universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Eine testtheoretische Analyse des Entwicklungsscreenings für den Einsatz im Hort – Altersgruppe 6 – 10-Jährige (ESH 6-10)

verfasst von

Nina Hasenöhrl

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat)

Wien, 2016

Studienkennzahl lt. Studienblatt:	A 298
Studienrichtung lt. Studienblatt:	Psychologie
Betreut von:	Ass.-Prof. Dr. Ursula Kastner-Koller

Danksagungen

Mein Dank gilt all denjenigen, die mich auf meinem Weg begleitet und unterstützt haben.

Besonders bedanken möchte ich mich bei meinen Betreuerinnen Ass.-Prof. Dr. Ursula Kastner-Koller und Ass.-Prof. Dr. Pia Deimann, zuallererst für die Möglichkeit, bei ihnen meine Diplomarbeit schreiben zu dürfen, und für das Vertrauen in meine Arbeit. Ihre unkomplizierte und unmittelbare Hilfsbereitschaft war eine beruhigende und ausgesprochen wertvolle Stütze für mich – herzlichen Dank!

Ebenfalls bedanken möchte ich mich an dieser Stelle bei meiner Familie für den Rückhalt.

Meinem Partner Matthias Gamisch, der mir immer verlässlich, geduldig und zuversichtlich den Rücken gestärkt hat, will ich schließlich meinen ganz speziellen Dank aussprechen: Ohne Dich wäre ich nicht, wo ich bin.

Zusammenfassung

Das in einem fortgeschrittenen Stadium der Entstehung vorliegende Entwicklungsscreening für Hortkinder im Alter von 6 bis 10 Jahren (ESH 6-10) wurde nach seiner ersten Vorgabe in Horten der kooperierenden Organisation *Kinder in Wien* und seiner darauffolgenden Überarbeitung ein zweites Mal zur Anwendung gebracht. Die Erfassung und Analyse dieser zweiten Stichprobe sowie weiterführende Testanalysen über alle bisher erhobenen Daten insgesamt waren Gegenstand der vorliegenden Arbeit. Es fanden sich für die 14 postulierten Skalen zufriedenstellende Reliabilitätswerte von α = .708 bis .906, wobei zur Optimierung des Verfahrens vereinzelt geringfügige Veränderungen bei der Itemselektion vorgenommen und mögliche Erweiterungen vorgeschlagen wurden. Eine zur Prüfung der Validität durchgeführte explorative Faktorenanalyse belegt die dem Verfahren zugrundegelegten theoretischen Annahmen über die Abgrenzbarkeit der entwicklungspsychologisch bedeutsamen Bereiche der Sprache und der Arbeitshaltungen. Die Bereiche Persönlichkeit und Soziale Interaktion sind den ermittelten Befunden zufolge dagegen nicht überzeugend voneinander zu trennen; diesbezüglich werden die Persönlichkeitsdimensionen Verträglichkeit und Extraversion als mögliche relevante Faktoren andiskutiert.

Abstract

Upon its first application and subsequent revision, emerging development screening procedure for children aged between six and ten "ESH 6-10" was implemented at after-school care clubs of cooperating organization *Kinder in Wien* for the second time. Entry and analysis of the newly collected data were subject of this thesis as well as further analyses of the merged whole of all data obtained so far. Thus satisfying reliability values between α = .708 and .906 were found concerning the 14 postulated scales, with slight modifications of the item selection so as to optimize the questionnaire. In addition, potential enhancements were suggested. To examine the test procedure's validity, an explorative factor analysis was conducted. The results support the underlying theory regarding the identifiability of both language and attitude to work in terms of essential topics in developmental psychology, whereas personality and social development appear not to be convincingly distinguishable. Referring to this, the possibly relevant personality dimensions agreeableness and extraversion are touched upon.

Inhaltsverzeichnis

Tabellenverzeichnis

I. Theoretischer Teil

1. Einleitung

Die vorliegende Arbeit befasst sich mit der Weiterentwicklung und Analyse des *Entwicklungsscreenings für Hortkinder im Alter von 6 bis 10 Jahren* (infolge genannt *ESH 6-10*), das zu Beginn dieses Projektabschnittes der Verfasserin in seiner mittlerweile zweiten Version zur Verfügung steht. Ziel der Verfahrensentwicklung des ESH 6-10 war und ist es, pädagogisch relevanten Personen ein Werkzeug zur Seite zu stellen, das verlässlich und effizient diejenigen Kinder ihres Betreuungsbereiches herausfiltern kann, deren Entwicklung einer genaueren Betrachtung bedarf. Das möglichst frühe Aufspüren etwaiger Entwicklungsprobleme gilt als entscheidend, bei Schwierigkeiten oder Defiziten wirksam intervenieren zu können; angesichts meist vorherrschender ungünstiger Betreuungsverhältnisse und der damit einhergehenden zeitlichen Einschränkungen erweist sich diese Aufgabe allerdings nicht selten als (zu) große Herausforderung. – Diesem skizzierten Bedarf entsprechend bot sich zu Beginn des Gesamtprojektes ESH 6-10 ein Screening-Verfahren als Methode der Wahl logisch an. Inhaltlich sollten auf Basis theoretischer Überlegungen die Entwicklungsbereiche Arbeitshaltungen, Persönlichkeit, Motivation, Sprache und Soziale Interaktion Berücksichtigung finden.

In einem ersten Entwicklungsschritt wurden also im Rahmen von Diplomarbeiten an der Universität Wien unter der Leitung von Frau Prof. Dr. Pia Deimann und Frau Prof. Dr. Ursula Kastner-Koller von Kunst (2014), Matschiner (2015) und Neugschwentner (2014) entsprechende Items entwickelt und selbige einem Expertinnenurteil unterzogen. Im Anschluss daran wurde das neue Verfahren ESH 6-10 von HortbetreuerInnen der kooperierenden Organisation „Kinder in Wien (KiWi)“ an einer ersten Stichprobe angewendet. Die Daten und Auswertungen dieser ersten Erhebung dienten Hasenhindl (in Vorbereitung) und Kremser (in Vorbereitung) als Grundlage für eine Überarbeitung des Itempools, woraufhin das ESH 6-10 in nun also adaptierter Form ein zweites Mal zur Anwendung gelangen konnte.

Im Rahmen der vorliegenden Arbeit sollen nun die Daten der zweiten Erhebung erfasst und ausgewertet werden. Zu den aus entwicklungspsychologischer Sicht zugrundeliegenden theoretischen Ansätzen sei auf die genannten Vorarbeiten zum Thema verwiesen; relevant ist an dieser Stelle ein testtheoretischer Zugang, und hier besonders die (Entwicklung der) Test- und Skalenkennwerte. Darüber hinaus von Interesse ist die Frage, ob sich die ursprünglich angenommenen Entwicklungsbereiche Arbeitshaltungen, Persönlichkeit, Motivation, Sprache und Soziale Interaktion auf Basis der Daten mittels

einer Faktorenanalyse als valide Faktoren belegen lassen und so das theoretische Fundament gestärkt werden kann.

2. Grundlagen der Testanalyse

Um die Qualität eines psychometrischen Verfahrens einheitlich beurteilen zu können, wurden in der Psychologischen Diagnostik Testgütekriterien etabliert. „Testgütekriterien und Itemkennwerte [...] sind von entscheidender Bedeutung für die Neukonstruktion und Veränderung eigener Tests“ (Bortz & Döring, 2002, S. 192). Als zentrale Kriterien gelten *Objektivität*, *Reliabilität* und *Validität*, die allesamt grundsätzlich in möglichst hohem Grade vorhanden sein sollen. Sie stellen nach Lienert & Raatz (1998, S. 7) die Hauptgütekriterien dar; als Nebengütekriterien werden demnach Normierung, Vergleichbarkeit, Ökonomie und Nützlichkeit betrachtet. Darüber hinaus werden von Kubinger (2006, S. 33) auf Basis der vom Testkuratorium der Föderation Deutscher Psychologenvereinigungen festgelegten Gütekriterien die Kriterien Zumutbarkeit, (Un-) Verfälschbarkeit sowie Fairness berücksichtigt. Rost (2004, S. 356) legt vor allem auf die *klassische Trias* von Testgütekriterien Augenmerk: „Die Qualität eines Tests wird traditionellerweise an den klassischen Gütekriterien der Objektivität, Reliabilität und Validität [...] festgemacht“.

2.1 Hauptgütekriterien

Weiters betont Rost (2004, S. 356), dass zwischen diesen drei Hauptgütekriterien logische Beziehungen bestehen; Demzufolge „ist die Objektivität eine logische Voraussetzung für die Reliabilität und diese wiederum ist logische Voraussetzung für die Validität“. Bühner (2011, S. 71) formuliert in diesem Zusammenhang beispielsweise: „Ein Test, der nicht objektiv ist, kann mit großer Wahrscheinlichkeit keine optimale Reliabilität erreichen“. Rost (2004, S. 392ff.) geht in seinen Ausführungen besonders auf die Verknüpfung zwischen Reliabilität und Validität ein.

Vor der nun folgenden Begriffsklärung zeichnet sich also bereits ab, dass Testgütekriterien keine „Check-List“ abzuarbeitender und einfach zu erfüllender Einzelforderungen darstellen. Vielmehr müssen im Rahmen einer Testkonstruktion immer mehrere Aspekte abgewogen werden – zum Teil auch gegeneinander.

2.1.1 Objektivität

„Unter ***Objektivität*** eines Tests verstehen wir den Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind.“ (Lienert & Raatz, 1998, S. 7).

Unterschieden werden diesbezüglich die *Durchführungsobjektivität* (Unabhängigkeit der Testergebnisse von Verhaltensvariationen des Testleiters während der Durchführung des Tests), die *Auswertungsobjektivität* bzw. *Verrechnungssicherheit* (Unabhängigkeit der Testergebnisse von der Person des Testleiters durch genaue Auswertungsvorschriften bzw. Reglementierungen) sowie die *Interpretationsobjektivität* bzw. *Interpretationseindeutigkeit* (Unabhängigkeit der Interpretation der Testergebnisse von der Person des Beurteilers).

2.1.2 Reliabilität

„Unter ***Reliabilität*** versteht man den Grad der Genauigkeit, mit dem ein Test ein bestimmtes Merkmal misst, unabhängig davon, was er zu messen beansprucht.“ (Bühner, 2011, S. 60).

Die Reliabilität „wird als Anteil der ‘wahren Varianz’ an der Varianz der beobachteten Testwerte bestimmt“ (Moosbrugger & Hartig, 2003b, S. 410), wobei sich der entsprechende Wertebereich zwischen null (es liegt keine Messung im eigentlichen Sinn vor, es werden lediglich Zufallszahlen realisiert) und eins (Test misst fehlerfrei) bewegt.

Der angesprochene Grad der Genauigkeit kann nun über unterschiedliche Ansätze ermittelt werden, was zur Folge hat, dass für ein und denselben Test mehrere – auch voneinander abweichende – Reliabilitätskoeffizienten vorliegen können. Den Ansätzen gemein ist, dass „sie Schätzwerte für den Anteil liefern, in dem ein einzelner Testwert fehlerbehaftet ist oder sein kann“ (Lienert & Raatz, 1998, S. 10); ermittelte Unterschiede wiederum begründen sich demnach aus unterschiedlichen Berücksichtigungen spezifischer Messfehlerarten und -anteile.

> *Exkurs*
>
> Grundsätzlich werden im Rahmen der *Klassischen Testtheorie* nur unsystematische Messfehler betrachtet. Der Umstand, dass in der Praxis allerdings auch systematische Messfehler auftreten, wird dabei nicht berücksichtigt. Dies bewirkt allerdings eine verminderte Präzision bei der Schätzung der Reliabilität – neben der Stichprobenabhängigkeit der Testwerte der Klassischen Testtheorie und anderen Aspekten ein Kritikpunkt an der Klassischen Testtheorie selbst. Obwohl die *Probabilistische Testtheorie* an dieser Stelle aufzeigen und alternative Ansätze zur Testkonstruktion

bieten kann, erfolgt eine Mehrheit der Testentwicklungen nach wie vor nach der Klassischen Testtheorie. Rost (1999, S. 141) spricht dabei von einem Anteil von über 95% und berichtet überdies, dass bei vergleichenden Analysen nach beiden Ansätzen die Testergebnisse häufig sehr gut übereinstimmen. Auch in Quellen jüngeren Datums wird bestätigt, dass sich die Klassische Testtheorie bei der Testentwicklung in der Praxis oft bewährt (z.B. Bühner, 2011, S. 54). Aus diesem Grund können in der praktischen Anwendung üblicherweise die Schwächen in den theoretischen Annahmen der Klassischen Testtheorie vernachlässigt werden; auf die Konzepte der Probabilistischen Testtheorie wird infolge im Rahmen der vorliegenden Arbeit nicht näher eingegangen.
Exkursende

Die erwähnten Ansätze zur Schätzung der Reliabilität sind nun die *Paralleltest-Methode* (Vorlage von zwei miteinander streng vergleichbaren Tests und anschließend Korrelation der Ergebnisse), die *Retest-Methode* (wiederholte Vorgabe eines Tests und danach Korrelation der Ergebnisreihen), die *Methode der Testhalbierung* bzw. *Split-half-Methode* (Teilung eines Tests nach einmaliger Vorgabe in zwei gleichwertige Hälften, Ermittlung des Testergebnisses für jede Testhälfte und nachfolgend Korrelation der Testergebnisse) sowie die *Methode der Konsistenzanalyse* (Teilung eines Tests in so viele Testelemente, wie er Items besitzt (Kubinger, 2006, S. 49), und infolge Ermittlung der Reliabilität über die Kennwerte Aufgabenschwierigkeits- und Trennschärfestatistiken). Die beiden letztgenannten Methoden beschreiben die *innere* bzw. *interne Konsistenz* eines Tests (Lienert & Raatz, 1998, S. 9); zur Methode der Konsistenzanalyse und insbesondere den Konzepten der *Aufgabenschwierigkeit* und der *Trennschärfe* vgl. unten. Als wichtiges Maß im Rahmen der inneren Konsistenz sei der *Alpha-Koeffizient* von *Cronbach* genannt. Dieser entspricht „der mittleren Testhalbierungs-Reliabilität eines Tests für alle möglichen Testhalbierungen“ (Bortz & Döring, 2002, S. 198); Werte von über .9 werden als hoch betrachtet.

2.1.3 Validität

„Unter der ***Validität*** eines Tests versteht man das Ausmaß, in dem der Test das misst, was er messen soll.“ (Rost, 2004, S. 34).

Unterschieden werden grundsätzlich drei Validitätsarten: die *Inhaltsvalidität* bzw. *inhaltliche Gültigkeit* (nicht quantifizierbare Bestimmung durch logische und fachliche Überlegungen von ExpertInnen, „*Experten-Rating*“), die *Kriteriumsvalidität* (Zusammenhang des Testergebnisses mit einem anderen, inhaltlich analogen Kriterium,

dem „*Außenkriterium*“) sowie die *Konstruktvalidität* (Abgleich des erwarteten Testkonstrukts mit anderen, theoretisch begründeten analogen Konstrukten). Für letztere stehen nach Bühner (2011, S. 64) folgende konkrete Strategien der Quantifizierung zur Verfügung:

- Zugang über konstruktverwandte Tests: Hierbei „werden Korrelationen mit Tests gleicher oder ähnlicher Gültigkeitsbereiche ermittelt“ (ebd.) und hohe Zusammenhänge erwartet. Man spricht in diesem Zusammenhang auch von *konvergenter Validität*.
- Zugang über konstruktfremde Tests: Hier „werden Korrelationen mit Tests anderer Gültigkeitsbereiche ermittelt“ (ebd.) und entsprechend niedrigere Zusammenhänge erwartet. Es handelt sich hier um eine *diskriminante* oder auch *divergente Validität*.
- Zugang über Faktorenanalysen: „Die so genannte *faktorielle Validität* dient zum einen dazu, homogene konstruktnahe Inhaltsbereiche zusammenzufassen, und zum anderen, diese von konstruktfremden Bereichen zu trennen“ (ebd.); man spricht von faktorieller Validität, „wenn Items eines Tests auch inhaltlich einen Faktor bilden“ (Bühner, 2011, S. 76). Mithilfe einer konfirmatorischen Faktorenanalyse können ganze Testmodelle geprüft werden. (Ausführlicher zum Thema Faktorenanalyse vgl. Kapitel unten.)

Es sei an dieser Stelle zur Konstruktvalidität hinzugefügt, dass es trotz der angesprochenen Quantifizierbarkeit durch die Korrelationen „*keine* verbindlichen Richtlinien oder festgelegte Grenzen für die *Höhe* der Korrelationen“ (Bühner, 2011, S. 67) gibt, ab denen vom Vorliegen einer Konstruktvalidität gesprochen werden kann; eine diesbezügliche Interpretation des ermittelten Korrelationsmusters obliegt also dem Testprüfer.

2.2 Nebengütekriterien

Neben den beschriebenen Hauptgütekriterien berücksichtigt die Psychologische Diagnostik bei der Testentwicklung und -analyse wie oben erwähnt mit den sogenannten *Nebengütekriterien* noch eine Reihe weiterer Aspekte, auf deren Ausführung im Rahmen der vorliegenden Arbeit zum Teil verzichtet wird; der interessierte Leser sei diesbezüglich auf die eingangs genannte Basisliteratur verwiesen. Näher eingegangen werden soll an dieser Stelle aber auf die Kriterien der *Ökonomie* und der *Nützlichkeit.*

2.2.1 Ökonomie

„Ein Test erfüllt das Gütekriterium ***Ökonomie***, wenn er, gemessen am diagnostischen Informationsgewinn, relativ wenig Ressourcen (Zeit und Geld) beansprucht“ (Kubinger, 2006, S. 94). Obwohl der finanzielle Faktor natürlich nicht unbeachtet bleiben darf – schließlich können bei einer Testung über Personal, Material, Gerätschaft, Räumlichkeiten, Lizenzgebühren usw. nicht zu vernachlässigende Kosten anfallen –, wird im Rahmen einer Testanalyse in der Regel dem Faktor Zeit größere Aufmerksamkeit geschenkt (Woike, 2003b, S. 406). Sowohl die getestete Person als auch der Testleiter sowie schließlich auch der Testauswerter werden üblicherweise daran interessiert sein, genügend Zeit, aber nicht mehr Zeit als notwendig mit dem Testprozedere zu verbringen. Entscheidend für die Ökonomie des Tests ist dann in Relation dazu der Informationsgewinn durch ihn, und infolge also der Kontext: Es muss geklärt werden, was der Zweck des Verfahrens ist.

Von Bedeutung sind an dieser Stelle die Konzepte der ***Sensitivität*** und der ***Spezifität***: Die *Sensitivität* gibt den Anteil von positiv klassifizierten Objekten im Verhältnis zur Gesamtheit der tatsächlich positiven Objekte (also die validen positiven Fälle) an, man spricht hier von einer richtig-positiv-Rate (analog zur Signalentdeckungstheorie: *Hit*) (Bortz & Döring, 2002, S. 164). Ergänzend dazu betrifft die falsch-negativ-Rate diejenigen Objekte, die ebenfalls tatsächlich als positiv zu klassifizieren gewesen wären, aber als negativ eingestuft (also nicht „entdeckt“) wurden (*Miss*). Die *Spezifität* beschreibt dagegen den Anteil von negativ klassifizierten Objekten im Verhältnis zur Gesamtheit der tatsächlich negativen Objekte (also die validen negativen Fälle), das ist die richtig-negativ-Rate (*Correct Rejection*). Die falsch-positiv-Rate meint schließlich diejenigen Objekte, die als positiv eingestuft wurden, aber eigentlich negativ sind (*False Alarm*).

2.2.2 Nützlichkeit

Ob nun im konkreten Fall der *Sensitivität* oder der *Spezifität* mehr Bedeutung beigemessen werden soll, hat wiederum mit Überlegungen zur ***Nützlichkeit*** eines Verfahrens zu tun: „Ein Test ist dann *nützlich*, wenn für das von ihm gemessene Merkmal praktische Relevanz besteht und die auf seiner Grundlage getroffenen psychologischen Entscheidungen (Maßnahmen) mehr Nutzen als Schaden erwarten lassen“ (Kubinger, 2006, S. 107).

Exkurs

Die skizzierten Zusammenhänge von *Ökonomie* und *Nützlichkeit* seien für die Praxis am Beispiel von ***Screening-Verfahren*** erläutert:

„Screening ist eine diagnostische Vorgehensweise, bei der Personen zunächst (relativ) oberflächlich erfasst werden, um zu entscheiden, ob ein [...] aufwendigeres diagnostisches Vorgehen im Anschluss angezeigt erscheint“, so Woike (2003a, S. 375), und weiter: „Charakteristisch ist die relativ Zeit sparende Vorgabemöglichkeit bei im Allgemeinen eingeschränkter Tiefe der Analyse“. Als Anwendungsgebiet kommen demnach einerseits mehrere Testungen bei ein und derselben Person infrage oder andererseits jeweils einmalige Testungen bei größeren Personengruppen, um Ausprägungen ganz bestimmter Merkmale grob erfassen zu können. – Die Beschreibung des Anwendungsgebiets macht bereits deutlich, dass im Rahmen eines Screening-Verfahrens mit einer größeren Anzahl an Testungen (ob nun je Person oder je Merkmal) gerechnet werden muss. Ein möglichst geringer Aufwand je Testung (beispielsweise ein möglichst geringer Zeitaufwand) dürfte sich also bezahlt machen. Gleichzeitig wird die Abwägung zwischen potentiellem Schaden und Nutzen maßgeblich: Screening-Verfahren sind darauf angelegt, „möglichst alle gesuchten Fälle zu identifizieren“ (ebd., S. 376); es soll also die Sensitivität hoch sein, wobei in Kauf genommen wird, dass ein Teil der Personen irrtümlich ebenfalls als positiv klassifiziert wird. Diese Konstellation wird bei Screening-Verfahren als nützlicher betrachtet als eine in Richtung Spezifität orientierte.

Exkursende

Zum Thema der Operationalisierung der beschriebenen *Hauptgütekriterien* bringen Lienert & Raatz (1998, S. 29) nun auf den Punkt: Die Qualitäten Objektivität, Reliabilität und Validität „kann ein Test nur dann besitzen, wenn auch die einzelnen Aufgaben objektiv, reliabel und valide sind“. Für die Reliabilität kann überhaupt „streng genommen lediglich für den Messwert eines Tests eine bestimmte Messgenauigkeit geschätzt werden“ (Bühner, 2011, S. 142), nicht jedoch für den Test als Ganzes.

Diese Überlegungen führen schließlich dazu, die einzelnen Items eines Tests und ihre Zusammenstellung zu Skalen bzw. Verfahren näher zu betrachten.

2.3 Analyse der Items

Neben den beschriebenen Gütekriterien können nun also auf Itemebene folgende Anhaltspunkte und Kennwerte zur Verfahrensanalyse (sowie später zur Itemselektion, und damit zur Verbesserung des Tests) herangezogen werden:

2.3.1 Rohwertverteilung

Zunächst ist es sinnvoll, sich mittels Histogrammen einen Überblick über das Antwortverhalten der in einer ersten Stichprobe untersuchten Testpersonen zu verschaffen. „Hinreichende Streuung, Symmetrie und Eingipfligkeit der ***Rohwertverteilung*** [...] sind zwar keine notwendigen Bedingungen eines guten Tests“, so Lienert & Raatz (1998, S. 58), aber dennoch wünschenswert. Beispielsweise setzen viele inferenzstatistische Verfahren normalverteilte Testwerte voraus (z.B. Bortz & Döring, 2002, S. 217). Eine Prüfung empirisch erhobener Daten auf das Vorliegen einer Normalverteilung (respektive auf das überzufällige Abweichen von derselben) kann mit dem Goodness-of-Fit-Chi-Quadrat-Test oder mit dem Kolmogorov-Smirnov-Test erfolgen.

Für den Fall einer nicht oder möglicherweise nicht erzielten Normalverteilung beschreiben Lienert & Raatz (1998, S. 147ff.) folgende Abweichungsmöglichkeiten als bedeutsam:

- die *Schiefe* der Rohwertverteilung: Eine Verteilung kann linksgipfelig (positiv-schief, rechts-asymmetrisch) oder rechtsgipfelig (negativ-schief, links-asymmetrisch) sein.
- den *Exzess* der Rohwertverteilung: Eine Verteilung kann hypexzessiv (breitgipflig) oder hyperexzessiv (schmalgipflig) sein.
- die *Unregelmäßigkeit* der Rohwertverteilung: Eine Verteilung kann auch bimodal (zweigipflig), U-förmig, J-förmig oder komplett unregelmäßig sein.

Als mögliche Ursachen für eine anomale Verteilung führen Lienert & Raatz (1998, S. 152ff.) einerseits Auffälligkeiten der Stichprobe an (heterogene Stichprobe, z.B. wenn unbemerkt zwei „Untergruppen“ existieren, deren Varianzen, Mittelwerte oder Größen stark abweichen), andererseits Mängel bei der Testkonstruktion (mittlere Aufgabenschwierigkeit zu hoch oder zu niedrig und infolge Boden- bzw. Deckeneffekte, ungünstiger Schwierigkeitsverlauf durch fehlerhafte Reihung der Aufgaben, oder auch eine unzweckmäßige Aufgabenbewertung). Letztlich kann demnach eine anormale Rohwertverteilung aber auch infolge eines nicht-normalverteilten Persönlichkeitsmerkmals zustande kommen, was sich dem Einfluss des Testentwicklers entzieht.

Mängel bei der Testkonstruktion sowie bei der Stichprobe hingegen lassen sich im Zuge einer Testrevision beheben oder zumindest mildern. Dieses Vorgehen sollte zumindest dann in Betracht gezogen werden, wenn man aufgrund inhaltlich-theoretischer

Überlegungen eigentlich normalverteilte Rohwerte erwartet hätte. Ist eine erhobene „Nicht-Normalverteilung der Testwerte theoriekonform, kann der Test unverändert bleiben“ (Bortz & Döring, 2002, S. 217), wobei allerdings bei der weiteren statistischen Auswertung auf diesen Umstand Rücksicht genommen werden muss (z.B. durch eine größere Stichprobe oder die Anwendung verteilungsfreier Techniken).

Darüber hinaus besteht auch die Möglichkeit, dass eine bestimmte, nicht-normalverteilte Häufigkeitsverteilung durchaus erwünscht ist: Dies kann der Fall sein, wenn es um die Differenzierungsfähigkeit eines Tests geht, denn schiefe „Items differenzieren in den Randbereichen der Fähigkeit oder Eigenschaftsausprägung“ (Bühner, 2011, S. 175). Wie Lienert & Raatz (1998, S. 159) dazu ausführen, ist es z.B. bei einem Test, der in erster Linie der Bestauslese dienen soll, erstrebenswert, dass „der rechte Schenkel der Häufigkeitsverteilung weit ausläuft“. Im Falle eines Tests, der dagegen für Leistungsschwache besonders gut differenzieren soll, „wird seine Rohwerteverteilung nach links weiter als nach rechts auslaufen müssen“ (ebd.). Analog sollte die Verteilung schmalgipflig sein, wenn in beiden Extrembereichen besser differenziert werden soll als im mittleren Merkmalsbereich, und breitgipflig, wenn besonders der Mittelbereich gut differenziert werden soll.

2.3.2 Itemschwierigkeit

Der *Schwierigkeitsindex* eines Items drückt die psychometrische ***Itemschwierigkeit*** aus. Er bezieht sich auf die beobachteten „richtigen“ Antworten im Verhältnis zur gesamten Stichprobe, wobei zu beachten ist, dass die Schwierigkeit „als Zustimmung zu einem Item *in Schlüsselrichtung* der Skala definiert“ ist (Bühner, 2011, S. 219). Je höher die Itemschwierigkeit, „desto leichter fällt es Personen, ein Item zu ‘lösen’ bzw. ‘symptomatisch’ zu beantworten“ (Moosbrugger & Hartig, 1998, S. 413), viele Personen der Stichprobe werden diesem Item zustimmen. Stimmen nur wenige Personen dem Item zu bzw. beantworten es „richtig“, ist die psychometrische Itemschwierigkeit niedrig.

Bei der Ermittlung der Schwierigkeitsindizes für jedes Item ist zwischen dichotomen und polytomen Antwortalternativen zu differenzieren: Bei zweistufigen Antwortformaten berechnet man nach Bortz & Döring (2002, S. 218) die Schwierigkeit (p_i), indem man die Anzahl der richtigen Lösungen (R_i) durch die Gesamtzahl der Antworten (N_i) dividiert:

$$p_i = \frac{R_i}{N_i}$$

Für mehrstufige Items gilt: die Summe der erreichten Punkte auf Item i (x_i) wird dividiert durch das Produkt der maximalen Punktzahl je Person (k_i) und der Anzahl der antwortenden Personen (n):

$$p_i = \frac{\sum_{m=1}^{n} x_{im}}{k_i \cdot n}$$

Der Wertebereich der Itemschwierigkeit liegt somit zwischen 0 (schwerstes Item) und 1 (leichtestes Item). Im Allgemeinen werden Werte zwischen .2 und .8 als mittelschwer betrachtet. Soll ein Test „in allen Bereichen des Fähigkeits- oder Eigenschaftsbereiches von Personen differenzieren“ (Bühner, 2011, S. 223), ist bei der Testkonstruktion auf eine breite Streuung der Schwierigkeitsindizes zu achten; der Test soll also unterschiedlich schwierige Items enthalten. Ist dagegen „eine Differenzierung in einem ganz bestimmten Bereich der Merkmalsskala erwünscht, so wird man hauptsächlich Aufgaben von dieser Schwierigkeitsstufe zu einem Test zusammenstellen“, führen Lienert & Raatz (1998, S. 107) aus. Die Gestaltung der Aufgabenschwierigkeiten ist also abhängig vom anvisierten *Geltungsbereich*.

2.3.3 Itemtrennschärfe

Das Konzept der ***Itemtrennschärfe*** oder „***Trennschärfe*** zielt darauf ab, wie *scharf* die Antworten auf ein Item zwischen hohen und niedrigen Eigenschaftsausprägungen *trennen*, also wie gut sie die Personenstichprobe *teilen*“ (Rost, 2004, S. 369). Die Trennschärfe gibt „an, wie gut ein Item die angestrebte Fähigkeit oder Eigenschaft misst“ (Bühner, 2011, S. 171). Sie wird ermittelt als „Korrelation eines Items mit dem Summenwert der *übrigen* Items einer Skala“ (ebd.), der Ergebniswert wird als *Trennschärfekoeffizient* eines Items bezeichnet. Wie Bortz & Döring (2002, S. 219) präzisieren, wird bei intervallskalierten Test-Scores „als Trennschärfe (r_{it}) die Produkt-Moment-Korrelation zwischen den Punktwerten pro Item i und dem korrigierten Gesamttestwert t“ gewählt:

$$r_{it} = \frac{\mathrm{cov}(i, t)}{s_i \cdot s_t}$$

Der Wertebereich des Trennschärfekoeffizienten reicht, korrelationstypisch, von –1 bis 1, wobei folgende Zuschreibungen gelten (zum Umgang mit den Werten vgl. unten):

- hohe positive Trennschärfe:
  Die entsprechende Aufgabe unterscheidet deutlich zwischen „guten“ und „schlechten“ Testpersonen; und zwar beantworten gute Testpersonen das Item meist richtig, schlechte dagegen meist falsch oder gar nicht (Lienert & Raatz, 1998, S. 78). In diesem Zusammenhang sei noch einmal festgehalten, dass „richtige“ Antworten bzw. „gelöste“ Items inhaltlich im Sinne der zu messenden Skala verrechnet werden (Bühner, 2011, S. 176) und umgekehrt. Eine hohe Korrelation zwischen den Ergebnissen einer Testperson bei dem jeweiligen Item und dem Gesamttestwert deutet darauf hin, „dass das Item etwas Ähnliches erfasst wie der Gesamttest“ (Moosbrugger & Hartig, 2003b, S. 413). Im Allgemeinen gelten für die Trennschärfe Werte von über .5 als hoch und Werte zwischen .3 und .5 als mittelmäßig.

- Trennschärfe um null:
  „Gute“ und „schlechte“ Testpersonen antworten ähnlich häufig richtig oder falsch; das Item „trennt“ die Gruppen also nicht gut, und es hat mit dem restlichen Test nicht viel gemeinsam. Lienert & Raatz (1998, S. 78) bezeichnen solche Aufgaben als unbrauchbar.

- hohe negative Trennschärfe:
  „Gute“ Testpersonen beantworten das Item meist falsch, „schlechte“ dagegen eher richtig. Dies könnte ein Hinweis darauf sein, dass die „Items von den untersuchten Personen im Sinne des zu messenden Konstrukts umgekehrt wie beabsichtigt verstanden wurden“ (Moosbrugger & Hartig, 2003b, S. 414). In dem Fall ist zunächst die Polung des betreffenden Items zu überprüfen, und gegebenenfalls auch die Eindeutigkeit der Formulierung zu hinterfragen.

2.4 Zusammenstellung der Items

Von Interesse bei einer Testanalyse ist weiters die Frage, inwieweit die einzelnen Items eines Tests respektive einer Skala im Durchschnitt miteinander korrelieren.

2.4.1 Homogenität

Dies kann über die ***Homogenität*** oder ***Item-Interkorrelationen*** zum Ausdruck gebracht werden: Hohe Homogenität bedeutet, dass die Items eines Tests ähnliche Informationen erfassen. Im Umkehrschluss sind bei eindimensionalen Instrumenten hohe Homogenitäten erstrebenswert. Nachdem die mittlere Item-Interkorrelation in den oben genannten *Alpha-Koeffizienten* von Cronbach eingeht, wird dieser Kennwert oft auch als *Homogenitätsindex* bezeichnet. Bühner (2011, S. 168) zeigt demgegenüber, dass auch bei einer mehrdimensionalen Skala ein hohes *Cronbach-α* ermittelt werden kann, womit aus seiner Sicht die Funktion der mittleren Itemkorrelation als Prüfgröße der Homogenität und *Cronbach-α* als ihr Index im Grunde zu hinterfragen ist. Obwohl diese Kritik in der Literatur gegenwärtig zunehmend Zustimmung zu erfahren scheint, findet das beschriebene Konzept der Homogenitätsbestimmung auf Basis der Item-Interkorrelationen in der Praxis weiterhin Verbreitung.

Zu beachten ist infolge jedenfalls, dass die Ermittlung von *Cronbach-α* bei von vornherein mehrdimensional konzipierten Tests nur je Skala und nicht für den Gesamttest sinnvoll ist.

Für das *Cronbach-α* gilt weiters, dass es sich „nicht nur mit wachsender Item-Interkorrelation, sondern auch mit steigender Itemzahl erhöht“ (Bortz & Döring, 2002, S. 220), und dass es bei heterogenen Tests die Reliabilität unterschätzt (ebd., S. 198). Wie für jedes korrelative Maß gilt nach Schendera (2007, S. 157) für das *Cronbach-α* die Forderung nach ausreichend vielen Messwertpaaren und möglichst wenigen Missings (vgl. auch unten).

2.4.2 Beurteilung der Kennwerte

Zur Bedeutung der Höhe der Itemkennwerte für die Beurteilung von Aufgaben und Testzusammensetzung kann folgendes gesagt werden:

Wenngleich es zunächst nachvollziehbar erscheint (und in der Literatur auch immer wieder so formuliert wird, vgl. z.B. bei Bortz & Döring, 2002, S. 219, oder auch bei Rost, 2004, S. 370), bei einer Testentwicklung grundsätzlich nach möglichst hohen Itemtrennschärfen zu streben, müssen hier immer auch inhaltliche Überlegungen eine Rolle spielen. Viele Items mit sehr hohen Trennschärfen und einander ähnlichen Schwierigkeiten würden bedeuten, dass alle diese Items etwas Ähnliches erfassen. Dies

kann gewünscht sein, etwa wenn nämlich nach Bühner (2011, S. 248) mit dem Test ein relativ enger Verhaltensausschnitt erfasst werden soll, es kann aber auch, über redundante Items, einer eigentlich gewünschten heterogenen Itemzusammensetzung im Wege stehen.

Items mit eher niedrigen Trennschärfen müssen nicht zwangsweise zu entfernen sein; Vielmehr ist bei einer Itemselektion darauf zu achten, dass neben den Trennschärfen auch die Streuungen und die Schwierigkeitsindizes der Items berücksichtigt werden sowie auch deren Zusammenhänge untereinander:

Zunächst ist noch einmal klar zu formulieren: Schwierigkeit und Trennschärfe sind nicht unabhängig voneinander. „Je extremer die Schwierigkeit, desto geringer die Trennschärfe“, so Bortz & Döring (2002, S. 219), und weiter: „bei sehr leichten und sehr schweren Items wird man deshalb Trennschärfeeinbußen in Kauf nehmen müssen.“ Aufgaben mittlerer Schwierigkeit haben bessere Voraussetzungen für eine hohe Trennschärfe. Nach Lienert & Raatz (1998, S. 107) „ist bei Tests mit schwierigkeitshomogenen Aufgaben von ausschließlich 50%iger Schwierigkeit die beste Differenzierung zu erwarten“, was besonders für mittlere Ausprägungen des untersuchten Merkmals gilt. Weiters wächst demnach mit einer steigenden Differenzierung der Items auch die Testreliabilität; durchschnittlich schwierige Tests sind also „im allgemeinen reliabler als solche von über- oder unterdurchschnittlicher Schwierigkeit“ (ebd.).

Für die Trennschärfe gilt: Als Korrelation ist sie grundsätzlich beeinflussbar durch die Varianz der korrelierten Variablen und besonders durch Ausreißerwerte (Bühner, 2011, S. 173f.). Aus diesem Grund ist es ratsam, die Verteilungen der Rohwerte zu prüfen, *bevor* Entscheidungen zur Itemselektion unter Berücksichtigung oder gar auf alleiniger Basis der Trennschärfe getroffen werden. In diesem Zusammenhang wird deutlich, dass auch die Zusammensetzung der Stichprobe eine Rolle spielt: ist eine Stichprobe eher homogen, ist die Varianz der Messwerte eingeschränkt, und so können in der Regel in homogenen Stichproben keine so hohen Trennschärfen erwartet werden wie in heterogenen Stichproben. Eine große Streuung muss allerdings nicht unbedingt zu hohen Trennschärfen führen; dies gilt nach Bühner (2011, S. 259) nur, „wenn es sich um systematische und nicht durch Messfehler bedingte Unterschiede handelt“. Analog dazu führt Bühner (2011, S. 178f.) aus: „Dieselben Einflussgrößen, die auf die Trennschärfe wirken, beeinflussen auch die Reliabilitätsschätzung“, und so kann eine hohe Streuung hohe Reliabilitätsschätzungen „begünstigen, wohingegen eine geringe Varianz zu geringen Reliabilitätsschätzungen führen kann“.

Trotz der beschriebenen Notwendigkeit, die Testkennwerte zueinander in Relation setzen und ihre inhaltliche Bedeutung bei der Testkonstruktion immer im Auge behalten zu müssen, können nun zusammenfassend folgende konkrete Werte zur Orientierung herangezogen werden:

- Die Gestaltung der *Itemschwierigkeiten* hat nach der intendierten Funktion des Verfahrens und dem damit gewünschten Differenzierungsbereich zu erfolgen; Items kleiner als .2 gelten als schwierig, solche ab .8 als leicht.
- Für Items mit *Trennschärfen* kleiner als .3 wird gemeinhin empfohlen, sie aus dem Test zu entfernen; Trennschärfen ab .5 gelten als hoch.
- Ein *Cronbach-α* größer als .9 wird als hoch betrachtet; im Rahmen psychologischer Konstrukte gelten allerdings auch Werte zwischen .9 und .7 (sowie bei heterogenen Skalen sogar darunter) als akzeptabel (Field, 2009, S. 675).

Wird nun entsprechend bedachtsam eine inhaltlich nur wenig zu ihrer Skala passende Aufgabe (also ein Item mit geringer Trennschärfe) aus dem Test entfernt, „führt dies zu einer Steigerung der Reliabilität“ (Bühner, 2011, S. 249, 259); man spricht bei diesem – vorzugsweise sukzessiv zu erfolgenden – Vorgehen von der *Alpha-Maximierung*. Diese Reliabilitätssteigerung könnte allerdings, wie oben angedeutet, auf Kosten der Testheterogenität und infolge der Inhaltsvalidität gehen: werden nach und nach immer mehr Items entfernt, bleiben am Ende nur noch sehr inhaltsähnliche Items über, und das ursprüngliche Konstrukt muss vermutlich enger gefasst werden. Zudem könnten auf diese Art besonders leichte oder besonders schwere Items ausgeschieden werden, was zur Folge hätte, dass der Test in den Randbereichen des Fähigkeits- oder Eigenschaftsspektrums nicht mehr differenzieren könnte (Bühner, 2011, S. 182).

2.4.3 Dimensionalität

Wie in diesem Zusammenhang erkennbar wird, mag sich die inhaltliche Auseinandersetzung mit der ***Dimensionalität*** eines Verfahrens nicht auf theoretische Überlegungen zu Beginn der Testentwicklungsarbeit beschränken lassen. Die Frage ist eben, ob ein Test nur *ein* Merkmal oder *mehrere* Konstrukte erfasst. Wie Bortz & Döring (2002, S. 220f.) berichten, erweisen sich eindimensional intendierte Tests „nicht selten bei späteren empirischen Dimensionalitätsüberprüfungen als mehrdimensional“. Geprüft

werden kann die Dimensionalität von Testverfahren im Rahmen einer Faktorenanalyse (vgl. dazu ausführlich Kapitel unten). Können dabei die Item-Interkorrelationen auf einen einzelnen Faktor reduziert werden, ist der Rückschluss auf Eindimensionalität zulässig. Wie im Folgenden ausgeführt wird, ist die Technik der Faktorenanalyse allerdings mit einem breiten Interpretationsspielraum verbunden, für dessen Handhabung Erfahrung auf dem Feld der Testkonstruktion zweifellos von großem Nutzen ist.

3. Weiterführende Testanalyse: Die Faktorenanalyse

In Datensätzen mit einer großen Anzahl an Variablen ist es praktisch nicht möglich, korrelative Zusammenhänge „per Augenschein" zu interpretieren. Eine Faktorenanalyse (als Oberbegriff für verschiedene Verfahren, die aus einer großen Zahl von beobachteten Variablen eine möglichst geringe Zahl von nicht beobachtbaren Faktoren bzw. Dimensionen zu extrahieren vermögen) erlaubt es, in so einem Variablengeflecht eine Ordnung zu bestimmen, „aus der sich die beobachtete Konstellation der Variableninterkorrelationen erklären lässt" (Bortz & Schuster, 2010, S. 387).

3.1 Grundlagen und Eigenschaften

Erste Ansätze einer Faktorenanalyse wurden bereits Anfang des 19. Jahrhunderts von Charles E. Spearman entwickelt; sein 1904 im *American Journal of Psychology* veröffentlichter Artikel „*General intelligence, objectively determined and measured*" über einen möglichen einzelnen Faktor allgemeiner Intelligenz gilt gleichsam als Geburtsstunde der Faktorenanalyse (Schendera, 2010, S. 180f.).

Ziel einer Faktorenanalyse ist es nun eben, Zusammenhänge von „Items untereinander durch eine geringere Anzahl dahinter liegender homogener Faktoren zu erklären" (Bühner, 2011, S. 296). Es sollen korrelierende Variablen in wenige, voneinander unabhängige Variablengruppen geordnet und auf höherer Abstraktionsebene zusammengefasst werden (Bortz & Döring, 2005, S. 383, und Bortz & Schuster, 2010, S. 386). Die Faktorenanalyse ist somit ein *datenreduzierendes* Verfahren.

Wie bereits erwähnt, geht es bei der Faktorenanalyse darum, *eine* Ordnung zu bestimmen – theoretisch gibt es aber unüberschaubar viele Möglichkeiten, wie diese

Ordnung aussehen kann. Die Faktorenanalyse führt zu *interpretativ mehrdeutigen* Ergebnissen, die zwar eine „Hypothesenbildung erleichtern, jedoch keine Überprüfung inhaltlicher Hypothesen über Variablenstrukturen gestatten“ (Bortz & Schuster, 2010, S. 388). Aufgrund der durch die formale Gleichwertigkeit verschiedener Lösungen bedingte Uneindeutigkeit des Verfahrens ist es nicht zulässig, von richtigen oder falschen Ergebnissen zu sprechen (ebd., S. 396). Das Ziel muss vielmehr sein, dasjenige Ordnungssystem zu identifizieren, das mit dem theoretischen Kontext der untersuchten Variablen *am besten* zu vereinbaren ist. So wählt man „diejenige Lösung, die nach dem jeweiligen Stand der Theoriebildung über die untersuchten Variablen *am plausibelsten* ist“ (ebd., S. 395). Es werden also Hypothesen über die den beobachteten Daten zugrundeliegenden Strukturen formuliert; die Faktorenanalyse ist folglich ein *heuristisches, hypothesengenerierendes* Verfahren (ebd., 2010, S. 387f.).

Des Weiteren ist die Faktorenanalyse „ein Verfahren zur Überprüfung der Dimensionalität komplexer Merkmale“ (ebd.). Lienert (1998, S. 227) beschreibt die Faktorenanalyse als probates Mittel, um Aussagen zur Konstruktvalidität eines Tests tätigen zu können (vgl. auch Kapitel oben).

3.2 Begriffsklärung

In der Literatur wird grundsätzlich zwischen zwei Klassen von Faktorenanalysen unterschieden: Während die *explorative Faktorenanalyse (EFA)* a priori keine bzw. keine konkreten Annahmen über die Struktur von empirisch beobachteten Korrelationen zwischen Variablen trifft und daher wie oben ausgeführt als hypothesen*generierendes* Verfahren betrachtet wird, dient die *konfirmatorische Faktorenanalyse* (*KFA* bzw. *CFA*) „dazu, theoretisch oder empirisch gut fundierte Modelle auf ihre empirische Passung mit den Daten hin zu testen oder mit alternativen Modellen zu vergleichen“ (Bühner, 2011, S. 380) und stellt somit ein hypothesen*prüfendes* Verfahren dar. „Im Gegensatz zur EFA ist hier [...] bereits bekannt, wie viele Faktoren erwartet werden, wie diese zueinander in Beziehung stehen und welche Items sie erklären.“ (ebd., S. 381); es werden bei der KFA, anders als bei der EFA, „gewöhnlich nicht alle Ladungen geschätzt, sondern nur die, die vorher als theoretisch relevant angenommen wurden.“ (ebd., S. 399) (zum Konzept der Ladungen siehe unten).

Wie Schendera (2010, S. 291f.) ausführt, sollen vor der Durchführung einer Faktorenanalyse inhaltsorientierte Vermutungen über zu bündelnde Dimensionen

vorliegen. „Eine Faktorenanalyse kann keine validen Faktoren oder Komponenten ermitteln, wenn diese nicht einmal von der Theorieseite her unterstellt werden können. Angesichts einer immer notwendigen Rückversicherung gegenüber diesem vernunftorientierten Minimalkriterium als sine non qua“, so Schendera weiter, „ist *jede* Faktorenanalyse eine konfirmatorische Faktorenanalyse.“ Obgleich im Rahmen einer Testentwicklung und ersten Testerprobung also üblicherweise freilich bereits zu Beginn inhaltliche Konzepte zu den Bedingungszusammenhängen existieren (müssen), erscheint der Einsatz einer EFA in diesen Zusammenhängen vertretbar, solange kein detailliert begründetes Testmodell vorliegt. In der Praxis werden explorative Faktorenanalysen „oft in Situationen angewendet, in denen man bereits Annahmen über die Anzahl der einem Datensatz zugrundeliegenden Faktoren, deren Interkorrelationen oder die Ladungen der manifesten Variablen auf den Faktoren hat“ (Werner, 2014, S. 2). Dies muss nicht auf ein inkonsequentes oder gar schlampiges Vorgehen hindeuten, sondern kann auch dem Umstand Rechnung tragen, dass inhaltliche Aspekte des zugrundeliegenden Konstrukts noch als bearbeit- und veränderbar betrachtet werden. Jöreskog (2007, S. 58) sieht bezüglich der Unterscheidung zwischen einem explorativen und einem konfirmativen Ansatz in der Praxis ebenfalls einen gewissen Spielraum: „Factor analysis need not be strictly exploratory or strictly confirmatory. Most studies are to some extent both exploratory and confirmatory because they involve some variables of known and other variables of unknown composition“.

Im Rahmen der vorliegenden Arbeit erscheint jedenfalls lediglich die Beschreibung und Anwendung des Konzepts der explorativen Faktorenanalyse relevant. Deshalb – und aus Komplexitätsgründen – wird an dieser Stelle auf weitere Ausführungen zum Thema konfirmatorische Faktorenanalyse verzichtet. Im Folgenden werden, wenn nicht anders angegeben, die Begriffe „Faktorenanalyse“ und „explorative Faktorenanalyse“ bzw. „EFA“ synonym verwendet.

3.3 Voraussetzungen und Kriterien der Durchführbarkeit

Grundsätzlich ist zu Beginn inhaltlich zu prüfen, ob es sich bei einer Faktorenanalyse um das ideale Modell zur Beantwortung der gegebenen Fragestellung handelt (Bühner, 2011, S. 343).

Nachdem die Güte der Ergebnisse einer Faktorenanalyse von der Qualität der Ausgangsdaten abhängt (Backhaus et al., 2011, S. 336), sind nun selbige noch vor Beginn

der eigentlichen Analyse kritisch zu betrachten. Als wichtig wird zunächst erachtet, dass es sich bei den erhobenen Informationen auch um für den Untersuchungsgegenstand relevante handelt, andernfalls sie auszusortieren sind.

3.3.1 Eignung der Korrelationsmatrix

Voraussetzung dafür, Faktoren zusammenfassen zu können, ist es nun, *substanzielle* Zusammenhänge zwischen (Paaren von) Variablen vorliegen zu haben; Backhaus et al. (2011, S. 336) sprechen in diesem Zusammenhang veranschaulichend von „bündelungsfähigen" Variablenpaaren. Das Erstellen einer Korrelationsmatrix erlaubt demnach eine erste Abschätzung, welche Variablen mit welchen anderen Variablen (in zunächst unbekannter Form) zusammenhängen. Viele kleine Werte in der Korrelationsmatrix können das Ergebnis einer zugrundeliegenden heterogenen Datenstruktur sein, was eine sinnvolle Anwendung einer Faktorenanalyse in Frage stellen würde: „Liegen keine Korrelationen über .30 vor, gibt es für eine Faktorenanalyse demnach auch nichts zu faktoranalysieren." (Schendera, 2010, S. 293f). Hohe Werte sind demgegenüber positiv zu bewerten; das Vorliegen sowohl kleiner als auch großer Werte lässt auf Basis der Korrelationsmatrix kein eindeutiges Urteil über die Eignung der Daten für eine Faktorenanalyse zu (Backhaus et al., 2011, S. 339).

Darüber hinaus muss die Frage gestellt werden, ob eine vorliegende Korrelationsmatrix nur zufällig von einer Einheitsmatrix (bedeutet es liegen keine Korrelationen zwischen den Variablen vor) abweicht oder ob ein systematischer Unterschied besteht (Bortz & Schuster, 2010, S. 417). Dies kann mithilfe des ***Bartlett-Tests*** auf Sphärizität überprüft werden, wobei vorausgesetzt wird, dass die Variablen in der Erhebungsgesamtheit normalverteilt sind (Backhaus et al., 2011, S. 341). „Der Bartlett-Test prüft die globale Nullhypothese, dass alle Korrelationen der Korrelationsmatrix gleich null sind." (Bühner, 2011, S. 347f). Erhält man ein signifikantes Ergebnis, kann man davon ausgehen, dass alle Korrelationen der Korrelationsmatrix größer als null sind und eine Faktorenanalyse durchgeführt werden kann. Wird der Test nicht signifikant, ist davon auszugehen, dass die Items unkorreliert und daher für eine Faktorenanalyse nicht geeignet sind.

Zu beachten ist in diesem Zusammenhang auch die Stichprobengröße: der Bartlett-Test wird umso eher signifikant, je größer die Stichprobe ist (für weitere Aspekte zum Thema Stichprobe siehe unten) (Bühner, 2011, S. 347f).

Als weiterer Indikator dafür, ob eine Faktorenanalyse sinnvoll erscheint oder nicht, gilt nun der ***Kaiser-Meyer-Olkin-Koeffizient (KMO)***. Dieser basiert auf der Idee, die Varianz einer Variablen aufzuteilen in einen Anteil, der durch die verbleibenden Variablen mithilfe einer multiplen Regressionsanalyse erklärt werden kann (auch genannt ***Image*** eines Items), und einen Anteil, der von den übrigen Variablen unabhängig ist (auch genannt ***Anti-Image***) (Backhaus et al., 2011, S. 341f). Variablen sind nur dann für eine Faktorenanalyse geeignet, wenn das Anti-Image möglichst gering ausfällt (d.h. der Teil, der von den anderen Variablen unabhängig ist), da man ja eben von zugrundeliegenden gemeinsamen Faktoren ausgeht. Als Richtwert kann nach Schendera (2010, S. 295) für die zugehörige ***Anti-Image-Korrelations-Matrix*** (AIC-Matrix) mit ihren Partialkorrelationen gelten: „Die Nicht-Diagonal-Elemente (Anti-Image-Korrelationen) müssen klein (um 0) sein; der Anteil der Nicht-Diagonal-Elemente, der größer als 0,09 ist, sollte unter 25% liegen".

Angezeigt wird beim KMO-Koeffizienten nun, in welchem Umfang die Ausgangsvariablen zusammengehören (Backhaus et al., 2011, S. 342), d.h. infolge, „ob die Itemauswahl für eine Faktorenanalyse geeignet ist" (Bühner, 2011, S. 346). Der Wertebereich des KMO-Koeffizienten liegt zwischen 0 und 1; als Anhaltspunkte für die Beurteilung der Werte können nach Bühner (2011, S. 347) folgende Zuschreibungen herangezogen werden:

< 0.50	inkompatibel mit der Durchführung
0.50 – 0.59	schlecht
0.60 – 0.69	mäßig
0.70 – 0.79	mittel
0.80 – 0.89	gut
≥ 0.90	sehr gut

Tabelle 1: Interpretation des KMO-Koeffizienten

Liegt der KMO-Koeffizient unter .5, sollte keine Faktorenanalyse durchgeführt werden (Bühner, 2011, S. 348). Field (2009, S. 647) empfiehlt für diesen Fall, entweder mehr Daten zu sammeln oder die Zusammenstellung der Items zu überdenken.

3.3.2 Eignung der einzelnen Items

Korrespondierend zum KMO-Koeffizienten steht bezüglich der einzelnen Items der ***MSA*(„measure of sampling adequacy“)-*Koeffizient*** zur Verfügung. Dieser berücksichtigt die Korrelationen bzw. Partialkorrelationen eines Items mit den übrigen Items, er „gibt an, ob ein Item eine hohe Einzigartigkeit besitzt“ (Bühner, 2011, S. 348). Die MSA-Koeffizienten finden sich auf der Diagonalen der Anti-Image-Korrelationsmatrix. Die Beurteilung der erhaltenen Werte erfolgt äquivalent zum KMO-Koeffizienten; Items mit niedrigen MSA-Werten bedürfen einer weiteren Inspektion. Bühner (2011, S. 347f) führt als mögliche Gründe für niedrige MSA-Werte an, dass Items eine niedrige Reliabilität besitzen könnten, eine geringe Itemvarianz aufgrund extremer Itemschwierigkeiten gegeben sein könnte, oder die zugehörige Skala Klumpen von ähnlich formulierten Items enthalten könnte, denen das betreffende Item allen nicht zuzuordnen ist. Schendera (2010, S. 293) formuliert bezüglich der geringen Itemvarianz etwas allgemeiner: „Eine Einschränkung des Ranges kann u.U. einen Bias in den Vorgängen des Samplings oder auch Messens andeuten, und kann u.U. ausgesprochen niedrige Korrelationskoeffizienten verursachen“, was wiederum die Extraktion brauchbarer Faktoren behindert. Jedenfalls sollten die erhobenen Werte der Items eine gewisse Streuung (*Range*) aufweisen. Backhaus et al. (2011, S. 339) fassen zusammen, dass „die Höhe der Korrelationskoeffizienten durch die Verteilung der Variablen in der Erhebungsgesamtheit (Symmetrie, Schiefe und Wölbung der Verteilung) beeinflusst wird“, weshalb eine Prüfung der Variablen auf Normalverteilung bzw. zumindest auf eine Gleichartigkeit der Verteilungen empfehlenswert ist.

Zusammenfassend ist es also ratsam, die Itemverteilungen bzw. Streudiagramme zu betrachten und die Relevanz der Items zu prüfen, und danach individuell über ein Beibehalten oder Entfernen des Items zu entscheiden. Ein besonderes Augenmerk ist bei der Prüfung der Itemverteilungen auf etwaige Ausreißer zu legen, da sich diese besonders gravierend auf die Höhe der Korrelationen auswirken können. Es wird diesbezüglich empfohlen, gegebenenfalls aus inhaltlichen Überlegungen (fehlendes Instruktionsverständnis, falsche Gruppenzugehörigkeit usw.) Ausschlüsse vorzunehmen (Bühner, 2011, S. 343).

3.3.3 Zum Zusammenhang von Itemanzahl, Ladungen und Stichprobengröße

Grundsätzlich ist neben den Eigenschaften der verwendeten Items auch deren Anzahl zu beachten: Die Anzahl der Beobachtungen *N* sollte zunächst einmal jedenfalls deutlich größer als die Anzahl der Variablen sein (Schendera, 2010, S. 294). Genauer sollten es pro (erwartetem) Faktor mindestens 4 Items sein, und für jeden Aspekt eines Konstrukts sollte die gleiche Itemanzahl verwendet werden, so laut Bühner (2011, S. 344ff), der außerdem für Testkennwerte und Items Reliabilitäten von mindestens .6 fordert. Als Mindestschätzung der Reliabilität gelte die Kommunalität eines Items (zum Konstrukt der Kommunalität siehe unten), die in Zusammenhang mit der Stichprobengröße als Kriterium dienen sollte (je höher die Kommunalitäten und je größer die Stichprobe, desto stabiler die Faktorenlösung).

Als vielzitierte (z.B. Bortz & Schuster, 2010, S. 396 und 422, Field, 2009, S. 647, Moosbrugger & Hartig, 2003a, S. 143ff) Anwendungsempfehlung können die von Guadagnoli & Velicer (1988, S. 274) formulierten Zusammenhänge bzw. Bedingungen herangezogen werden: „If components possess four or more variables with loadings above .60, the pattern may be interpreted whatever the sample size used“. Bei einer größeren Variablenanzahl (mindestens 10 bis 12 Variablen je Faktor) mit allerdings geringeren Ladungen (um .40, zum Konzept der Ladung siehe unten) müsse für eine Interpretierbarkeit ein Stichprobenumfang von mindestens N=150 gegeben sein. Im Falle weniger Variablen je Faktor *und* geringen Ladungen „...the pattern should not be interpreted unless a sample size of 300 ore more observations has been used.” (ebd.). Bei N<300 sei demnach eine Studienreplikation angeraten.

3.3.4 Eignung der Stichprobe

Von Vorteil ist zusammenfassend eine möglichst homogene Stichprobe, da wie oben erwähnt die Höhe der Korrelationen zwischen den Variablen durch den Homogenitätsgrad der Befragungsstichprobe beeinflusst wird (Backhaus et al., 2011, S. 336). Die untersuchte Stichprobe sollte möglichst groß und repräsentativ sein (Bortz & Schuster, 2010, S. 396).

3.3.5 Skalenniveau und Antwortformat

Während Bühner (2011, S. 343) dem Skalenniveau keine praktische Bedeutung zuschreibt, empfehlen Bortz & Schuster (2010, S. 397) die Durchführung einer Faktorenanalyse

möglichst für intervallskalierte Merkmale. Bezüglich der Anzahl der Intervalle auf den Skalen weist Bühner (2011, S. 343) darauf hin, dass „die Faktorisierung von Items mit dichotomem oder mehrstufigem ordinalen Antwortformat“ in SPSS umständlich sei. Auch Schendera (2010, S. 293) sieht diesen Aspekt problematisch und rät von der Faktorisierung gemischter Skalenniveaus überhaupt ab.

3.4 Durchführung

Die Faktorenanalyse geht „von der grundlegenden Annahme aus, dass jeder Beobachtungswert einer Ausgangsvariablen x_j oder der standardisierten Variablen z_j sich als eine *Linearkombination* mehrerer (hypothetischer) Faktoren beschreiben lässt“ (Backhaus et al., 2011, S. 344).

3.4.1 Ermittlung der Faktoren

Diese Faktoren sind aber „weder beobachtbar noch inhaltlich eindeutig, sondern ergeben sich indirekt aus den Koeffizienten sämtlicher Linearkombinationen“ (Kubinger, 2006, S. 53). Wie viel ein einzelner Faktor nun mit einer Ausgangsvariablen zu tun hat, bzw. „wie gut eine Variable zu einer Variablengruppe passt“ (Bortz & Schuster, 2010, S. 386), gibt dabei die ***Faktorladung*** λ_{iq} an. „Die standardisierte Ausprägung von Personen (z-Wert) auf einem Item ergibt sich aus einer Kombination aus gewichteten Ausprägungen der Personen auf den Faktoren (***Faktorwerten***), zu der ein Fehler addiert wird.“ (Bühner, 2011, S. 299), wobei dieser Fehler für jede Person unterschiedlich ausfallen kann. Demzufolge wird an der Stelle wie folgt formuliert:

$$z_{vi} = \lambda_{i1} \cdot \xi_{v1} + \lambda_{i2} \cdot \xi_{v2} + ... + \lambda_{iq} \cdot \xi_{vq} + \varepsilon_{vi}$$

wobei gilt:

λ_{iq} = Ladung des Items *i* auf Faktor *q* (Gewichtung)

ξ_{vq} = Faktorwert (= Ausprägung) der Person *v* auf Faktor *q*

ε_{vi} = Fehler der Person *v* bei der Messung des Items *i*

In Matrizenschreibweise kann diese Definitionsgleichung folgendermaßen dargestellt werden (Bühner, 2011, S. 302):

$$Z = L' \cdot F + E$$

wobei gilt:
Z = Matrix der z-Werte von *v* Personen auf *i* Items
L' = transponierte Ladungsmatrix der *i* Items auf den *q* Faktoren
F = Matrix der Faktorwerte von *v* Personen bei *q* Faktoren
E = Matrix der Fehlerkomponenten von *v* Personen bei *i* Items

Nachdem nur die z-Werte der Personen (also deren standardisierte Itemantworten) direkt beobachtet werden können, müssen Ladungen, Faktorwerte und Fehler geschätzt werden. Dies geschieht durch Berücksichtigung von Formeln für Korrelationskoeffizienten über Strukturgleichungen, auf deren Ausführung aus Komplexitätsgründen im Rahmen der vorliegenden Arbeit verzichtet wird. Als Ergebnis ergibt sich nach Vereinfachung die ***Fundamentalgleichung der Faktorenanalyse*** (Bühner, 2011, S. 306f):

$$R = L \cdot L' + V$$

wobei gilt:
R = Korrelationsmatrix
L = Ladungsmatrix
L' = transponierte Ladungsmatrix
V = Rest (Fehlervarianz)

Backhaus et al. (2011, S. 345 + 354) betonen an dieser Stelle die Prämisse der Unabhängigkeit der Faktoren voneinander: So ergibt sich als Fundamentalgleichung vielmehr

$$R = A \cdot C \cdot A' + U$$

wobei gilt:
R = Korrelationsmatrix
A = Ladungsmatrix
C = Matrix der Korrelationen zwischen den Faktoren
A' = transponierte Ladungsmatrix
U = Rest (potenzielle Messfehler und spezifische Varianz)

Unter der Voraussetzung, dass man von unabhängigen (orthogonalen) Faktoren ausgeht, „entspricht C einer Einheitsmatrix (einer Matrix, die auf der Hauptdiagonalen nur Einsen und sonst Nullen enthält)“ (Backhaus et al., 2011, S. 345), wodurch schließlich erst zu

$$R = A \cdot A' + U$$

vereinfacht werden kann.

Das ***Fundamentaltheorem der Faktorenanalyse*** besagt zusammenfassend, dass sich die ursprüngliche „Korrelationsmatrix durch die Faktorladungen (Matrix A) und die Korrelationen zwischen den Faktoren (Matrix C) reproduzieren lässt“ (ebd.).

Durch die Quadrierung der Faktorladungen in Bezug auf ein Item bzw. eine Variable und deren anschließende Summation wird (bei unkorrelierten Faktoren) „der durch die Faktoren wiedergegebene Varianzerklärungsanteil der betrachteten Variablen“ (Backhaus et al., 2011, S. 353) ermittelt. Der Varianzerklärungsanteil entspricht also dem Bestimmtheitsmaß, das den Anteil der erklärten Streuung an der Gesamtstreuung darstellt. Würden alle möglichen Faktoren ermittelt werden, d.h. würde die gesamte Streuung erklärt werden, wäre der Wert des Bestimmtheitsmaßes gleich 1 (Backhaus et al., 2011, S. 75 + 353).

In der Regel ist allerdings zu erwarten, dass eben nicht alle möglichen Faktoren ermittelt werden können, und dass also nicht die gesamte Varianz erklärt werden kann. Ein Grund dafür kann in einer fehlenden Linearität liegen: Mit einer Faktorenanalyse werden nur diejenigen Merkmalsvarianzen erfasst, „die sich aufgrund linearer Beziehungen aus den Faktoren vorhersagen lassen“ (Bortz & Schuster, 2010, S. 394). „Der systematische Varianzanteil eines Items, der nicht durch andere Faktoren erklärt werden kann“ (Bühner, 2011, S. 300), wird als ***Spezifität*** (spezifische Faktoren) bezeichnet; zusammen mit dem potenziellen ***Messfehler*** ergibt sich die Einzigartigkeit bzw. ***Uniqueness*** $(1–h^2)$ eines Items, d.h. „die Varianz, die sich dieses Item mit keinem anderen teilt“ (ebd.).

Demgegenüber geben die ***Kommunalitäten*** h_i^2 an, wie gut ein Item durch die ermittelten *gemeinsamen* Faktoren repräsentiert wird.

3.4.2 Bestimmung der Kommunalitäten

Ein wichtiger Schritt bei der Faktorenanalyse besteht nun darin, eben diese Kommunalitäten zu schätzen. Der Umstand, dass man sie einerseits zur Berechnung der Ladungen benötigt, sie aber andererseits a priori nicht kennt und daher schätzen muss, wird als das ***Kommunalitätenproblem*** bezeichnet (Bühler, 2011, S. 320). Wie erwähnt handelt es sich bei der Kommunalität um den Teil der Gesamtvarianz einer Variablen, der durch

die gemeinsamen Faktoren erklärt werden soll; Der restliche Teil der Gesamtvarianz besteht demzufolge aus einer spezifischen Varianz und einem potenziellen Messfehler (vgl. oben). Entscheidend ist an dieser Stelle, dass die Höhe der Kommunalitäten subjektiv und aus inhaltlichen Überlegungen festzulegen ist. So kann beispielsweise von der Annahme ausgegangen werden, dass die gesamte Varianz der Ausgangsvariablen (100 %) durch die Faktorenanalyse erklärt werden soll, und daher die Kommunalitäten auf 1 gesetzt werden. Genauso kann aber auch vermutet werden, dass ein bestimmter Schätzwert für die Erklärbarkeit der Ausgangsvarianz adäquat sei (z.B. 80%), und daher die Kommunalität auf einen bestimmten anderen Wert (in diesem Fall .8) gesetzt wird (Backhaus et al., 2011, S. 355). Bühner (2011, S. 312) unterscheidet und empfiehlt für die a-priori-Schätzung der Kommunalitäten wie folgt:

- Einsetzen von Einsen: In diesem Fall würde man (vgl. oben) davon ausgehen, „dass die Items messfehlerfrei gemessen werden können, also die gesamte Itemvarianz durch die Faktoren erklärbar ist.“ Dies erscheint unrealistisch.
- Verwendung der höchsten Korrelation einer Variable mit eine anderen (möglichst parallelen, der selben Skala zugeordneten) Variable: Dahinter steht die Idee, dass die Korrelation zweier Variablen (bzw. Items) eine Mindestschätzung der Reliabilität darstellt. Allerdings wird in diesem Fall nur wenig Information berücksichtigt (nämlich nur die zweier Items).
- Verwendung der quadrierten multiplen Korrelation eines Items mit den restlichen Items: Auf diese Art wird mehr Information verwendet, da angenommen wird, „dass die Varianz, die ein Item mit *allen* anderen Items teilt, durch die Faktoren erklärbar ist.“ (Bühner, 2011, S. 312). Dieser Ansatz entspricht auch „der ursprünglichen Annahme, dass die Faktoren die Korrelationen zwischen den Items erklären sollen.“ (ebd.). Es handelt sich dabei ebenso um eine Mindestschätzung der Itemreliabilität, da nicht anzunehmen ist, dass wirklich alle möglichen Items berücksichtigt wurden. – Nachdem diese Schätzmethode unter den angeführten die meiste Information verwendet, solle sie nach Möglichkeit zum Einsatz kommen.

Backhaus et al. (2011, S. 355) weisen darauf hin, dass die Bestimmung der Kommunalitäten nun viel mit der Wahl des Faktorenextraktionsverfahrens zu tun hat:

3.4.3 Methoden

„Die ***Hauptkomponentenanalyse*** geht davon aus, dass die Varianz einer Ausgangsvariablen *vollständig* durch die Extraktion von Faktoren erklärt werden kann, d.h. sie unterstellt, dass *keine Einzelrestvarianz* (= spezifische Varianz + Messfehlervarianz) in den Variablen existiert.“ (Backhaus et al., 2011, S. 356). Demzufolge wird bei der Kommunalitätenschätzung zunächst der Wert 1 vorgegeben. Werden ebensoviele Faktoren extrahiert wie Variablen vorhanden sind, wird eine Kommunalität von 1 auch immer reproduziert werden können. Ergeben sich durch eine geringere Anzahl extrahierter Faktoren Kommunalitäten von kleiner 1, so wird „der ‘nicht erklärte’ Varianzanteil (1 – Kommunalität) jedoch nicht als Einzelrestvarianz, sondern als durch die Faktoren nicht reproduzierter Varianzanteil und damit als (bewusst in Kauf genommener) Informationsverlust deklariert.“ (ebd.). Bühner (2011, S. 314) formuliert, dass sich nach der Komponentenextraktion „die Varianz eines Items in die durch die Komponenten aufgeklärte Varianz ($h^2 = \lambda^2_{i1} + \lambda^2_{i2} + ... + \lambda^2_{iq}$) und die nicht durch die Komponenten aufgeklärte Varianz“ (auch: Einzigartigkeit, $1–h^2$, vgl. oben) teilt. Ziel ist also eine möglichst umfassende Reproduktion der Datenstruktur mit möglichst wenigen Faktoren, es wird nicht zwischen Kommunalitäten und Einzelrestvarianz unterschieden und rein beschreibend vorgegangen (Backhaus et al., 2011, S. 356).

Demgegenüber unterstellt die ***Hauptachsenanalyse***, „dass sich die Varianz einer Variablen immer in die Komponenten Kommunalität und Einzelrestvarianz aufteilt“ (ebd.). Ziel ist es in dem Fall, „lediglich die Varianzen der Variablen *in Höhe der Kommunalitäten* zu erklären“ (ebd.), was bedeutet, dass bei der Kommunalitätenschätzung zunächst immer ein Wert kleiner 1 vorgegeben wird. Wie groß nun dieser Wert ist, kann einerseits aufgrund inhaltlicher Überlegungen festgesetzt oder andererseits durch einen Iterationsprozess sukzessiv geschätzt werden. *Differenzen* zwischen den vorgegebenen und den sich ergebenden Kommunalitätenwerten werden als nicht reproduzierter Varianzanteil und infolge als Informationsverlust betrachtet. Das Ziel bei einer Hauptachsenanalyse ist also eine Erklärung der Varianz der Variablen durch Faktoren, und es wird zwischen Kommunalitäten und Einzelrestvarianz unterschieden (ebd.).

Während bei der Hauptkomponentenanalyse und der Hauptachsenanalyse die Berechnung der Ladungen über einen Kleinste-Quadrate-Ansatz erfolgt („die quadrierten Abweichungen zwischen der beobachteten Korrelationsmatrix und der aus den Ladungen reproduzierten Korrelationsmatrix [sollen] möglichst gering ausfallen bzw. ein Minimum

ergeben“ (Bühner, 2011, S. 315)), besteht auch die Möglichkeit, einen Maximum-Likelihood-Ansatz anzuwenden; es handelt sich dann um eine „***Maximum-Likelihood-Faktorenanalyse (ML)***“ (Bühner, 2011, S. 316ff).

Bühner (2011, S. 318) fasst zusammen, dass sich die Hauptkomponentenanalyse, die Hauptachsenanalyse und die Maximum-Likelihood-Methode kaum unterscheiden, wenn „die Kommunalitäten der Items hoch sind, die Items normalverteilt sind und diese Intervalldatenniveau aufweisen“. Da die Hauptkomponentenmethode nicht zwischen Kommunalitäten und Einzelrestvarianz unterscheidet, stellt sie keine faktorenanalytische Methode im eigentlichen Sinn dar (Bühner, 2011, S. 309 und Backhaus, 2011, S. 356). Vorteil der Hauptkomponentenanalyse ist zwar, dass sie immer zu einer Lösung führt; Üblicherweise sind nach Bühner (2011, S. 318) aber die Hauptachsenanalyse oder die ML-Methode vorzuziehen, wobei diejenige Methode gewählt werden sollte, „die am besten die beobachtete Korrelationsmatrix reproduziert“.

3.4.4 Zahl der Faktoren: Extraktionskriterien

Die Frage nach der Anzahl der Faktoren, die im Rahmen einer Faktorenanalyse herausgearbeitet werden sollen, führt zum ***Extraktionsproblem*** der Faktorenanalyse. Dieses „besteht darin, dass vor der Schätzung der Ladungen und Fehlervarianzen geklärt sein muss, wie viele Faktoren extrahiert werden müssen. Dazu gibt es aber kein allgemeingültiges bzw. anerkanntes Abbruchkriterium“ (Bühner, 2011, S. 320). Grundsätzlich sollte einerseits die inhaltliche Plausibilität der Faktoren überlegt und andererseits berücksichtigt werden, wie gut die Faktorladungen die beobachtete Korrelationsmatrix reproduzieren können. Folgende Kriterien können zur Beurteilung der Anzahl bedeutsamer Faktoren herangezogen werden:

1. Theoriebasierter Zugang
   Verfügt man über theoriebegründete Annahmen über eine den vorliegenden Daten zugrundeliegende Struktur, bietet sich an, zunächst von der demnach postulierten Faktorenanzahl auszugehen und zu prüfen, ob die Items wie angenommen auf den entsprechenden Faktoren laden. Streng genommen müsste laut Bühner (2011, S. 321) dann eigentlich eine Konfirmatorische Faktorenanalyse durchgeführt werden

(vgl. dazu aber auch die Ausführungen zu explorativen bzw. konfirmatorischen Faktorenanalysen zu Beginn des Abschnitts).

2. Minimum-Prozent-Kriterium

   Auch bei diesem Ansatz kommen zunächst inhaltliche Überlegungen zum Tragen: *Vor* der Faktorenanalyse wird festgelegt, wie viel Prozent der Gesamtvarianz die extrahierten Faktoren gemeinsam *mindestens* erklären sollen; es werden infolge so viele Faktoren beibehalten, bis diese mindestens die geforderte Höhe der Gesamtvarianz erklären. Schendera (2010, S. 211) unterstützt die Anwendung dieses Kriteriums: „Erst wenn die Varianz hinreichend erklärt wurde, macht eine Beurteilung der Anzahl von Faktoren Sinn." Als erster Anhaltspunkt für eine „gute" Lösung können an dieser Stelle „75% bei 2 Faktoren- und ca. 85% bei 3-Faktorenlösungen" (Schendera, 2010, S. 185) genannt werden.

3. Techniken auf Basis der Eigenwerte

   Der Eigenwert eines Faktors entspricht bei unkorrelierten Faktoren der „Summe der quadrierten Faktorladungen *eines* Faktors über alle Variablen." (Backhaus et al., 2011, S.359). Er gibt also an, wie viel von der Gesamtvarianz aller Items durch diesen Faktor erklärt wird, d.h. welche Bedeutung bzw. welche Wichtigkeit der Faktor hat (Backhaus et al., 2011, S. 339 + 359, Bortz & Schuster, 2010, S. 393). Bühner (2011, S. 316) unterscheidet genauer den Eigenwert als „relativen Anteil, den ein Faktor an der *gemeinsamen Varianz aller Items"* (bei faktorenanaltischen Methoden) vs. „an der *Gesamtvarianz aller Items"* (die Hauptkomponentenmethode betreffend) erklärt. Bortz & Schuster (2010. S. 393) fassen zusammen, dass „der Eigenwert desjenigen Faktors, der am meisten Varianz erklärt, [...] umso größer [ist], je höher die Variablen miteinander korrelieren", und sie halten fest: „Je höher die Variablen (absolut) miteinander korrelieren, desto weniger Faktoren werden zur Aufklärung der Gesamtvarianz benötigt".

   a. Kaiser-Guttmann-Kriterium

   Ist nun „der Eigenwert eines Faktors größer als 1, klärt ein Faktor mehr Varianz auf, als ein standardisiertes Item besitzt" (Bühner, 2011, S. 321); ist er kleiner als 1, erklärt er weniger Varianz als eine einzelne Variable. Demzufolge ist es sinnvoll, in einer Faktorenanalyse nur Faktoren zu interpretieren, deren *Eigenwerte größer als*

1 sind. Bortz & Schuster (2010, S. 415) weisen in dem Zusammenhang allerdings darauf hin, dass vor allem bei großer Itemanzahl die Anzahl der bedeutsamen Faktoren tendenziell überschätzt wird. Außerdem ist zu beachten, dass es sich bei an einer Stichprobe gewonnenen Eigenwerten lediglich um Schätzungen der wahren Eigenwerte handle, „sodass korrekterweise für jeden Eigenwert ein Konfidenzintervall zu bestimmen“ und dieses bei der Interpretation zu berücksichtigen ist. Bühner (2011, S. 321) berichtet von der Anwendungsempfehlung, der zufolge das Eigenwertkriterium dann angewendet werden sollte, „wenn eine besonders differenzierte Aufgliederung eines Merkmalsbereichs angestrebt wird und besonders reliable Messwerte vorliegen (Kommunalität bzw. Reliabilität nahe eins)“.

b. Scree-Test nach Cattell

Bei dieser sehr veranschaulichenden Methode werden die Eigenwerte der Größe nach in einem Koordinatensystem angeordnet. Anhand des entstehenden „Scree-Plots“ (von engl. „scree“ = „Geröll“) wird dann nach einem bedeutsamen Eigenwertabfall gesucht: Dort, wo die Differenz der Eigenwerte eines und des darauffolgenden Faktors am größten ist, lässt sich graphisch ein „Knick“ im Eigenwerteverlauf nachvollziehen; Faktoren vor bzw. links von diesem Knick werden infolge als bedeutsam angesehen, Faktoren danach bzw. rechts davon als unbedeutsam. Backhaus et al. (2011, S. 359f) geben zu bedenken, dass dieses Verfahren allerdings nicht immer eindeutige Lösungen liefert, da mögliche ähnliche Differenzen der Eigenwerte einen nicht eindeutigen Knick bewirken würden. Nach Bühner (2011, S. 322) hat sich der Scree-Test zwar bewährt, wird wegen seiner Subjektivität aber kritisiert, weshalb im Grunde objektivere Methoden der Faktorenextraktion vorzuziehen sind.

c. Parallelanalyse nach Horn

Dieser Ansatz basiert auf einem Vergleich des Eigenwerteverlaufs „der empirisch ermittelten Korrelationsmatrix mit dem Eigenwerteverlauf der Korrelationen zwischen normalverteilten Zufallsvariablen“ (Bortz & Schuster, 2010, S. 416). Es sollen nur diejenigen Faktoren extrahiert werden, deren beobachtete Eigenwerte größer als die entsprechenden Eigenwerte von Zufallswerten sind: Zu erwarten ist folglich, dass die empirischen Eigenwerte „vor allem bei den ersten Faktoren stark

von den Eigenwerten aus Zufallszahlen“ (Bühner, 2011, S. 323) abweichen. Bei der Parallelanalyse ist unter anderem zu beachten, dass eine große Stichprobe zu einer Extraktion von mehr Faktoren führt als eine kleinere Stichprobengröße, und dass die wirkliche Anzahl der Faktoren in wenigen Fällen überschätzt wird. Eine Anwendung im Rahmen einer Hauptkomponentenanalyse führt in der Regel zu weniger Faktoren als eine Anwendung bei einer Hauptachsenanalyse (wobei die dort zusätzlich ermittelten Faktoren meist nur von geringer Bedeutung sind) (Bühner, 2011, S. 324f).

4. Minimum-Average-Partial-Test (MAP-Test)
   Hier handelt es sich um ein Verfahren, bei dem – nach der Durchführung einer Hauptkomponentenanalyse – schrittweise so lange Komponenten aus der Korrelationsmatrix auspartialisiert werden, „bis keine systematischen Zusammenhänge der Items in der verbleibenden Korrelationsmatrix mehr bestehen“ (Bühner, 2011, S. 328). Wenn dann also die systemische Varianz der Korrelationsmatrix ausgeschöpft ist, steht die Anzahl der (nämlich bis dahin extrahierten) Komponenten fest (Bühner, 2011, S. 325).

5. Modelltest der ML-Faktorenanalyse
   Mit einem Likelihood-Quotienten-Test kann geprüft werden, ob eine extrahierte Faktorenanzahl zur Erklärung der Itemkorrelationen ausreicht. Zu beachten ist nach Bühler (2011, S. 326ff), dass dieser Test stichprobenabhängig ist und bei großen Stichproben sowie auch bei nicht vorliegender multivariater Normalverteilung der Items zu einer Überschätzung der Faktorenanzahl führt. – Ist das Ziel des Tests nun die Bestimmung der Anzahl der Faktoren, muss er für unterschiedliche Faktorenanzahlen wiederholt ausgeführt werden, bis sein Ergebnis nicht mehr signifikant ausfällt.

3.4.5 Interpretierbarkeit: Rotation

Im Anschluss an die Bestimmung der Anzahl der Faktoren, die zu diesem Zeitpunkt „indeterminiert, also zwar numerisch, aber nicht inhaltlich bestimmt“ (Schendera, 2010, S. 202) sind, sollen selbige nun (besser) interpretierbar gemacht werden.

Das ***Rotationsproblem der Faktorenanalyse*** besteht darin, dass es angesichts der Fundamentalgleichung der Faktorenanalyse $R = L \cdot L' + V$ (vgl. oben) viele verschiedene Möglichkeiten gibt, eine Korrelationsmatrix zu reproduzieren, weil die Ladungsmatrizen nicht eindeutig bestimmt sind (*Problem der Identifizierbarkeit*). Dies bedeutet, dass man viele qualitativ unterschiedlich zu interpretierende Faktoren findet, die alle das Zustandekommen der beobachteten Korrelationsmatrix bedingt haben könnten. „Um eindeutig zu bestimmende Faktorladungen zu erhalten, müssen daher gewisse Festlegungen getroffen werden“ (Bühner, 2011, S. 330).

Durch die sukzessive Aufklärung maximaler Varianzen entsteht im Rahmen einer Faktorenanalyse typischerweise die schwierig zu interpretierende Situation, dass auf dem ersten Faktor viele Items hoch und auf den übrigen Faktoren viele Items in mittelmäßigem oder niedrigem Ausmaß laden. Das ***Einfachstrukturkriterium*** („*simple structure*“) nach Thurstone fordert demgegenüber, „dass pro Faktor einige Variablen möglichst hoch und andere möglichst niedrig laden, was mit der Forderung gleichzusetzen ist, dass die Varianz der Faktorladungen pro Faktor möglichst groß sein soll“ (Bortz & Schuster, 2010, S. 419). Außerdem sollen demnach auf verschiedenen Faktoren verschiedene Items möglichst hoch laden. Weil ja im Rahmen psychologischer Tests Testergebnisse inhaltlich nach Möglichkeit auf einen einzelnen Faktor zurückgeführt werden sollen, so hält Bühner (2011, S. 329) fest, sollten dementsprechend also auf methodischer Ebene die Unterschiede in der Itembeantwortung möglichst nur durch einen Faktor erklärt werden.

Ausgehend von den vorläufigen Faktorladungen wird nun geprüft, ob und mit welcher nonsingulären linearen *Transformation* eine Repräsentation der Faktorladungen erreicht werden kann, bei der eben möglichst nur ein Faktor (und nicht mehrere) als Erklärung für Beantwortungsunterschiede herangezogen werden muss (und so also das Kriterium der Einfachstruktur erfüllt werden kann). Diese Transformationen werden als ***Rotation*** bezeichnet.

Differenziert werden bei Bortz & Schuster (2010, S. 418ff) folgende Rotationstechniken und -arten:

1. Graphische Rotation

   Die Items werden in das durch (zunächst) zwei Faktoren „aufgespannte“ Koordinatensystem eingetragen, wobei die Koordinaten den Ladungen der Items auf den jeweiligen Faktoren entsprechen. Wenn die Itempunkte nicht direkt auf bzw. nahe an einer der Faktoren (Achsen) liegen, nimmt man eine *Drehung des Achsensystems in*

seinem Ursprung vor, mit Augenmerk darauf, dass nun möglichst viele der Punkte auf der Achse liegen (d.h. durch die Achse repräsentiert werden). Es wird also „versucht, den Abstand zwischen allen Items und der jeweiligen Achse zu minimieren“ (Bühner, 2011, S. 339); Items, die zuvor von beiden Faktoren mittelmäßig repräsentiert waren, sollen nun eindeutig einem der Faktoren zugeordnet werden können. Im Anschluss an eine erste Rotation werden für weitere Faktoren(paare) weitere Rotationen durchgeführt (Bortz & Schuster, 2010, S. 419).

Bei größeren Faktoren- bzw. Itemanzahlen erscheint die graphische Rotation sehr umständlich und ungeeignet.

2. Analytische Rotation

Als Alternative zur potenziell mühsamen graphischen Rotation kann eine Rotation auch rechnerisch bewerkstelligt werden. Bortz & Schuster (2010, S. 419), Bühner (2011, S. 336) sowie Backhaus et al. (2011, S. 363) betonen an dieser Stelle die Bedeutung des ***Varimax-Kriteriums*** nach Kaiser, das die Erreichung einer möglichst guten *Einfachstruktur* (siehe oben) auf analytischem Wege ermöglichen soll. Dabei „werden die Faktoren so rotiert, dass die *Varianz der quadrierten Ladungen* [der Items] pro Faktor *maximiert* wird“ (Bortz & Schuster, 2010, S.420); Ladungen werden entweder unbedeutender oder extremer. Es handelt sich bei diesem Verfahren um eine ***orthogonale Rotationstechnik***, was bedeutet, dass die *Unabhängigkeit* der Faktoren erhalten bleibt.

Im Gegensatz dazu erhält man bei einer ***obliquen Rotation*** *korrelierte Faktoren*. Faktorenstrukturen, die inhaltlich begründet Korrelationen zwischen den Faktoren zulassen, können in der Regel besser interpretiert werden. Gleichzeitig ist zu beachten, dass diese Faktoren eben aufgrund ihrer Interkorrelationen zum Teil redundante Informationen enthalten, womit das ursprüngliche Ziel der Datenreduktion hintangestellt wird (Bortz & Schuster, 2010, S. 418). Bühner (2011, S. 338) empfiehlt in diesem Zusammenhang die ***Promax-Rotation***, in deren Rahmen ursprünglich orthogonale Faktoren im Winkel verändert werden, sodass sie korrelieren können. Die Ladungen werden demnach mit den Exponenten 2, 4 oder 6 potenziert, wodurch sowohl geringe als auch hohe Ladungen reduziert werden. Der Vorteil dabei ist, „dass moderate oder kleine Ladungen fast null werden, hohe Ladungen aber nur geringfügig reduziert werden.“ (Bühner, 2011, S. 338).

3. Kriteriumsrotation

   Diese Methode ermöglicht es, zwei oder mehrere Faktorstrukturen miteinander zu vergleichen. Eine gegebene Vergleichsstruktur soll über eine zu ermittelnde Transformationsmatrix so rotiert werden, dass ihre Ähnlichkeit mit einer vorgegebenen (empirisch oder theoretisch begründeten) Zielstruktur maximal wird (Bortz & Schuster, 2010, S. 424). – Auf diesen Ansatz soll im Rahmen der vorliegenden Arbeit nicht näher eingegangen werden.

Zusammenfassend ist zu sagen, dass unabhängig davon, ob Faktoren nach der Extraktion noch rotiert werden, „sich der Anteil der aufgeklärten Varianz *aller* Faktoren an der Gesamtvarianz *nicht* [ändert]. Es ändert sich jedoch die Verteilung der aufgeklärten Varianz über die Faktoren hinweg. Sie verteilt sich nun ‘gleichmäßiger’ als vorher“ (Bühner, 2011, S. 316, siehe dazu auch Bortz & Schuster, 2010, S. 405 + 422). Es ist daher sinnvoll, die aufgeklärte Varianz erst *nach* der Rotation anzugeben.

Darüber hinaus weist Schendera (2010, S. 204 + 209) darauf hin, dass eine Rotation eine Faktorlösung nicht notwendigerweise immer verbessert, sondern sie im Gegenteil sogar verschlechtern kann; diese Gefahr besteht besonders im Falle zuwenig extrahierter Faktoren („*underextraction*“).

3.4.6 Faktorwerte (auch: Faktorenwerte)

„Ein Faktorwert ist ein gewichteter Wert, der den Ausprägungsgrad einer Person auf einem Faktor darstellt.“ (Bühner, 2011, S. 340). Die Gewichtung wird dabei auf Basis der Ladungen umgesetzt: Items mit höheren Faktorladungen haben mehr Gewicht, da die Unterschiede in ihrer Beantwortung durch das Konstrukt besser vorhergesagt werden können. „Daher tragen solche Items zur Feststellung der Ausprägung einer Person auf dem Konstrukt in höherem Ausmaß als Items mit geringeren Ladungen bei“ (ebd.). Berücksichtigt werden dabei die Werte einer Person bei *allen* Items, die auf dem Faktor laden (d.h. auch denjenigen Items, die nur geringe Ladungen auf dem betreffenden Faktor aufweisen, also inhaltlich eigentlich wenig mit ihm zu tun haben).

Im Vergleich dazu wird bei einem *Summenwert* jedes Item gleich gewichtet, und darüber hinaus werden lediglich diejenigen Items berücksichtigt, die dem entsprechenden Faktor zugeordnet werden.

Bühner (2011, S. 341f) warnt an dieser Stelle vor einer unreflektierten Verwendung von Faktorwerten: So komme es bei kleinen Stichproben zu ungenauen Schätzungen der Faktorladungen; zudem könnten sich bei obliquen Rotationen die Faktorkorrelationen teils erheblich unterscheiden, was für die Interpretation nicht unerheblich ist: Unterschiedliche Rotationen führen zu unterschiedlichen Faktoren führen zu unterschiedlichen Faktorwerten!

Summenwerte ihrerseits ignorieren die Idee der unterschiedlichen Beiträge jedes Items zum Konstrukt, d.h. sie verzichten auf einen Teil der Information.

Mit der also gebotenen Bedachtsamkeit können Faktorwerte nach Bühner (ebd.) nun wie folgt ermittelt werden:

- Bei der *Regressionsmethode* wird ein Kleinste-Quadrate-Schätzer verwendet. Dieser besitzt die geringste Fehlervarianz und ist somit am effizientesten; Die Schätzung variiert je nach gezogener Stichprobe am geringsten.
- Ein Maximum-Likelihood-Schätzer kommt bei der *Bartlettmethode* zum Einsatz. Man erhält konsistente und erwartungstreue Schätzungen der Faktorwerte, die mit zunehmendem Stichprobenumfang gegen die wahren Populationsparameter streben. Es wird von einer Normalverteilung für jedes Item ausgegangen.
- Mit der *Anderson-Rubin-Methode*, einer Modifikation des Bartlett-Tests (Field, 2009, S. 635), werden stets unkorrelierte Faktorwerte geschätzt; die Einsatzmöglichkeiten für diese Technik sind entsprechend eingeschränkt.

Grundsätzlich lassen sich die Faktorwerte im Rahmen einer Hauptkomponentenanalyse *eindeutig* und *exakt* berechnen, alle angeführten Methoden führen zum gleichen Ergebnis. Bei einer Hauptachsen- oder ML-Faktorenanalyse dagegen können die Faktorwerte zwar *exakt*, aber (aufgrund des enthaltenen Fehlerterms) *nicht eindeutig* bestimmt werden. Bühner (2011, S. 342) empfiehlt daher zusammenfassend, Faktorwerte wenn möglich mittels Hauptkomponentenanalyse zu ermitteln; für ML- oder Hauptachsenanalysen ist demnach die Bartlettmethode vorzuziehen.

4. Zum Umgang mit Daten

Bevor nun schließlich zum empirischen Teil der vorliegenden Arbeit übergegangen werden soll, an dieser Stelle noch einige Ausführungen zur Problematik unvollständiger Datensätze.

Schendera (2007, S. 121f.) unterscheidet folgende potentielle Ursachen für fehlende Werte in erhobenen Daten: So können Datenlücken sowohl vor oder während (organisations-, prozess- oder technisch bedingt) des Erhebungsvorgangs als auch danach (beispielsweise durch unprofessionelles Datenmanagement, versehentliches Überschreiben, Hardwareprobleme usw.) entstehen. Des Weiteren besteht die Möglichkeit, dass Daten nicht geliefert werden, weil seitens der Informationsbereitsteller Zweifel an der Relevanz der Umfrage bzw. Studie oder sogar gegenläufige Interessen bestehen. Bei Längsschnittstudien können sowohl Änderungen des Gegenstands (institutionelle Veränderungen wie Unternehmensumstrukturierungen oder individuelle Veränderungen wie Berufswechsel usw.) als auch Nichterreichbarkeit (Umzug, Mortalität) Ursache für fehlende Daten sein. Schließlich kann es bei der Datenerhebung auch zu individuellen bzw. punktuellen Problemen kommen: Beispiele hierfür wären ein versehentliches Überspringen einzelner Items, sachliches oder sprachliches Nichtverstehen von Fragen oder auch eine Antwortverweigerung aufgrund von Aversion, Überforderung o.ä.

Der Umstand, dass die Psychologische Diagnostik angesichts dieser vielfältigen und großteils nachvollziehbaren Möglichkeiten der Datenunvollständigkeit an ***Missings*** gewöhnt ist, entbindet allerdings nicht grundsätzlich von Überlegungen, was die Konsequenzen von Lücken in den Daten sein können. Schendera (2007, S. 126ff.) weist diesbezüglich unter anderem darauf hin, dass entstehende Verzerrungen (*Bias*) ein massives Problem darstellen können. Eine weitere Herausforderung ist die Verringerung des *N*, was für viele Verfahren einen Verstoß gegen inferenzstatistische Verfahrensvoraussetzungen bedeuten kann; so gilt beispielsweise für Faktorenanalysen, dass die Anzahl der Beobachtungen deutlich größer als die Anzahl der Variablen sein soll. Darüber hinaus „basieren Analyseergebnisse unterschiedlicher (v.a.) multivariater Verfahren immer auch auf verschiedenen Fallzahlen. Die erzielten Ergebnisse können nur ausnahmsweise direkt miteinander verglichen werden, da die konkret zugrundeliegende Datenbasis genau betrachtet jeweils eine andere ist“ (ebd., S. 130f.).

Für den Umgang mit Missings ist es nun zunächst einmal sinnvoll, zwischen den verschiedenen Ursachen für vorliegende Datenlücken zu differenzieren und dies auch über die Datenkodierung nachvollziehbar zu dokumentieren (ebd., S. 135). In weiterer Folge ist zu entscheiden, ob *mit* den Missings weitergerechnet werden soll (was sich in der Regel aufwendig gestaltet und umfassendes Verständnis der Materie voraussetzt), ob Missings *ersetzt* oder ob sie (respektive Variablen oder Personen mit Missings) *gelöscht* werden sollen.

> Rost (2004, S. 365) weist in diesem Zusammenhang darauf hin, dass das Löschen von *Items* prinzipiell dem Entfernen von *Personen* vorzuziehen ist, da Items ja „von Menschenhand gemacht und [...] mit allen Fehlern behaftet sein" könnten. Eine Eliminierung von Variablen kann daher gerechtfertigt sein, wohingegen das Entfernen „unpassender" Personen eher als Datenmanipulation gewertet werden könnte. In Ausnahmefällen (wenn z.B. offensichtlich fehlende Testmotivation, verfälschende Absicht, mangelnde Konzentration und dgl. zu de facto unbrauchbaren Testprotokollen geführt haben), kann aber ein Ausschließen einzelner Testbögen von der Analyse durchaus vertretbar sein.

Für das Rekonstruieren und infolge *Ersetzen* von fehlenden Werten stehen nach Schendera (2007, S. 141) nun verschiedene Möglichkeiten der Imputation zur Verfügung: die Cold deck-Imputation (Ersetzen der Missings durch eine Konstante), zufallsbasiertes, logisches oder stereotypengeleitetes Vorgehen, die univariate Schätzung, die Hot deck-Imputation (Ergänzen der lückenhaften Datenreihen über ähnliche, vollständige Datenreihen) und die multivariate Schätzung; auf eine genauere Beschreibung dieser Verfahren wird im Rahmen der vorliegenden Arbeit aus Komplexitätsgründen verzichtet.

Das *Entfernen* von Variablen oder Testpersonen mit fehlenden Werten betreffend wird zwischen dem *paarweisen* und dem *listenweisen Löschen* unterschieden:

Beim *listenweisen* Löschen werden „alle Fälle, die einen Missing in einer Variablen aufweisen [...] als komplette Datenzeile aus der Analyse ausgeschlossen" (Schendera, 2007, S. 136); das bedeutet, dass auch Variablen, in denen eigentlich Werte vorliegen, betroffen sind. Der Vorteil dieses Vorgehens liegt in seiner Einfachheit: alle Variablen weisen danach dasselbe *N* auf, was die Vergleichbarkeit separater, auch multivariater

Analysen erleichtert (vgl. oben). Der Nachteil besteht im kumulativen Informationsverlust und einem umso wahrscheinlicheren Bias, je mehr Fälle ausgeschlossen werden.

Beim *paarweisen* Löschen „wird eine Analyse nur mit den bei jeder Variablen jeweils verfügbaren Werten vorgenommen“ (ebd.). Bei bivariaten Analysen wie z.B. Korrelationen werden dann „Wertepaare, die Missings einschließen, aus der Analyse ausgeschlossen [..], die Ergebnisse [basieren] also nur noch auf dem paarweise gemeinsamen *N*“ (ebd.). Da je nach Analyse ein anderes paarweises *N* zur Anwendung kommen kann, ist die Vergleichbarkeit separater bivariater Analysen erschwert; bei einer größeren Variablenanzahl werden dazu noch unter Umständen multivariate Interaktionen übergangen. Als Vorteil des paarweisen Löschens kann demgegenüber betrachtet werden, dass der unmittelbare Informationsverlust durch das Entfernen vorhandener Daten minimiert werden kann.

Grundsätzlich ist nach Schendera (2007, S. 141) das Ersetzen von Missings dem Löschen vorzuziehen, da es so zu keinem Verwerfen von vorhandenen Informationen kommt und man mit einem vollständigen Datensatz weiterrechnen kann. Voraussetzung dafür ist allerdings, dass die Rekonstruktion der fehlenden Angaben inhaltlich plausibel ist.

Das Entstehen von Missings „bereits im Ansatz so weit wie möglich zu vermeiden“ (ebd., S. 160) sollte angesichts der angeführten Einschränkungen jedenfalls erstrebenswertes Ziel sein.

II. Empirischer Teil

5. Ziele

Die vorliegende Arbeit hat als Teil des Entwicklungsprozesses des *Entwicklungsscreenings für Hortkinder im Alter von 6 bis 10 Jahren (ESH 6-10)* zum Ziel, nach einer erfolgten ersten Revision des Verfahrens den aktuellen Stand der Verfahrenseigenschaften zu beurteilen sowie darauf basierend Aussagen zur möglichen Weiterentwicklung zu treffen.
Die Aufgabe besteht einerseits darin, die Daten einer zweiten Stichprobe zu erfassen und hinsichtlich ihrer Testkennwerte zu analysieren; Andererseits sollen diese Daten mit der bereits bearbeiteten Stichprobe zur ersten Version des Verfahrens (vgl. dazu Kunst (2014), Matschiner (2014) und Neugschwentner (2015)) verknüpft und infolge ebenfalls ausgewertet werden. Darüber hinaus soll für den entstehenden Gesamtdatensatz eine Faktorenanalyse gerechnet werden, um einen Vergleich der dem Verfahren zugrundegelegten inhaltlichen Dimensionen mit der testanalytisch erhobenen Struktur der Daten zu ermöglichen.

Insgesamt sollen die beschriebenen Schritte dazu beitragen, dass schließlich ein reliables und valides Entwicklungsscreeningsverfahren zu einem breiten Einsatz bei 6- bis 10-jährigen Hortkindern kommen kann.

6. Methode

6.1 Ablauf

Anschließend an die erste Überarbeitung des ESH 6-10 wurden von Hasenhindl (in Vorbereitung) und Kremser (in Vorbereitung) die nunmehr neuen Erhebungsbögen an die teilnehmenden Horte der Organisation „Kinder in Wien“ (KiWi) verteilt.

Nachdem die dort arbeitenden PädagogInnen zwischen November 2014 und Februar 2015 ihre Einschätzungen für jeweils einige ihrer Schützlinge vorgenommen hatten, wurden die ausgefüllten Bögen von Hasenhindl (in Vorbereitung) und Kremser (in Vorbereitung) wieder eingesammelt und der Autorin zur Erfassung und weiteren Auswertung der Daten vorgelegt. Zur Anwendung gelangte in der Folge *SPSS* (Brosius, 2011) in der Version *22*, es wurden Reliabilitätsanalysen sowie eine explorative Faktorenanalyse durchgeführt.

6.2 Instrument

6.2.1 Allgemeines

Beim ESH 6-10 handelt es sich um ein Einschätzverfahren, mit dessen Hilfe Hortkinder im Alter zwischen sechs und zehn Jahren hinsichtlich möglicher Entwicklungsstörungen beurteilt werden können. Von Interesse ist bei diesem *Screening-Verfahren* die grundlegende Frage, welche Kinder einer weiterführenden entwicklungspsychologisch-diagnostischen Untersuchung unterzogen werden sollten und welche Kinder demgegenüber als nicht gefährdet bzw. unproblematisch betrachtet werden dürfen. Das Ziel ist es, möglichst alle Kinder zu identifizieren, bei denen sich entwicklungspsychologische Defizite abzeichnen könnten; der Fokus liegt also auf einer *hohen Sensitivität* des Verfahrens. Als Beurteiler fungieren dabei diplomierte HortpädagogInnen, die das jeweilige Kind gut kennen, wodurch die Einschätzungen nicht auf einmaliger Beobachtung basieren und außerdem das Ausfüllen des Bogens auch unabhängig von der Anwesenheit des Kindes erfolgen kann. Zeitlich soll das Verfahren möglichst *wenig Ressourcen* beanspruchen; veranschlagt ist eine Ausfülldauer von etwa 10 bis 15 Minuten.

Das ESH 6-10 umfasst in seiner aktuellen Form insgesamt 74 Items. Davon verfügen 67 über ein 5-stufiges Antwortformat, 7 Fragen sind dichotom gestaltet. Die Beurteilungsmöglichkeiten erstrecken sich von „*sehr selten*“ über „*selten*“, „*manchmal*“, „*oft*“ bis „*sehr oft*“ bzw. „*ja*“ oder „*nein*“. Zusätzlich wird bei jedem Item die Kategorie „*nicht beurteilbar*“ angeboten.

Neben den Itembeurteilungen werden beim ESH 6-10 Geburtsdatum, Alter, Geschlecht und Schulstufe des Kindes erfasst; darüber hinaus seine Gruppen- bzw. Hortzugehörigkeit und der Name der ausfüllenden Person sowie das Datum der Beurteilung. Weiters werden bei der aktuellen Verfahrensversion erstmals zusätzlich auch die Erstsprache und – für den Fall, dass es sich dabei *nicht* um Deutsch handelt – eine Einschätzung der Deutschkenntnisse des Kindes (auf einer Notenskala von ein bis fünf) erhoben.

Am Beginn des Bogens werden die PädagogInnen im Rahmen einer kurzen schriftlichen Einleitung direkt angesprochen und mit den Antwortkategorien vertraut gemacht; am Ende des Einschätzbogens finden sie noch Raum für eigene Anmerkungen.

Vor der nun folgenden Beschreibung der Items und ihrer Zusammenstellung zu den Skalen ein Hinweis zur Orientierung:

Nach der ersten Erhebung und der im Anschluss stattgefundenen Itemselektion wurden die Items des ESH 6-10 für die zweite Vorgabe neu durchnummeriert; im Sinne einer einfacheren Vergleichbarkeit der Testanalysen auf Itemebene werden im Rahmen der vorliegenden Arbeit allerdings die *ursprünglichen* Nummern verwendet. Demgegenüber werden zwischenzeitlich erfolgte Adaptionen der Formulierung *übernommen*, sofern das Item in seiner inhaltlichen Dimension nicht verändert worden ist. – Zu Details und Hintergründen der Überarbeitungen vgl. Hasenhindl (in Vorbereitung) und Kremser (in Vorbereitung). Eventuelle in der Kategorie *Notizen und Anmerkungen* vorhandene Rückmeldungen der PädagogInnen (vgl. Anhang B) können sich freilich nur auf die ihnen vorliegende, also die *veränderte* Nummerierung beziehen.

6.2.2 Items und Skalen

Zu Konzeption und Entwicklung der Skalen zum Entwicklungsbereich *Arbeitshaltungen* vgl. Matschiner (2015), zu den Bereichen *Persönlichkeit* und *Motivation* vgl. Neugschwentner (2014) sowie zu den Bereichen *Sprache* und *Soziale Interaktion* vgl. Kunst (2015). – Die einzelnen Skalen und ihre Items stellen sich in ihrer aktuellen Version nun wie folgt dar:

- **Entwicklungsbereich Arbeitshaltungen**

  - Skala Aufmerksamkeit:

Item 1	Kann sich auf die Hausaufgaben gut konzentrieren
Item 2	Arbeitet genau und sorgfältig
Item 3	Macht Flüchtigkeitsfehler
Item 6	Verliert oder vergisst Schulunterlagen oder Stifte
Item 79	Ist unruhig, zappelig, kann nicht lange stillsitzen

  - Skala Exekutivfunktionen:

Item 7	Beginnt mit dem Erledigen der Aufgaben von alleine

Item 8	Plant im Vorfeld die Durchführung der Aufgaben
Item 9	Teilt sich die Erledigung ihrer/seiner Aufgaben ein
Item 12	Wird mit den Aufgaben in der vorgenommenen Zeit fertig

- Skala Ablenkbarkeit:

Item 13	Lässt sich beim Erledigen der Aufgaben leicht ablenken
Item 14	Wechselt von einer Tätigkeit zu einer neuen, ohne die erste beendet zu haben (z. B. ein Spiel)

- Skala Anstrengungsbereitschaft:

Item 16	Möchte den Arbeitsaufwand möglichst gering halten
Item 17	Löst gerne schwierige Fragen oder Aufgaben
Item 18	Erledigt die Aufgabe erneut, nachdem auf Fehler aufmerksam gemacht wurde
Item 19	Zeigt Durchhaltevermögen, bleibt auch dran, wenn etwas schwierig wird

- Skala Ausdauer:

Item 24	Bleibt auch bei längeren Aufgaben daran sitzen
Item 25	Erledigt ihre/seine Aufgabe auf einmal
Item 26	Wird mit den Aufgaben fertig

- Skala Selbstständigkeit:

Item27	Weiß darüber Bescheid, was in der Schule aufgegeben wurde
Item28	Weiß über den Tagesablauf Bescheid (wann evt. Kurse sind, wann sie/er abgeholt wird etc.)
Item29	Teilt sich selbst ein, wann die Hausaufgaben gemacht werden
Item30	Benötigt Hilfe bei Freizeitaktivitäten (Basteln, Spielanleitung, etc.)
Item76	Findet von selbst Beschäftigung

- **Entwicklungsbereich Persönlichkeit**

  - Skala Internalisierende Störungen:

Item 35	Zieht sich zurück, wenn andere Kinder auf sie/ihn zugehen
Item 84	Ist ein/e Einzelgänger/in, spielt lieber alleine
Item 86	Hat viele Sorgen, erscheint häufig bedrückt/unglücklich, weint häufig
Item 88	Hat übermäßige Ängste, fürchtet sich schnell
Item 87	Hat wenigstens eine gute Freundin oder einen guten Freund

  - Skala Externalisierende Störungen:

Item 40	Behandelt andere Kinder schlecht, hänselt oder mobbt andere Kinder
Item 51	Schwindelt absichtlich, um etwas zu bekommen oder zu bewirken
Item 79	Ist unruhig, zappelig, kann nicht lange stillsitzen
Item 80	Neigt zu Wutanfällen
Item 81	Lügt, stiehlt, macht Sachen kaputt
Item 82	Entschuldigt sich, wenn sie/er sich schlecht benommen hat
Item 83	Ist impulsiv, handelt ohne zu überlegen

- **Entwicklungsbereich Motivation**

  - Skala Leistungsmotivation:

Item 16	Möchte den Arbeitsaufwand möglichst gering halten
Item 17	Löst gerne schwierige Fragen oder Aufgaben
Item 18	Erledigt die Aufgabe erneut, nachdem auf Fehler aufmerksam gemacht wurde
Item 19	Zeigt Durchhaltevermögen, bleibt auch dran, wenn etwas schwierig wird
Item 23	Investiert ausreichend Zeit oder Anstrengungen in die Erledigung der Hausaufgaben
Item 67	Zeigt Freude beim Erledigen der Hausaufgaben
Item 68	Ist ruhig und fokussiert bei den Hausaufgaben

Item 69	Sucht sich Aufgaben/Spiele, die zu einfach sind
Item 71	Lässt die eigene Hausaufgabe nur ungern kontrollieren
Item 72	Sucht sich Aufgaben/Spiele, die zu schwierig sind
Item 73	Macht sich übertriebene Vorwürfe, wenn ihr/ihm etwas nicht gelungen ist
Item 74	Freut sich ganz offensichtlich, wenn sie/er gelobt wird
Item 75	Schiebt die Schuld bei schlechten Leistungen auf andere
Item 76	Findet von selbst Beschäftigung

- **Entwicklungsbereich Sprache**

  - Skala Grammatik und Schriftspracherwerb:

Item 57	Die Grammatik ist fehlerhaft in Bezug auf:
a	verwechselt Artikel, Pronomen oder Präpositionen (z.B. "das Stift"; "Ich gehe zu Hause")
b	macht Fehler bei Wortendungen (z.B. "Meine Freunden warten schon auf mich")
c	die Wortstellung ist fehlerhaft (z.B. "Essen fertig das ist")
Item 58	Das Schreiben der (Deutsch-) Hausübung gelingt ähnlich gut wie bei Klassenkolleginnen und Klassenkollegen in Bezug auf:
a	Wortschatz
b	Rechtschreibung
c	Dauer der Hausübung

  - Skala Entwicklung der Aussprache:

Item 59	Hat Probleme mit der Sprachflüssigkeit (z. B. stottert, verhaspelt sich, Worte bleiben stecken, verwendet Füllworte wie “äh” zwischen Worten und Sätzen
Item 60	Die Aussprache ist korrekt, dem Dialekt entsprechend
Item 61	Hat Schwierigkeiten bei der Aussprache einzelner Laute (z. B.: verwechselt “s” mit “sch”, lässt Buchstaben aus)
Item 63	Betont Sätze richtig, spricht nicht monoton

- Skala Entwicklung der Sprachpragmatik:

Item 48	Kann ihre/seine Bedürfnisse ausdrücken
Item 64	Beherrscht Höflichkeitsformen und Rituale (z. B.: Bitte, Danke)
Item 65	Grüßt und verabschiedet sich in angemessener Form
Item 66	Kann bei einem Thema bleiben, ohne abzuschweifen, und erzählt dabei die relevanten Dinge
Item 82	Entschuldigt sich, wenn sie/er sich schlecht benommen hat

- **Entwicklungsbereich Soziale Interaktion**

- Skala Anpassung an Gruppenregeln:

Item 51	Schwindelt absichtlich, um etwas zu bekommen oder zu bewirken
Item 52	Tut, worum sie/er gebeten wird
Item 53	Hört auf die Pädagoginnen und Pädagogen
Item 55	Bricht Regeln, wenn sie/er glaubt, nicht gesehen zu werden
Item 56	Hält sich von sich aus an (für sie/ihn verständliche) Regeln

- Skala Soziale Interaktion:

Item 31	Spielt mit anderen Kindern, wenn die Freunde nicht da sind
Item 32	In der Freundschaft steht vor allem das gemeinsame Spielen im Vordergrund
Item 33	In der Freundschaft spielen Gemeinsamkeiten und Reden eine wichtige Rolle
Item 35	Zieht sich zurück, wenn andere Kinder auf sie/ihn zugehen
Item 36	Wird von anderen Kindern ignoriert
Item 37	Wird von anderen Kindern freundlich/gut behandelt
Item 38	Gibt und hilft anderen Kindern, worum sie/er gebeten wird
Item 39	Wird von anderen Kindern schlecht behandelt/gehänselt oder gemobbt
Item 40	Behandelt andere Kinder schlecht, hänselt oder mobbt andere Kinder
Item 41	Nimmt an sportlichen Freizeitaktivitäten in der Gruppe teil
Item 43	Bevorzugt ältere Kinder als Spielpartner
Item 44	Bevorzugt jüngere Kinder als Spielpartner

Item 47	Sucht die Aufmerksamkeit der Pädagogin/des Pädagogen, wenn diese/dieser Zeit mit anderen Kindern verbringt
Item 50	Weicht der Pädagogin/dem Pädagogen aus, vermeidet Kontakt
Item 84	Ist ein/e Einzelgänger/in, spielt lieber alleine
Item 85	Ist in eine Gruppe von Kindern, eine "Clique" integriert
Item 87	Hat wenigstens eine gute Freundin oder einen guten Freund

6.3 Stichprobe

6.3.1 Aktuelle Erhebung

Von den zuletzt ausgegebenen Einschätzbögen wurden 150 Stück retourniert und gelangten zur Auswertung. Zwei der Testbögen waren allerdings zum Teil (Tp 122) bzw. komplett (Tp 150) unausgefüllt, weshalb sie aus den folgenden Analysen ausgeschlossen wurden. In einem weiteren Fall (Tp 43) fehlten die Angaben zu sowohl Geburtsdatum als auch Alter; nachdem als Schulstufe „Vorschule" eingetragen war, konnte nicht als sicher erachtet werden, dass das betreffende Kind das geforderte Mindestalter von sechs Jahren bereits erreicht hatte, weshalb auch dieser Bogen ausgeschlossen wurde. Die Stichprobe der aktuellen, zweiten Erhebung umfasst also 147 Personen.

Mit 69 Mädchen und 77 Buben (bei einer fehlenden Angabe) stellt sich das Geschlechterverhältnis in der Stichprobe relativ ausgewogen dar. Die Altersbereiche der 7- und 8-Jährigen sind besonders stark vertreten (32,7% bzw. 29,9%), bei den 10-Jährigen finden sich dagegen nur fünf Kinder (3,4%). In Tabelle 2 ist die Verteilung ersichtlich:

Verteilung der aktuellen Stichprobe

		Alter					
		6	7	8	9	10	Gesamtsumme
Geschlecht	weiblich	7	27	20	11	4	69
	männlich	20	21	24	11	1	77
Gesamtsumme		27	48	44	22	5	146

Tabelle 2: Verteilung der aktuellen Stichprobe

Betreffend die erstmals erfassten Sprachkenntnisse der Mädchen und Buben wurde für 98 Kinder (66,7%) Deutsch als Erstsprache angegeben. Bei den restlichen Kindern wurden 20 verschiedene andere Sprachen bzw. Sprachkombinationen (zweisprachige Varianten) genannt, wobei die größte Gruppe auf Türkisch entfiel (neun Kinder bzw. 6,1%), gefolgt von Polnisch und Russisch (jeweils fünf Kinder bzw. 3,4%). Die komplette Auflistung findet sich in Anhang A. Die Deutschkenntnisse der Mädchen und Buben mit nicht deutscher Erstsprache wurden mehrheitlich zwischen sehr gut und befriedigend eingeschätzt.

6.3.2 Gesamtstichprobe

Die Gesamtstichprobe setzt sich zusammen aus den zwischen Februar und März 2014 nach der ersten Version des ESH 6-10 beurteilten Kindern und den Mädchen und Buben der oben beschriebenen zweiten Stichprobe aus dem Zeitraum von November 2014 bis Februar 2015. Aus der ersten Stichprobe hatten aufgrund von Unvollständigkeit drei Einschätzbögen entfernt werden müssen, darüber hinaus konnte ein Datensatz aufgrund des Alters des Kindes (elf Jahre) nicht miteinbezogen werden. Mit den verbleibenden 160 Bögen resultiert damit eine Gesamtstichprobe von 307 Kindern.

Auch in diesem Fall liegt ein ausgewogenes Geschlechterverhältnis vor: 150 Mädchen stehen 155 Buben gegenüber (bei nunmehr zwei fehlenden Angaben). Analog zum oben beschriebenen Teil der Stichprobe sind auch hier wieder die 7- (32,6%) und 8-Jährigen (27,7%) am stärksten vertreten. Tabelle 3 beschreibt die Verteilung der Gesamtstichprobe.

Verteilung der Gesamtstichprobe

		Alter					
		6,00	7,00	8,00	9,00	10,00	Gesamtsumme
Geschlecht	weiblich	20	52	40	25	13	150
	männlich	27	47	45	28	8	155
Gesamtsumme		47	99	85	53	21	305

Tabelle 3: Verteilung der Gesamtstichprobe

7. Ergebnisse

Vor den weiterführenden Auswertungsschritten wurden zunächst die Items der neuen Stichprobe entsprechend ihrer inhaltlichen Ausrichtung im gegebenen Fall umgepolt.

Parallel zur Analyse der zweiten Erhebung (N = 147) wurde danach, wie angesprochen, ein Gesamtdatensatz (N = 307) gebildet und dieser separat ebenfalls einer Testanalyse unterzogen. Bei der Verbindung der beiden Stichproben (respektive der Itempaare) wurde davon ausgegangen, dass zwischenzeitlich erfolgte Formulierungsänderungen bei einzelnen Items vernachlässigbar sind; die inhaltliche Dimension der betreffenden Items sollte im Zuge der Überarbeitung schließlich nicht verändert werden. Als Ausnahme wurde bei der Zusammenführung der Datensätze das Item 87 betrachtet, dessen Antwortformat geändert worden war – vgl. Hasenhindl (in Vorbereitung) und Kremser (in Vorbereitung); dieses Item wurde infolge im Gesamtdatensatz nicht berücksichtigt. Bei Item 57b war im Zuge der Überarbeitung die inhaltliche Ausrichtung geändert worden; um die Stichprobenteile sinnvoll zusammenfügen zu können, wurde Item 57b zuvor im Datensatz der ersten Stichprobe umgepolt.

Zunächst erfolgte nun für beide Stichproben mittels Kolmogorov-Smirnov-Anpassungstest eine Prüfung der erhobenen Daten auf Normalverteilung, wobei wenig überraschend für jedes der Items ein signifikantes Ergebnis ermittelt wurde. Nähere Betrachtungen zeigten bei vielen Items rechtssteile Verteilungen, die das Konzept des ESH 6-10 wiederspiegeln: Deckeneffekte sind im Rahmen eines Entwicklungsscreening-Verfahrens zu erwarten; im oberen Entwicklungsbereich kann und muss nicht so gut differenziert werden. Die deskriptiven Statistiken für einerseits die aktuelle Stichprobe und andererseits die Gesamtstichprobe finden sich in tabellarischer Form in Anhang A; die Histogramme zur Veranschaulichung werden an selber Stelle aus Platzgründen nur für den Gesamtdatensatz dargestellt.

Den Testkennwerten der zweiten Erhebung werden nun – skalenweise – die Ergebnisse der Gesamtstichprobe gegenüber gestellt. Zu beachten ist in diesem Zusammenhang, dass die beiden Stichproben nicht unabhängig voneinander betrachtet werden können, macht die „neue“ Stichprobe doch rund die Hälfte der Gesamtstichprobe aus. Von besonderem

Interesse sind infolge etwaige größere Veränderungen, die Rückschlüsse auf erfolgreiche Itembearbeitungen zwischen erster und zweiter Stichprobe zulassen würden sowie die Gesamtergebnisse an sich. Gegebenenfalls wird auf einzelne Items näher eingegangen.

Ebenso berücksichtigt werden Hinweise aus den Datensätzen (bzw. auch direkt von den PädagogInnen, vgl. auch die Kommentare in Anhang B) zur Frage der Beurteilbarkeit einzelner Items. Die als „nicht beurteilbar" klassifizierten Items werden ebenso wie die kommentarlos nicht beantworteten Items *nicht* als Bestandteil der im Folgenden jeweils angegebenen gültigen Itemanzahl betrachtet.

Zur Einstufung der Kennwerte vgl. die Ausführungen im Theorieteil der vorliegenden Arbeit; eine Übersicht über alle Skalenreliabilitäten (Tabelle 18) findet sich im Anschluss an die nachstehende detaillierte Ergebnisdarstellung. Die entsprechenden Ergebnisse der ersten Stichprobe sind bei Kunst (2014), Matschiner (2015) und Neugschwentner (2014) nachzulesen.

- Entwicklungsbereich Arbeitshaltungen

Tabelle 4 zeigt die Kennwerte der Skala *Aufmerksamkeit*. Für sämtliche Items wurden mittlere bis tendenziell leichte Itemschwierigkeiten und hohe Trennschärfen ermittelt, und zwar gleichermaßen bei der aktuellen Stichprobe wie bei der Gesamtstichprobe. Die innere Konsistenz der Skala ist mit $\alpha = .851$ bzw. gesamt .873 befriedigend, alle fünf Items dieser Skala werden infolge behalten.

		aktuelle Stichprobe				**Gesamtstichprobe**			
Skala Aufmerksamkeit	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 1	143	.758	.748	–	303	.734	.784	–
	Item 2	143	.783	.823	–	303	.762	.840	–
	Item 3	140	.663	.672	–	297	.638	.682	–
	Item 6	142	.832	.597	–	302	.832	.607	–
	Item 79	145	.703	.516	–	304	.685	.623	–
Cronbach Alpha:		vor Selektion: **.851**		nach Selektion:		vor Selektion: **.873**		nach Selektion:	

Tabelle 4: Kennwerte der Skala Aufmerksamkeit

Ein ähnliches Bild zeichnet sich bei der Skala *Exekutivfunktionen* ab (Tabelle 5): Sämtliche Itemschwierigkeiten bewegen sich im mittleren bzw. tendenziell leichten Bereich, die Trennschärfen sind hoch. Sowohl für die aktuelle Stichprobe (α = .883) als auch gesamt (α = .890) ist die Skalenreliabilität hoch.

Auffällig ist bei dieser Skala allerdings das in Tabelle 5a + b wiedergegebene Antwortverhalten für Item 8 und Item 9: Während bei Item 8 die Häufigkeiten der Antwort „nicht beurteilbar" konstant bleiben (15,6% bzw. 14,7%), weichen die Werte für Item 9 zwischen der ersten und zweiten Stichprobe voneinander ab; So berichtet Matschiner (2015) in ihrer Stichprobe von 5,6% fehlenden Antworten, in der aktuellen Stichprobe sind es dagegen 12,2%.

Nun sollen grundsätzlich beide Items die Dimension *Handlungsplanung* erfassen (vgl. ebd.), und zwar über die Beobachtung des Umgangs mit Hausaufgaben. Die Verteilung der Antworten aufgeschlüsselt nach Alter (vgl. in Anhang A) zeigt, dass bei beiden Items umso öfter nicht beurteilt werden konnte, je jünger die Kinder waren. Dies könnte dadurch erklärt werden, dass jüngere Schulkinder vermutlich weniger Hausaufgaben zu bearbeiten haben als ältere und für sie infolge die Beurteilung nicht möglich ist.

Eine Begründung für den Anstieg der Nicht-Beurteilbarkeit von Item 9 wird dadurch allerdings nicht unmittelbar ersichtlich; ob die geringfügige Umformulierung von „Teilt sich die Erledigung *der* Aufgaben ein" hin zu „Teilt sich die Erledigung *ihrer/seiner* Aufgaben ein" eine Rolle spielen könnte, sei an dieser Stelle dahin gestellt.

Insgesamt sind die Eigenschaften der Skala jedenfalls zufriedenstellend, es werden alle vier Items behalten.

		aktuelle Stichprobe				**Gesamtstichprobe**			
Skala Exekutivfunktionen	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 7	135	.778	.770	–	293	.790	.767	–
	Item 8	124	.682	.798	–	261	.677	.784	–
	Item 9	127	.729	.764	–	278	.733	.795	–
	Item 12	139	.820	.653	–	297	.789	.696	–
Cronbach Alpha:		vor Selektion: **.883**		nach Selektion:		vor Selektion: **.890**		nach Selektion:	

Tabelle 5: Kennwerte der Skala Exekutivfunktionen

Item 8		**aktuelle Stichprobe**		**Gesamt-stichprobe**	
		Häufigkeit	**Prozent**	Häufigkeit	**Prozent**
gültig	sehr selten	13	8,8	32	10,4
	selten	24	16,3	41	13,4
	manchmal	22	15,0	49	16,0
	oft	29	19,7	73	23,8
	sehr oft	36	24,5	66	21,5
	Gesamtsumme	124	84,4	261	85,0
fehlend	**nicht beurteilbar**	23	**15,6**	45	**14,7**
	nicht beantwortet			1	,3
	Gesamtsumme			46	15,0
Gesamtsumme		147	100,0	307	100,0

Tabelle 5a: Antwortverhalten bei Item 8

Item 9		**aktuelle Stichprobe**		**Gesamt-stichprobe**	
		Häufigkeit	**Prozent**	Häufigkeit	**Prozent**
gültig	sehr selten	10	6,8	20	6,5
	selten	19	12,9	36	11,7
	manchmal	17	11,6	43	14,0
	oft	41	27,9	97	31,6
	sehr oft	40	27,2	82	26,7
	Gesamtsumme	127	86,4	278	90,6
fehlend	**nicht beurteilbar**	18	**12,2**	27	**8,8**
	nicht beantwortet	2	1,4	2	,7
	Gesamtsumme	20	13,6	29	9,4
Gesamtsumme		147	100,0	307	100,0

Tabelle 5b: Antwortverhalten bei Item 9

Die Skala *Ablenkbarkeit* (Tabelle 6) verfügt mit Item 13 und Item 14 über zwei unterschiedlich schwierige Items, die beide eine hohe Trennschärfe aufweisen und inhaltlich gut zusammen passen; Cronbach $\alpha = .718$ bzw. gesamt .780. Eine Itemreduktion ist weder möglich (im Sinne der Skalenerhaltung) noch vonnöten, die Skala kann unverändert bestehen bleiben.

		aktuelle Stichprobe				**Gesamtstichprobe**			
Skala Ablenkbarkeit	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 13	143	.599	.563	–	303	.587	.642	–
	Item 14	146	.760	.563	–	305	.723	.642	–
Cronbach Alpha:		vor Selektion: **.718**		nach Selektion:		vor Selektion: **.780**		nach Selektion:	

Tabelle 6: Kennwerte der Skala Ablenkbarkeit

Vier unterschiedlich schwierige Items mit durchwegs hohen Trennschärfen und eine Reliabilität von α = .807 bzw. .806 können bei der Skala *Anstrengungsbereitschaft* (Tabelle 7) als zufriedenstellend gesehen werden; auch diese Skala muss nicht verändert werden.

Eine Betrachtung lohnt allerdings die Entwicklung des Item 19 (Tabelle 7a): Während in der ersten Stichprobe hier noch 17,4% auf die Kategorie „nicht beantwortbar" entfielen (vgl. Matschiner, 2015), sind es in der aktuellen Stichprobe nur mehr 4,1%. Die zwischenzeitlich erfolgte Umformulierung von „*Übt eine neue oder schwere Aufgabe so lange, bis sie beherrscht wird*" hin zu „*Zeigt Durchhaltevermögen, bleibt auch dran, wenn etwas schwierig wird*" (vgl. Hasenhindl, in Vorbereitung) scheint für die einschätzenden PädagogInnen eine Erleichterung darzustellen.

		aktuelle Stichprobe				**Gesamtstichprobe**			
Skala Anstrengungsbereitschaft	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 16	141	.626	.537	–	285	.603	.558	–
	Item 17	136	.593	.682	–	287	.586	.626	–
	Item 18	139	.886	.536	–	296	.861	.549	–
	Item 19	140	.743	.798	–	270	.692	.792	–
Cronbach Alpha:		vor Selektion: **.807**		nach Selektion:		vor Selektion: **.806**		nach Selektion:	

Tabelle 7: Kennwerte der Skala Anstrengungsbereitschaft

Item 19		**aktuelle Stichprobe**		**Gesamt-stichprobe**	
		Häufigkeit	**Prozent**	Häufigkeit	**Prozent**
gültig	sehr selten	14	9,5	25	8,1
	selten	14	9,5	36	11,7
	manchmal	24	16,3	66	21,5
	oft	34	23,1	76	24,8
	sehr oft	54	36,7	67	21,8
	Gesamtsumme	140	95,2	270	87,9
fehlend	**nicht beurteilbar**	6	**4,1**	34	**11,1**
	nicht beantwortet	1	,7	3	1,0
	Gesamtsumme	7	4,8	37	12,1
Gesamtsumme		147	100,0	307	100,0

Tabelle 7a: Antwortverhalten bei Item 19

Tabelle 8 präsentiert die Kennwerte zur Skala *Ausdauer*. Es finden sich eher leichte Items bei gleichzeitig hohen Trennschärfen; die Skala kann belassen werden, wie sie ist. Die innere Konsistenz ist sowohl für die aktuelle Stichprobe (α = .824) als auch insgesamt (α = .861) sehr zufriedenstellend.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Ausdauer	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 24	138	.786	.645	–	296	.776	.665	–
	Item 25	141	.800	.720	–	297	.780	.787	–
	Item 26	143	.863	.698	–	302	.834	.765	–
Cronbach Alpha:		vor Selektion: **.824**		nach Selektion:		vor Selektion: **.861**		nach Selektion:	

Tabelle 8: Kennwerte der Skala Ausdauer

Mit Item 29 („Teilt sich selbst ein, wann die Hausaufgaben gemacht werden“) beinhaltet die Skala *Selbstständigkeit* (Tabelle 9) ein Item, das einen über alle Stichproben konstant eher großen Anteil an Nicht-Beantwortbarkeit aufweist (vgl. Tabelle 9a). Der Grund dafür erschließt sich aus einigen Rückmeldungen der PädagogInnen zum Thema, wonach sich die Kinder nicht aussuchen können, wann die Hausübungen zu erledigen sind; dementsprechend kann in einigen Fällen über dieses Item keine Information zur Selbstständigkeit gewonnen werden. Im Gegensatz zu den oben betrachteten Items 8 und 9 können bei Item 29 allerdings keine altersspezifischen Zusammenhänge ausgemacht werden, es scheint hier vielmehr Unterschiede zwischen den Horten zu geben (vgl. für die Gesamtstichprobe Anhang A). Angesichts der Tatsache, dass doch bei etwa 85% der eingeschätzten Kinder gültige Angaben erhalten wurden, wird Item 29 wie die anderen Items der Skala beibehalten. Insgesamt wird die innere Konsistenz der Skala mit α = .746 für die aktuelle Stichprobe bzw. mit α = .771 für die Gesamtstichprobe als ausreichend hoch betrachtet; die Trennschärfen können als passabel bis hoch bezeichnet werden, die Itemschwierigkeiten als tendenziell leicht.

		aktuelle Stichprobe				Gesamtstichprobe			
	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Skala Selbstständigkeit	Item27	141	.896	.624	–	300	.880	.566	–
	Item28	142	.913	.571	–	298	.893	.613	–
	Item29	126	.781	.500	–	266	.774	.517	–
	Item30	146	.803	.532	–	306	.787	.552	–
	Item76	147	.860	.428	–	306	.835	.511	–
Cronbach Alpha:		vor Selektion: **.746**		nach Selektion:		vor Selektion: **.771**		nach Selektion:	

Tabelle 9: Kennwerte der Skala Selbstständigkeit

Item 29		**aktuelle Stichprobe**		**Gesamt-stichprobe**	
		Häufigkeit	**Prozent**	Häufigkeit	**Prozent**
gültig	sehr selten	10	6,8	19	6,2
	selten	13	8,8	23	7,5
	manchmal	16	10,9	40	13,0
	oft	27	18,4	76	24,8
	sehr oft	60	40,8	108	35,2
	Gesamtsumme	126	85,7	266	86,6
fehlend	**nicht beurteilbar**	21	**14,3**	41	**13,4**
	nicht beantwortet				
	Gesamtsumme				
Gesamtsumme		147	100,0	307	100,0

Tabelle 9a: Antwortverhalten bei Item 29

- Entwicklungsbereich Persönlichkeit

Die Skala *Internalisierende Störungen* (Tabelle 10) besteht aus zunächst fünf Items, die alle über einen relativ geringen Schwierigkeitsgrad verfügen. Besonders das Item 87 sticht mit einem Wert von p = .932 heraus; die für solch einen extremen Wert zu befürchtende niedrige Trennschärfe fällt mit r = .192 so gering aus, dass das Item an dieser Stelle nicht mehr berücksichtigt wird. Es ergibt sich für die Skalenreliabilität infolge ein akzeptabler Wert von α = .758 für die aktuelle Stichprobe. Für die Analyse der Gesamtstichprobe war Item 87 (aufgrund ungleicher Antwortformate, vgl. oben) bereits zuvor ausgeschieden worden; für die verbliebenen vier Items beträgt Cronbach α hier .727. Sowohl bei der aktuellen Stichprobe als auch bei der Gesamtstichprobe bewegen sich die Trennschärfen schließlich im mittleren bis hohen Bereich.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Internalisierende Störungen	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 35	145	.833	.519	.529	304	.826	.537	–
	Item 84	146	.862	.515	.491	306	.825	.459	–
	Item 86	144	.858	.650	.653	300	.856	.595	–
	Item 87	147	.932	.192	–	–	–	–	–
	Item 88	139	.865	.550	.556	294	.858	.486	–
Cronbach Alpha:		vor Selektion: .723		nach Selektion: **.758**		vor Selektion: **.727**		nach Selektion:	

Tabelle 10: Kennwerte der Skala Internalisierende Störungen

Tabelle 11 zeigt die Kennwerte der Skala *Externalisierende Störungen*. Wir finden durchwegs mittelschwere bis leichte Items vor bei gleichzeitig hohen Trennschärfen und einem Cronbach α von .867 für die aktuelle bzw. .896 für die gesamte Stichprobe – alle sieben Items werden infolge beibehalten.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Externalisierende Störungen	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 40	141	.850	.686	–	299	.827	.724	–
	Item 51	141	.823	.634	–	298	.811	.716	–
	Item 79	145	.703	.673	–	304	.685	.686	–
	Item 80	140	.860	.684	–	299	.843	.725	–
	Item 81	140	.903	.725	–	298	.887	.747	–
	Item 82	142	.765	.465	–	295	.761	.564	–
	Item 83	143	.775	.690	–	302	.753	.768	–
Cronbach Alpha:		vor Selektion: **.867**		nach Selektion:		vor Selektion: **.896**		nach Selektion:	

Tabelle 11: Kennwerte der Skala Externalisierende Störungen

- **Entwicklungsbereich Motivation**

Die Skala *Leistungsmotivation* besteht aus zunächst 14 Items mittleren bis geringen Schwierigkeitsgrades. Wie Tabelle 12 deutlich macht, bestehen allerdings für die Items 72, 73 und 74 sehr niedrige Trennschärfen. Diese Items hatten bereits bei der ersten Stichprobe

auffällige Trennschärfen aufgewiesen (vgl. Neugschwentner, 2014), weshalb sie einer Umformulierung unterzogen worden waren. Diese hat sich nun zwar offensichtlich positiv auf die erlebte Anwendbarkeit von Item 73 ausgewirkt (aktuell 4,1% als nicht beurteilbar Klassifizierte im Gegensatz zu 12,2% bei der ersten Stichprobe, vgl. Tabelle 12a sowie Neugschwentner, 2014), dieses und die anderen beiden Items trennen die Gruppen aber immer noch nicht gut, weshalb sie vorläufig aus der Skala entfernt werden. – Die Beurteilbarkeit betreffend zeigt sich bei Item 75 eine ähnliche Entwicklung wie bei Item 73 (aktuell 4,8% nicht beurteilbar im Vergleich zu 10,4% bei der ersten Stichprobe, vgl. Tabelle 12b sowie Neugschwentner, 2014). Obwohl die Trennschärfe dieses Items mit aktuell .434 (nach Selektion) deutlich geringer ist als bei der ersten Stichprobe – nach Matschiner (2014) lag der entsprechende Wert bei $r = .596$ –, ist sie doch groß genug, um das Item im Test belassen zu können.

Die Itemschwierigkeiten bewegen sich auch bei der Skala Leistungsmotivation im mittleren bis leichten Bereich, Item 17 ist mit einem Wert von $p = .593$ bzw. gesamt .586 noch am schwierigsten. Für die innere Konsistenz der Skala ergibt sich der – mit der eben erfolgten Testreduktion gestiegene – bemerkenswert hohe Wert von $\alpha = .900$ bzw. .904.

		aktuelle Stichprobe				**Gesamtstichprobe**			
	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Skala Leistungsmotivation	Item 16	141	.626	.503	.525	285	.603	.571	.600
	Item 17	136	.593	.619	.690	287	.586	.589	.663
	Item 18	139	.886	.686	.686	296	.861	.615	.627
	Item 19	140	.743	.806	.840	270	.692	.783	.809
	Item 23	138	.801	.820	.848	298	.772	.788	.810
	Item 67	141	.667	.639	.664	301	.641	.687	.719
	Item 68	139	.704	.757	.780	298	.683	.753	.788
	Item 69	135	.748	.604	.591	289	.655	.641	.634
	Item 71	137	.862	.539	.488	291	.867	.486	.455
	Item 72	134	.803	**-.164**	–	285	.731	-.342	–
	Item 73	141	.809	**.162**	–	281	.757	.076	–
	Item 74	147	.867	**.100**	–	307	.850	.090	–
	Item 75	139	.863	.481	.434	282	.818	.547	.534
	Item 76	147	.860	.465	.461	306	.835	.439	.442
	Cronbach Alpha:	vor Selektion: .858		nach Selektion: **.900**		vor Selektion: .841		nach Selektion: **.904**	

Tabelle 12: Kennwerte der Skala Leistungsmotivation

Item 73		aktuelle Stichprobe		Gesamt-stichprobe	
		Häufigkeit	**Prozent**	Häufigkeit	Prozent
gültig	sehr selten	6	4,1	11	3,6
	selten	12	8,2	35	11,4
	manchmal	19	12,9	62	20,2
	oft	37	25,2	69	22,5
	sehr oft	67	45,6	104	33,9
	Gesamtsumme	141	95,9	281	91,5
fehlend	**nicht beurteilbar**	6	**4,1**	26	8,5
	nicht beantwortet				
	Gesamtsumme				
Gesamtsumme		147	100,0	307	100,0

Tabelle 12a: Antwortverhalten bei Item 73

Item 75		aktuelle Stichprobe		Gesamt-stichprobe	
		Häufigkeit	**Prozent**	Häufigkeit	Prozent
gültig	sehr selten	1	,7	11	3,6
	selten	10	6,8	26	8,5
	manchmal	19	12,9	41	13,4
	oft	23	15,6	52	16,9
	sehr oft	86	58,5	152	49,5
	Gesamtsumme	139	94,6	282	91,9
fehlend	**nicht beurteilbar**	7	**4,8**	24	7,8
	nicht beantwortet	1	,7	1	,3
	Gesamtsumme	8	5,4	25	8,1
Gesamtsumme		147	100,0	307	100,0

Tabelle 12b: Antwortverhalten bei Item 75

- Entwicklungsbereich Sprache

Die Kennwerte der Skala *Grammatik und Schriftspracherwerb* sind in Tabelle 13 ablesbar. Für die aktuelle Stichprobe verfügen alle Items über einen geringen Schwierigkeitsgrad und – bis auf Item 58c – hohe Trennschärfen; im Vergleich zur ersten Stichprobe konnten die Trennschärfen teils erheblich verbessert werden (vgl. Kunst, 2014). Als möglicher Grund dafür kann die veränderte Formulierung des Items 57b genannt werden, durch die die Items 57a-c nun semantisch alle in die gleiche Richtung weisen, was infolge für die PädagogInnen mit weniger Missverständnissen verbunden zu sein scheint. Das Cronbach α beträgt für die aktuelle Stichprobe .875. In der Gesamtstichprobe fällt Item 58c mit einer niedrigen Trennschärfe auf; wird es entfernt, verringern sich allerdings auch die Trennschärfen von Item 58a und b – bei gleichzeitig nur marginaler Erhöhung von Cronbach α. Item 58c wird unter diesen Umständen trotz seiner geringen Trennschärfe vorläufig beibehalten.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Grammatik und Schriftspracherwerb	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 57a	145	.759	.773	–	296	.476	.538	.615
	Item 57b	145	.759	.702	–	290	.525	.581	.641
	Item 57c	143	.888	.681	–	296	.483	.488	.565
	Item 58a	138	.833	.759	–	292	.864	.339	.264
	Item 58b	135	.756	.773	–	288	.777	.378	.247
	Item 58c	142	.754	.442	–	298	.766	.251	–
Cronbach Alpha:		vor Selektion: **.875**		nach Selektion:		vor Selektion: .698		nach Selektion: **.708**	

Tabelle 13: Kennwerte der Skala Grammatik und Schriftspracherwerb

Tabelle 14 präsentiert die Skala *Entwicklung der Aussprache*. Alle vier Items verfügen sowohl in der aktuellen Stichprobe als auch insgesamt über hohe Trennschärfen; die Itemschwierigkeiten sind durchwegs gering. Insgesamt wird bei der Skala eine innere Konsistenz von α = .810 bzw. gesamt .786 erzielt, woraus also kein Veränderungsbedarf abgeleitet werden kann.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Entwicklung der Aussprache	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 59	144	.786	.625	–	303	.799	.611	–
	Item 60	145	.846	.704	–	305	.856	.651	–
	Item 61	143	.877	.593	–	303	.893	.532	–
	Item 63	147	.835	.610	–	304	.849	.603	–
Cronbach Alpha:		vor Selektion: **.810**		nach Selektion:		vor Selektion: **.786**		nach Selektion:	

Tabelle 14: Kennwerte der Skala Entwicklung der Aussprache

Die fünf Items der Skala *Entwicklung der Sprachpragmatik* (Tabelle 15) haben ebenfalls eine hohe Lösungswahrscheinlichkeit, besonders das Item 65 fällt mit einem Wert von p = .900 bei der aktuellen Stichprobe als sehr einfach auf. Zugleich sind die Trennschärfen hoch, einzig Item 48 weist insgesamt eine zwar im Vergleich geringere, aber mit r = .369 absolut betrachtet ausreichende Trennschärfe auf. Cronbach α beträgt .787 bzw. .763; die Skala kann unverändert bleiben.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Entwicklung der Sprachpragmatik	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 48	146	.799	.501	–	306	.785	.369	–
	Item 64	147	.860	.726	–	306	.865	.687	–
	Item 65	144	.900	.621	–	299	.870	.565	–
	Item 66	146	.784	.514	–	304	.783	.540	–
	Item 82	142	.765	.514	–	295	.761	.536	–
Cronbach Alpha:		vor Selektion: **.787**		nach Selektion:		vor Selektion: **.763**		nach Selektion:	

Tabelle 15: Kennwerte der Skala Entwicklung der Sprachpragmatik

- Entwicklungsbereich Soziale Interaktion

Fünf Items mit Schwierigkeitswerten von p = .756 bis .859, dazu hohe Trennschärfen und eine stabil hohe innere Konsistenz: an der Skala *Anpassung an Gruppenregeln* (Tabelle 16) wird ebenfalls nichts verändert.

		aktuelle Stichprobe				Gesamtstichprobe			
Skala Anpassung an Gruppenregeln	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
	Item 51	141	.823	.659	–	298	.811	.702	–
	Item 52	147	.859	.811	–	306	.839	.801	–
	Item 53	147	.854	.846	–	307	.843	.821	–
	Item 55	144	.765	.760	–	303	.756	.775	–
	Item 56	146	.812	.791	–	304	.807	.776	–
Cronbach Alpha:		vor Selektion: **.906**		nach Selektion:		vor Selektion: **.906**		nach Selektion:	

Tabelle 16: Kennwerte der Skala Anpassung an Gruppenregeln

		aktuelle Stichprobe				Gesamtstichprobe			
	Item	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion	Anzahl gültiger Antworten	Item-Schwierigkeit	Trennschärfe vor Selektion	Trennschärfe nach Selektion
Skala Soziale Interaktion	Item 31	147	.765	.506	.505	304	.739	.530	.514
	Item 32	144	.857	.334	.343	301	.839	.374	.387
	Item 33	145	.768	.433	.459	304	.766	.400	.411
	Item 35	145	.833	.453	.456	304	.826	.486	.490
	Item 36	145	.854	.805	.798	304	.833	.756	.747
	Item 37	147	.869	.754	.757	307	.847	.705	.703
	Item 38	147	.829	.617	.635	307	.803	.549	.544
	Item 39	142	.890	.648	.650	300	.881	.668	.654
	Item 40	141	.850	.437	.456	299	.827	.439	.436
	Item 41	140	.817	.484	.464	296	.784	.403	.389
	Item 43	140	.543	.137	–	291	.597	.163	–
	Item 44	139	.646	.335	.301	288	.580	.049	–
	Item 47	146	.749	.204	–	306	.735	.237	–
	Item 50	142	.880	.523	.580	301	.859	.450	.463
	Item 84	146	.862	.587	.587	306	.825	.635	.616
	Item 85	146	.814	.673	.657	304	.774	.631	.618
	Item 87	147	.932	.332	.338	–	–	–	–
	Cronbach Alpha:	vor Selektion: .855		nach Selektion: **.874**		vor Selektion: .837		nach Selektion: **.863**	

Tabelle 17: Kennwerte der Skala Soziale Interaktion

Bei der Skala *Soziale Interaktion* weisen die Items 43 und 47 für die aktuelle Stichprobe nur sehr geringe Itemtrennschärfen auf (vgl. Tabelle 17); betrachtet man die Werte über beide Stichproben, muss auch das Item 44 als zu wenig trennscharf erkannt werden. Bei genauerer Betrachtung wird deutlich, dass hier zwei Items als inhaltlich zusammenhängend wahrgenommen werden könnten: „*Bevorzugt ältere Kinder als Spielpartner*“ (Item 43) und „*Bevorzugt jüngere Kinder als Spielpartner*“ (Item 44) werden von BeurteilerInnen möglicherweise als Ausprägungen auf ein und dem selben Spektrum wahrgenommen, und infolge nicht unabhängig voneinander beantwortet – was aus testtheoretischer Sicht nicht optimal erscheint. Die Items 43, 44 und 47 werden an dieser Stelle jedenfalls nicht weiter berücksichtigt.

Das Item 87 verfügt mit r = .332 ebenfalls über eine grenzwertig niedrige Trennschärfe, was angesichts des leichten Schwierigkeitsgrads von p = .932 allerdings zu erwarten war. Für die Gesamtstichprobe kann an dieser Stelle keine Angabe gemacht werden, da das Item 87 wie oben beschrieben aufgrund der Änderung seines Antwortformats in diesem Fall nicht berücksichtigt wurde. Die typischerweise leicht zu

beantwortende Frage „Hat wenigstens eine gute Freundin oder einen guten Freund“ scheint aber einen positiven Abschluss für einen Entwicklungseinschätzbogen zu bilden, weshalb das Item grundsätzlich vorerst weiterhin Berücksichtigung finden kann.

Die verbleibenden Items sind durchwegs wieder von geringer Schwierigkeit, die Trennschärfen ausreichend bis hoch. Für die innere Konsistenz der Skala können sowohl für die aktuelle Stichprobe (α = .874) als auch insgesamt (α = .863) sehr zufriedenstellende Werte erreicht werden.

Übersicht über die Entwicklung der Reliabilitäten:

[Ergebnisse der 1. Stichprobe nach Kunst (2014), Matschiner (2015) bzw. Neugschwentner (2014)]

1. Stichprobe Cronbach α	2. Stichprobe Cronbach α	Gesamt	
.914	.851	.873	Skala Aufmerksamkeit
.900	.883	.890	Skala Exekutivfunktionen
.823	.718	.780	Skala Ablenkbarkeit
.804	.807	.806	Skala Anstrengungsbereitschaft
.883	.824	.861	Skala Ausdauer
.781	.746	.771	Skala Selbstständigkeit
.764	.758	.727	Skala Internalisierende Störungen
.913	.867	.896	Skala Externalisierende Störungen
.885	.900	.904	Skala Leistungsmotivation
.774	.875	.708	Skala Grammatik und Schriftspracherwerb
.751	.810	.786	Skala Entwicklung der Aussprache
.786	.787	.763	Skala Entwicklung der Sprachpragmatik
.914	.906	.906	Skala Anpassung an Gruppenregeln
.864	.874	.863	Skala Soziale Interaktion

Tabelle 18: Übersicht über die Skalenreliabilitäten

Der Gesamtdatensatz wurde nun in weiterer Folge einer *explorativen Faktorenanalyse* unterzogen, wobei die Items 43, 44, 47, 72, 73, 74 und 87 an dieser Stelle keine Berücksichtigung mehr fanden. Bezüglich der teils als schlecht beurteilbar wahrgenommenen und daher lückenhaft beantworteten Items 8, 19 und 29 (vgl. Ausführungen oben) wurde die Entscheidung getroffen, weder die Items zu entfernen noch

durch Imputationsverfahren diese Information der fehlenden Beurteilbarkeit zu überdecken; dem Problem der vielen Missings wurde anstattdessen durch paarweises Löschen begegnet, wodurch das *N* möglichst groß gehalten werden konnte (vgl. deskriptive Statistiken in Anhang A). Als Methode wurde die Hauptachsenanalyse gewählt.

Die zunächst ausgegebene Korrelationsmatrix (ohne Abbildung) weist eine gemischte Struktur auf, es finden sich Werte von rund um null genauso wie hohe Korrelationen um $r = .8$ und vereinzelt darüber. Die Frage der Eignung der Zusammenhänge für eine Faktorenanalyse kann aufgrund der Matrix also nicht eindeutig beantwortet werden. Wie der signifikante Bartlett-Test (unter Vernachlässigung seiner Forderung nach multivariater Normalverteilung der Items) zeigt, handelt es sich aber jedenfalls um eine Matrix, die sich systematisch von einer Einheitsmatrix unterscheidet, weshalb die Durchführung einer Faktorenanalyse vertretbar ist. Der Kaiser-Meyer-Olkin-Koeffizient KMO belegt mit einem sehr zufriedenstellenden Wert von .928 ebenfalls die Eignung der Itemzusammenstellung.

In weiterer Folge betrachtet (aus Platzgründen aber ebenfalls nicht abgebildet) werden nun die Anti-Image-Korrelationsmatrix und mit ihr die MSA-Koeffizienten, die Aufschluss über die Eignung der Daten auf Itemebene geben: dabei finden sich durchwegs sehr gute Werte, einzig die Items 57a-c fallen mit Werten zwischen .559 und .680 aus der Reihe. Nachdem sie das MSA-Mindestmaß von .5 aber doch überschreiten, wurden sie bei der Analyse zunächst beibehalten.

Für die Entscheidung über die Anzahl der zu extrahierenden Faktoren wurden sowohl die numerischen Ergebnisse der Eigenwerte (vgl. Anhang A) als auch das Scree-Plot betrachtet (Abbildung 1): Während erstere für zwölf Faktoren Eigenwerte > 1 zeigen, legt die grafische Darstellung eine Fünf-Faktoren-Lösung nahe. Nachdem dies auch inhaltlich – angesichts der dem ESH 6-10 zugrundeliegenden Theorie – eine plausible Variante darstellen würde, wurden in einem ersten Schritt also zunächst fünf Faktoren extrahiert. Tabelle 19 zeigt die rotierte (Varimax) Faktorenmatrix für diese Lösung (für die unrotierte Matrix vgl. Anhang A).

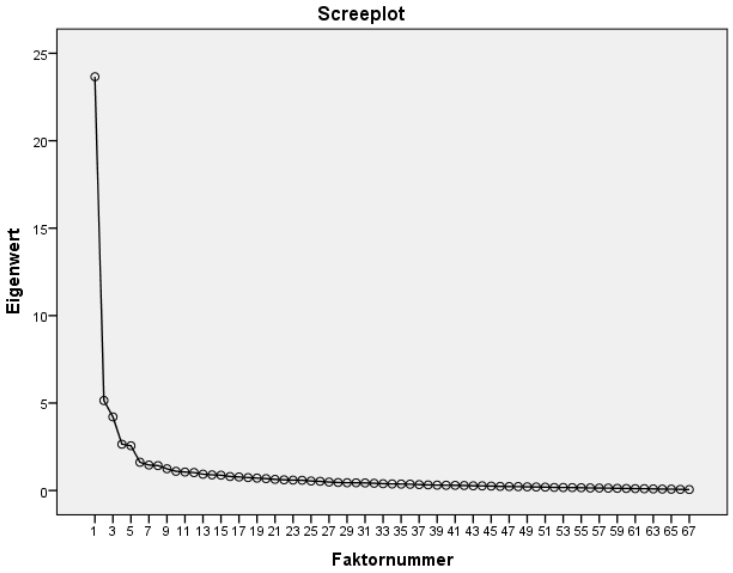


Abbildung 1: Screeplot Fünf-Faktoren-Lösung

Bei der Betrachtung der Ladungen fällt auf, dass Faktor fünf praktisch ausschließlich die drei Items 57abc erklärt; demnach haben diese Items mit den anderen Items nichts gemeinsam. Dies ist inhaltlich allerdings nur schwer nachvollziehbar, finden sich die restlichen dazupassenden (Sprach-)Items doch erwartungsgemäß gruppiert und passabel hoch ladend auf Faktor vier. Nachdem für diesen Befund keine inhaltliche Erklärung gefunden werden, ein möglicher Fehler bei der Datenverarbeitung dagegen nicht ausgeschlossen werden konnte, wurden die betreffenden Items entfernt. In Anbetracht der ansonsten schlüssigen Ladungsverteilungen wurde infolge eine Struktur mit vier extrahierten Faktoren ins Auge gefasst.

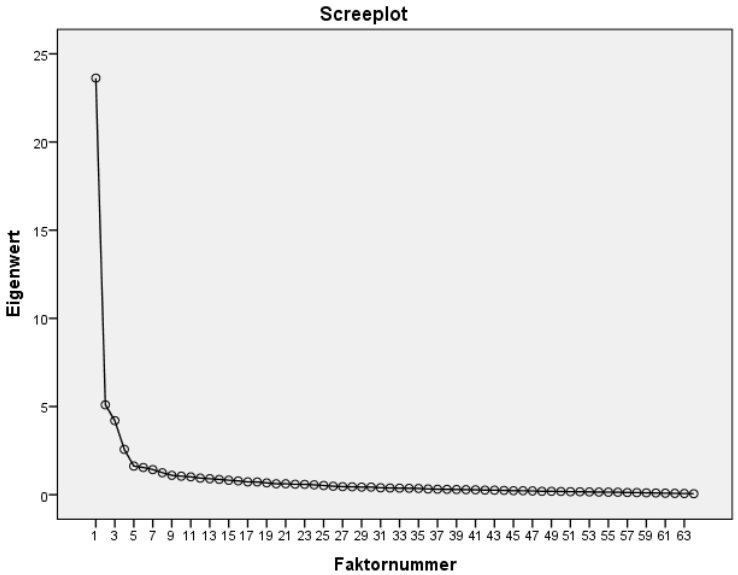


Abbildung 2: Screeplot Vier-Faktoren-Lösung

Rotierte Faktorenmatrix[a]

	Faktor				
	1	2	3	4	5
Item 26	,845	,080	,156	-,001	-,024
Item 12	,827	,084	,154	,001	-,021
Item 1	,812	,330	,120	,103	-,050
Item 68	,766	,452	,015	,117	-,027
Item 9	,760	,213	,154	,178	-,040
Item 25	,751	,224	,139	,055	-,013
Item 2	,740	,347	,092	,240	,049
Item 8	,739	,252	,033	,159	,017
Item 23	,726	,393	,080	,119	,102
Item 7	,722	,293	,064	,110	-,013
Item 13	,706	,347	,076	,120	-,049
Item 19	,699	,327	,115	,291	,210
Item 58c	,696	-,014	,121	,201	-,017
Item 3	,661	,210	-,002	,219	,065
Item 24	,656	,422	,080	,214	,024
Item 67	,645	,297	,112	,165	,076
Item 29	,627	,193	,072	,270	-,025
Item 27	,599	,256	,101	,305	,036
Item 17	,591	,135	,053	,357	,065
Item 6	,564	,285	,105	,206	-,018
Item 14	,542	,486	,119	,134	,018
Item 16	,512	,305	,053	,188	,090
Item 18	,502	,279	,233	,164	,049
Item 69	,463	,216	,219	,245	,210
Item 71	,358	,301	,184	,016	-,121
Item 53	,255	,834	,047	,063	-,002
Item 52	,317	,766	,076	,024	-,022
Item 55	,258	,762	-,022	,170	,057
Item 40	,187	,748	,099	,083	,110
Item 56	,254	,736	,055	,188	-,008
Item 51	,289	,693	,033	,161	,078
Item 83	,384	,676	,106	,032	,067
Item 64	,192	,669	,162	,107	-,120
Item 81	,271	,666	,089	,144	,129
Item 38	,295	,665	,292	,105	,140
Item 79	,462	,650	,062	,104	-,008
Item 82	,126	,631	,171	,112	-,140
Item 80	,358	,585	,251	,043	,066
Item 75	,315	,559	,209	,020	,020
Item 65	,192	,487	,267	-,006	-,066
Item 37	,175	,462	,410	,360	,174
Item 50	,068	,351	,328	,079	,064
Item 84	,171	,036	,734	,050	,104
Item 36	,129	,262	,685	,281	,118
Item 35	-,037	,013	,682	,084	-,135
Item 31	-,055	,133	,644	-,006	,079
Item 85	,182	,111	,634	,176	,145
Item 41	,085	-,018	,563	,058	,045
Item 86	,216	,163	,531	,135	-,093
Item 76	,198	,228	,486	,307	,006
Item 48	,116	,087	,481	,385	,090
Item 32	,005	,070	,467	,015	,038
Item 88	,097	,122	,452	,171	-,117
Item 39	,176	,355	,419	,399	,116
Item 30	,228	,132	,397	,382	-,011
Item 60	,090	,070	,162	,721	,038
Item 58a	,179	,068	,117	,695	,029
Item 63	,199	,068	,207	,624	-,070
Item 59	,184	,083	,199	,615	-,005
Item 61	,194	,116	-,006	,597	,068
Item 58b	,322	,033	,134	,553	,035
Item 66	,395	,287	,263	,488	-,087
Item 33	,142	,127	,337	,368	,065
Item 28	,313	,222	,218	,328	-,025
Item 57a	,033	,014	,014	,016	,856
Item 57c	,055	,058	,119	-,067	,824
Item 57b	-,018	,022	,072	,151	,714

Extraktionsmethode:

Hauptachsenfaktorenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Tabelle 19: Rotierte Faktorenmatrix Fünf-Faktoren-Lösung

Rotierte Faktorenmatrix[a]

	Faktor			
	1	2	3	4
Item 26	,846	,083	,155	,003
Item 12	,828	,087	,154	,005
Item 1	,810	,333	,118	,108
Item 68	,764	,455	,015	,122
Item 9	,757	,215	,154	,183
Item 25	,751	,227	,140	,058
Item 2	,735	,352	,095	,250
Item 8	,735	,257	,034	,168
Item 23	,720	,401	,085	,133
Item 7	,720	,297	,068	,113
Item 13	,702	,349	,075	,125
Item 58c	,696	-,012	,123	,201
Item 19	,691	,338	,124	,308
Item 3	,656	,216	,001	,233
Item 24	,651	,426	,083	,223
Item 67	,639	,303	,117	,180
Item 29	,623	,195	,073	,274
Item 27	,593	,259	,103	,315
Item 17	,585	,140	,057	,367
Item 6	,560	,287	,107	,210
Item 14	,537	,488	,119	,142
Item 16	,506	,310	,058	,200
Item 18	,500	,282	,236	,166
Item 69	,455	,225	,223	,266
Item 71	,357	,298	,181	,012
Item 53	,250	,835	,048	,064
Item 52	,314	,766	,076	,022
Item 55	,251	,765	-,019	,178
Item 40	,181	,752	,105	,092
Item 56	,248	,736	,056	,191
Item 51	,282	,697	,037	,172
Item 83	,380	,680	,109	,039
Item 81	,263	,671	,096	,156
Item 38	,288	,670	,298	,118
Item 64	,191	,662	,159	,098
Item 79	,458	,651	,061	,108
Item 82	,126	,622	,169	,101
Item 80	,353	,589	,255	,049
Item 75	,310	,561	,209	,027
Item 65	,191	,483	,263	-,010
Item 37	,166	,467	,417	,374
Item 50	,065	,352	,330	,082
Item 84	,169	,040	,738	,054
Item 36	,123	,264	,693	,284
Item 35	-,036	,007	,672	,072
Item 31	-,058	,134	,646	-,002
Item 85	,177	,116	,638	,187
Item 41	,083	-,017	,564	,060
Item 86	,216	,159	,529	,124
Item 48	,112	,088	,487	,385
Item 76	,193	,227	,484	,311
Item 32	,004	,070	,467	,018
Item 88	,097	,117	,448	,160
Item 39	,170	,358	,425	,403
Item 30	,222	,131	,396	,385
Item 60	,084	,069	,164	,719
Item 58a	,173	,067	,122	,692
Item 63	,195	,064	,207	,615
Item 59	,178	,080	,199	,615
Item 61	,188	,117	-,001	,601
Item 58b	,318	,034	,137	,555
Item 66	,389	,284	,260	,486
Item 33	,136	,129	,342	,374
Item 28	,309	,221	,215	,329

Extraktionsmethode: Hauptachsenfaktorenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Tabelle 20: Rotierte Faktorenmatrix Vier-Faktoren-Lösung

Um die Unabhängigkeit der Faktoren beizubehalten, wurde wieder mit Varimax rotiert. Tabelle 20 zeigt die rotierte Faktorenmatrix für die vier-Faktoren-Lösung: Auf jedem Faktor finden sich einige hoch ladende Items, die gleichzeitig auf den jeweils anderen Faktoren im Großen und Ganzen niedrige Ladungen aufweisen – das Einfachstrukturkriterium scheint erfüllt, das Ergebnis ist zufriedenstellend. Allerdings ist zu beachten, dass der Anteil der von vier Faktoren erklärten Varianz nur bei 52,6% liegt (vgl. Anhang A), was eine Interpretierbarkeit der Lösung in Frage stellt.

Dennoch seien an dieser Stelle inhaltliche Überlegungen angestellt:

Faktor 1 repräsentiert sehr deutlich den aus inhaltlichen Vorüberlegungen zusammengestellten Bereich der *Arbeitshaltungen*: Items 1 bis 27, 29 sowie die ebenfalls auf die Arbeitsweise beziehbaren Items 58c, 67 und 68.

Auf Faktor 4 finden sich mit den Items 58a, 58b, 59, 60, 61 und 63 klar die *sprachbezogenen* Items des ESH 6-10 wieder. Darüber hinaus mag den Ergebnissen bezüglich der Items 57abc ein zum gegenwärtigen Zeitpunkt nicht identifizierter Fehler in der Datenstruktur zugrundeliegen; es wäre gut vorstellbar, dass diese Items bei einer neuerlichen Untersuchung wie die anderen Sprach-Items ebenfalls auf Faktor 4 hoch laden würden.

Für Faktor 2 und 3 scheinen die Zuordnungen allerdings tatsächlich nicht den Erwartungen zu entsprechen, weshalb auf Itemebene neue Überlegungen angestellt wurden: Den ursprünglich in die Bereiche *Persönlichkeit* und *Soziale Interaktion* eingeteilten Items könnte demnach auch eine andere Faktorenstruktur zugrunde liegen. Die auf Faktor 2 eher hoch ladenden Items 38, 40, 51 bis 56, 64, 75 sowie 79 bis 83 umfassen inhaltlich Bereiche wie Kooperationsbereitschaft, Fähigkeit zu Anpassung und Integration; dies könnte auf die Persönlichkeitsdimension *Verträglichkeit* als erklärenden Faktor hindeuten.

Faktor 3 beinhaltet mit den Items 31, 35, 36, 41, 84 und 85 Aspekte, die sich auf das Verhalten in Gruppensituationen beziehen; erklärender Faktor könnte hier die *Extraversion* sein.

Der ursprünglich angenommene Entwicklungsbereich *Leistungsmotivation* kann mit den vorliegenden Ergebnissen nicht als eigener Faktor identifiziert werden; er ist möglicherweise vom Faktor Arbeitshaltungen nur schwer zu trennen bzw. fließt in selbigen ein.

8. Diskussion und Ausblick

Das ESH 6-10 stellt sich zusammenfassend als vielversprechendes Instrument zur groben Erfassung von Entwicklungsauffälligkeiten dar. Seine Skalen verfügen über sehr zufriedenstellende Reliabilitätswerte, die Itemkennwerte entsprechen durchwegs den Anforderungen an ein Screeningverfahren, das im unteren Leistungsbereich differenzieren soll. Die wenigen – im Ergebnisteil identifizierten und bei den weiterführenden Analysen ausgeschlossenen – auffälligen Items skizzieren bereits die aus Sicht der Verfasserin nunmehr angebrachte Itemselektion:

Die Items 47, 72, 73 und 74 genügen hinsichtlich ihrer Kennwerte nicht den an sie gestellten Ansprüchen und sollten daher aus dem Verfahren entfernt werden.

Die Items 43 und 44 konnten ebenfalls nicht überzeugen, wobei hier, wie im Ergebnisteil bereits angedeutet, die Unabhängigkeit der beiden Items voneinander (in der Wahrnehmung der RaterInnen) in Frage gestellt wird. Gleiches gilt für die zwar ergebnisunauffälligen Items 32 („*In der Freundschaft steht vor allem das gemeinsame Spielen im Vordergrund*“) und 33 („*In der Freundschaft spielen Gemeinsamkeiten und Reden eine wichtige Rolle*“), deren Formulierungen aber ebenfalls als Ausprägungen auf ein und demselben Spektrum wahrgenommen werden könnten. Für beide Itempaare könnte in der Folge überlegt werden, jeweils ein adäquates *einzelnes* Item zu generieren.

Ebenfalls neu erarbeitet werden könnten ein oder zwei zusätzliche Items zur Erfassung der Handlungsplanung, die im Gegensatz zu den Items 8 und 9 *nicht* auf dem Umgang mit Hausaufgaben basieren. Auf diese Weise könnte man auch entsprechende Informationen zu den jüngeren Kindern, die nur keine oder wenig Hausübungen zu erledigen haben, bekommen.

Das Item 87 sollte aufgrund seiner Testkennwerte aus der Skala Internalisierende Störungen entfernt werden; als Abschlussitem kann es für die Skala Soziale Interaktion aber beibehalten werden.

Eine Reihe weiterer Items war ebenfalls bei der ursprünglichen Testkonstruktion jeweils zwei unterschiedlichen Skalen zugerechnet worden, was im Grunde bereits die Eindimensionalität der betreffenden Items sowie die Unabhängigkeit der zugrundeliegenden Faktoren in Frage gestellt hat. Auf Basis der durchgeführten Faktorenanalyse lassen sich diese Doppel-Zuordnungen allerdings weitgehend auflösen: So laden die Items 16, 17, 18 und 19 eindeutig am höchsten (.500 bis .691) auf Faktor 1, der

den Bereich der Arbeitshaltungen repräsentiert; die ursprünglich getroffene Unterscheidung zwischen Anstrengungsbereitschaft und Leistungsmotivation kann hier nicht mehr nachvollzogen werden.

Die Items 40, 51, 79 und 82 laden allesamt auf Faktor 2 am höchsten (.622 bis .752); Aspekte der Persönlichkeit sowie der Sozialen Interaktion können demnach in einen einzelnen Faktor (Kooperationsbereitschaft bzw. Verträglichkeit, vgl. Ergebnisteil) integriert werden.

Die Items 35 (.672) und 84 (.738) sowie (wenn auch mit einer geringeren Ladungshöhe von .484) das Item 76 finden sich schließlich eindeutig auf Faktor 3 wieder, der wiederum Inhalte der Sozialen Interaktion und der Internalisierenden Störungen (über das beobachtbare Verhalten in Gruppensituationen, vgl. Ergebnisteil) zum Faktor Extraversion zusammenfassen könnte. – Sämtliche Schlüsse sind allerdings unter Vorbehalt zu ziehen, nachdem der von vier Faktoren erklärte Varianzanteil wie erwähnt bei lediglich 52,6% liegt (vgl. erneut Ergebnisteil).

Für das weitere Vorgehen im Zuge der Verfahrensentwicklung könnte nun von Interesse sein, inwieweit sich die über die Ergebnisse der Faktorenanalyse angeregte und wie oben beschriebene Differenzierung der „Persönlichkeits-Items“ als relevant erweisen kann. Aus der Sicht der Verfasserin scheint es möglich, dass die Inhalte der bestehenden Items zur Erfassung von Entwicklungsauffälligkeiten zum Teil mit basalen Persönlichkeitseigenschaften konfundiert sein könnten. Um diesen Zusammenhängen auf den Grund gehen zu können, wäre als weiterer Schritt eine neue Stichprobe wünschenswert, der einerseits das ESH 6-10 in seiner aktuellen Form und andererseits ein etabliertes Testverfahren zur Erfassung von Persönlichkeitseigenschaften vorgelegt werden könnte. – Der entstehende Mehraufwand wäre freilich für ein einmaliges Forschungsprojekt, nicht aber für eine breite Anwendung in Zusammenhang mit einem Screeningverfahren gerechtfertigt.

Über die Testzusammensetzung hinaus wäre es nach Ansicht der Verfasserin aus testtheoretischen Überlegungen weiters sinnvoll, dem Einfluss durch die BeurteilerInnen als mögliche Störvariable vermehrt Aufmerksamkeit zu schenken. Während Halo- und andere Testleitereffekte im vorliegenden Setting schwer vermeidbar erscheinen, könnte zumindest darauf geachtet werden, die RaterInnen in gleichem Maße mit Information zu Sinnhaftigkeit und Bedeutung ihrer Aufgabe auszustatten. Das Ziel sollte schließlich sein, möglichst testleiterunabhängige Angaben und weitgehend vollständige Datensätze zu erhalten. Bis zum gegenwärtigen Zeitpunkt dürften die BeurteilerInnen sowohl über

interne Kanäle als auch über die schriftliche Einleitung zu Beginn des Beurteilungsbogens instruiert worden sein – offen bleibt dabei, ob, wann und was auch wirklich gelesen wurde und wie mit eventuellen Vorbehalten bezüglich des (nicht aus eigenem Antrieb anzuwendenden) Verfahrens umgegangen wurde. Nach Kenntnisstand der Verfasserin konnte beim Austeilen der Bögen durch die Kolleginnen Hasenhindl und Kremser in einigen Fällen persönlicher Kontakt mit den PädagogInnen aufgenommen werden, in anderen dagegen nicht. Es liegt die Vermutung nahe, dass die BeurteilerInnen infolge möglicherweise nicht auf dem gleichen Informations-, und wahrscheinlich nicht auf dem gleichen Motivationsstand waren. Nachdem für zukünftige Erhebungen nicht davon auszugehen ist, dass im Vorfeld flächendeckend persönliche Gespräche mit den RaterInnen stattfinden werden können, sollte das Augenmerk daher auf eine einheitliche, prägnante und vor allem motivierende schriftliche Vorinformation gelegt werden. Eventuell könnten für das fertige Verfahren sogar Werbemaßnahmen in Betracht gezogen werden, um das Commitment der PädagogInnen zu erhöhen, noch bevor sie sich konkret mit der Bearbeitung auseinandersetzen müssen.

Im Zusammenhang mit den TestleiterInnen soll an dieser Stelle noch auf die mögliche Problematik einer für sie anderen Erstsprache als Deutsch eingegangen werden: Wie aus einigen Anmerkungen am Ende der Testbögen deutlich wird, kann es vorkommen, dass Hortkinder von PädagogInnen mit nicht deutscher Erstsprache eingeschätzt werden sollen. Daraus ergibt sich die Frage, inwieweit es für die betreffenden PädagogInnen zumutbar ist, bei vielen Items verbal sehr fein differenzieren zu sollen und gegebenenfalls sogar Aussagen über die Sprachkompetenz ihrer ebenfalls nicht deutsch(erst-)sprachigen Schützlinge treffen zu müssen. Diese möglicherweise unangenehme Situation kann seitens der Testentwickler nicht vermieden werden; zumindest könnte aber jedenfalls die Erstsprache der RaterInnen zukünftig ebenso miterhoben werden wie die der Hortkinder.

Insgesamt bieten das Ziel und die voranschreitende Verwirklichung der Testentwicklung des ESH 6-10 eine ausgesprochen erfreuliche Perspektive auf die Möglichkeiten der Früherkennung von entwicklungspsychologischen Auffälligkeiten bei Volksschulkindern. Jedes einzelne Kind, das von der Arbeit an diesem Konzept profitieren kann, mag auch für künftige DiplomandInnen bzw. ProjektmitarbeiterInnen eine schöne Motivation sein!

9. Literaturverzeichnis

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2011). Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Heidelberg: Springer.

Bortz, J. & Döring, N. (2002). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer.

Bortz, J. & Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler. Heidelberg: Springer.

Brosius, F. (2011). SPSS 19. Heidelberg: mitp.

Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion. München: Pearson.

Field, A. (2009). Discovering statistics using SPSS. London: Sage.

Guadagnoli, E. & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. In: *Psychological Bulletin, Volume 103(2),* 265-275.

Hasenhindl, L. M. (2015). Validierung einer Ratingskala für HortpädagogInnen: Die Entwicklungsdimensionen Arbeitsverhalten und Sprache. Unveröffentlichte Diplomarbeit, Universität Wien.

Jöreskog, K. G. (2007). Factor analysis and its extensions. In: Cudeck, R. & MacCallum, R. C. (Hrsg.), *Factor analysis at 100: Historical developments and future directions* (S. 47-77). Mahwah, New Jersey: Lawrence Erlbaum.

Kremser, G. (2015). Validierung einer Ratingskala für den Einsatz im Hort (Altersgruppe 6 bis 10 Jahre) mit einem besonderen Fokus auf die sozial-emotionale Entwicklung. Unveröffentlichte Diplomarbeit, Universität Wien.

Kubinger, K. D. (2006). Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens. Göttingen: Hogrefe.

Kunst, M. C. (2014). Entwurf eines Entwicklungsscreenings für den Hort. Diplomarbeit, Universität Wien.

Lienert, G. A. & Raatz, U. (1998). Testaufbau und Testanalyse. Weinheim: Psychologie Verlags Union.

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

Moosbrugger, H. & Hartig, J. (2003b). Testtheorie, Klassische. In: Kubinger, K. D. & Jäger, R. (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 408-415). Weinheim: Beltz.

Matschiner, F. (2015). Entwicklung und Erprobung eines Entwicklungseinschätzungsbogens für Hortbetriebe. Diplomarbeit, Universität Wien.

Neugschwentner, S. (2014). Konzipierung des Entwicklungsscreenings für Horte für 6-10 Jährige zu den Bereichen Persönlichkeit und Motivation. Diplomarbeit, Universität Wien.

Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? In *Psychologische Rundschau, 50,* 140-156.

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

Schendera, C. F. G. (2007). Datenqualität mit SPSS. München: Oldenbourg.

Schendera, C. F. G. (2010). Clusteranalyse mit SPSS: mit Faktorenanalyse. München: Oldenbourg.

Werner, C. (2014). Explorative Faktorenanalyse: Einführung und Analyse mit R. 12.5.2014. http://www.psychologie.uzh.ch/fachrichtungen/methoden/team/christinawerner/faktorenanalyse/explorative_faktorenanalyse_mit_r_cswerner.pdf [Zugriff: 27.11.2015].

Woike, J. K. (2003a). Screening. In: Kubinger, K. D. & Jäger, R. (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 375-376). Weinheim: Beltz.

Woike, J. K. (2003b). Testökonomie. In: Kubinger, K. D. & Jäger, R. (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 405-407). Weinheim: Beltz.

III. Anhang

Anhang A: detaillierte Ergebnisse

Erstsprache des Kindes

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	deutsch	98	66,7	68,1	68,1
	deutsch/persisch/englisch	1	,7	,7	68,8
	polnisch	5	3,4	3,5	72,2
	rumänisch	2	1,4	1,4	73,6
	slowakisch	1	,7	,7	74,3
	finnisch/deutsch	1	,7	,7	75,0
	armenisch	3	2,0	2,1	77,1
	englisch/deutsch	2	1,4	1,4	78,5
	türkisch	9	6,1	6,3	84,7
	ungarisch	3	2,0	2,1	86,8
	Tagalog (philippinisch)	1	,7	,7	87,5
	chinesisch	1	,7	,7	88,2
	albanisch/deutsch	1	,7	,7	88,9
	englisch	1	,7	,7	89,6
	russisch	5	3,4	3,5	93,1
	deutsch/slowenisch	1	,7	,7	93,8
	serbisch	2	1,4	1,4	95,1
	amharisch (Äthiopien, Eritrea)	1	,7	,7	95,8
	italienisch	1	,7	,7	96,5
	spanisch	1	,7	,7	97,2
	arabisch	4	2,7	2,8	100,0
	Gesamtsumme	144	98,0	100,0	
Fehlend	System	3	2,0		
Gesamtsumme		147	100,0		

Deskriptive Statistiken der aktuellen Stichprobe

	N	Minimum	Maximum	Mittelwert	Standardabw.	Schiefe		Kurtosis	
	Statistik	Statistik	Statistik	Statistik	Statistik	Statistik	Standard-fehler	Statistik	Standard-fehler
Item 1	143	1,00	5,00	3,7902	1,18591	-,766	,203	-,291	,403
Item 2	143	1,00	5,00	3,9161	1,20735	-,811	,203	-,450	,403
Item 3	140	1,00	5,00	3,3143	1,16351	,029	,205	-,962	,407
Item 6	142	1,00	5,00	4,1620	1,14630	-1,240	,203	,604	,404
Item 7	135	1,00	5,00	3,8889	1,35309	-,914	,209	-,470	,414
Item 8	124	1,00	5,00	3,4113	1,36156	-,331	,217	-1,167	,431
Item 9	127	1,00	5,00	3,6457	1,28191	-,663	,215	-,707	,427
Item 12	139	1,00	5,00	4,1007	1,21163	-1,312	,206	,795	,408
Item 13	143	1,00	5,00	2,9930	1,34529	,066	,203	-1,114	,403
Item 14	146	1,00	5,00	3,8014	1,22963	-,855	,201	-,182	,399
Item 16	141	1,00	5,00	3,1277	1,34085	-,182	,204	-1,164	,406
Item 17	136	1,00	5,00	2,9632	1,33004	,068	,208	-1,107	,413
Item 18	139	2,00	5,00	4,4317	,84313	-1,475	,206	1,443	,408
Item 19	140	1,00	5,00	3,7143	1,33727	-,744	,205	-,635	,407
Item 23	138	1,00	5,00	4,0072	1,14303	-1,056	,206	,280	,410
Item 24	138	1,00	5,00	3,9275	1,22408	-,975	,206	-,099	,410
Item 25	141	1,00	5,00	4,0000	1,11484	-1,098	,204	,613	,406
Item 26	143	1,00	5,00	4,3147	,99590	-1,536	,203	1,964	,403
Item 27	141	1,00	5,00	4,4823	,85023	-1,748	,204	2,686	,406
Item 28	142	2,00	5,00	4,5634	,82912	-1,984	,203	3,096	,404
Item 29	126	1,00	5,00	3,9048	1,31714	-,953	,216	-,340	,428
Item 30	146	1,00	5,00	4,0137	1,09536	-,953	,201	,071	,399
Item 31	147	1,00	5,00	3,8231	1,17450	-,781	,200	-,306	,397
Item 32	144	2,00	5,00	4,2847	,81647	-1,039	,202	,588	,401
Item 33	145	1,00	5,00	3,8414	1,11602	-,684	,201	-,440	,400
Item 35	145	1,00	5,00	4,1655	,95753	-,964	,201	,362	,400
Item 36	145	1,00	5,00	4,2690	,98075	-1,236	,201	,804	,400
Item 37	147	2,00	5,00	4,3435	,78906	-1,124	,200	,820	,397
Item 38	147	1,00	5,00	4,1429	,98621	-,900	,200	-,116	,397
Item 39	142	2,00	5,00	4,4507	,83859	-1,379	,203	,903	,404
Item 40	141	1,00	5,00	4,2482	1,07674	-1,347	,204	,955	,406
Item 41	140	1,00	5,00	4,0857	1,00707	-1,117	,205	,771	,407
Item 43	140	1,00	5,00	2,7143	1,13977	,227	,205	-,452	,407
Item 44	139	1,00	5,00	3,2302	1,04477	-,166	,206	-,085	,408
Item 47	146	1,00	5,00	3,7466	1,23630	-,726	,201	-,459	,399
Item 48	146	1,00	5,00	3,9932	1,04053	-,917	,201	,303	,399
Item 50	142	1,00	5,00	4,3979	,93691	-1,470	,203	1,271	,404
Item 51	141	1,00	5,00	4,1135	1,11543	-,979	,204	-,139	,406
Item 52	147	1,00	5,00	4,2925	,95236	-1,245	,200	,730	,397

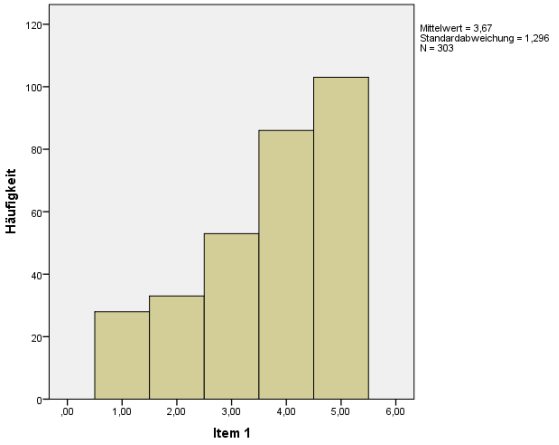
Item 53	147	1,00	5,00	4,2721	,96910	-1,212	,200	,783	,397
Item 55	144	1,00	5,00	3,8264	1,30266	-,693	,202	-,772	,401
Item 56	146	1,00	5,00	4,0616	1,07782	-,928	,201	-,048	,399
Item 57a	145	,00	1,00	,7586	,42940	-1,221	,201	-,515	,400
Item 57b	145	,00	1,00	,7586	,42940	-1,221	,201	-,515	,400
Item 57c	143	,00	1,00	,8881	,31634	-2,489	,203	4,252	,403
Item 58a	138	,00	1,00	,8333	,37404	-1,809	,206	1,289	,410
Item 58b	135	,00	1,00	,7556	,43136	-1,203	,209	-,562	,414
Item 58c	142	,00	1,00	,7535	,43249	-1,189	,203	-,595	,404
Item 59	144	1,00	5,00	3,9306	1,30439	-,828	,202	-,714	,401
Item 60	145	1,00	5,00	4,2276	1,09128	-1,473	,201	1,473	,400
Item 61	143	1,00	5,00	4,3846	1,02038	-1,679	,203	1,970	,403
Item 63	147	1,00	5,00	4,1769	1,03831	-1,367	,200	1,286	,397
Item 64	147	1,00	5,00	4,2993	,98908	-1,279	,200	,631	,397
Item 65	144	1,00	5,00	4,5000	,90839	-2,242	,202	5,140	,401
Item 66	146	1,00	5,00	3,9178	,99313	-,690	,201	-,149	,399
Item 67	141	1,00	5,00	3,3369	1,23128	-,359	,204	-,790	,406
Item 68	139	1,00	5,00	3,5180	1,25890	-,461	,206	-,833	,408
Item 69	135	1,00	5,00	3,7407	1,12610	-,713	,209	-,242	,414
Item 71	137	1,00	5,00	4,3102	1,01845	-1,464	,207	1,422	,411
Item 72	134	2,00	5,00	4,0149	,87561	-,507	,209	-,546	,416
Item 73	141	1,00	5,00	4,0426	1,15803	-1,092	,204	,259	,406
Item 74	147	1,00	5,00	4,3333	,88622	-1,428	,200	1,977	,397
Item 75	139	1,00	5,00	4,3165	1,00748	-1,276	,206	,497	,408
Item 76	147	2,00	5,00	4,2993	,86337	-1,075	,200	,349	,397
Item 79	145	1,00	5,00	3,5172	1,42937	-,433	,201	-1,243	,400
Item 80	140	1,00	5,00	4,3000	1,11852	-1,587	,205	1,600	,407
Item 81	140	1,00	5,00	4,5143	,94058	-2,094	,205	3,866	,407
Item 82	142	1,00	5,00	3,8239	1,28989	-,812	,203	-,529	,404
Item 83	143	1,00	5,00	3,8741	1,29390	-,910	,203	-,280	,403
Item 84	146	1,00	5,00	4,3082	1,07360	-1,491	,201	1,426	,399
Item 85	146	1,00	5,00	4,0719	1,20200	-1,065	,201	-,024	,399
Item 86	144	1,00	5,00	4,2917	1,06354	-1,598	,202	1,853	,401
Item 88	139	1,00	5,00	4,3237	1,01593	-1,571	,206	1,936	,408
Item 87	147	,00	1,00	,9320	,25265	-3,467	,200	10,156	,397
Gültige Anzahl (listen-weise)	66								

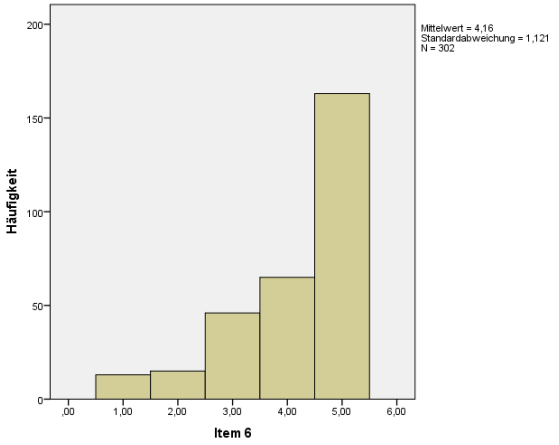
Deskriptive Statistiken der Gesamtstichprobe

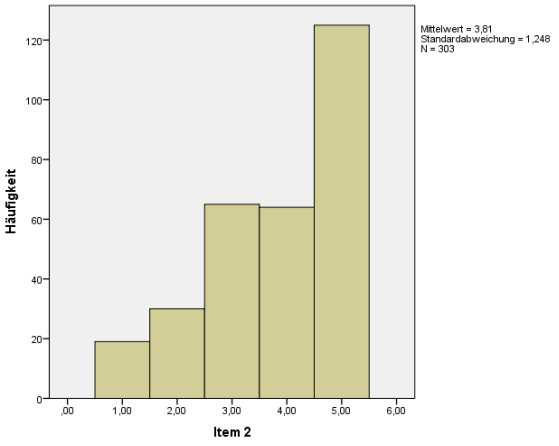
	N	Minimum	Maximum	Mittelwert	Standardabw.	Schiefe		Kurtosis	
	Statistik	Statistik	Statistik	Statistik	Statistik	Statistik	Standard-fehler	Statistik	Standard-fehler
Item 1	303	1,00	5,00	3,6700	1,29568	-,701	,140	-,617	,279
Item 2	303	1,00	5,00	3,8119	1,24788	-,730	,140	-,539	,279
Item 3	297	1,00	5,00	3,1919	1,15088	-,020	,141	-,769	,282
Item 6	302	1,00	5,00	4,1589	1,12118	-1,271	,140	,798	,280
Item 7	293	1,00	5,00	3,9488	1,31148	-1,042	,142	-,144	,284
Item 8	261	1,00	5,00	3,3831	1,34119	-,399	,151	-1,034	,300
Item 9	278	1,00	5,00	3,6655	1,22826	-,705	,146	-,510	,291
Item 12	297	1,00	5,00	3,9428	1,31008	-1,028	,141	-,143	,282
Item 13	303	1,00	5,00	2,9373	1,41867	,020	,140	-1,280	,279
Item 14	305	1,00	5,00	3,6131	1,29821	-,587	,140	-,782	,278
Item 16	285	1,00	5,00	3,0140	1,34262	-,052	,144	-1,202	,288
Item 17	287	1,00	5,00	2,9303	1,26906	,090	,144	-1,006	,287
Item 18	296	1,00	5,00	4,3074	,90390	-1,284	,142	1,091	,282
Item 19	270	1,00	5,00	3,4593	1,25407	-,443	,148	-,783	,295
Item 23	298	1,00	5,00	3,8624	1,16538	-,886	,141	-,051	,281
Item 24	296	1,00	5,00	3,8818	1,21413	-,962	,142	-,057	,282
Item 25	297	1,00	5,00	3,8990	1,22056	-,973	,141	-,025	,282
Item 26	302	1,00	5,00	4,1689	1,17333	-1,413	,140	1,063	,280
Item 27	300	1,00	5,00	4,4000	,92168	-1,703	,141	2,587	,281
Item 28	298	1,00	5,00	4,4631	,86475	-1,741	,141	2,584	,281
Item 29	266	1,00	5,00	3,8684	1,23837	-,937	,149	-,135	,298
Item 30	306	1,00	5,00	3,9346	1,11722	-,921	,139	,075	,278
Item 31	304	1,00	5,00	3,6941	1,16968	-,656	,140	-,417	,279
Item 32	301	2,00	5,00	4,1960	,82348	-,811	,140	,057	,280
Item 33	304	1,00	5,00	3,8289	1,07952	-,732	,140	-,199	,279
Item 35	304	1,00	5,00	4,1283	,99503	-1,110	,140	,886	,279
Item 36	304	1,00	5,00	4,1645	1,03372	-1,074	,140	,361	,279
Item 37	307	2,00	5,00	4,2329	,79746	-,796	,139	,012	,277
Item 38	307	1,00	5,00	4,0130	1,01612	-,797	,139	-,163	,277
Item 39	300	1,00	5,00	4,4067	,87014	-1,382	,141	1,130	,281
Item 40	299	1,00	5,00	4,1338	1,14787	-1,122	,141	,189	,281
Item 41	296	1,00	5,00	3,9189	1,04485	-,842	,142	,116	,282
Item 43	291	1,00	5,00	2,9863	1,21193	,026	,143	-,730	,285
Item 44	288	1,00	5,00	2,8993	1,12321	,021	,144	-,487	,286
Item 47	306	1,00	5,00	3,6732	1,22718	-,695	,139	-,503	,278
Item 48	306	1,00	5,00	3,9248	1,07318	-,891	,139	,209	,278
Item 50	301	1,00	5,00	4,2973	,95680	-1,185	,140	,459	,280
Item 51	298	1,00	5,00	4,0537	1,17943	-,985	,141	-,145	,281

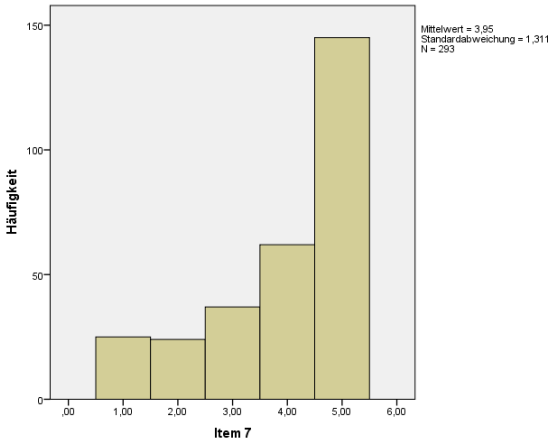
Item 52	306	1,00	5,00	4,1928	,97447	-1,144	,139	,765	,278
Item 53	307	1,00	5,00	4,2150	,98321	-1,148	,139	,596	,277
Item 55	303	1,00	5,00	3,7822	1,33425	-,648	,140	-,898	,279
Item 56	304	1,00	5,00	4,0362	1,05712	-,866	,140	-,095	,279
Item 57a	296	,00	1,00	,4764	,50029	,095	,142	-2,005	,282
Item 57b	290	,00	1,00	,5379	,49942	-,153	,143	-1,990	,285
Item 57c	296	,00	1,00	,4730	,50011	,109	,142	-2,002	,282
Item 58a	292	,00	1,00	,8562	,35153	-2,040	,143	2,178	,284
Item 58b	288	,00	1,00	,7778	,41646	-1,343	,144	-,197	,286
Item 58c	298	,00	1,00	,7483	,43471	-1,150	,141	-,682	,281
Item 59	303	1,00	5,00	3,9934	1,22878	-,947	,140	-,295	,279
Item 60	305	1,00	5,00	4,2820	,99299	-1,503	,140	1,844	,278
Item 61	303	1,00	5,00	4,4653	,94446	-1,929	,140	3,195	,279
Item 63	304	1,00	5,00	4,2467	,97239	-1,401	,140	1,555	,279
Item 64	306	1,00	5,00	4,3268	,95010	-1,411	,139	1,383	,278
Item 65	299	1,00	5,00	4,3512	1,00690	-1,662	,141	2,167	,281
Item 66	304	1,00	5,00	3,9145	1,05895	-,869	,140	,195	,279
Item 67	301	1,00	5,00	3,2043	1,25656	-,236	,140	-,913	,280
Item 68	298	1,00	5,00	3,4128	1,34869	-,377	,141	-1,073	,281
Item 69	289	1,00	5,00	3,2768	1,16942	-,173	,143	-,762	,286
Item 71	291	1,00	5,00	4,3351	,96950	-1,470	,143	1,554	,285
Item 72	285	1,00	5,00	3,6561	1,08819	-,488	,144	-,346	,288
Item 73	281	1,00	5,00	3,7829	1,18285	-,602	,145	-,674	,290
Item 74	307	1,00	5,00	4,2476	,91677	-1,305	,139	1,570	,277
Item 75	282	1,00	5,00	4,0922	1,18352	-1,088	,145	,054	,289
Item 76	306	1,00	5,00	4,1765	,93119	-1,143	,139	1,087	,278
Item 79	304	1,00	5,00	3,4243	1,45591	-,372	,140	-1,291	,279
Item 80	299	1,00	5,00	4,2174	1,20035	-1,446	,141	,955	,281
Item 81	298	1,00	5,00	4,4362	,99711	-1,782	,141	2,397	,281
Item 82	295	1,00	5,00	3,8068	1,23991	-,749	,142	-,506	,283
Item 83	302	1,00	5,00	3,7649	1,35704	-,747	,140	-,710	,280
Item 84	306	1,00	5,00	4,1275	1,13662	-1,250	,139	,809	,278
Item 85	304	1,00	5,00	3,8701	1,27830	-,893	,140	-,361	,279
Item 86	300	1,00	5,00	4,2800	,99913	-1,436	,141	1,517	,281
Item 88	294	1,00	5,00	4,2891	,99216	-1,408	,142	1,474	,283
Gültige Anzahl (listenweise)	113								

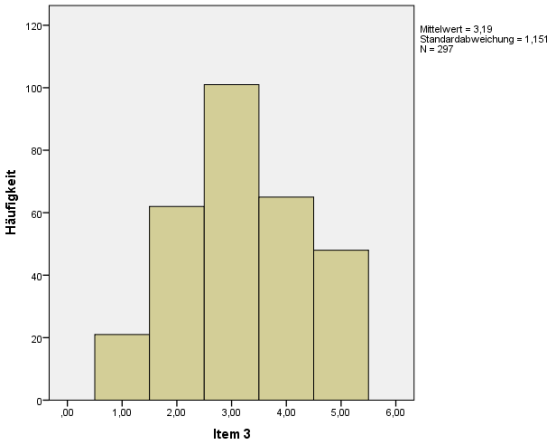
Histogramme Gesamtdatensatz

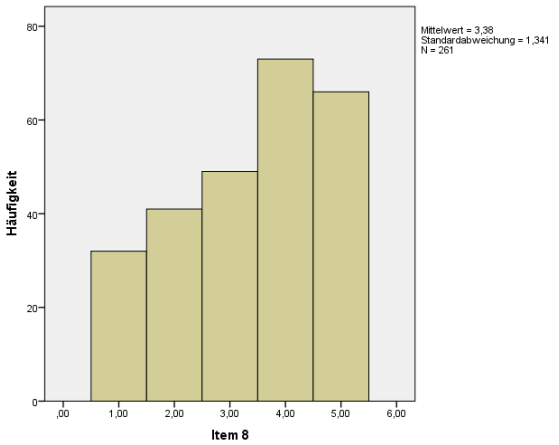


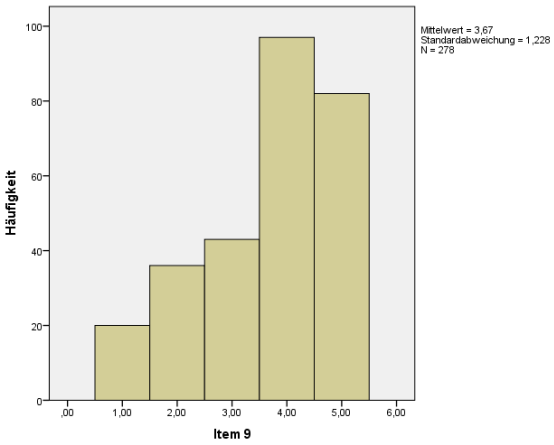


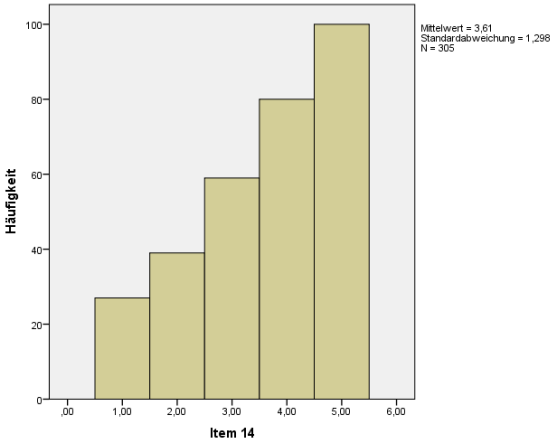


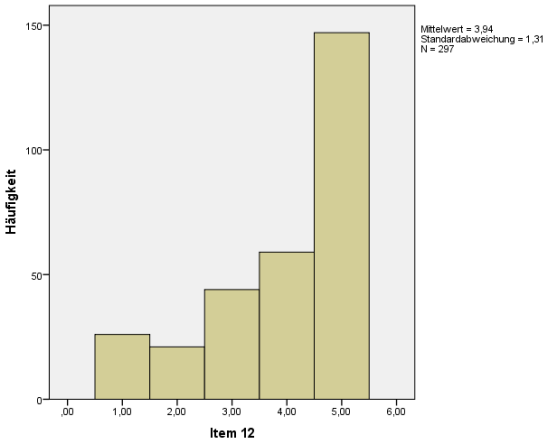


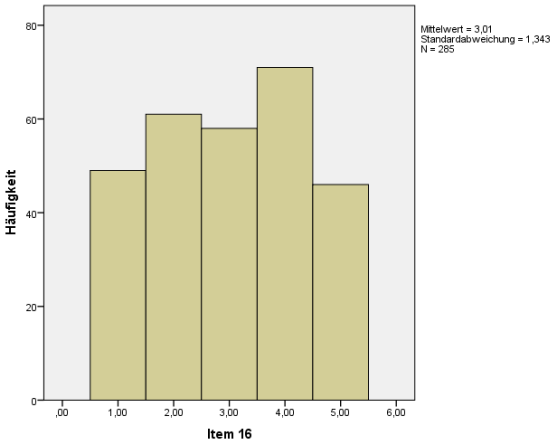


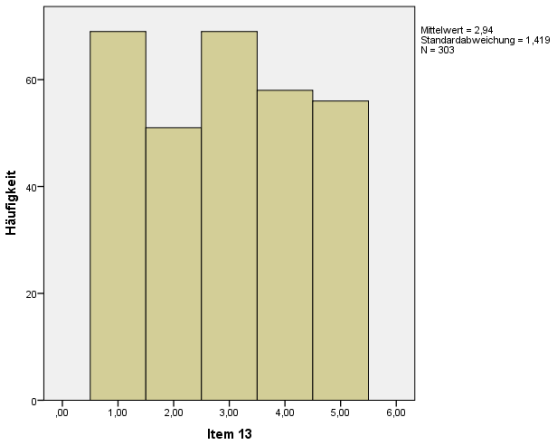


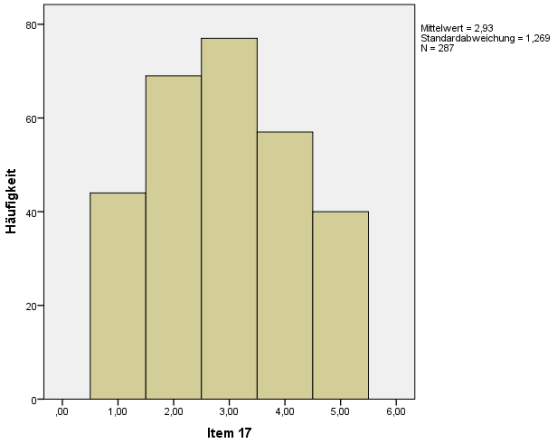

Mittelwert = 3,67
Standardabweichung = 1,228
N = 278
Häufigkeit
Item 9

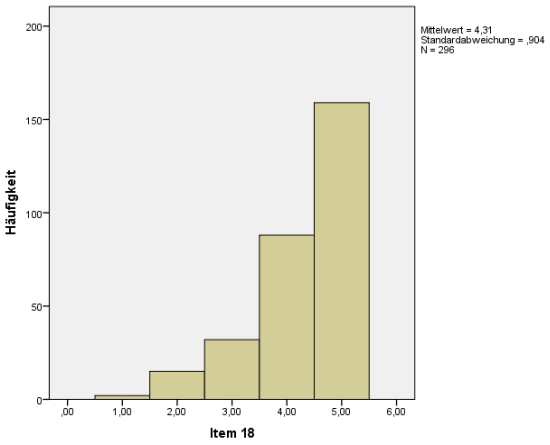

Mittelwert = 3,61
Standardabweichung = 1,298
N = 305
Häufigkeit
Item 14

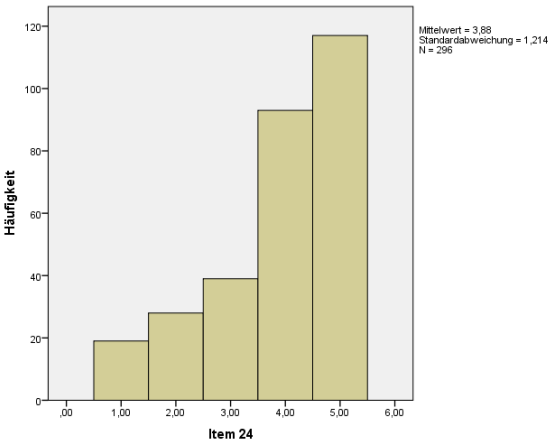

Mittelwert = 3,94
Standardabweichung = 1,31
N = 297
Häufigkeit
Item 12

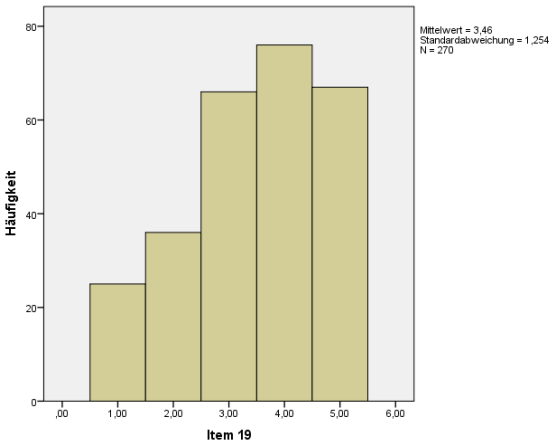

Mittelwert = 3,01
Standardabweichung = 1,343
N = 285
Häufigkeit
Item 16

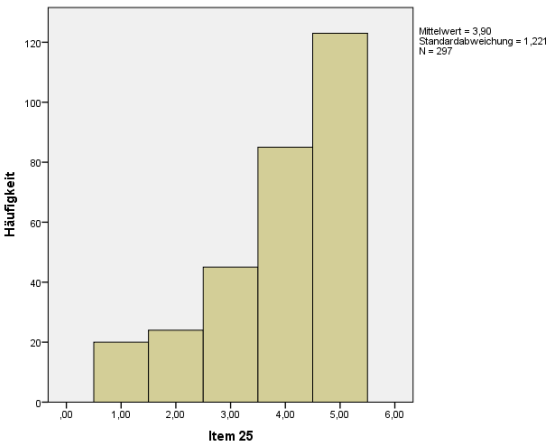

Mittelwert = 2,94
Standardabweichung = 1,419
N = 303
Häufigkeit
Item 13

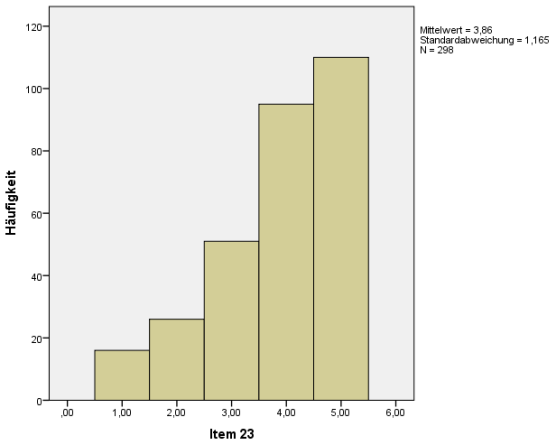

Mittelwert = 2,93
Standardabweichung = 1,269
N = 287
Häufigkeit
Item 17

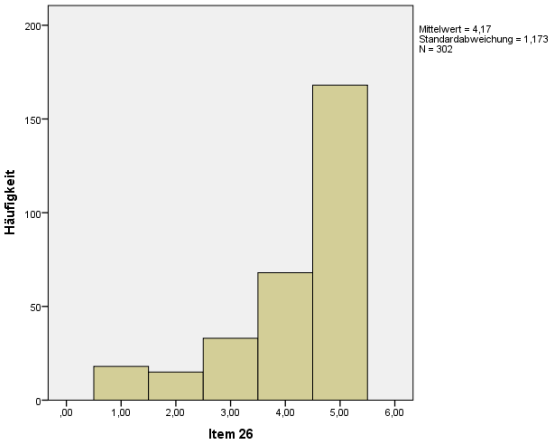

Mittelwert = 4,31
Standardabweichung = ,904
N = 296
Häufigkeit
Item 18

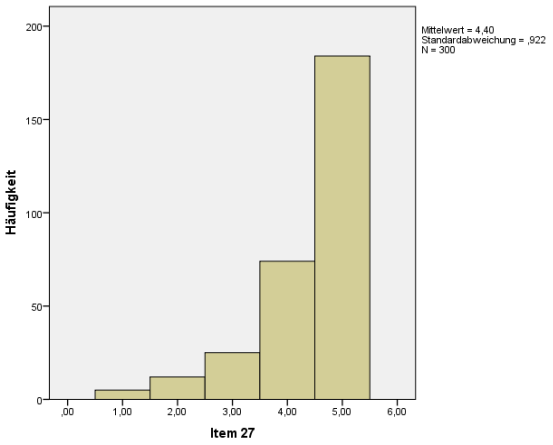

Mittelwert = 3,88
Standardabweichung = 1,214
N = 296
Häufigkeit
Item 24

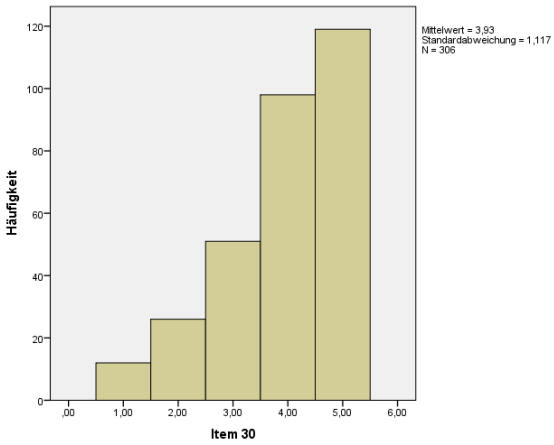

Mittelwert = 3,46
Standardabweichung = 1,254
N = 270
Häufigkeit
Item 19

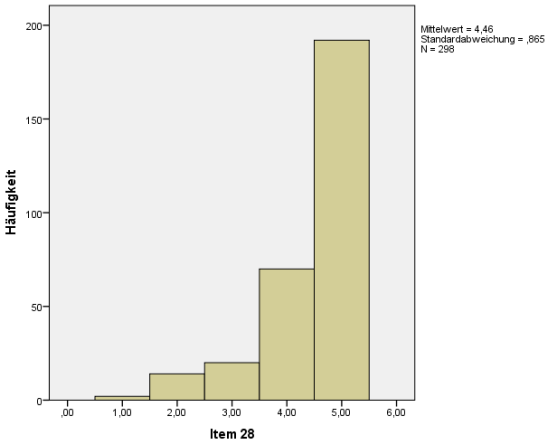

Mittelwert = 3,90
Standardabweichung = 1,221
N = 297
Häufigkeit
Item 25

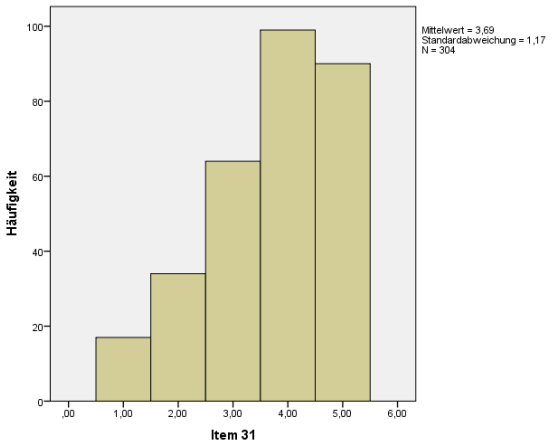

Mittelwert = 3,86
Standardabweichung = 1,165
N = 298
Häufigkeit
Item 23

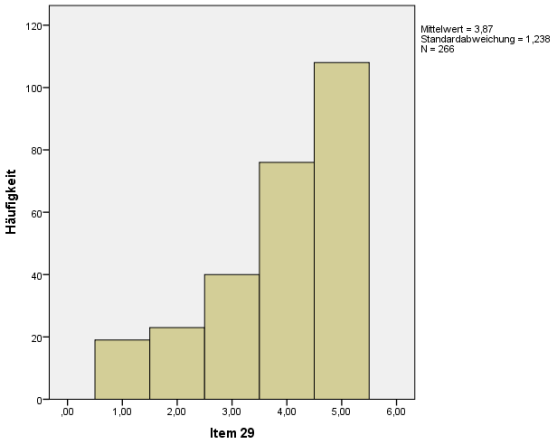

Mittelwert = 4,17
Standardabweichung = 1,173
N = 302
Häufigkeit
Item 26

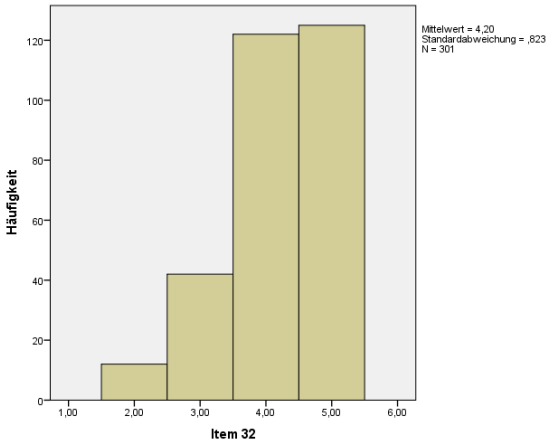

Mittelwert = 4,40
Standardabweichung = ,922
N = 300
Häufigkeit
Item 27

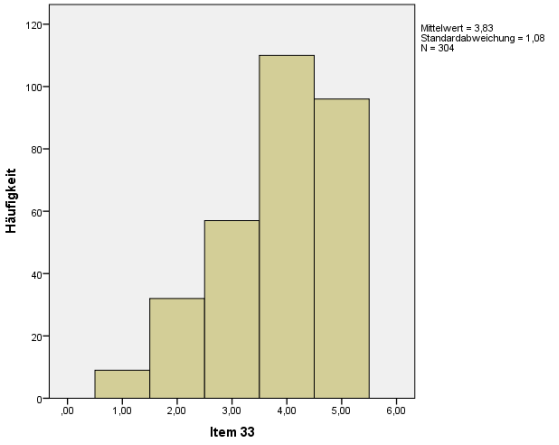

Mittelwert = 3,93
Standardabweichung = 1,117
N = 306
Häufigkeit
Item 30

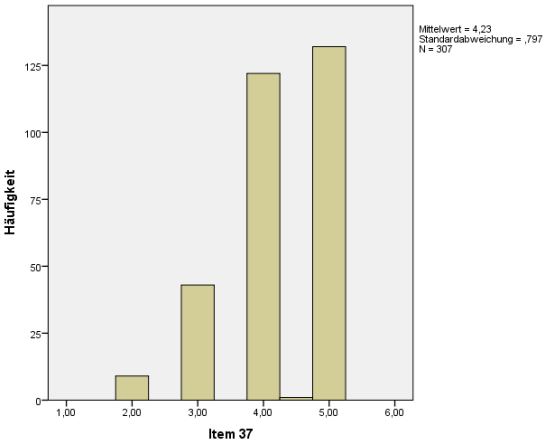

Mittelwert = 4,46
Standardabweichung = ,865
N = 298
Häufigkeit
Item 28

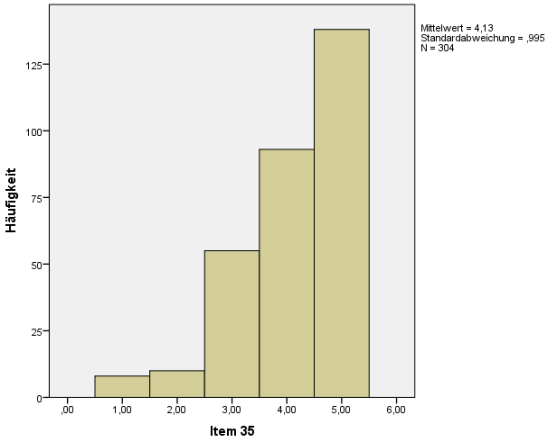

Mittelwert = 3,69
Standardabweichung = 1,17
N = 304
Häufigkeit
Item 31

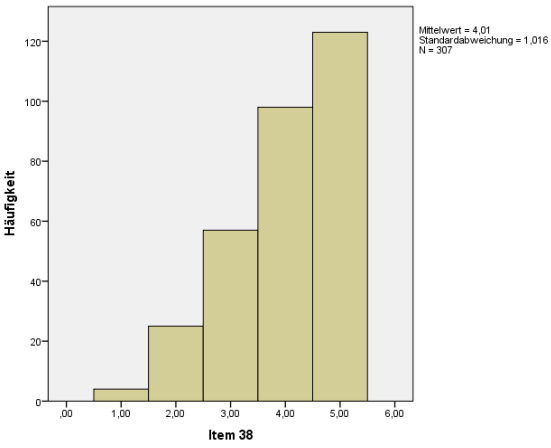

Mittelwert = 3,87
Standardabweichung = 1,238
N = 266
Häufigkeit
Item 29

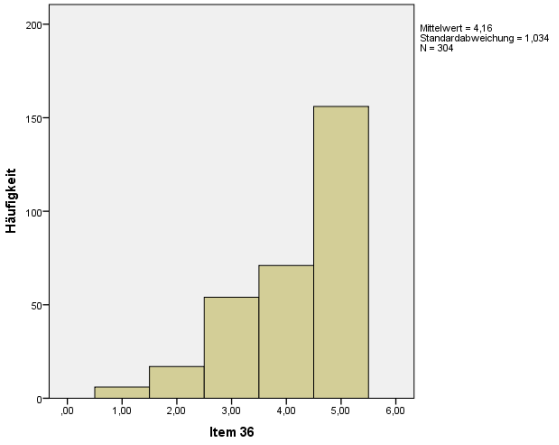

Mittelwert = 4,20
Standardabweichung = ,823
N = 301
Häufigkeit
Item 32

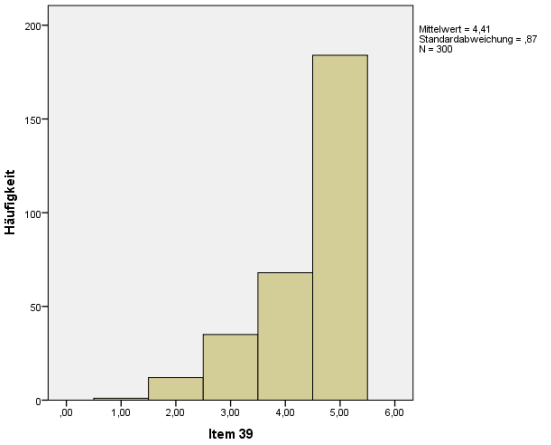

Mittelwert = 3,83
Standardabweichung = 1,08
N = 304
Häufigkeit
Item 33

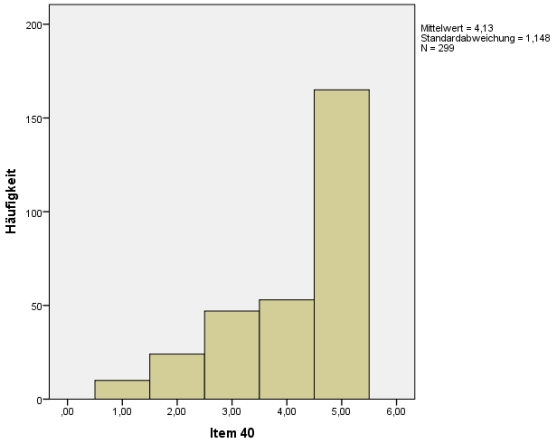

Mittelwert = 4,23
Standardabweichung = ,797
N = 307
Häufigkeit
Item 37

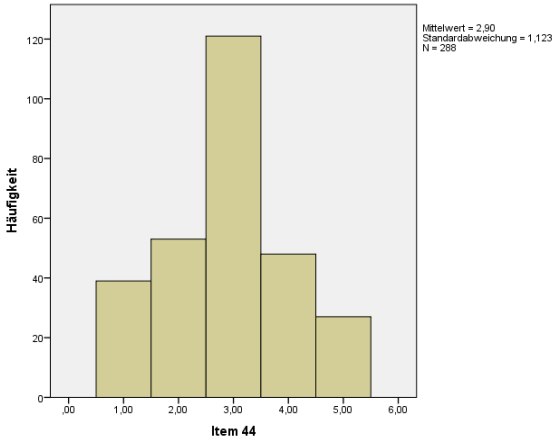

Mittelwert = 4,13
Standardabweichung = ,995
N = 304
Häufigkeit
Item 35

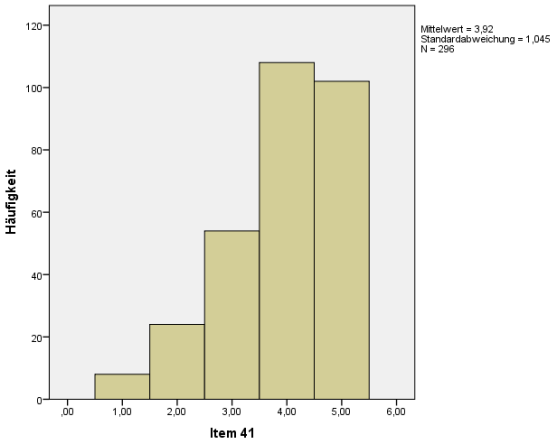

Mittelwert = 4,01
Standardabweichung = 1,016
N = 307
Häufigkeit
Item 38

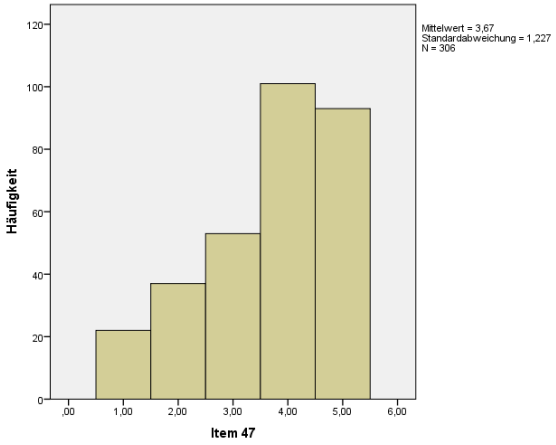

Mittelwert = 4,16
Standardabweichung = 1,034
N = 304
Häufigkeit
Item 36

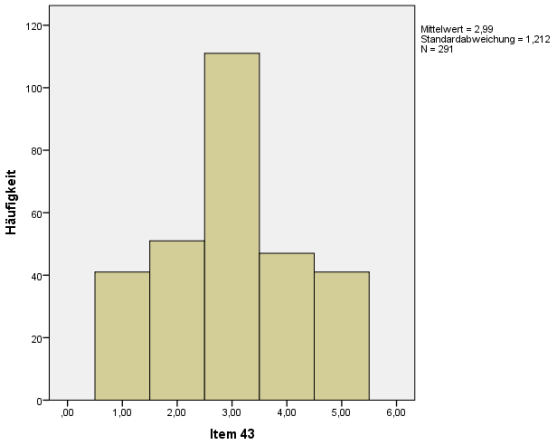

Mittelwert = 4,41
Standardabweichung = ,87
N = 300
Häufigkeit
Item 39

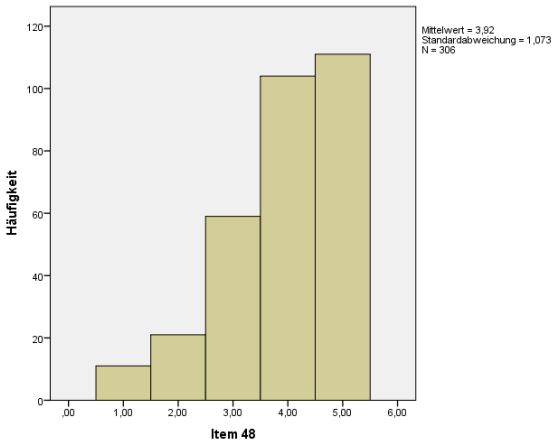

Mittelwert = 4,13
Standardabweichung = 1,148
N = 299
Häufigkeit
Item 40

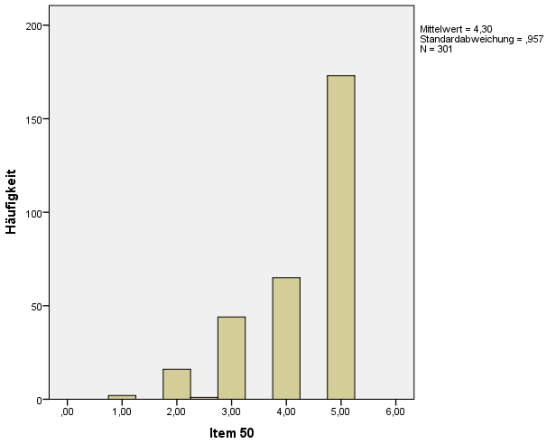

Mittelwert = 2,90
Standardabweichung = 1,123
N = 288
Häufigkeit
Item 44

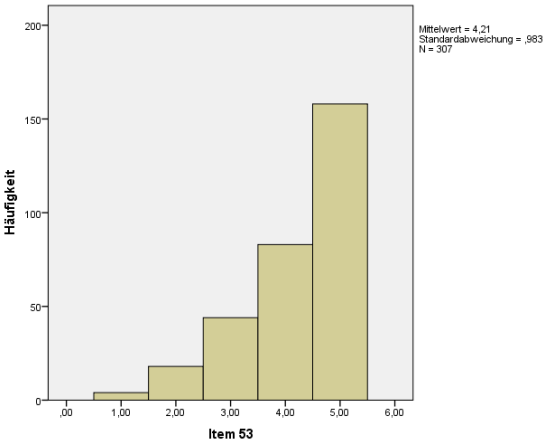

Mittelwert = 3,92
Standardabweichung = 1,045
N = 296
Häufigkeit
Item 41

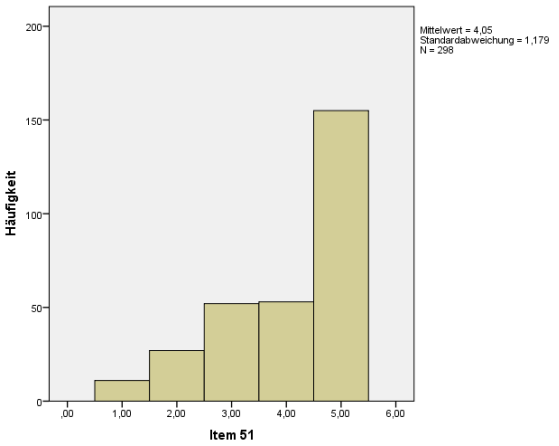

Mittelwert = 3,67
Standardabweichung = 1,227
N = 306
Häufigkeit
Item 47

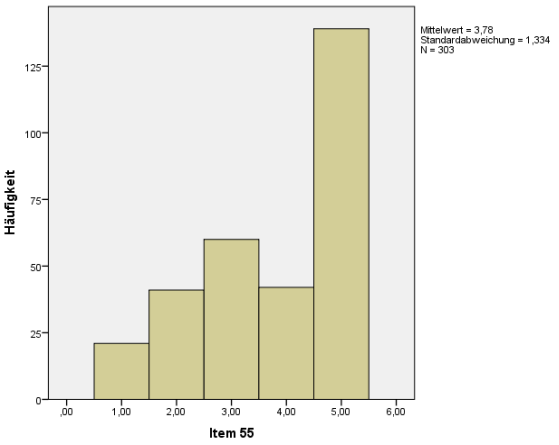

Mittelwert = 2,99
Standardabweichung = 1,212
N = 291
Häufigkeit
Item 43

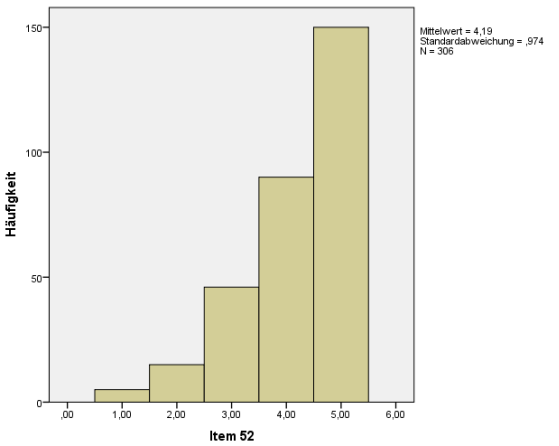

Mittelwert = 3,92
Standardabweichung = 1,073
N = 306
Häufigkeit
Item 48

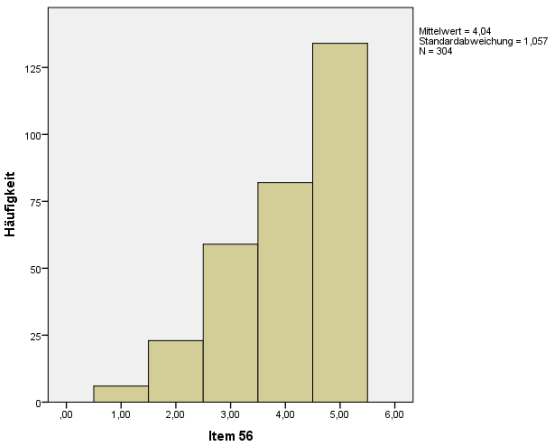

Mittelwert = 4,30
Standardabweichung = ,957
N = 301
Häufigkeit
Item 50

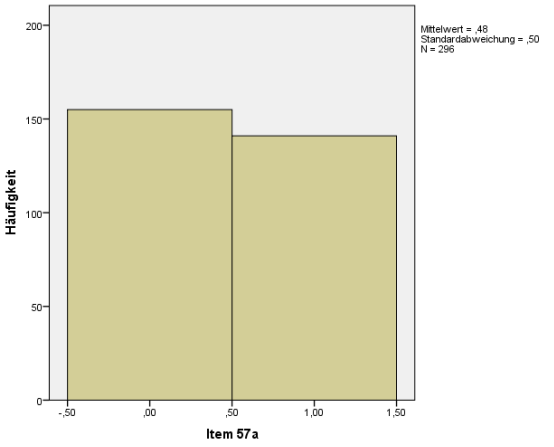

Mittelwert = 4,21
Standardabweichung = ,983
N = 307
Häufigkeit
Item 53

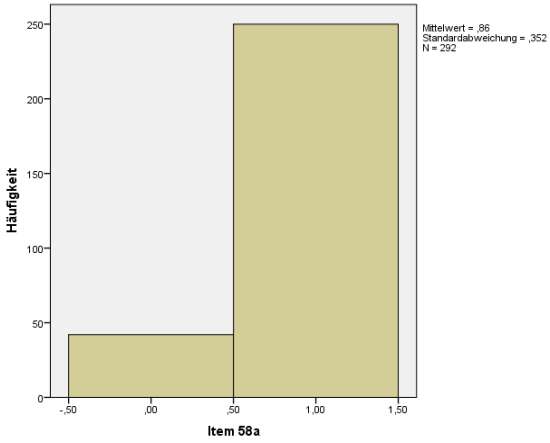

Mittelwert = 4,05
Standardabweichung = 1,179
N = 298
Häufigkeit
Item 51

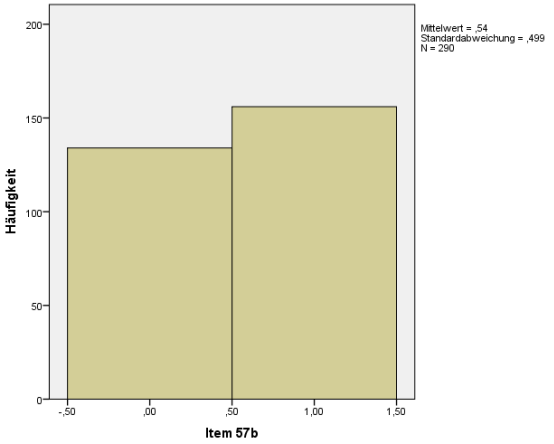

Mittelwert = 3,78
Standardabweichung = 1,334
N = 303
Häufigkeit
Item 55

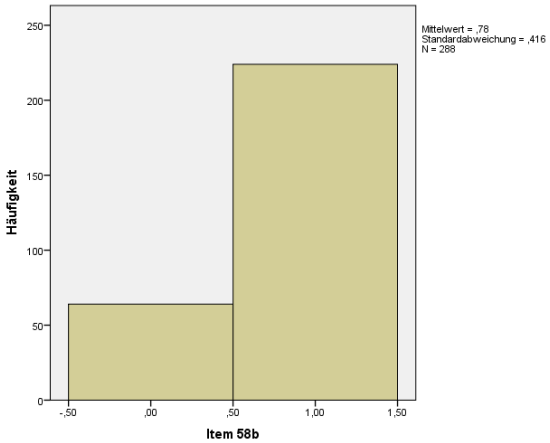

Mittelwert = 4,19
Standardabweichung = ,974
N = 306
Häufigkeit
Item 52

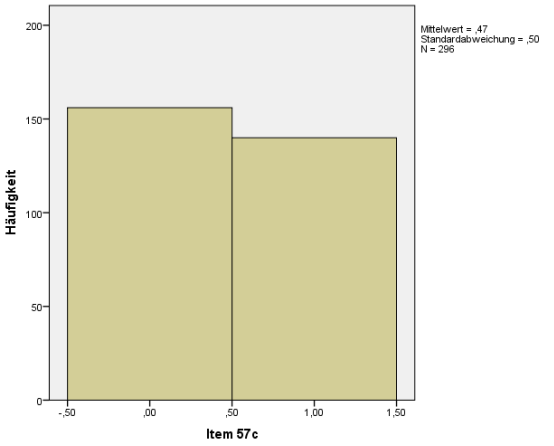

Mittelwert = 4,04
Standardabweichung = 1,057
N = 304
Häufigkeit
Item 56

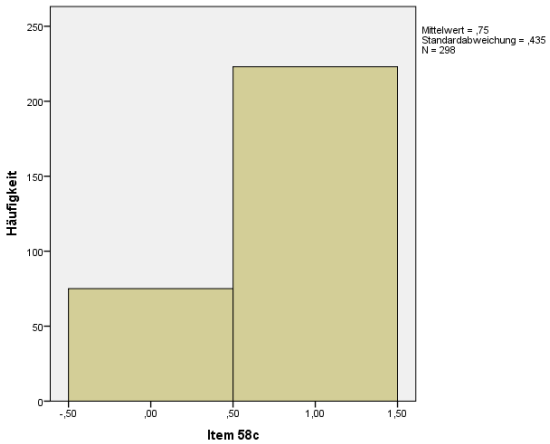

Mittelwert = ,48
Standardabweichung = ,50
N = 296
Häufigkeit
Item 57a

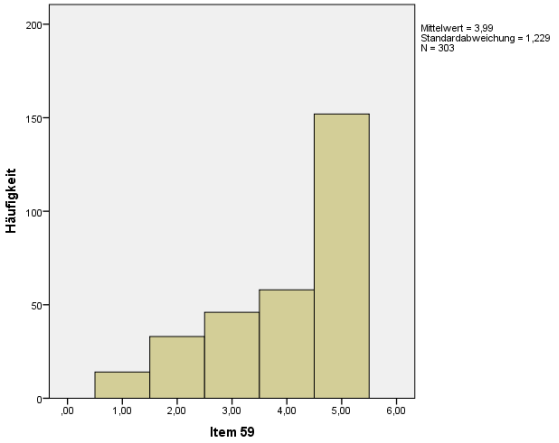

Mittelwert = ,86
Standardabweichung = ,352
N = 292
Häufigkeit
Item 58a

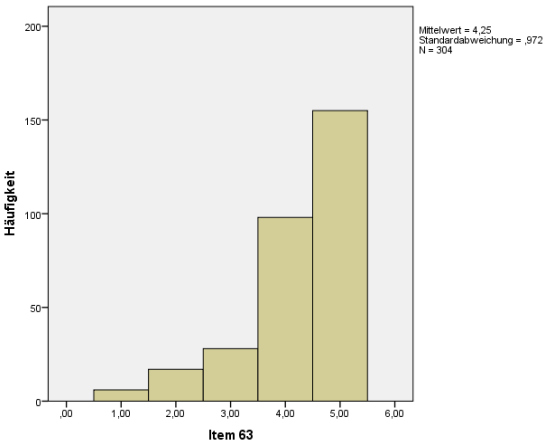

Mittelwert = ,54
Standardabweichung = ,499
N = 290
Häufigkeit
Item 57b

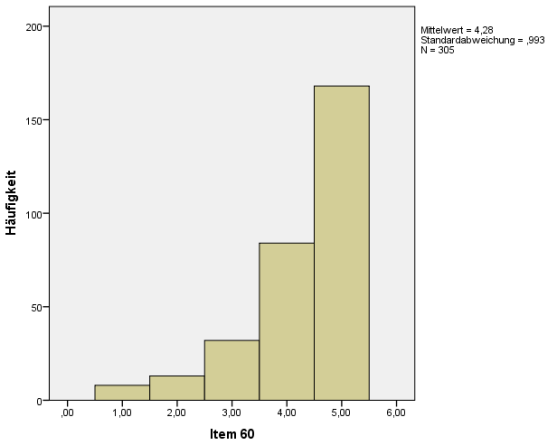

Mittelwert = ,78
Standardabweichung = ,416
N = 288
Häufigkeit
Item 58b

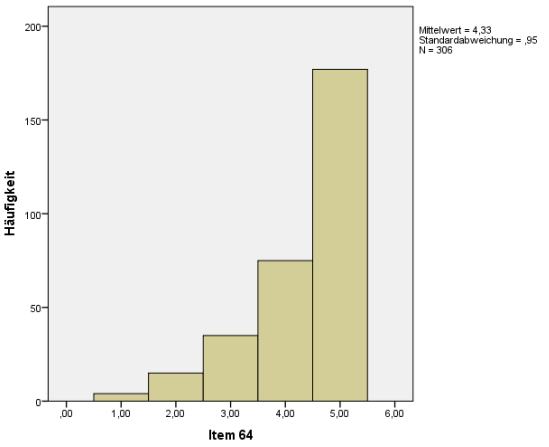

Mittelwert = ,47
Standardabweichung = ,50
N = 296
Häufigkeit
Item 57c

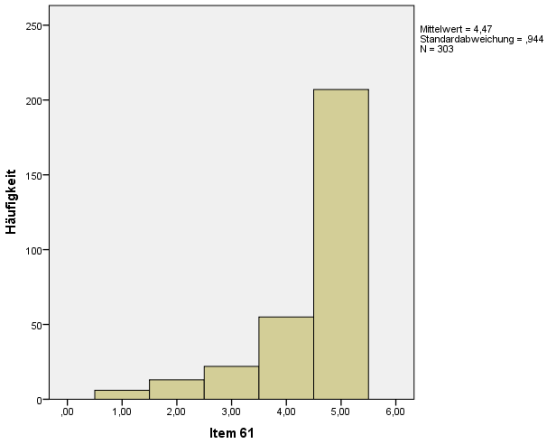

Mittelwert = ,75
Standardabweichung = ,435
N = 298
Häufigkeit
Item 58c

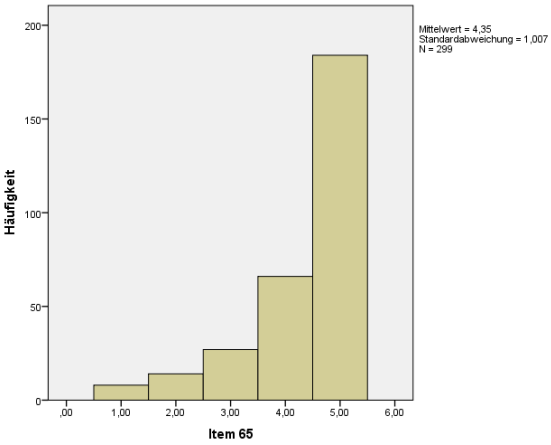

Mittelwert = 3,99
Standardabweichung = 1,229
N = 303
Häufigkeit
Item 59

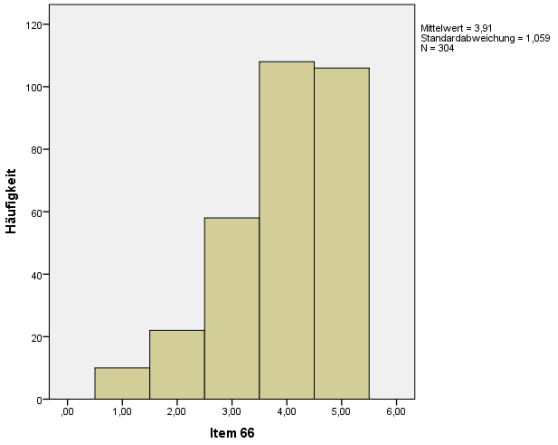

Mittelwert = 4,25
Standardabweichung = ,972
N = 304
Häufigkeit
Item 63

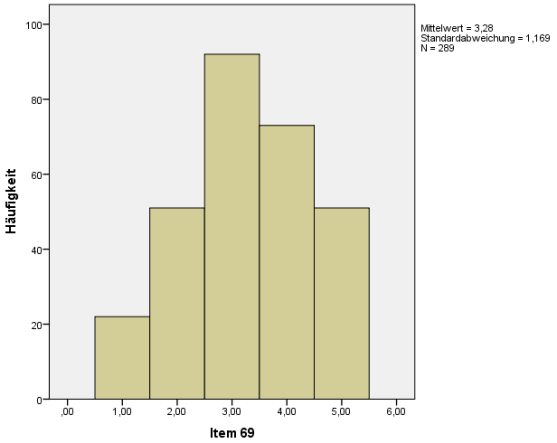

Mittelwert = 4,28
Standardabweichung = ,993
N = 305
Häufigkeit
Item 60

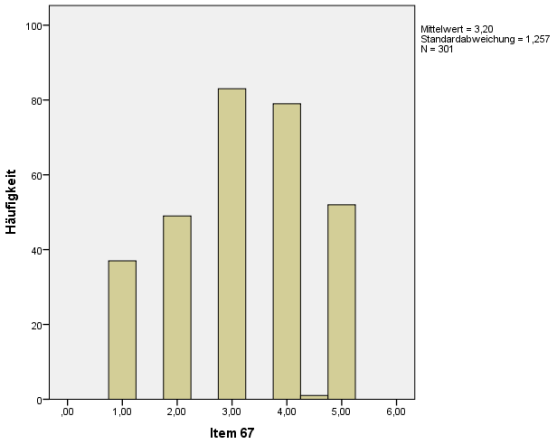

Mittelwert = 4,33
Standardabweichung = ,95
N = 306
Häufigkeit
Item 64

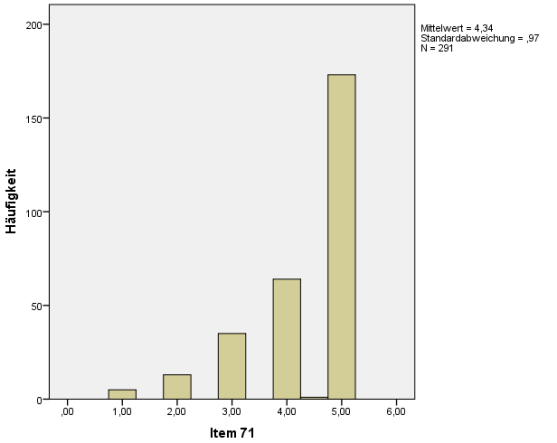

Mittelwert = 4,47
Standardabweichung = ,944
N = 303
Häufigkeit
Item 61

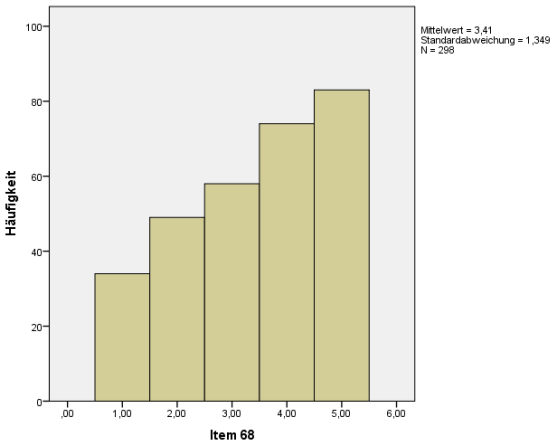

Mittelwert = 4,35
Standardabweichung = 1,007
N = 299
Häufigkeit
Item 65

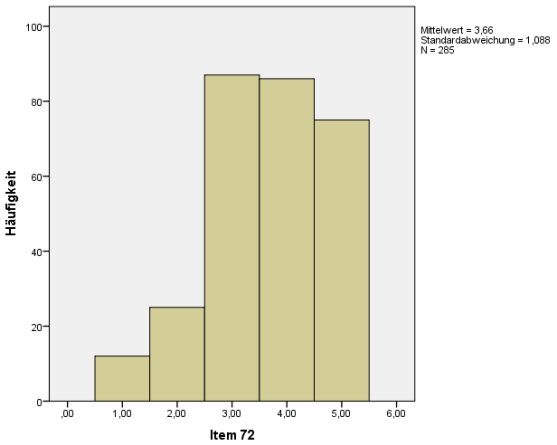

Mittelwert = 3,91
Standardabweichung = 1,059
N = 304
Häufigkeit
Item 66

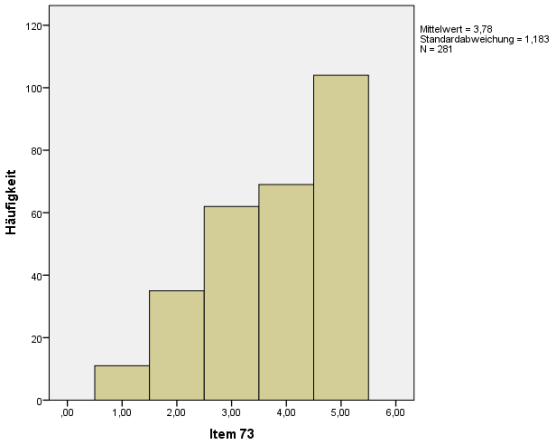

Mittelwert = 3,28
Standardabweichung = 1,169
N = 289
Häufigkeit
Item 69

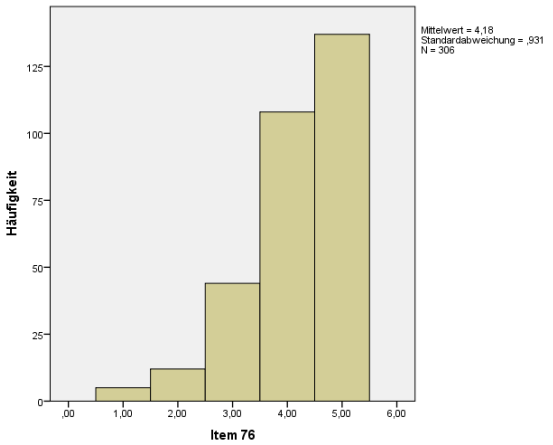

Mittelwert = 3,20
Standardabweichung = 1,257
N = 301
Häufigkeit
Item 67

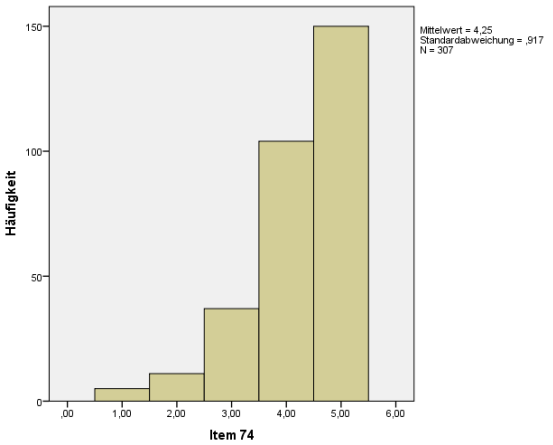

Mittelwert = 4,34
Standardabweichung = ,97
N = 291
Häufigkeit
Item 71

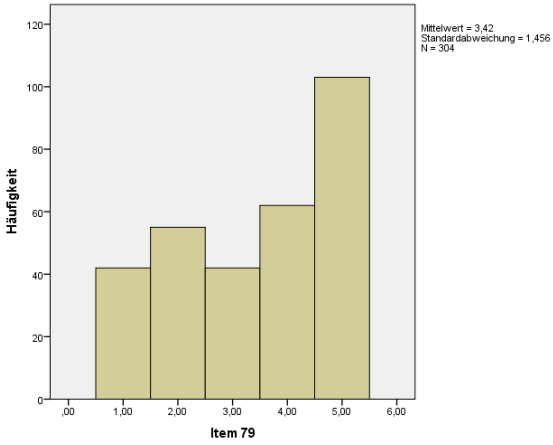

Mittelwert = 3,41
Standardabweichung = 1,349
N = 298
Häufigkeit
Item 68

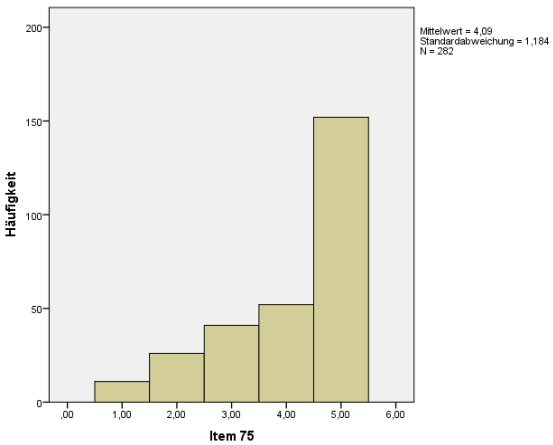

Mittelwert = 3,66
Standardabweichung = 1,088
N = 285
Häufigkeit
Item 72

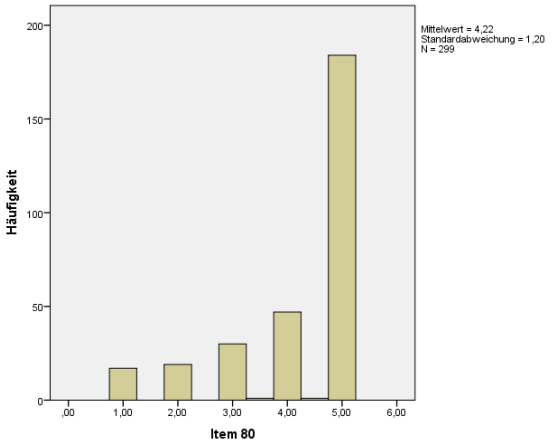

Mittelwert = 3,78
Standardabweichung = 1,183
N = 281
Häufigkeit
Item 73

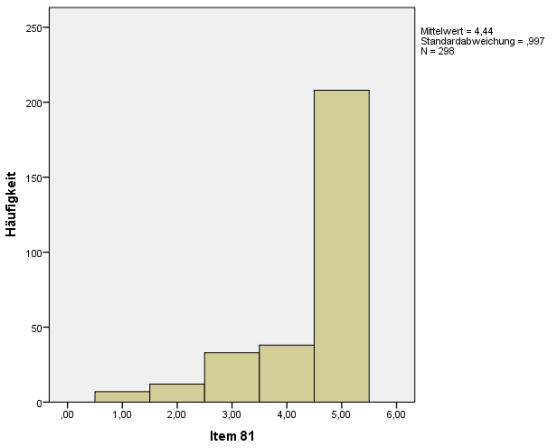

Mittelwert = 4,18
Standardabweichung = ,931
N = 306
Häufigkeit
Item 76

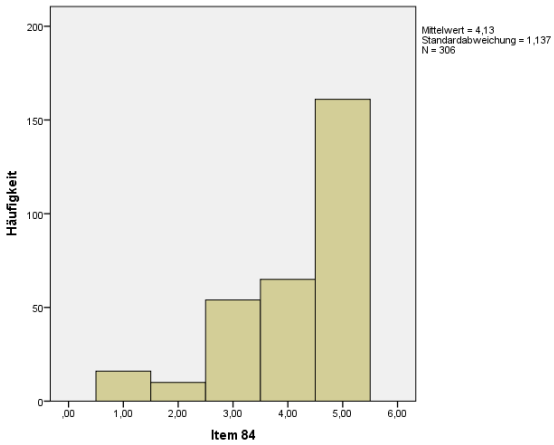

Mittelwert = 4,25
Standardabweichung = ,917
N = 307
Häufigkeit
Item 74

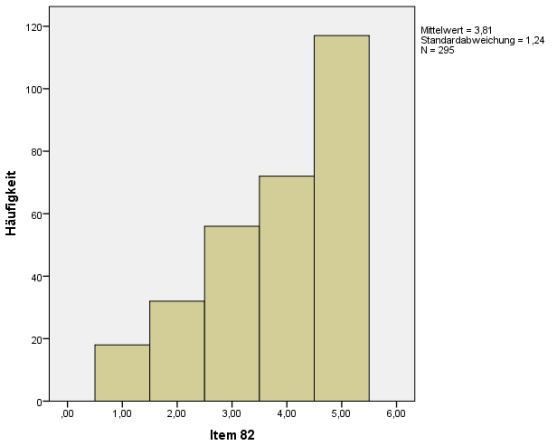

Mittelwert = 3,42
Standardabweichung = 1,456
N = 304
Häufigkeit
Item 79

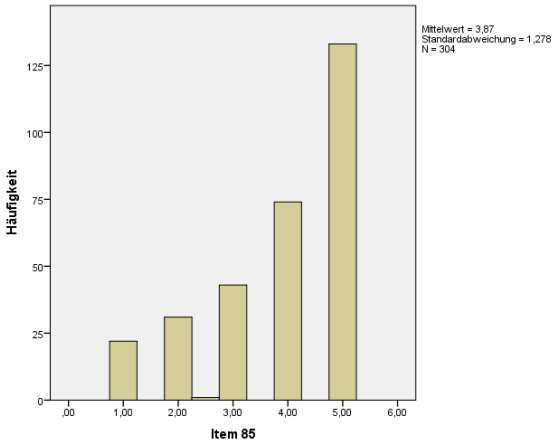

Mittelwert = 4,09
Standardabweichung = 1,184
N = 282
Häufigkeit
Item 75

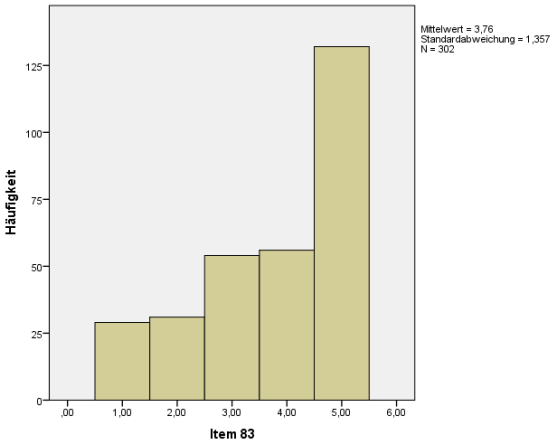

Mittelwert = 4,22
Standardabweichung = 1,20
N = 299
Häufigkeit
Item 80

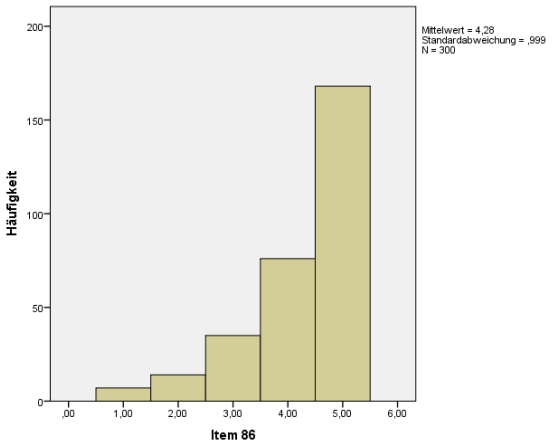

Mittelwert = 4,44
Standardabweichung = ,997
N = 298
Häufigkeit
Item 81

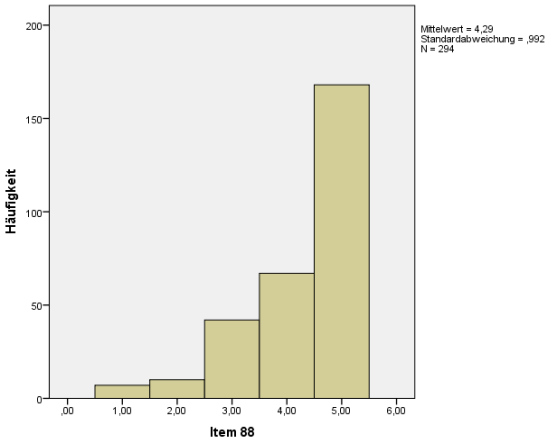

Mittelwert = 4,13
Standardabweichung = 1,137
N = 306
Häufigkeit
Item 84


Mittelwert = 3,81
Standardabweichung = 1,24
N = 295
Häufigkeit
Item 82


Mittelwert = 3,87
Standardabweichung = 1,278
N = 304
Häufigkeit
Item 85


Mittelwert = 3,76
Standardabweichung = 1,357
N = 302
Häufigkeit
Item 83


Mittelwert = 4,28
Standardabweichung = ,999
N = 300
Häufigkeit
Item 86



Beurteilbarkeit aufgeschlüsselt nach Alter (aktuelle Stichprobe)

		Fälle					
		Gültig		Fehlend		Gesamtsumme	
	Alter	H	Prozent	H	Prozent	H	Prozent
Item 8	6	23	82,1%	5	17,9%	28	100,0%
	7	39	81,3%	9	18,8%	48	100,0%
	8	37	84,1%	7	15,9%	44	100,0%
	9	20	90,9%	2	9,1%	22	100,0%
	10	5	100,0%	0	0,0%	5	100,0%
Item 9	6	21	75,0%	7	25,0%	28	100,0%
	7	42	87,5%	6	12,5%	48	100,0%
	8	39	88,6%	5	11,4%	44	100,0%
	9	20	90,9%	2	9,1%	22	100,0%
	10	5	100,0%	0	0,0%	5	100,0%

Beurteilbarkeit aufgeschlüsselt nach Alter (Gesamtstichprobe)

		Fälle					
		Gültig		Fehlend		Gesamtsumme	
	Alter	H	Prozent	H	Prozent	H	Prozent
Item 8	6	38	79,2%	10	20,8%	48	100,0%
	7	81	81,0%	19	19,0%	100	100,0%
	8	75	88,2%	10	11,8%	85	100,0%
	9	48	90,6%	5	9,4%	53	100,0%
	10	19	90,5%	2	9,5%	21	100,0%
Item 9	6	41	85,4%	7	14,6%	48	100,0%
	7	90	90,0%	10	10,0%	100	100,0%
	8	77	90,6%	8	9,4%	85	100,0%
	9	50	94,3%	3	5,7%	53	100,0%
	10	20	95,2%	1	4,8%	21	100,0%

Beurteilbarkeit aufgeschlüsselt nach Hort (Gesamtstichprobe)

		Fälle					
		Gültig		Fehlend		Gesamtsumme	
	Hort	H	Prozent	H	Prozent	H	Prozent
Item 29	Ziegelofengasse	13	100,0%	0	0,0%	13	100,0%
	Lange Gasse	45	95,7%	2	4,3%	47	100,0%
	Eibengasse	14	77,8%	4	22,2%	18	100,0%
	Donaucity	15	88,2%	2	11,8%	17	100,0%
	Reisnerstraße	11	100,0%	0	0,0%	11	100,0%
	Clemens-Holzmeister-Straße	18	85,7%	3	14,3%	21	100,0%
	Hertha-Firnbergstraße	20	100,0%	0	0,0%	20	100,0%
	Kabelwerk	8	100,0%	0	0,0%	8	100,0%
	Alma-Seidler-Weg	38	100,0%	0	0,0%	38	100,0%
	Hietzing	19	55,9%	15	44,1%	34	100,0%
	Kalvarienberggasse	20	74,1%	7	25,9%	27	100,0%
	Cottagegasse	16	100,0%	0	0,0%	16	100,0%
	Scheibenbergstraße	21	95,5%	1	4,5%	22	100,0%
	Huschkagasse	8	66,7%	4	33,3%	12	100,0%

Faktorenanalyse

Deskriptive Statistiken

	Mittelwert	Standard abweichung	Analyse N	Fehlendes N
Item 1	3,6700	1,29568	303	4
Item 2	3,8119	1,24788	303	4
Item 3	3,1919	1,15088	297	10
Item 6	4,1589	1,12118	302	5
Item 7	3,9488	1,31148	293	14
Item 8	3,3831	1,34119	261	46
Item 9	3,6655	1,22826	278	29
Item 12	3,9428	1,31008	297	10
Item 13	2,9373	1,41867	303	4
Item 14	3,6131	1,29821	305	2
Item 16	3,0140	1,34262	285	22
Item 17	2,9303	1,26906	287	20
Item 18	4,3074	,90390	296	11
Item 19	3,4593	1,25407	270	37
Item 23	3,8624	1,16538	298	9
Item 24	3,8818	1,21413	296	11
Item 25	3,8990	1,22056	297	10
Item 26	4,1689	1,17333	302	5
Item 27	4,4000	,92168	300	7
Item 28	4,4631	,86475	298	9
Item 29	3,8684	1,23837	266	41
Item 30	3,9346	1,11722	306	1
Item 31	3,6941	1,16968	304	3
Item 32	4,1960	,82348	301	6
Item 33	3,8289	1,07952	304	3
Item 35	4,1283	,99503	304	3
Item 36	4,1645	1,03372	304	3
Item 37	4,2329	,79746	307	0
Item 38	4,0130	1,01612	307	0
Item 39	4,4067	,87014	300	7
Item 40	4,1338	1,14787	299	8
Item 41	3,9189	1,04485	296	11
Item 48	3,9248	1,07318	306	1
Item 50	4,2973	,95680	301	6
Item 51	4,0537	1,17943	298	9
Item 52	4,1928	,97447	306	1
Item 53	4,2150	,98321	307	0
Item 55	3,7822	1,33425	303	4
Item 56	4,0362	1,05712	304	3
Item 57a	,4764	,50029	296	11
Item 57b	,5379	,49942	290	17
Item 57c	,4730	,50011	296	11
Item 58a	,8562	,35153	292	15
Item 58b	,7778	,41646	288	19
Item 58c	,7483	,43471	298	9
Item 59	3,9934	1,22878	303	4
Item 60	4,2820	,99299	305	2
Item 61	4,4653	,94446	303	4
Item 63	4,2467	,97239	304	3
Item 64	4,3268	,95010	306	1
Item 65	4,3512	1,00690	299	8
Item 66	3,9145	1,05895	304	3
Item 67	3,2043	1,25656	301	6
Item 68	3,4128	1,34869	298	9
Item 69	3,2768	1,16942	289	18
Item 71	4,3351	,96950	291	16
Item 75	4,0922	1,18352	282	25
Item 76	4,1765	,93119	306	1
Item 79	3,4243	1,45591	304	3
Item 80	4,2174	1,20035	299	8
Item 81	4,4362	,99711	298	9
Item 82	3,8068	1,23991	295	12
Item 83	3,7649	1,35704	302	5
Item 84	4,1275	1,13662	306	1
Item 85	3,8701	1,27830	304	3
Item 86	4,2800	,99913	300	7
Item 88	4,2891	,99216	294	13

KMO und Bartlett-Test

Kaiser-Meyer-Olkin-Maß der Stichprobeneignung.		,928
Bartlett-Test auf Sphärizität	Näherungsweise Chi-Quadrat	12417,810
	df	2211
	Sig.	,000

Kommunalitäten

	Anfänglich	Extraktion
Item 1	,875	,797
Item 2	,835	,737
Item 3	,703	,534
Item 6	,616	,453
Item 7	,828	,623
Item 8	,800	,636
Item 9	,838	,680
Item 12	,848	,715
Item 13	,753	,641
Item 14	,739	,562
Item 16	,662	,401
Item 17	,702	,502
Item 18	,648	,413
Item 19	,890	,738
Item 23	,808	,712
Item 24	,813	,661
Item 25	,777	,637
Item 26	,865	,745
Item 27	,725	,528
Item 28	,637	,303
Item 29	,757	,509
Item 30	,649	,373
Item 31	,633	,441
Item 32	,416	,224
Item 33	,530	,290
Item 35	,641	,491
Item 36	,741	,647
Item 37	,760	,572
Item 38	,739	,645
Item 39	,731	,505
Item 40	,780	,623
Item 41	,519	,330
Item 48	,598	,409
Item 50	,483	,245
Item 51	,765	,597
Item 52	,818	,694
Item 53	,836	,767
Item 55	,759	,680
Item 56	,761	,644
Item 57a	,774	,734
Item 57b	,641	,538
Item 57c	,754	,704
Item 58a	,643	,534
Item 58b	,661	,430
Item 58c	,723	,539
Item 59	,591	,458
Item 60	,658	,560
Item 61	,527	,412
Item 63	,619	,482
Item 64	,753	,536
Item 65	,615	,350
Item 66	,681	,553
Item 67	,735	,550
Item 68	,887	,806
Item 69	,659	,413
Item 71	,585	,268
Item 75	,675	,456
Item 76	,648	,422
Item 79	,789	,651
Item 80	,752	,540
Item 81	,760	,563
Item 82	,622	,476
Item 83	,803	,621
Item 84	,704	,583
Item 85	,678	,500
Item 86	,596	,382
Item 88	,606	,272

Extraktionsmethode:

Hauptachsenfaktorenanalyse.

Erklärte Gesamtvarianz

	Anfängliche Eigenwerte			Extrahierte Summen von quadrierten Ladungen			Rotierte Summen von quadrierten Ladungen		
Faktor	Gesamt summe	% der Varianz	Kumulativ %	Gesamt summe	% der Varianz	Kumulativ %	Gesamt summe	% der Varianz	Kumulativ %
1	23,658	35,311	35,311	23,243	34,691	34,691	13,355	19,933	19,933
2	5,145	7,679	42,990	4,651	6,942	41,633	9,952	14,854	34,787
3	4,207	6,279	49,269	3,793	5,661	47,293	5,502	8,211	42,998
4	2,648	3,952	53,220	2,253	3,362	50,656	4,871	7,271	50,269
5	2,557	3,816	57,036	2,097	3,130	**53,785**	2,356	3,516	53,785
6	1,614	2,409	59,445						
7	1,456	2,172	61,618						
8	1,424	2,125	63,743						
9	1,241	1,853	65,596						
10	1,098	1,638	67,234						
11	1,055	1,575	68,809						
12	1,024	1,529	70,338						
13	,933	1,392	71,730						
14	,896	1,337	73,067						
15	,875	1,305	74,372						
16	,800	1,194	75,566						
17	,772	1,152	76,718						
18	,738	1,102	77,820						
19	,709	1,059	78,879						
20	,677	1,010	79,889						
21	,634	,947	80,836						
22	,612	,914	81,749						
23	,594	,887	82,636						
24	,583	,871	83,507						
25	,543	,811	84,318						
26	,526	,785	85,103						
27	,484	,722	85,825						
⋮	⋮	⋮	⋮						
66	,066	,098	99,920						
67	,053	,080	100,000						

Extraktionsmethode: Hauptachsenfaktorenanalyse.

Unrotierte Faktorenmatrix[a]

	Faktor				
	1	2	3	4	5
Item 68	,832	-,327	,064	-,033	,042
Item 1	,821	-,259	,178	-,079	,134
Item 2	,820	-,182	,171	,036	,009
Item 19	,812	-,101	,177	,191	,002
Item 23	,797	-,234	,085	,079	,095
Item 24	,791	-,178	,061	,012	-,018
Item 9	,752	-,163	,270	-,069	,101
Item 13	,748	-,243	,118	-,066	,068
Item 79	,741	-,188	-,252	-,025	-,039
Item 14	,732	-,140	-,074	-,009	,021
Item 7	,724	-,251	,165	-,029	,090
Item 8	,721	-,254	,221	,012	,053
Item 25	,710	-,219	,215	-,050	,190
Item 67	,709	-,146	,137	,052	,065
Item 38	,709	,084	-,357	,070	,054
Item 83	,696	-,142	-,340	,033	,027
Item 27	,692	-,079	,195	,028	-,062
Item 53	,683	-,149	-,517	-,018	-,102
Item 52	,683	-,163	-,445	-,048	-,031
Item 26	,681	-,253	,358	-,068	,292
Item 80	,672	-,001	-,279	,002	,099
Item 12	,671	-,246	,346	-,065	,284
Item 56	,669	-,080	-,398	-,015	-,180
Item 66	,667	,190	,130	-,110	-,208
Item 55	,662	-,143	-,423	,063	-,196
Item 51	,659	-,106	-,357	,071	-,138
Item 3	,655	-,211	,235	,072	-,016
Item 29	,654	-,128	,250	-,027	-,042
Item 6	,651	-,112	,123	-,033	-,007
Item 81	,645	-,051	-,354	,108	-,087
Item 18	,632	,007	,078	,001	,085
Item 37	,628	,361	-,174	,101	-,078
Item 17	,623	-,072	,305	,073	-,103
Item 40	,615	-,045	-,481	,082	-,074
Item 16	,611	-,125	,072	,083	-,017
Item 75	,605	-,028	-,286	-,035	,082
Item 69	,600	,077	,130	,168	,045
Item 64	,587	-,010	-,400	-,152	-,089
Item 39	,581	,389	-,074	,047	-,096
Item 58c	,578	-,121	,425	-,034	,093
Item 82	,523	,030	-,401	-,172	-,103
Item 28	,512	,141	,086	-,051	-,106
Item 76	,509	,395	,011	-,081	,018
Item 65	,482	,056	-,297	-,132	,094
Item 71	,473	-,056	-,054	-,165	,103
Item 30	,472	,353	,129	-,071	-,063
Item 58b	,462	,195	,315	,046	-,277
Item 86	,437	,347	,017	-,201	,173
Item 63	,434	,316	,252	-,065	-,355
Item 59	,430	,319	,227	-,001	-,346
Item 33	,392	,347	,084	,016	-,091
Item 50	,371	,226	-,228	-,008	,072
Item 36	,535	,581	-,067	-,018	,133
Item 35	,201	,565	-,009	-,275	,238
Item 84	,377	,533	,056	-,060	,387
Item 31	,228	,520	-,151	-,067	,302
Item 85	,440	,498	,044	,012	,236
Item 48	,399	,488	,105	,010	-,028
Item 41	,236	,437	,064	-,077	,270
Item 32	,187	,366	-,058	-,066	,217
Item 88	,318	,352	,004	-,202	,076
Item 57a	,086	,119	-,020	,829	,157
Item 57c	,130	,149	-,073	,767	,266
Item 57b	,108	,221	-,003	,690	,039
Item 60	,382	,380	,225	,059	-,464
Item 58a	,421	,296	,263	,057	-,443
Item 61	,394	,153	,202	,113	-,425

Extraktionsmethode: Hauptachsenfaktorenanalyse.

a. 5 Faktoren extrahiert. 6 Iterationen erforderlich.

KMO und Bartlett-Test (ohne Items 57abc)

Kaiser-Meyer-Olkin-Maß der Stichprobeneignung.		,934
Bartlett-Test auf Sphärizität	Näherungsweise Chi-Quadrat	11867,527
	df	2016
	Sig.	,000

Erklärte Gesamtvarianz

	Anfängliche Eigenwerte			Extrahierte Summen von quadrierten Ladungen			Rotierte Summen von quadrierten Ladungen		
Faktor	Gesamt summe	% der Varianz	Kumulativ %	Gesamt summe	% der Varianz	Kumulativ %	Gesamt summe	% der Varianz	Kumulativ %
1	23,625	36,915	36,915	23,204	36,257	36,257	13,161	20,564	20,564
2	5,097	7,964	44,878	4,592	7,175	43,432	10,027	15,668	36,232
3	4,203	6,567	51,445	3,783	5,911	49,343	5,523	8,629	44,861
4	2,571	4,017	55,462	2,077	3,246	**52,589**	4,946	7,728	52,589
5	1,625	2,538	58,001						
6	1,541	2,408	60,409						
7	1,425	2,226	62,635						
8	1,246	1,947	64,582						
9	1,099	1,717	66,299						
10	1,052	1,643	67,943						
11	1,011	1,579	69,522						
12	,939	1,467	70,989						
13	,907	1,416	72,405						
14	,870	1,359	73,764						
15	,821	1,283	75,047						
16	,784	1,226	76,273						
17	,726	1,134	77,407						
18	,722	1,128	78,535						
19	,671	1,048	79,583						
20	,622	,971	80,554						
21	,618	,966	81,521						
22	,591	,923	82,444						
23	,587	,917	83,361						
24	,558	,872	84,232						
25	,521	,815	85,047						
26	,480	,750	85,797						
27	,458	,715	86,512						
⋮	⋮	⋮	⋮						
63	,067	,105	99,909						
64	,058	,091	100,000						

Extraktionsmethode: Hauptachsenfaktorenanalyse.

Anhang B: Notizen und Anmerkungen der BeurteilerInnen

Wiedergegeben werden hier die *itembezogenen* Bemerkungen der Rater; etwaige Äußerungen zur jeweiligen Testperson selbst scheinen an dieser Stelle weder von Relevanz noch – hinsichtlich der den Kindern und ihren Eltern zugesicherten Anonymität – angebracht und werden daher nicht veröffentlicht.

Tp 49: zu 17 [Anm. Item 25]: Kinder bekommen eine Wochenhausübung -> selbstorganisatorisches Einteilen!!

Tp 50: zu 17 [Anm. Item 25]: Kinder bekommen eine Wochenaufgabe -> selbstorganisatorische Einteilung

Tp 98: Zu Fragen 5 [Anm. Item 7], 13 [Anm. Item 18], 14 [Anm. Item 19], 15 [Anm. Item 23], 16 [Anm. Item 24]: „nicht beurteilbar“ angekreuzt, „muss er (aber)“ ergänzt
Zu Frage 21 [Anm. Item 29]: „nicht beurteilbar“ angekreuzt, „kann er nicht“ ergänzt

Tp 99: Zu Fragen 5 [Anm. Item 7], 13 [Anm. Item 18], 14 [Anm. Item 19], 15 [Anm. Item 23], 16 [Anm. Item 24], 59 [Anm. Item 71]: „nicht beurteilbar“ angekreuzt, „muss er“ ergänzt
Zu Frage 5 [Anm. Item 7]: „fixer Tagesablauf“ ergänzt
Zu Frage 21 [Anm. Item 29]: „nicht beurteilbar“ angekreuzt, „kann er nicht“ ergänzt

Tp 108: zu Frage 7 [Anm. Item 9]: „müssen alle Aufgaben im Hort erledigen“ ergänzt
Zu Frage 13 [Anm. Item 18]: „sie müssen die Aufgaben erledigen, bis keine Fehler mehr sind“
Zu Frage 21 [Anm. Item 29]: „müssen die Aufgabe machen“
Zu Frage 59 [Anm. Item 71]: „nicht beurteilbar“ angekreuzt, „werden kontrolliert“ ergänzt

Curriculum vitae

Angaben zur Person

Name	Nina Hasenöhrl
Geburtsort	Mödling, NÖ
Staatsbürgerschaft	österreichisch

Angaben zur Ausbildung

BG + BRG Baden Frauengasse (Matura mit Auszeichnung)
Universität Wien, Studium der Psychologie (berufsbegleitend, Schwerpunkt Diagnostik)

Berufserfahrung im psychologischen Bereich

2010/2011	HILL International GmbH, 1030 Wien, Praktikum in der Abteilung Forschung & Entwicklung
2010/2011	Universität Wien, Validierungsstudie AID englisch, Testleitertätigkeit
	Rudolf Steiner School, Kings Langley
	Sir John Cass's Foundation School, London
seit 2011	Gesundheitspraxis Webgasse, 1060 Wien, Mitarbeit in der Abteilung Marketing