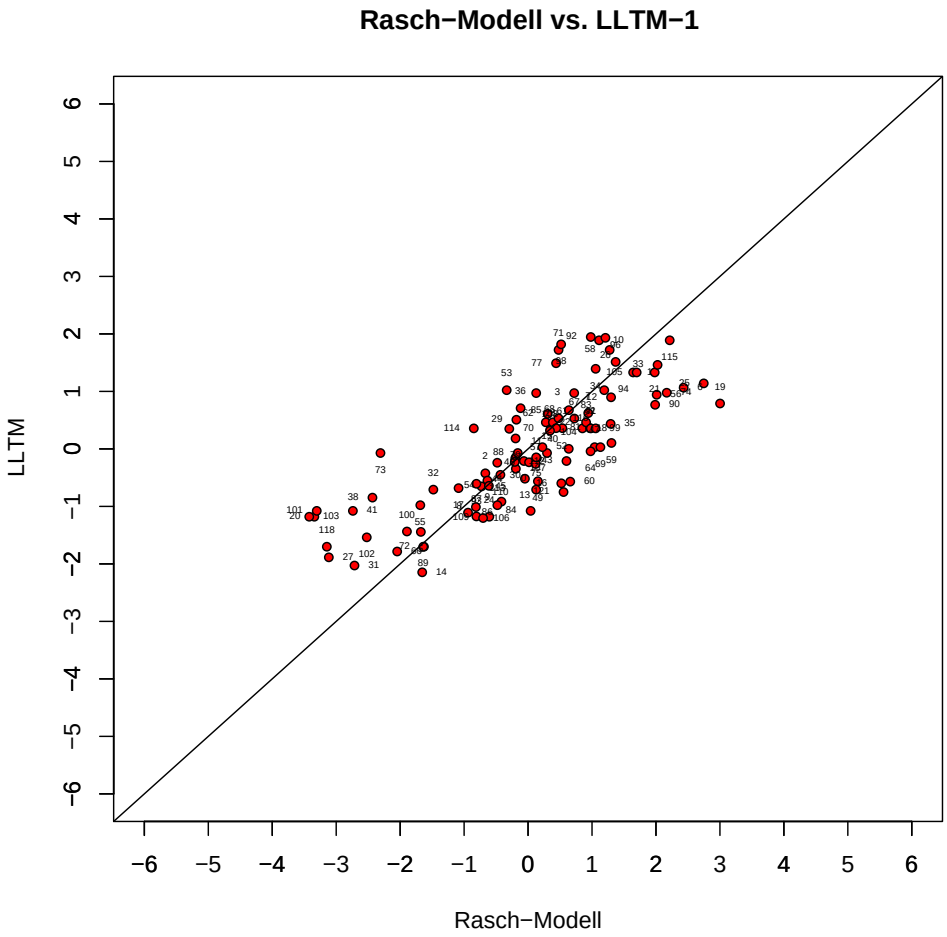


Abbildung 3.9.: Rasch-Modell vs. LLTM-1

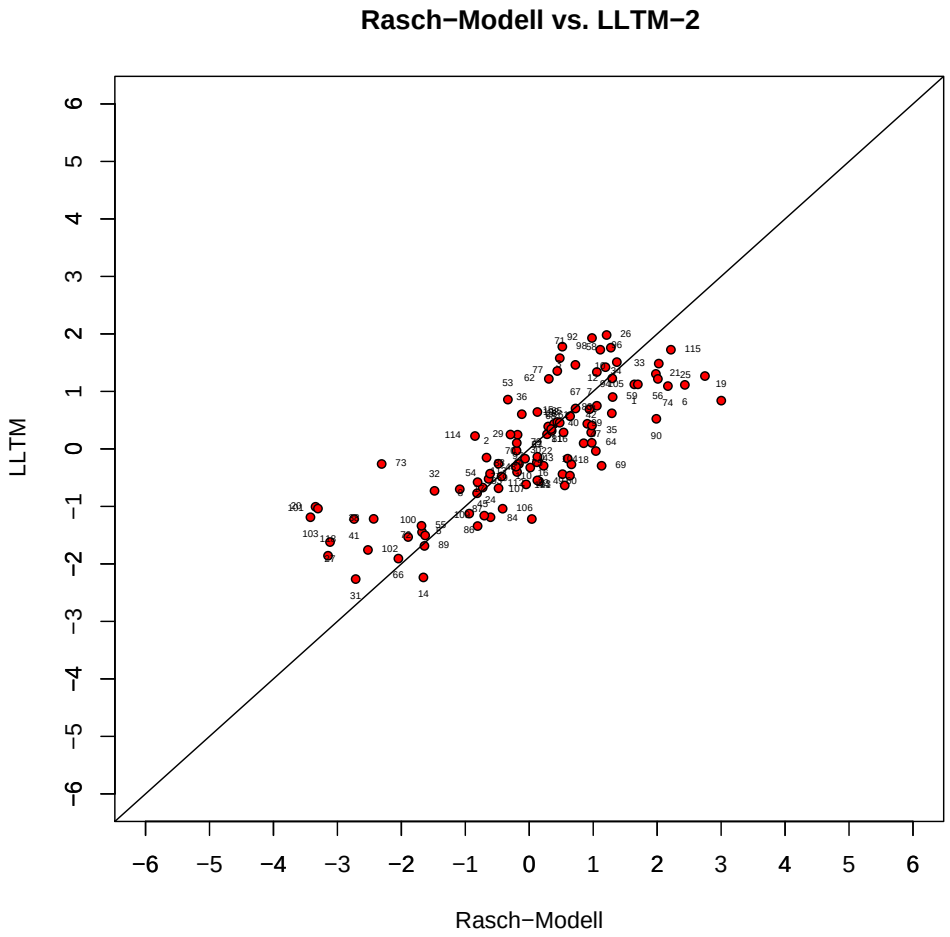


Abbildung 3.10.: Rasch-Modell vs. LLTM-2

Tabelle 3.6.: Likelihood-Quotienten Test: Rasch-Modell vs. LLTM-2

$logL_{lltm}$	$logL_{rm}$	$-2\log(\frac{L_{lltm}}{L_{rm}})$	$\chi^2_{\alpha_{1\%}}$	df	
-14314.16	-12970.38	2687.55	113.51	81	signifikant

Abweichungen zum Rasch-Modell sehr groß. Vor allem dürften die Randbereiche, also die besonders schwierigen bzw. die besonders leichten Items, vom LLTM systematisch unter- bzw. überschätzt werden. Im Rasch-Modell werden die Parameter quasi „extremer“ geschätzt als im LLTM. Im Grafischen Modelltest 3.10 ist das eindeutig erkennbar; der Punkteschwarm macht an seinen Enden jeweils einen deutlichen Knick.

In Tabelle 3.7 sind nun jene Items aufgelistet, deren beide Itemparameter (Rasch-Modell und LLTM) sich um mindestens 1,5 Einheiten unterscheiden.[4] Wie man sieht sind bei den Items 20, 41, 73, 101 und 103, die mittels des LLTM errechneten Parameter deutlich näher bei null als die durch das Rasch-Modell geschätzten Parameter. Betrachtet man diese Items nun inhaltlich, so stellt man fest, dass es sich hierbei um Items handelt, die eine sehr extreme Komponente beinhalten. Konkret sind es jeweils Situationen die nicht nur einen geringen Aufforderungscharakter aufweisen, sondern man könnte eine Person, die in dieser Weise agiert, auch als eine *extrem* Extravertierte mit hoher Impulsivität bei der Befriedigung ihrer Extraversionsbedürfnisse bezeichnen. Das verleiht diesen Items eine besondere Qualität, welcher in der folgenden Strukturmatrix Rechnung getragen wird.

LLTM-3

Die dritte Strukturmatrix basiert auf einem leicht modifizierten Modell gegenüber 3.2. Der Teil „Zweck der Situation“ des Konstruktionsrationals wurde um eine Sub-Dimension erweitert:

1. Die dargestellte Situation dient der Verrichtung einer Alltagstätigkeit.
2. Die dargestellte Situation dient der Interaktion mit anderen Menschen.
3. Die dargestellte Situation dient der Ausbildung oder der Ausübung des Berufs.
4. Die dargestellte Situation ist global formuliert, und lässt keine Rückschlüsse auf den eigentlichen Zweck zu.
5. NEU: *Die dargestellte Person benutzt eine unpassende Situation, um unmittelbar ihre Extraversionsbedürfnisse zu befriedigen.*

Um die Anzahl der verwendeten Basisparameter, im Sinne der Sparsamkeit des verwendeten Modells möglichst gering zu halten, wurde in der dritten Version der Strukturmatrix nur auf jenen neu eingeführten Basisparameter aus LLTM-1 zurückgegriffen

[4] Selbstverständlich sind beide Item-Parameter Schätzungen „Summe-null“ normiert.

Tabelle 3.7.: Items die die größten Abweichungen zwischen den Parameterschätzungen des LLTM und des Rasch-Modells aufwiesen.

	Nummer	Inhalt	β_{rm}	β_{lltm}
1	19	In ? von 5 Fällen zeige ich mir bekannten Personen wenn ich mich freue.	3.00	0.84
2	20	In ? von 5 Fällen beginne ich beim Einkaufen mit einer mir unbekannten Person ein Gespräch.	−3.34	−1.01
3	41	In ? von 5 Fällen komme ich während des Wartens in einer Arztpraxis mit einer mir unbekannten Person ins Gespräch.	−2.74	−1.22
4	73	In ? von 5 Fällen suche ich einen Ort auf an dem mir unbekannte Personen anwesend sind, anstatt alleine zuhause zu sitzen.	−2.31	−0.26
5	101	In ? von 5 Fällen komme ich beim Einkaufen mit einer mir unbekannten Person ins Gespräch.	−3.30	−1.04
6	103	In ? von 5 Fällen beginne ich während des Wartens in einer Arztpraxis mit einer mir unbekannten Person ein Gespräch.	−3.42	−1.19

(siehe: 3.4.2), dessen Wert sich am deutlichsten von null unterschied, und damit den stärksten Einfluss hatte. Dies war der Parameter *Entscheidungsfrage*. Die modifizierte Strukturmatrix beinhaltet nun in ihrer dritten Version 20 Basisparameter.

Tabelle 3.8.: Likelihood-Quotienten Test: Rasch-Modell vs. LLTM-3

$logL_{lltm}$	$logL_{rm}$	$-2\log(\frac{L_{lltm}}{L_{rm}})$	$\chi^2_{\alpha_{1\%}}$	df	
-14010.9	-12970.38	2081.03	115.88	83	signifikant

Im Likelihood-Quotienten-Test wird wieder ein signifikantes Ergebnis erzielt, was dafür spricht dass das angenommene Modell die Daten nicht ausreichend gut erklären kann. Man kann allerdings diese modifizierte Variante als die bisher Beste bezeichnen. Nicht nur die Korrelation zwischen den Schwierigkeitsparametern ist mit 0,86 nun höher, sondern im Grafischen Modelltest (Abbildung 3.11) zeigt der Punkteschwarm insbesondere in den Extrembereichen geringere Abweichungen von der 45° Geraden. Vergleicht man nun die drei Modelle mittels Informationsindizes miteinander, so wie in Tabelle 3.9 dargestellt, so sieht man die Überlegenheit des dritten, leicht modifizierten Modells deutlich. Sowohl der AIC als auch der BIC zeigen, dass das dritte Modell mit dem neu eingeführten Basisparameter die Daten im Vergleich zu den beiden Anderen am besten erklärt.

Trotzdem ist klar, dass keines der drei Modelle die Daten so gut erklärt wie das Rasch-Modell. Für die zukünftige Itemkonstruktion kann das Konstruktionsrational vielleicht

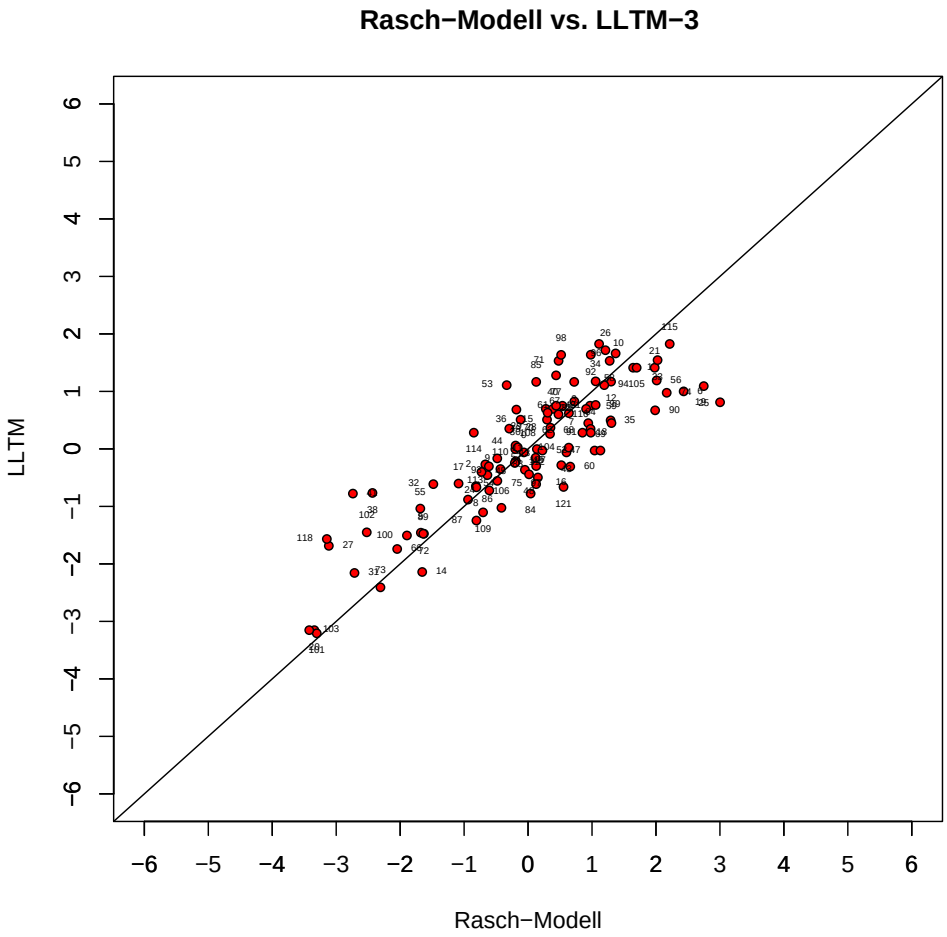


Abbildung 3.11.: Rasch-Modell vs. LLTM-3

Tabelle 3.9.: Vergleich zwischen den drei unterschiedlichen Modellen mittels Informationsindizes

	AIC	BIC
$LLTM_1$	28879.02	29069.91
$LLTM_2$	28672.31	28905.63
$LLTM_3$	28061.80	28273.90

eine grobe Abschätzung der zu erwarteten Itemschwierigkeit liefern, jedoch sicherlich nicht die Entscheidungsprozesse bei der Bearbeitung eines Items befriedigend genau abbilden.

4. Diskussion

Diese empirische Arbeit setzt sich mit der üblichen Vorgehensweise bei der Konstruktion von Fragebogenitems auseinander und stellt einen kleinen Schritt Richtung kontrollierter Itementwicklung in diesem Bereich dar. Die Konstruktion und Kategorisierung der Items fanden innerhalb strenger Grenzen statt und für die Kalibrierung und Validierung der Items wurden ebenfalls strengste Modellannahmen getroffen, und diese auch überprüft. Es konnte auch für einen reduzierten Itempool post-hoc Modellgeltung bezüglich des Rasch-Modells erreicht werden. Vom Gesamtitempool (insgesamt 122 Items zur Extraversion) mussten 18 Items (dies sind weniger als 15%) ausgeschlossen werden, um Modellgültigkeit zu erzielen. Dies liegt knapp über dem Anteil von 10% den Kubinger und Draxler (2007) als maximal noch verträgliche Menge an auszuscheidenden Items vorschlagen. Allerdings kann man den Anteil von ca. 15% als durchaus befriedigend erachten, hinsichtlich der Tatsache, dass es sich hier um einen Fragebogen handelt, und nicht um einen Intelligenz – oder Leistungstest.

Die anschließenden Versuche, die Itemkonstruktionsprinzipien mittels des LLTM adäquat abzubilden, scheiterten jedoch insofern, als das Rasch-Modell die Daten immer noch deutlich besser erklärte als jedes der geprüften LLTM-Modelle. Für die zukünftige Itemkonstruktion in der Dimension Extraversion, kann das Konstruktionsrational eine grobe Abschätzung der zu erwartenden Itemschwierigkeiten liefern, jedoch sicherlich nicht die Entscheidungsprozesse bei der Bearbeitung der Items befriedigend genau abbilden.

Offensichtlich werden gerade jene Items, die in den Extrembereichen des Kontinuums liegen, mangelhaft durch das Konstruktionsrational abgebildet. Die Vermutung liegt nahe, dass Items die extreme bzw. ungewöhnliche Verhaltensweisen darstellen eine besondere „Qualität“ aufweisen. Dieser Vermutung wurde Rechnung getragen durch die Berücksichtigung eines weiteren Basis-Parameters in der Strukturmatrix des LLTM-3 (siehe Kapitel 3.4.2). Dieser Basisparameter war all jenen Items gemein, welche eben jene „extreme“ Form von Extraversion beinhalteten. Das Ergebnis war eine merkliche Besserung bezüglich der Informationsindizes (siehe Kapitel 3.4.2), jedoch blieb der Unterschied zum Rasch-Modell immer noch signifikant.

Ein weiterer Ansatz wäre nun, unter anderem, spezielle Interaktionen zwischen den Basisparametern anzunehmen und diese in weiteren Modelldefinitionen Rechnung zu tragen. Eine Möglichkeit wäre bspw. die Itemcharakteristika (siehe Kapitel 3.1) nicht mehr einzeln zu modellieren, sondern das gemeinsame Auftreten der Ausprägungen bestimmter Charakteristika (z. B. die Ausprägungen von „Bekanntheitsgrad“ und „Aktivität“). Man würde damit den Kombinationen also quasi eine eigene Qualität zuschreiben und jede Kombination als neuen Basisparameter definieren. Problematisch wäre das insofern, als dadurch die Anzahl der Basisparameter steigen würde (bei der Kombination

von bspw. „Bekanntheitsgrad“ und „Aktivität“ würde die Anzahl von 7 Basisparameter auf 12 steigen.). Das Resultat wäre ein weniger sparsames Modell.

Jedenfalls wäre ein naives „Austesten“ von vielen unterschiedlichen Modellen besonders kritisch zu sehen, da bei mehrfacher Anwendung von inferenzstatistischen Tests auf denselben Datensatz bei konstantem α (bspw.: 1%) das fälschliche Verwerfen der Nullhypothese (Fehler 1.Art) sehr wahrscheinlich wird.

Ein kritischer Punkt der vorliegenden Arbeit ist sicherlich die Dichotomisierung der sechs Antwortkategorien. Dies war, wie in Kapitel 3.4 beschrieben, aufgrund der mangelhaften Besetzung vor allem der extremen Antwortkategorien nicht anders möglich. Problematisch ist dies insofern, als für einen Datensatz, für den Modellgültigkeit bezüglich des „Partial-Credit-Models“ angenommen werden kann, nach einer Dichotomisierung dies nicht automatisch auch hinsichtlich des dichotomen Rasch-Modells gilt, ja dies gar nicht gelten kann (Jansen & Roskam, 1986). Somit ist wohl das Hauptziel für zukünftige Untersuchungen in jedem Fall, genügend Testpersonen zu akquirieren, so dass die Kalibrierung mittels des „Partial-Credit-Model“ und eine anschließende Überprüfung des Konstruktionsrationals mittels des „Linear-Partial-Credit-Model“ (LPCM) möglich ist.

Ein weiteres zukünftiges Ziel muss auch sein, die erreichte (vorläufige) a-posteriori Geltung des Rasch-Modells für den Itempool auch an einem zweiten, unabhängigen Itempool zu „kreuzvalidieren“ wie das Kubinger (2005) gefordert hat. Kubinger und Draxler (2007) schlagen ebenfalls eine Überprüfung des Itempools mittels des Martin-Löf-Tests (vgl. Kubinger, 1989) vor, um die Itemhomogenität zu prüfen, was zum Zeitpunkt der Untersuchung leider nicht möglich war, da kein Softwarepaket diesen für ein, wie hier vorgegebenes, unvollständiges Design anbietet.

Im Sinne einer Modellüberprüfung bzw. einer genaueren Inspektion des Antwortverhaltens der Personen wäre auch eine Anwendung des Mixed-Rasch-Modells auf den Itempool wünschenswert, wie das Rost et al. (1999) durchführten. Leider fehlt auch hier eine Software-Lösung für die Anwendung des Mixed-Rasch-Modells auf unvollständige Designs.

Kritisch zu betrachten ist auch die Erhebung der Stichprobe über das Medium Internet, da zwar in kurzer Zeit und ohne großen Aufwand viele Personen erreicht werden können, es sich hierbei aber um eine „anfallende“ Stichprobe handelt (Rasch & Kubinger, 2006). Dies bedeutet, dass die Ergebnisse nicht ohne weitere Überprüfungen auf die gesamte Bevölkerung zu generalisieren sind. Auch problematisch bezüglich der Erhebung via Internet ist anzumerken, dass hier die Testsituation vom Testleiter nicht kontrolliert werden kann, d.h. die Kontrolle von etwaigen Störvariablen nicht möglich ist.

Ein weiterer zu beachtender Punkt die Stichprobe betreffend, ist deren Zusammensetzung. Es sind ausschließlich Personen, die den Fragebogen freiwillig, anonym und ohne persönliche Konsequenzen bearbeitet haben. Interessant wäre daher vor allem die Erprobung des Fragebogens an einer Stichprobe, bei der das Ergebnis des Fragebogens an persönliche Konsequenzen geknüpft ist, da bei jenen Testpersonen eher mit einer Verfälschung (siehe Kapitel 2.3.3) in Richtung soziale Erwünschtheit zu rechnen ist. Dieser Effekt Richtung „soziale Erwünschtheit“, könnte bei einer derartigen Stichprobe mittels Person-fit Indizes nachzuweisen versucht werden.

Auch die Weiterentwicklungen des Itempools durch die Konstruktion neuer Items,

sowie die Entwicklung weiterer „Konstruktionsrationale“ für neue Persönlichkeitsdimensionen ist geplant, um diesen Ansatz, Items zu konstruieren und zu validieren, auf weitere Dimensionen auszudehnen.

5. Zusammenfassung

Das Ziel der vorliegenden Arbeit war, ein Konstruktionsrational zu erstellen, auf dessen Basis ein Fragebogen zur Persönlichkeitsdimension „Extraversion“ konstruiert werden sollte, um dieses dann anhand der erhobenen Daten empirisch zu überprüfen.

Nach der Wahl eines geeigneten „Itemtyps“ (siehe Kapitel 2.3.1 und 3.2) wurde ein Konstruktionsrational erstellt, das die mögliche Variation des Iteminhalts abbildet. Diese Variationen sollen die Itemschwierigkeit beeinflussen bzw. im Idealfall vollständig erklären.

Der Fragebogen sollte alle notwendigen Voraussetzungen erfüllen, um in weiterer Folge auch in einer adaptiven Form anwendbar zu sein. Die Voraussetzungen hierfür sind einerseits ein großer Itempool (zumindest 60 – 70 Items), genügend viele Items auch in den extremen Bereichen (hohe bzw. niedrige Itemschwierigkeit), um auch hier eine hohe Messgenauigkeit zu erzielen und die Geltung des Rasch-Modells für den Itempool.

Bei der Prüfung des Itempools hinsichtlich Rasch-Modell Konformität mussten – aufgrund mangelnder Modellpassung – 18 der insgesamt 122 Items ausgeschieden werden. Durch das Entfernen dieser Items konnte eine a-posteriori Modellpassung erzielt werden. Dadurch dass nur 18 Items ausgeschieden werden mussten, ist der Itempool immer noch groß genug für eine potentielle adaptive Vorgabe.

Im nächsten Schritt wurde der Rasch-Modell-konforme Itempool mittels des LLTM analysiert. Den Strukturmatrizen, die das LLTM zur Schätzung der Basisparameter benötigt, wurde das Konstruktionsrational zugrundegelegt. So war eine empirische Überprüfung möglich, inwiefern die im Konstruktionsrational festgelegten Strukturmerkmale der Items die Schwierigkeit derselben erklären.

Diese Überprüfung der Itemstruktur wurde mit drei unterschiedlichen Strukturmatrizen durchgeführt. Die erste Überprüfung verlief unbefriedigend, da einerseits der Likelihood-Ratio-Test eine signifikante Abweichung zum Rasch-Modell ergab und andererseits auch im Grafischen Modelltest deutliche Abweichungen zu beobachten waren.

Die zweite Analyse nach dem LLTM mit einer Strukturmatrix, die nun auch drei weitere Merkmale, nämlich „Polung“, „Entscheidungsfrage“ und „Häufigkeit der Situation“ berücksichtigte, zeigte eine leichte Verbesserung aber immer noch starke Abweichungen im Grafischen Modelltest und einen signifikanten Likelihood-Ratio-Test.

Bei der dritten Analyse wurden jene Items inhaltlich inspiziert, die besonders hohe Abweichung zwischen den beiden Parameterschätungen (Rasch-Modell vs. LLTM) aufwiesen. Zur Ursache der Abweichungen wurde eine Hypothese generiert, auf deren Basis ein neues Strukturmerkmal eingeführt wurde. Diese Version zeigt zwar eine deutlich bessere Anpassung als die beiden ersten Modell wie den Informationsindizes zu entnehmen ist, jedoch erklärt dieses Modell immer noch die Daten signifikant schlechter als das Rasch-Modell.

Obwohl die Analysen nach dem LLTM nicht bestätigen, dass die Itemschwierigkeiten sich ausschließlich auf die postulierte Struktur im Konstruktionsrational zurückführen lassen, lässt die moderate bis hohe Korrelation zwischen den Parameterschätzungen (Rasch-Modell und LLTM) immerhin den Schluss zu, dass das Konstruktionsrational doch einen gewichtigen Beitrag zur Erklärung der Item-Schwierigkeit leistet.

Literaturverzeichnis

Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6. Aufl.). Stuttgart: Kohlhammer.

Angleitner, A., John, O. P. & Löhr, F.-J. (1986). It´s What You Ask and How You You Ask It: An Itemmetric Analysis of Personality Questionnairs. In A. Angleitner & J. Wiggins (Hrsg.), *Personality Assessment via Questionnairs* (S. 61–108). Berlin: Springer.

Asendorpf, J. B. (2007). *Psychologie der Persönlichkeit* (4. Aufl.). Heidelberg: Springer.

Becker, P. (2002). Das Trierer Integrierte Persönlichkeitsinventar. *Diagnostica*, *48*(2), 68–79.

Becker, P. (2003a). Persönlichkeitsdimensionen. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 325–331). Weinheim: Beltz.

Becker, P. (2003b). Persönlichkeitsfragebogen. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 332–337). Weinheim: Beltz.

Becker, P. (2003c). *Trierer Integriertes Persönlichkeitsinventar (TIPI).* Göttingen: Hogrefe.

Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation.* Berlin: Springer.

Briggs, K. C. & Myers, I. B. (1991). *Myers-Briggs Typenindikator (MBTI).* Weinheim: Beltz-Test.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson.

Buss, D. M. & Craik, K. H. (1983a, April). The Act Frequency Approach to Personality. *Psychological Review*, *90*(2), 105–126.

Buss, D. M. & Craik, K. H. (1983b). Act Prediction and the Conceptual Analysis of Personality Scales: Indices of Act Density, Bipolarity, and Extensity. *Journal of Personality and Social Psychology*, *45*(5), 1081–1095.

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists* (1. Aufl.). New Jersey: Lawrence Erlbaum Associates.

Eysenck, H., Wilson, G. & Jackson, C. (1998). *Eysenck Personality Profiler (EPP–D).* Frankfurt: Swets–Test Services.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests.* Bern: Huber.

Fischer, G. H. & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied psychological measurement*, *6*(4), 397–416.

Hohensinn, C., Kubinger, K. D., Reif, M., Holocher-Ertl, S., Khorramdel, L. & Frebort, M. (2008). Examining Item-Position Effects in Large-Scale Assessment using linear logistic test model. *Psychology Science Quarterly*, *50*.

Holocher-Ertl, S., Kubinger, K. D. & Menghin, S. (2003). Big Five Plus One Per-

sönlichkeitsinventar (B5PO) [Software und Manual]. Mödling: Dr. G. Schuhfried GmbH.

Hornke, L. F. & Rettig, K. (1989). Regelgeleitete Itemkonstruktion unter Zuhilfenahme kognitionsypsychologischer Überlegungen. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 140–162). Weinheim: Beltz.

Irle, M. & Allehoff, W. (1984). *Berufs-Interessen-Test II (BIT II).* Göttingen: Hogrefe.

Jankisz, E. & Moosbrugger, H. (2007). Planung und Entwicklung von psychologischen Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 27–73). Heidelberg: Springer.

Jansen, P. G. & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika*, *51*(1), 69–91.

Khorramdel, L. & Kubinger, K. D. (2006). The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure. *Psychology Science*, *48*(3), 378–397.

Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 19–83). Weinheim: Beltz.

Kubinger, K. D. (2000). Replik auf Jürgen Rost "Was ist aus dem Rasch-Modellgeworden?":Und für die Psychologische Diagnostik hat es doch revolutionäre Bedeutung. *Psychologische Rundschau*, *51*, 1–7. Available from `http://www.univie.ac.at/Psychologie/diagnostik/files/Rostrep.pdf` (Langversion)

Kubinger, K. D. (2003a). Adaptives Testen. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 1–9). Weinheim: Beltz.

Kubinger, K. D. (2003b). Testtheorie, Probabilistische. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415–423). Weinheim: Beltz.

Kubinger, K. D. (2005). Psychological Test Calibration Using the Rasch Model–Some Critical Suggestions on Traditional Approaches. *International Journal of Testing*, *5*(4), 377–394.

Kubinger, K. D. (2006). *Psychologische Diagnostik.* Göttingen: Hogrefe.

Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly*, *50*(3), 311–327.

Kubinger, K. D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, *53*(3), 131-143.

Mair, P. & Hatzinger, R. (2007a). erm: Extended rasch modeling. [Computer software manual]. (R package version 0.9-5)

Mair, P. & Hatzinger, R. (2007b, 2 22). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, *20*(9), 1–20. Available from `http://www.jstatsoft.org/v20/i09`

Mischel, W., Shoda, Y. & Ayduk, O. (2008). *Introduction to personality* (8. Aufl.). New Jersey: Wiley.

Moosbrugger, H. & Hartig, J. (2002). Factor analysis in personality research: Some arte-

facts an their consequences for psychological assessment. *Psychologische Beiträge*, *44*, 136–158.

Moosbrugger, H. & Hartig, J. (2003). Faktorenanalyse. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 137–145). Weinheim: Beltz.

Müller, H. (1999). *Probabilistische Testmodelle für diskrete und kontinuierliche Ratingskalen* (1. Aufl.). Bern: Huber.

Ortner, T. M. (2005). *Möglichkeiten und Grenzen adaptiver Persönlichkeitsfragebogen.* Lengerich: Pabst.

Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae (NEO–PI–R).* Göttingen: Hogrefe.

Podsakoff, P. M. & MacKenzie, S. B. (2003). Common Method Biases in Behavioral Research: A Vritical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *88*, 879–903.

Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling).* Lengerich: Pabst.

R Development Core Team. (2007). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Rasch, D. & Kubinger, K. D. (2006). *Statistik für das Psychologiestudium* (1. Aufl.). München: Elsevier.

Rost, J. (2002). When personality questionnaires fail to be unidimensional. *Psychologische Beiträge*, *44*, 108–125.

Rost, J. (2004). *Testtheorie–Testkonstruktion.* Bern: Hans Huber.

Rost, J., Carstensen, C. H. & Davier, M. von. (1999). Sind die Big Five Rasch-skalierbar? Eine Reanalyse der NEO-FFI-Normierungsdaten. *Diagnostica*, *45*(3), 119-127.

Schaarschmidt, U. & Fischer, A. W. (2003). *Arbeitsbezogenes Verhaltens- und Erlebensmuster (AVEM)* (2. Aufl.). Frankfurt/M: Swets Test Services.

Seiwald, B. B. (2003a). Antwortformat. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 23–28). Weinheim: Beltz.

Seiwald, B. B. (2003b). Antworttendenzen (response set). In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 29–32). Weinheim: Beltz.

Seiwald, B. B. (2003c). Lügenskala. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 271–273). Weinheim: Beltz.

Sonnleitner, P. (2007). *Versuch eines Konstruktionsrationals zum Leseverständnis.* Unveröffentlichte Diplomarbeit, Universität Wien – Fakultät für Psychologie, Wien.

Westhoff, K. & Kluck, M.-L. (2003). *Psychologische Gutachten* (4. Aufl.). Berlin: Springer.

Wiggins, J. S. (1979). A psychological taxonomy of trait descriptive terms: I. The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395–412.

Tabellenverzeichnis

Abbildungsverzeichnis

A. Anhang

ai[1]	ab	ag	gg	aa	ex	ub	ak	pt	ve	nv	di	kf	mo	fd	pg	pe	ne
0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0
0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1
0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0

[1] . . . von links nach rechts: Zweck:Interaktion, Zweck:Beruf, Zweck:Gefühlsoffenbarung, Zweck:Ganzglobal, Zweck:Alltagstätigkeit, Aufforderungscharakter:Extraversionsgeladen, Bekanntheitsgrad:unbekannt, Aktivität:aktiv, Aktivität:passiv2, Aktivität:vermeidend, Kommunikation:nonverbal, Stil:Diskussion, Stil:Konfrontation, Stil:Monolog, Stil:Dominanz, Stil:persönliches Gespräch, Emotionen: positiv, Emotionen:negativ

0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0
1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0
0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0
1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1
1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
0	1	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1
0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0
0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0

Tabelle A.1.: Strukturmatrix 1

ai[2]	ab	ag	gg	aa	ex	ub	ak	pt	ve	nv	di	kf	mo	fd	pg	pe	ne	se	ha	ef	ng
0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1
0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1
0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	1
0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0
0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1
0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1

[2] . . . von links nach rechts: Zweck:Interaktion, Zweck:Beruf, Zweck:Gefühlsoffenbarung, Zweck:Ganzglobal, Zweck:Alltagstätigkeit, Aufforderungscharakter:Extraversionsgeladen, Bekanntheitsgrad:unbekannt, Aktivität:aktiv, Aktivität:passiv2, Aktivität:vermeidend, Kommunikation:nonverbal, Stil:Diskussion, Stil:Konfrontation, Stil:Monolog, Stil:Dominanz, Stil:persönliches Gespräch, Emotionen: positiv, Emotionen:negativ, Häufigkeit:selten, Häufigkeit:häufig, Entscheidungsfrage:ja, Polung:negativ

1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0
0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0
1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1
0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0
0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1
0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1
0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0	1	1
1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1
0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0

0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0
0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	1
1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1	0	0	1
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1
1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1
0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0

0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0
1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0
0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0
0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0

Tabelle A.2.: Strukturmatrix 2

ai[3]	ab	ag	gg	aa	us	ex	ub	ak	pt	ve	nv	di	kf	mo	fd	pg	pe	ne	ef
0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1
0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1
1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0

[3] . . . von links nach rechts: Zweck:Interaktion, Zweck:Beruf, Zweck:Gefühlsoffenbarung, Zweck:Ganzglobal, Zweck:Alltagstätigkeit, Zweck:unmittelbare Befriedigung von Extraversionsbedürfnisse ,Aufforderungscharakter:Extraversionsgeladen, Bekanntheitsgrad:unbekannt, Aktivität:aktiv, Aktivität:passiv2, Aktivität:vermeidend, Kommunikation:nonverbal, Stil:Diskussion, Stil:Konfrontation, Stil:Monolog, Stil:Dominanz, Stil:persönliches Gespräch, Emotionen: positiv, Emotionen:negativ, Entscheidungsfrage:ja

1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0
1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0
0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1
1	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1		
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0			
0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0					
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0			
1	0	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	
0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0			
0	1	1	0	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1	0	1	1	0	0	0	0	1	1	0	
0	1	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	1		
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0			
0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0		

0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	0

Tabelle A.3.: Strukturmatrix 3

Curriculum Vitae

Ausbildung

1986 – 1990	Volksschule Rothenburggasse 1, 1120 Wien
1990 – 1994	AHS Rosasgasse 1-3, 1120 Wien
1994 – 1995	HTL Leberstrasse 4c, 1030 Wien
1995 – 1999	AHS Anton-Baumgartnerstrasse 123, 1230 Wien
1999 – 2000	Ableistung des Zivildienstes im AKH, 1090 Wien
seit 2000	Diplomstudium Psychologie, Universität Wien
seit 2007	Bakkalaureatsstudium Statistik, Universität Wien

Berufliche Tätigkeiten

03.2004 – 09.2004	6-Wochen Praktikum am Arbeitsbereich Psychologische Diagnostik
10.2004 – 01.2008	Studienassistent am Arbeitsbereich Psychologische Diagnostik
seit 03.2006	Projektmitarbeiter an der „Test- und Beratungsstelle“ mit dem Schwerpunkt der Aufbereitung und statistischen Analyse der Daten zu den „Bildungsstandards“

Besondere Kenntnisse

Routinemäßiger Umgang mit einschlägiger Statistiksoftware: R, eRm, lpcm-win, Acer-Conquest, SPSS, CADEMO sowie dem Textsatzprogramm LaTeX

Wien, 20. November 2008