



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Konstruktion eines neuen Syllogismen-Tests und dessen
erste testtheoretische Prüfung anhand des Rasch-
Modells

Verfasserin

Dipl.-Jur. Sarah Treiber

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, im April 2011

Studienkennzahl: A298

Studienrichtung: Psychologie

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

Danksagung

Ich danke Herrn Univ.-Prof. Dr. Mag. Klaus Kubinger für seine sehr engagierte und hilfreiche Betreuung meiner Diplomarbeit und die Einführung in die Psychologische Diagnostik.

Auch Herrn Mag. Poinstingl, der mich als Subbetreuer während der Anfangsphase der Diplomarbeit unterstützt und die von ihm entwickelte Software für die Testungen zur Verfügung gestellt hat, möchte ich danken.

Außerdem möchte ich an dieser Stelle Herrn Dr. Brugger danken, der als Leiter des Heerespsychologischen Dienstes die Testungen bei der Stellungskommission Wien und die Umsetzung der Tests auf das dortige System ermöglicht hat, sowie bei dem gesamten Team des Heerespsychologischen Dienstes in der Vorgartenstraße, besonders auch bei den Testleitern.

Meinen Eltern, Rose Metzger-Treiber und Günter Treiber, und meinem Freund Zeljko Uzelac danke ich für ihre Unterstützung während des gesamten Studiums. Außerdem möchte ich mich bei meinem Bruder und meinen Freunden und Kollegen bedanken, die mich unterstützt und mir hilfreiche Anregungen gegeben haben, vor allem bei Maria Hlauschka, Sarah Funk und Susanna Jahn.

Jan Steinfeld danke ich für die Hilfe beim Erstellen der Testsoftware und Ines Hoffmann für die gute Zusammenarbeit bei den Diplomarbeitstestungen.

Abstract (deutsch)

Im Rahmen der vorliegenden Arbeit wurde der Test *Syllogismen 2010* zur Messung der Reasoning-Fähigkeit im verbalen Bereich konstruiert und auf Konformität mit dem (dichotomen) logistischen Testmodell von Rasch geprüft.

Ausgehend von den 19 schlüssigen Formen von Syllogismen der klassischen formalen Logik wurden 55 Items erstellt. Diese unterscheiden sich in ihrer Schwierigkeit aufgrund der Figur des Syllogismus, der Bekanntheit der enthaltenen Begriffe, der Verwendung des Konjunktivs sowie der Variation der verwendeten Formulierung der Beziehungen zwischen den Begriffen (Quantoren). Weitere Items für den Test wurden durch Hoffmann (in Vorbereitung) im Rahmen ihrer Diplomarbeit nach ähnlichen Kriterien erstellt.

In der empirischen Untersuchung zur ersten testtheoretischen Prüfung bearbeiteten 245 Personen im Rahmen der Stellung bei der Stellungskommission Wien eine Auswahl von zehn der durch die Autorin neu konstruierten Items sowie zehn von Hoffmann konstruierte Items. Zusätzlich wurden den Teilnehmern zwölf Items aus dem verbalen Reasoningtest *FRRT* (Poinstingl, Kubinger, Skoda, & Schechtner, in Vorbereitung) zur Prüfung der Übereinstimmungsvalidität vorgegeben.

Die Analysen in Bezug auf die Geltung des dichotomen logistischen Modells von Rasch ergaben a-posteriori Modellgültigkeit nach Ausscheiden von zwei der zwanzig Items der *Syllogismen 2010*. Für den *FRRT* konnte Modellgültigkeit hinsichtlich aller vorgegebenen Items festgestellt werden. Der Korrelationskoeffizient nach Pearson für die Korrelation zwischen den Rasch-Modell-konformen Items der *Syllogismen 2010* und den vorgegebenen Items des *FRRT* beträgt $r = 0,525$.

In zukünftigen Untersuchungen wird eine weitere Prüfung derjenigen Items, für die vorliegend Modellgeltung angenommen werden konnte, durchzuführen sowie die Rasch-Modell-Konformität der hier nicht untersuchten Items zu prüfen sein.

Abstract (English)

In this study, a new verbal reasoning test, *Syllogismen 2010*, has been developed and tested for fitting to the Rasch model.

55 items were created based on 19 different forms of valid syllogisms from the classical formal logic. The item difficulty varies due to (i) the figure of the syllogism, (ii) the familiarity of the terms used, (iii) the use of conjunctive in the statements, and (iv) the variation in expressing the relationships between the terms in the syllogisms. Hoffmann (in process) designed additional items for the test according to similar criteria.

In the empirical study, ten of the items created by the author as well as ten items designed by Hoffmann (in process), were presented to 245 male subjects during the medical and psychological examination for military service. Additionally, 12 items of the verbal reasoning test *FFRT* (Poinstingl, Kubinger, Skoda, & Schechtner, in process) were presented to the participants for the purpose of testing concurrent validity.

The results show that after two of the 20 items were removed, the items of *Syllogismen 2010* fit to the Rasch model (*a posteriori* model fit). The model check for the *FFRT* items used in the present study showed that all items fit the Rasch model. Pearson's correlation coefficient for the correlation between the Rasch model fitting items of *Syllogismen 2010* and *FFRT* items equals $r = .525$.

Future research will further analyze the Rasch model fit for the items investigated in this study and conduct analysis of conformity with the Rasch model for the remaining items.

Inhaltsverzeichnis

Danksagung	I
Abstract (deutsch)	III
Abstract (English)	V
Inhaltsverzeichnis	VII
I. Einleitung	1
II. Theoretischer Teil	2
1. Reasoning und Syllogismen	2
1.1 Reasoning	2
1.2 Syllogismen	3
1.3 Der Test <i>Syllogismen</i> (Srp, 1994)	9
1.4 Der <i>Family-Relation-Reasoning-Test (FRRT)</i> als Test der verbalen Reasoning-Fähigkeit	10
1.5 Zusammenhang zwischen <i>Syllogismen</i> und <i>FRRT</i>	11
2. Die Item-Response-Theorie und das Rasch-Modell als testtheoretische Grundlage	12
2.1 Item-Response-Theorie	12
2.2 Das dichotome logistische Testmodell von Rasch	13
3. Verschiedene Multiple-Choice-Formate und das Problem der Ratewahrscheinlichkeit	17
III. Empirischer Teil	22
4. Konstruktion der <i>Syllogismen 2010</i>	22
4.1 Itemkonstruktion	22
4.2 Instruktion	26
5. Ziel der Untersuchung und Hypothesen	26
6. Untersuchungsplanung und Durchführung der Untersuchung	27
7. Stichprobe	29
8. Ergebnisse	31
8.1 Ergebnisse der Rasch-Modell-Analysen	31
8.2 Deskriptive Ergebnisse	51
8.3 Korrelation zwischen <i>Syllogismen 2010</i> und <i>FRRT</i>	53
9. Diskussion	54
10. Zusammenfassung	58

11. Abbildungsverzeichnis.....	61
12. Tabellenverzeichnis.....	62
13. Literaturverzeichnis	63
14. Anhang.....	66
Curriculum Vitae	87

I. Einleitung

Die Syllogismen aus der klassischen Logik stellen ein beliebtes Material zur Untersuchung des logischen Denkens innerhalb der Kognitionspsychologie dar; auch gibt es bereits psychologische Tests, die Syllogismen als Aufgaben einsetzen. Ihre Verwendung als Aufgabenmaterial für einen Test zum schlussfolgenden Denken im sprachlichen Bereich ist attraktiv, weil es sich im Vergleich zum figuralen oder numerischen Material, dessen sich Reasoningtests sonst oft bedienen, um relativ ansprechendes Material handelt. Nachdem Srp (1994) eine computerbasierte Form eines Syllogismen-Tests mit Rasch-Modell-konformen Items entwickelt hatte, welcher das Prinzip des adaptiven Testens verwirklicht, sollte in Anlehnung an diesen eine neue Form eines Syllogismen-Test entwickelt werden, wobei zwischenzeitlich sichtbar gewordene Fehler korrigiert und ein neuer, größerer Itempool mit abwechslungsreicheren Items, geschaffen werden sollte.

Im theoretischen Teil dieser Arbeit werden zunächst einige Grundlagen in Bezug auf das zu messende Konstrukt Reasoning sowie die Syllogismen der klassischen Logik als Aufgabenmaterial vorgestellt, anschließend der Test *Syllogismen* von Srp (1994) sowie der *FRRT* (Poinstingl, Kubinger, Skoda, & Schechtner, in Vorbereitung). Im Anschluss daran werden kurz einige wichtige Grundlagen zur Item-Response-Theorie und insbesondere zum dichotomen logischen Testmodell von Rasch dargestellt. Im empirischen Teil wird zunächst die Konstruktion des Tests *Syllogismen 2010* beschrieben, dann die Planung und der Ablauf der empirischen Untersuchung sowie die verwendete Stichprobe. Anschließend werden die Ergebnisse der Untersuchung in Bezug auf die Geltung des Rasch-Modells für die beiden eingesetzten Tests, die deskriptiven Ergebnisse und zuletzt die Ergebnisse zur Übereinstimmungsvalidität (Korrelation zwischen *Syllogismen 2010* und *FRRT*) dargestellt.

II. Theoretischer Teil

1. Reasoning und Syllogismen

In diesem Abschnitt wird die Bedeutung des Konstrukts Reasoning beschrieben sowie kurz auf unterschiedliche Methoden zur Messung von Reasoning eingegangen. Darauf folgt eine Darstellung der theoretischen Grundlagen zu den Syllogismen der klassischen Logik sowie zum logischen Denken bei Syllogismen.

1.1 Reasoning

Reasoning stellt einen der sieben Primärfaktoren der Intelligenz nach dem Intelligenzmodell von Thurstone dar. Man versteht darunter „die Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge erkennen und zweckentsprechend verwerten zu können“ (Kubinger, 2009, S. 206). Es handelt sich also um die Fähigkeit des logisch-schlussfolgernden Denkens, mithilfe welcher aus vorgegebenem Material Folgerungen hinsichtlich logischer Zusammenhänge gezogen werden und diese beispielsweise zur sinnvollen Ergänzung von Mustern oder zum Erkennen von Beziehungen zwischen verschiedenen Objekten genutzt werden können.

Die Fähigkeit des logischen Schlussfolgerns kann anhand sehr unterschiedlichen Materials - verbal, numerisch und figural - untersucht werden.

Zur psychologisch-diagnostischen Messung der Reasoning-Fähigkeit steht eine große Anzahl an Verfahren zur Verfügung. Insbesondere die bekannten Matrizentests, beispielsweise der *Wiener Matrizen-Test (WMT)*; Formann & Piswanger, 1979) oder der *Adaptive Matrizentest (AMT)*; Hornke, Etzel, & Rettig, 1997) stellen typische sprachfreie Vertreter der Reasoning-Tests dar, die figurales Material verwenden. In Bezug auf die Messung des schlussfolgernden Denkens im sprachlichen Bereich ist die Anzahl der zur Verfügung stehenden einschlägigen Verfahren allerdings sehr begrenzt; sie kommen vor allem als Untertests in Intelligenztestbatterien vor. Ein Beispiel hierfür ist der Untertest *Analogien* im *IST 2000-R (Intelligenz-Struktur-Test 2000 R)*; Liepmann, Beauducel, Brocke, & Amthauer, 2001).

Ein noch nicht publiziertes Verfahren zur Messung von verbalem Reasoning, welches ebenfalls am Arbeitsbereich Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien entwickelt wurde, ist der *FRRT (Family Relation Reasoning Test)*; Poinstingl, Kubinger, Skoda, & Schechtner, in Vorbereitung). Hierbei werden verschiedene Familienbeziehungen zwischen mehreren Personen in Textform vorgegeben, aus welchen eine Schlussfolgerung auf die verwandtschaftliche Beziehung zwischen zwei bestimmten Personen zu ziehen ist.

1.2 Syllogismen

Ein Syllogismus besteht aus zwei Prämissen, die jeweils eine Aussage in Bezug auf die Beziehung zwischen einem in beiden Prämissen vorkommenden Begriff, dem Mittelbegriff (M), sowie einem unterschiedlichen zweiten Begriff (Subjektbegriff S bzw. Prädikatbegriff P) enthalten und einer Konklusion, die eine Aussage über die Beziehung zwischen S und P trifft, welche sich im Fall der gültigen (schlüssigen) Syllogismen logisch notwendig aus den Prämissen ergibt.

Die Syllogistik geht auf Aristoteles zurück; er entwarf ein System des logischen Schließens in Syllogismen, klassifizierte und benannte verschiedene Formen von Syllogismen und entwickelte Methoden zum Beweis ihrer Gültigkeit (vgl. Wolff, 2009).

Die kategorische Syllogistik beschäftigt sich (anders als die hypothetische und disjunktive Syllogistik) mit Regeln des Schließens lediglich aus logisch nicht zusammengesetzten Aussagen. Die Schlussregeln innerhalb dieser Art der Syllogistik beruhen auf Definitionen der Bedeutung nicht von Satzverknüpfungen, sondern von zur Verknüpfung von Begriffen verwendeten Ausdrücken (Wolff, 2009).

Kategorische Syllogismen sind Argumente, die aus zwei Prämissen und einer Konklusion bestehen und sich ausschließlich aus kategorialen Aussagen zusammensetzen.

1.2.1 Arten von Urteilen

In kategorischen Syllogismen kommen die folgenden vier Arten von Urteilen, das sind Aussagen über die Beziehungen zwischen zwei Begriffen, vor:

A-Urteil	universell bejahend	Alle A sind B.
I-Urteil	partikulär bejahend	Einige A sind B
E-Urteil	universell verneinend	Alle A sind nicht B. (Kein A ist B.)
O-Urteil	partikulär verneinend	Einige A sind nicht B.

Tabelle 1.1: Arten von Urteilen

Verknüpfungen, über welche eine Aussage in Bezug auf die Beziehung zwischen zwei Begriffen getroffen wird, wobei Objekten eine bestimmte Eigenschaft oder Zugehörigkeit zu einer Gruppe zu- oder abgesprochen wird, werden Quantoren genannt.

I- und E-Urteil sind symmetrisch, das heißt die beiden verknüpften Begriffe können vertauscht werden, ohne dass sich an der logischen Gültigkeit der Aussage etwas ändert. Für die übrigen beiden Urteile gilt dies nicht.

Aus dem universellen Urteil lässt sich das partikuläre Urteil jeweils ableiten, d. h. wenn das A-Urteil (bzw. E-Urteil) gilt, gilt zugleich das I-Urteil (bzw. O-Urteil).

Bei Aristoteles kommen außerdem singuläre Sätze vor, in welchen eine Aussage, die sich auf nur genau ein in Rede stehendes Objekt bezieht, enthalten ist (Wolff, 2009). Hierunter fällt beispielsweise das bekannte Argument:

Prämisse 1 (P_1) Alle Menschen sind sterblich.
 Prämisse 2 (P_2) Sokrates ist ein Mensch.
 Konklusion (K) Sokrates ist sterblich.

Es handelt es sich um einen gültigen Schluss, jedoch nicht um einen kategorialen Syllogismus, da sowohl die zweite Prämisse als auch die Konklusion keine kategorialen Aussagen darstellen (Salmon, 1983).

Außerdem wäre ein solches Argument, da als Konklusion nur eine Aussage, die dem Objekt die in der ersten Prämisse genannte Eigenschaft entweder zu- oder abspricht, in Frage kommt, offensichtlich keine geeignete Grundlage für Testaufgaben und wird daher im Folgenden nicht weiter berücksichtigt.

1.2.2 Figur des Syllogismus

Aus der Kombination der verschiedenen Urteilsarten in den beiden Prämissen eines Syllogismus ergibt sich die sogenannte Begriffslage, aus der Stellung des Mittelbegriffs (M) innerhalb der Prämissen die Figur des Syllogismus. In der Standardform ist der Prädikatbegriff (P) immer in der ersten, der Subjektbegriff (S) in der zweiten Prämisse enthalten.

	Figur 1	Figur 2	Figur 3	Figur 4
P₁	M P	P M	M P	P M
P₂	S M	S M	M S	M S
K	S P	S P	S P	S P

Tabelle 1.2: Figuren des Syllogismus (Standardform)

Nach Aristoteles sind vollkommene von unvollkommenen Syllogismen zu unterscheiden. Wie diese Unterscheidung genau zu verstehen ist, ist umstritten. Teilweise wird der Unterschied darin gesehen, dass die Gültigkeit vollkommener im Gegensatz zu derjenigen der unvollkommenen Syllogismen „evident“ sei bzw. keines Beweises bedürfe, diese Sichtweise entspricht aber nicht den Ausführungen im Originaltext. Aristoteles beweist die unvollkommenen Syllogismen durch *reductio ad impossibile* (wobei zusätzlich zu den Prämissen die Annahme des kontradiktorischen Gegenteils der Konklusion als zu widerlegende Annahme in Betracht gezogen werden muss), durch Exposition oder durch Konversion, welche jeweils ebenfalls des Hinzuziehens einer weiteren Annahme bedürfen. Diese Syllogismen sind also in der Hinsicht „unvollkommen“, dass zu deren Beweis noch eine weitere Annahme (außer den in den Prämissen enthaltenen Annahmen) hinzuzufügen ist (Wolff, 2010). Da diese Unterscheidung zwar in Bezug auf die Beurteilung der Schwierigkeit eines Syllogismus eine Rolle spielen könnte, dahingehend, dass vollkommene Syllogismen leichter zu lösen sein könnten als unvollkommene, die Unterscheidung aber nicht notwendig ist, weil dies über die Betrachtung der Figur des Syllogismus geklärt werden kann, wird hierauf nicht weiter eingegangen.

1.2.3 Schlüssige Syllogismen und Schlussregeln

Insgesamt sind 19 der sich aus den Kombinationen der verschiedenen Begriffslagen und Figuren ergebenden Syllogismen nach der klassischen Logik schlüssig¹; die Konklusion folgt bei diesen logisch notwendig aus den Prämissen.

Zum Ziehen der logischen Folgerung werden nach Freytag-Löringhoff (1961) die folgenden Schlussregeln angewandt: *Dictum de omni, dictum de nullo* und „Vergessbarkeitsprinzip“. Diese sind bei Heinrich (1973) und Srp (1993) beschrieben.

Für die Konstruktion der Items des Tests *Syllogismen* verwendete Srp (1993) (bzw. schon Heinrich (1973), von welcher Srp einen Großteil der Items übernahm) 18 verschiedene Formen von gültigen Syllogismen. Nicht verwendet wurde der ebenfalls gültige Syllogismus, der *Bamalip* genannt wird – die Vokale bezeichnen die vorkommenden Urteile in der jeweiligen Reihenfolge, im vorliegenden Fall ein A-Urteil in den beiden Prämissen und ein I-Urteil in der Konklusion:

Alle P sind M.

Alle M sind S.

Einige S sind P.

Srp (1993), die sich auf Heinrich (1973) bezieht, begründet dies mit einem Verweis auf Freytag-Löringhoff (1961), welcher diesen Syllogismus als eine aus zwei Syllogismen zusammengesetzte Form betrachtet und somit als eine Sonderform, die aus dem System der Syllogismen auszuschneiden ist. Außerdem werde hierbei ein in der Syllogistik sonst nicht verwendetes Prinzip, das der *conversio per accidens*, angewendet. Dieses kann allerdings beispielweise auch bei dem bei Srp (1993) mit der Nummer 2 bezeichneten Syllogismus (A-A-I, 3. Figur) angewendet werden. Dort kann die Konklusion zwar auch über das „Vergessbarkeitsprinzip“ (nach Freytag-Löringhoff, 1961), gefunden werden, ebenso aber durch Überführung der A-Aussage in der zweiten Prämisse in eine I-Aussage unter Vertauschung von Subjekt und Prädikat (*conversio per accidens*). Der Ausschluss des *Bamalip* aus dem System der Syllogismen erscheint somit jedenfalls nicht als notwendig. Diese Form ist außerdem

¹ Zum Teil wird auch von 24 gültigen Syllogismenformen ausgegangen; dies beruht darauf, dass sich aus den fünf Syllogismen mit universeller Konklusion (sogenannte starke Form) jeweils ein gültiger Syllogismus mit der entsprechenden partikulären Konklusion (sogenannte schwache Form) ableiten lässt.

besonders interessant, da hierbei die Konklusion tatsächlich nur über logisches Schließen gefunden werden kann und nicht allein über die Betrachtung bestimmter Merkmale der Prämissen, wie in der Atmosphärentheorie von Woodworth und Sells (1935, zitiert nach Srp, 1993, S. 22) beschrieben. Hiernach wird nämlich aus zwei universell bejahenden Prämissen ein universell bejahender Schluss gezogen. In diesem Fall ist aber infolge der Stellung des Mittelbegriffs (Figur 3) und der Vorgabe der Reihenfolge von Subjekt- und Prädikatbegriff in der Konklusion nur eine partikuläre affirmative Konklusion möglich. Daher wurde im Gegensatz zu den Arbeiten von Heinrich (1973) und Srp (1993) auch *Bamalip* als eine gültige Syllogismenform im Rahmen der Itemkonstruktion berücksichtigt.

Es werden also die 18 bei Srp (1993) beschriebenen Formen von Syllogismen und zusätzlich, als 19. Form, der Syllogismus des Typs *Bamalip* (in der oben dargestellten Form) als Grundlage der Itemkonstruktion verwendet².

1.2.4 Theorien zum logischen Denken bei Syllogismen

Zum logischen Denken bei Syllogismen wurde eine Vielzahl von Theorien entwickelt, unter anderem die klassische Atmosphären- und die Konversionstheorie, sowie verschiedene Informationsverarbeitungstheorien (vgl. Srp, 1993).

Das gegenwärtig vorherrschende Modell zur Erklärung der mentalen Prozesse in Bezug auf das logische Denken bei Syllogismen ist die Theorie der Mentalen Modelle nach Johnson-Laird (1991, zitiert nach Ball & Quayle, 2009). Hiernach wird beim logischen Schließen in Syllogismen ein mentales Modell der Prämissen konstruiert, im welchem die Begriffe und ihre kategorische Beziehung zueinander in abstrakter Form repräsentiert sind und zwar in einer zweidimensionalen räumlichen Anordnung (Ball & Quayle, 2009). Logische Schlussfolgerungen bei Syllogismen werden nach dieser Theorie in der Weise gezogen, dass, nachdem ein mentales Modell der Prämissen konstruiert wurde, eine vorläufige Schlussfolgerung gezogen wird, die mit dem Modell in Einklang steht. Anschließend wird versucht, alternative Modelle zu finden (als Gegenbeispiele), in welchen die Prämissen gelten, die Konklusion jedoch nicht. Wenn solche Modelle nicht gefunden werden, wird die Konklusion akzeptiert (Morley, Evans, & Handley, 2004). Fehler beim Ziehen von Schlussfolgerungen

² Eine Auflistung der 19 verwendeten Formen von Syllogismen findet sich im Anhang unter A.

geschehen oft aus dem Grund, dass Personen nur einige oder sogar nur ein einziges Modell der Prämissen in Betracht ziehen (Johnson-Laird, 2001). Bei der Konstruktion der Mentalen Modelle fließen allerdings nicht nur die Annahmen aus den Prämissen mit ein, sondern ebenfalls in indirekter Weise das Wissen einer Person. So wird, wenn eine Konklusion dem Wissen einer Person widerspricht, eher nach weiteren Mentalen Modellen gesucht, innerhalb welcher die Prämissen gelten und die Konklusion nicht (Johnson-Laird, 2010). Die Theorie der mentalen Modelle kann somit den vielfach beobachteten Effekt, dass glaubwürdig erscheinende Konklusionen, also solche, die mit dem Wissen einer Person in Einklang stehen, eher als gültige Schlussfolgerungen akzeptiert werden, gut erklären (Ball & Quayle, 2009).

Dickstein (1978) zeigte, dass die Figur eines Syllogismus einen starken Einfluss darauf hat, wie häufig Syllogismen-Aufgaben richtig gelöst werden und führte dies darauf zurück, dass die in den Prämissen enthaltene Information unterschiedlich verarbeitet wird. Bei Syllogismen der ersten Figur, welche in der ersten Prämisse eine Beziehung der Form M-P enthalten und in der zweiten Prämisse eine Aussage der Form S-M, wird nach Dickstein (1978) eine nach vorne gerichtete Verarbeitung (*forward processing*) begünstigt, d. h. in Richtung der in der Konklusion zu bildenden Aussage S-P. Dadurch wird das Auffinden der korrekten Konklusion erleichtert. Bei der vierten Figur (P-M, M-S) dagegen wird nach Dickstein (1978) stark eine rückwärtsgerichtete Informationsverarbeitung (*backward processing*) von P nach S begünstigt. Versuchspersonen neigen in diesem Fall eher dazu, eine Konklusion P-S zu bilden und die gefundene Beziehung dann fälschlicherweise auf die Beziehung S-P, welche für die Konklusion vorgegeben ist, anzuwenden. Im Fall der Symmetrie der Aussage in der Konklusion (I- oder E-Urteil) wird dieser Fehler nicht sichtbar, da dann die über die Konklusion P-S gefundene Beziehung der korrekten Konklusion S-P entspricht. Im Fall der anderen beiden Aussagen (A und O), die nicht symmetrisch sind, kommt es jedoch zu Fehlern. Bei Syllogismen der zweiten und dritten Figur wird eher als bei der ersten Figur, aber in geringerem Ausmaß als bei der vierten Figur, *backward processing* begünstigt, daher werden dort die beschriebenen Fehler häufiger als bei der ersten aber weniger häufig als bei der vierten Figur gemacht. Aus diesem Grund sind nach Dickstein (1978) Aufgaben, die Syllogismen der ersten Figur enthalten, am leichtesten und solche, die Syllogismen der vierten Figur enthalten am schwierigsten zu lösen.

Im Rahmen der Theorie der Mentalen Modelle wird der Einfluss der Figur in Syllogismen-Aufgaben anders betrachtet. Aus den Prämissen von Syllogismen der ersten und vierten Figur können hiernach leichter Modelle gebildet werden, da bei diesen die Mittelbegriffe „aneinander angrenzen“. Bei der ersten Figur ist das der Fall, nachdem die Reihenfolge der Prämissen vertauscht wurde (S-M, M-P), bei der vierten Figur bereits in der vorgegebenen Reihenfolge (P-M, M-S). Bei den anderen beiden Figuren ist dieses Aneinandergrenzen nicht gegeben und kann nur hergestellt werden, indem aus einer der beiden Prämissen eine Aussage mit umgekehrter Reihenfolge der Begriffe gebildet wird. Der Effekt der Richtung, in welcher eine Konklusion gebildet wird, wird als sich beim Ziehen von Schlussfolgerungen natürlich ergebende Richtung betrachtet (Stupple & Ball, 2007), entsprechend der Leserichtung von links nach rechts (in Richtung von S-P bei der ersten Figur, in Richtung von P-S bei der vierten Figur). Hiernach sind also die Syllogismen der zweiten und dritten Figur am schwierigsten zu lösen.

1.3 Der Test *Syllogismen* (Srp, 1994)

Der von Srp (1993) im Rahmen ihrer Diplomarbeit entwickelte Test *Syllogismen* (Srp, 1994) beinhaltet 74 Items für die Rasch-Modell-Konformität angenommen werden kann. Es handelt sich dabei um einen computergestützten Test, der adaptives Testen in Form des sogenannten *tailored testing* (vgl. Kubinger, 1996) verwirklicht. Das verwendete Antwortformat ist eine besondere Form des Multiple-Choice-Antwortformats, wobei die 4 Antwortmöglichkeiten sequentiell vorgegeben werden.

Hansmann (2010) verwendete in ihrer Diplomarbeitsstudie eine Auswahl von Items aus den *Syllogismen* (Srp, 1994), mit 5 Antwortmöglichkeiten pro Item, welche über ein Softwareprogramm³ im „klassischen“ MC-Format einerseits und im MC-Format mit sequentieller Vorgabe der Antwortmöglichkeiten⁴ andererseits vorgegeben werden können. Hierbei zeigte sich, dass bei Vorgabe im „klassischen“ Multiple-Choice-Format signifikant mehr Aufgaben gelöst wurden.

³ Dieses wurde durch Poinstingl am Arbeitsbereich Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien entwickelt.

⁴ zu den verschiedenen Antwortformaten siehe unten, unter 3.

1.4 Der *Family-Relation-Reasoning-Test (FRRT)* als Test der verbalen Reasoning-Fähigkeit

Skoda (2005) entwickelte in ihrer Diplomarbeit nach einer Idee von Kubinger den *VRT (Verwandtschaften-Reasoning-Test)*, ein Verfahren, mittels welchem logisches Schlussfolgern im sprachlichen Bereich anhand von Aussagen zu Verwandtschaftsbeziehungen gemessen werden soll. Es sollte ein neuartiges Verfahren als Alternative zu den *Syllogismen* (Srp, 1994) und der begrenzten Anzahl an anderen Verfahren zur Messung von Reasoning im sprachlichen Bereich, die als Untertests in verschiedenen Intelligenztestbatterien enthalten sind, geschaffen werden. Es werden hierbei im Itemstamm Sätze dargeboten, in welchen Informationen über die Verwandtschaftsbeziehungen verschiedener Personen enthalten sind. Gefragt wird nach der verwandtschaftlichen Beziehung zwischen zwei der im Itemstamm erwähnten Personen. Die Testperson soll aus den vorgegebenen Informationen auf die Beziehung zwischen diesen beiden Personen schlussfolgern. Skoda (2005) wählte als Antwortformat eine Auswahlliste mit einer Vielzahl verschiedener Familienbeziehungen, in welcher die Testperson die ihrer Meinung nach korrekte Beziehung markiert.

Die Konstruktion der Items erfolgte mehr oder weniger intuitiv. Als die Schwierigkeit beeinflussende Faktoren wurden die Komplexität der Familienbeziehung (4 Stufen, von Kernfamilie bis „Patchwork-Familie“) sowie die Anzahl der insgesamt genannten Familienbeziehungen variiert. Poinstingl (2009) prüfte die Konstruktvalidität mittels Analyse nach dem linear-logistischen Testmodell (LLTM) nach Fischer, das eine Erweiterung des Rasch-Modells⁵ darstellt, wobei die Itemparameter des Rasch-Modells als Summe von gewissen Basisparametern modelliert werden. Es zeigte sich, dass bei Verwendung der beiden bei der Testkonstruktion angenommenen Faktoren keine Modellanpassung gegeben war, unter Anwendung anderer Faktoren jedoch eine zufriedenstellende Modellgeltung erreicht werden konnte.

Schechtner (2009) wendete diese im Sinne von itemgenerierenden Regeln in ihrer Diplomarbeit an und entwickelte so eine veränderte Fassung. Der Name des Tests wurde später auf *FRRT (Family-Relation-Reasoning-Test)* geändert.

⁵ zum Rasch-Modell siehe 2.2

Während im *VRT* von Skoda noch eine Vielzahl von besonderen Verwandtschaftsbeziehungen, wie Halbbruder oder Stiefmutter vorkamen, reduzierte Schechtner (2009) die beinhalteten Familienbeziehungen auf drei Generationen inklusive Cousinen und Cousins, wobei nur (noch) verheiratete Paare und deren eigene Kinder, somit auch nur „echte“ Geschwister, vorkommen. Es wurden folgende, die Schwierigkeit beeinflussende Faktoren, welche jeweils in verschiedenen Abstufungen vorkommen, eingeführt: Anzahl der Generationen, Anzahl der Relationen und Stufen von Redundanz (d.h. Anzahl der im Itemstamm angegebenen überflüssigen Informationen, die nicht zur Lösung der Aufgabe benötigt werden). Auch wurde statt einer Liste mit Verwandtschaftsbeziehungen zur Auswahl der Lösung nun ein freies Antwortformat verwendet.

Fiktives Beispiel-Item zum *FRRT* (ohne Antwortmöglichkeiten)

Nina ist die Tochter von Susi. Susi ist die Tochter von Rosa.

Rosa ist die _____ von Nina.

Hansmann (2010) entwarf in ihrer Diplomarbeit, zusätzlich zu den 50 von Schechtner (2009) entwickelten Items, zehn weitere Items nach den bei Schechtner beschriebenen Konstruktionsregeln. Es wurden pro Item fünf Antwortmöglichkeiten konstruiert, welche im „klassischen“ MC-Format einerseits und im MC-Format mit sequentieller Vorgabe der Antwortmöglichkeiten andererseits vorgegeben werden können.

1.5 Zusammenhang zwischen *Syllogismen* und *FRRT*

Hansmann (2010) gab in ihrer Diplomarbeitsstudie allen Teilnehmern sowohl Items der *Syllogismen* (Srp, 1994) als auch Items des *FRRT* (Poinstingl, Kubinger, Skoda, & Schechtner, in Vorbereitung) vor, teilweise im klassischen MC, und teilweise mit sequentieller Vorgabe der Antwortmöglichkeiten. Es ergab sich eine signifikante Korrelation ($p = 0,001 < 0,05$) zwischen den relativen Lösungshäufigkeiten der Testpersonen (bezogen auf die Rasch-Modell-konformen Items) in *Syllogismen* und *FRRT* mit einem Korrelationskoeffizienten nach Spearman von $r = 0,227$ (insgesamt für alle Gruppen der österreichischen Stichprobe).

Hauser (2011) gab in ihrer Diplomarbeitstesting den Teilnehmern unter anderem Items des *FRRT* (teilweise unter einer Stressbedingung) sowie 10 der im Rahmen der vorliegenden Arbeit neu erstellten Items der *Syllogismen 2010* vor⁶, allerdings im „klassischen“ MC-Format. Die Anzahl gelöster Aufgaben in den beiden Tests korrelierte insgesamt mit einem Korrelationskoeffizienten nach Pearson von $r = 0,307$ (dass die Korrelation nicht signifikant war, ist auf die sehr geringe Stichprobengröße von $n = 16$ zurückzuführen). Die eher geringe Stärke der Korrelation kann dadurch erklärt werden, dass die Aufgaben der *Syllogismen 2010* nicht wie vorgesehen (siehe dazu unter 4.1.3) in einem MC-Format mit sequentieller Vorgabe der Antwortmöglichkeiten angewendet wurde, sondern mit „klassischem“ MC-Antwortformat. Dies führte in der verwendeten Stichprobe, die sich ausschließlich aus Personen mit hohem Bildungsniveau, nämlich Studenten (der Medizin, Rechtswissenschaft, Mathematik Chemie und Informatik) zusammensetzte, dazu, dass die Aufgaben zu leicht waren und nur eine geringe Varianz auftrat. Die relative Lösungshäufigkeit über alle Aufgaben der *Syllogismen 2010* betrug 78,8 %.

2. Die Item-Response-Theorie und das Rasch-Modell als testtheoretische Grundlage

Im Folgenden werden kurz einige wichtige Grundlagen zur Item-Response-Theorie und insbesondere zum dichotomen logistischen Testmodell von Rasch (Rasch-Modell) dargestellt.

2.1 Item-Response-Theorie

Nach der *Item-Response-Theorie* (IRT), teilweise auch Probabilistische Testtheorie genannt, liegen den beobachteten Reaktionen von Testpersonen auf Items aus einem Verfahren (manifeste Verhaltensweisen) nicht beobachtbare, latente Eigenschaften zugrunde. Diese latenten Eigenschaften stehen in einem bestimmten wahrscheinlichkeitsfunktionalen Zusammenhang, der jeweils modellspezifisch definiert ist (Kubinger, 1989).

⁶ Die verwendeten Items sind diejenigen, welche in der vorliegenden Arbeit den Items 1 bis 10 entsprechen und im Anhang unter B. mit syl_1 bis syl_10 bezeichnet sind.

Im Gegensatz zur Klassischen Testtheorie, die ein nicht prüfbares („deterministisches“) Modell in Bezug auf den Zusammenhang zwischen dem Testwert einer Person und ihrer tatsächlichen Eigenschaftsausprägung darstellt, werden nach der Item-Response-Theorie wahrscheinlichkeitstheoretische Modelle aufgestellt (Kubinger, 2003).

Probabilistische Testmodelle beschreiben beispielsweise im Fall eines Leistungstests die Beziehung zwischen dem Lösen oder Nichtlösen einer Testaufgabe durch eine Person und der zu messenden Eigenschaft als wahrscheinlichkeitsfunktionalen Zusammenhang. Die Ausprägung der latenten Eigenschaft wird hierbei in Bezug auf die Personen über Personenparameter, in Bezug auf die Items über Itemparameter quantifiziert. Abhängig von diesen Parametern ist die Wahrscheinlichkeit für eine bestimmte beobachtbare Reaktion (Antwortverhalten) einer Person unterschiedlich groß (Kubinger, 1989).

Die Mehrzahl der Testmodelle nach der Item-Response Theorie stellen Spezialfälle oder Verallgemeinerungen des dichotomen logistischen Testmodells von Rasch dar (Kubinger, 1989).

2.2 Das dichotome logistische Testmodell von Rasch

Das dichotome logistische Testmodell von Rasch (Rasch-Modell) beschreibt die Lösungswahrscheinlichkeit als Funktion der Fähigkeit einer Person und der Schwierigkeit einer Aufgabe, welche jeweils über einen einzigen Parameter modelliert werden. Es handelt sich also um ein Modell, das nur einen Itemparameter beinhaltet, das Rasch-Modell wird daher auch 1-PL-Modell (1-parametric logistic model) genannt.

Die Wahrscheinlichkeit, dass eine Person v mit der Fähigkeit ξ_v ein Item i der Schwierigkeit σ_i löst wird folgendermaßen modelliert:

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

2.2.1 Bedeutung des Rasch-Modells in der Psychologischen Diagnostik

Dem dichotomen logistischen Testmodell von Rasch kommt besondere Bedeutung im Rahmen der Testkonstruktion und Psychologischen Diagnostik zu. Viele Tests verwenden als Maß für die Ausprägung einer Personeneigenschaft die Anzahl von Items, die gelöst (im Fall von Leistungstest) bzw. bejahend beantwortet (im Fall von Verfahren zur Persönlichkeitsmessung) wurden. Dieser Verrechnungsmodus bildet die empirischen Verhaltensrelationen jedoch nur dann adäquat ab und ein Test erfüllt nur dann das Gütekriterium der Skalierung, wenn das Rasch-Modell (oder eine monotone Transformation von diesem Modell) gilt (Kubinger, 2009). Nur dann ist „lokale stochastische Unabhängigkeit“ gegeben, d.h. die Reaktion einer Person ist lediglich von ihrer Fähigkeit und der Schwierigkeit des bearbeiteten Items abhängig. Dies kann mittels eines mathematischen Beweises gezeigt werden (Kubinger, 1989).

Der Itempool eines Tests, bei dem die Anzahl gelöster Aufgaben als Testwert verwendet werden soll, muss also notwendigerweise mit dem Rasch-Modell konform sein.

Gleichzeitig werden bei Geltung des Rasch-Modells „spezifisch-objektive Vergleiche“ möglich. Einerseits kann unabhängig davon, welche Items zwei Personen bearbeitet haben, die Ausprägung ihrer Fähigkeit (oder einer anderen Eigenschaft) über den Personenparameter ξ_v verglichen werden. Andererseits kann, was noch wichtiger ist, die Schwierigkeit von zwei Items (bzw. der Grad der Herausforderung bei anderen Verfahren als Leistungstests) verglichen werden, unabhängig davon in welcher Stichprobe sie vorgegeben wurden (Eigenschaft der Stichprobenunabhängigkeit) (Kubinger, 1989).

Items aus einem Itempool, der Rasch-Modell-konform ist, messen eindimensional, also alle nur genau eine und zwar dieselbe Eigenschaft.

Im Gegensatz zu anderen Modellen handelt es sich um ein prüfbares Modell, das heißt, es kann für jeden Itempool überprüft werden, ob die Voraussetzungen der Modellgeltung erfüllt sind (Kubinger, 2009).

2.2.2 Prüfung der Geltung des Rasch-Modells

Der (bedingte) Likelihood-Ratio-Test (LRT) von Andersen (1973) bedient sich zur Prüfung der Geltung des Raschmodells dessen Eigenschaft der Stichprobenunabhängigkeit. Er basiert auf dem Vergleich zwischen den Itemparameterschätzungen in verschiedenen Subgruppen der Stichprobe mit unterschiedlicher Anzahl an gelösten Aufgaben (Score) und in der Gesamtstichprobe. Dieser Test wird in der Regel nicht wie ursprünglich vorgesehen für alle möglichen Scoregruppen, sondern in Verbindung mit einer Teilung der Stichprobe in zwei Gruppen angewandt, und zwar auch für den Vergleich von Parameterschätzungen in zwei Teilstichproben, die nach anderen Kriterien als der Testleistung (internes Teilungskriterium) geteilt wurden (Kubinger, 1989). Insbesondere Alter und Geschlecht werden häufig als sogenannte externe Teilungskriterien verwendet (Kubinger, 2005).

Über den Graphischen Modell-„Test“ (Graphische Modellkontrolle), bei welchem die in den beiden Teilstichproben geschätzten Itemparameter in einem Koordinatenkreuz gegeneinander aufgetragen werden, lässt sich die Modellgeltung (deskriptiv) am einfachsten beurteilen. Wenn alle Punkte auf oder nahe der 45°-Geraden liegen, unterscheiden sich die geschätzten Parameter nicht (bzw. nur zufällig) und Modellgeltung ist gegeben. Auch der „z-Test“ nach Fischer und Scheiblechner (1970, zitiert nach Kubinger, 1989) zur Beurteilung der Modellgeltung pro Item ist, obwohl eigentlich als Inferenzstatistisches Verfahren gedacht, eher der deskriptiven Statistik zuzuordnen, da es aufgrund der Vielzahl der durchgeführten Signifikanztests zu einer sehr starken Überhöhung des Risikos für einen Fehler erster Art kommt (Kubinger, Rasch & Yanagida, in Druck). Die asymptotisch normalverteilte Testgröße ergibt sich hierbei aus der Differenz zwischen den Parameterschätzungen innerhalb der beiden Teilstichproben, bezogen auf deren Standardabweichung als Quadratwurzel der Summe der Kehrwerte der Informationsfunktionen (Kubinger, Rasch, & Yanagida, in Druck; Kubinger, 1989).

Neben dem LRT nach Andersen gibt es aber noch eine Vielzahl anderer globaler Modelltests sowie auch Tests auf Itemebene zur Prüfung der Geltung des Rasch-Modells. Neben der Eigenschaft der Stichprobenunabhängigkeit werden dabei die Modelleigenschaften der Eindimensionalität oder der Parallelität der Item-Response-

Funktionen zur Prüfung der Modellgeltung herangezogen (vgl. Glas & Verhelst, 1995; Kubinger, 1989).

2.2.3 Gütekriterien bei Rasch-Modell-konformen Tests

In diesem Abschnitt werden einige der wichtigsten Gütekriterien für psychologisch-diagnostische Verfahren in Bezug auf Rasch-Modell-konforme Tests angesprochen.

2.2.3.1 Validität

Ein Test ist valide, wenn er „tatsächlich jenes psychische Merkmal misst, welches er zu messen behauptet“ (Kubinger, 2009, S. 55). Hierbei gibt es keine Besonderheiten im Vergleich zu den nach der klassischen Testtheorie konstruierten Tests.

Die sogenannte Übereinstimmungsvalidität wird mittels Korrelation mit anderen Verfahren, die dasselbe Konstrukt messen, erfasst (Kubinger, 2009).

2.2.3.2 Reliabilität

Die innere Konsistenz ist bei Verfahren, die dem Rasch-Modell entsprechen, gegeben, da alle Items eine und zwar dieselbe Dimension messen (Kubinger, 2009).

Die Bestimmung der Messgenauigkeit ist über den Standardschätzfehler der Personenparameter (als Umkehrwert der Informationsfunktion) für jede Testperson individuell bestimmbar (Kubinger, 2009). Im Gegensatz zur klassischen Testtheorie wird keine Messgenauigkeit des Tests als solchem, die für alle Testpersonen gelten soll, angenommen und berechnet.

2.2.3.3 Skalierung

Das Gütekriterium der Skalierung ist, wie oben dargelegt, erfüllt, wenn ein Test dem Rasch-Modell entspricht; die Summe der gelösten Aufgaben bildet dann die empirisch gegebenen Verhaltensrelationen adäquat ab (Kubinger, 2009).

3. Verschiedene Multiple-Choice-Formate und das Problem der Ratewahrscheinlichkeit

Das Multiple-Choice(MC)-Antwortformat ist aufgrund der Ökonomie in der Auswertung, der großen Verrechnungssicherheit sowie der Ermöglichung der Vorgabe als Gruppentestung sehr beliebt. Außerdem sind manche Tests auch mit einem offenen Antwortformat nur schwer vorstellbar, beispielsweise Matrizentests (wobei allerdings zurzeit auch versucht wird, einen Matrizentest mit einem mehr oder weniger freien Antwortformat zu konstruieren) oder auch ein Test, der Syllogismen als Aufgabenmaterial verwendet. Wenn hierbei nur die beiden Prämissen mit einem offenen Antwortformat vorgegeben würden, wäre die Möglichkeit gegeben, dass die Testperson eine Schlussfolgerung aus einer der beiden Prämissen zieht, beispielsweise aus einem A-Urteil ein I-Urteil ableitet oder aus einem I-Urteil wieder ein I-Urteil unter Vertauschung der beiden Begriffe. Beide Schlussfolgerungen sind korrekt, entsprechen jedoch nicht der erwarteten Leistung. Auch wenn man die Testpersonen über die Instruktion anweisen würde, beide Prämissen zu verwenden bzw. festlegen würde, dass der Mittelbegriff in der Konklusion nicht auftauchen darf, könnte immer noch ein Schluss gezogen werden, in welchem der Prädikatbegriff an erster und der Subjektbegriff an zweiter Stelle steht. Eine solche Schlussfolgerung zu ziehen kann je nach der Figur des Syllogismus wesentlich leichter sein. Werden dagegen sowohl die zu verwendenden Begriffe als auch deren Reihenfolge innerhalb der Konklusion festgelegt, so handelt es sich eigentlich schon wieder um ein MC-Format, da auch die zu verwendenden Verknüpfungen (die vier Quantoren) festgelegt sind und nur noch aus diesen ausgewählt werden muss. Dem entspricht das sogenannte *Fill-in*-Format, wobei die Begriffe vorgegebenen sind und die Testperson über eine Auswahlliste die Quantität (einige/alle) und die Qualität (sind/sind nicht) der Quantoren festlegt.

Den Vorteilen des MC-Formats stehen verschiedene Nachteile gegenüber, insbesondere derjenige, dass solche Tests in der Regel diagnostisch weniger aussagekräftig sind als Tests mit offenem Antwortformat. Zum Teil ist sogar fraglich, ob überhaupt die zu messen beabsichtigte Eigenschaft erfasst wird. Ein weiteres Problem ist das des Rateeffekts (Kubinger, 2009).

Wenn bei einem Test kein offenes Antwortformat verwendet wird, also die Lösung eines Items nicht in irgend einer Weise von der Testperson selbst produziert werden muss, sondern verschiedene Antwortmöglichkeiten - die Lösung(en) und eine bestimmte Anzahl von sogenannten Distraktoren, die eine weniger fähige Person von der Lösung „ablenken“ sollen, indem sie ebenfalls als plausible Antwortoptionen erscheinen - zur Auswahl vorgegeben werden, besteht für die Testperson immer die Möglichkeit, zufällig die richtige(n) Antwortmöglichkeit(en) zu wählen, also durch Raten unabhängig von ihrer Fähigkeit eine vorgegebene Aufgabe zufällig zu lösen. Diese Ratewahrscheinlichkeit, also die Wahrscheinlichkeit, bei zufälliger Auswahl einer (oder mehrerer) Antwortmöglichkeiten eine Aufgabe zu lösen, ist besonders hoch, wenn nur eine aus fünf („1 aus 5“-MC-Format) oder sogar eine aus nur vier möglichen Antworten zu wählen ist. Die a-priori Ratewahrscheinlichkeit, d. h. die Wahrscheinlichkeit mit der eine Aufgabe aufgrund des Antwortformats durch rein zufällige Auswahl einer Antwortoption gelöst wird, beträgt dann 20% (1/5) beziehungsweise im letzteren Fall sogar 25% (1/4). Genau dies ist aber beim klassischen MC-Format die Regel. Dadurch, dass auch Personen mit eher geringer Fähigkeit außerdem oft noch einen oder zwei der Distraktoren als unplausibel erkennen und ausschließen können, ist die faktische Ratewahrscheinlichkeit meist noch weit höher (Kubinger, 2009).

Eine hohe Ratewahrscheinlichkeit bedroht die Reliabilität (Messgenauigkeit), aber auch die Validität eines Tests: Einerseits wird die Messung ungenau, dadurch dass es zufällige Schwankungen abhängig davon, ob eine Person mehr oder weniger Glück beim Raten hat, gibt, andererseits misst ein Test dann nicht mehr (nur) die zu messen beabsichtigte Fähigkeit. Außerdem wird auch die Fairness eines Tests beeinträchtigt, da nicht alle Personen gleichermaßen zum Raten neigen (Kubinger, 2009).

Eine Möglichkeit zur Reduktion der Ratewahrscheinlichkeit ist die Erhöhung der Zahl der Distraktoren, hierbei stößt man jedoch schnell an Grenzen. Erstens wird, wenn sehr viele Distraktoren angeboten werden, die Aufmerksamkeit der Testperson stark beansprucht (und die Testleistung damit auch in hohem Maße abhängig von Merk- und Konzentrationsfähigkeit sowie Leistungsmotivation der Testperson) und auch die Bearbeitungszeit pro Item deutlich erhöht, zweitens ist es in vielen Fällen auch

schwierig, eine große Anzahl von plausibel erscheinenden Distraktoren zu konstruieren, die nicht ohne Weiteres auch von Testpersonen mit eher geringer Fähigkeit von vornherein ausgeschlossen werden können.

Eine weitere Möglichkeit, den Rateeffekt zu reduzieren, stellt die Verwendung von Aufgaben mit mehr als einer korrekten Antwortmöglichkeit dar. So kann die Ratewahrscheinlichkeit durch die Verwendung eines „2 aus 5“-Formats, bei welchem zwei der fünf Antwortmöglichkeiten richtig sind, und eine Aufgabe nur dann als gelöst bewertet wird, wenn diese beiden Antwortoptionen gewählt wurden und gleichzeitig keine der anderen, deutlich reduziert werden. Die a-priori Ratewahrscheinlichkeit liegt in diesem Fall nur noch bei 10%. Es konnte gezeigt werden, dass dieses Antwortformat dem MC-Format „1 aus 6“ in Bezug auf die Minimierung von Rateeffekten deutlich überlegen ist (Kubinger, Holoher-Ertl, Reif, Hohensinn, & Frebort, 2010).

Eine noch weitreichendere Möglichkeit ist der Einsatz eines „x aus n“-MC-Formats. Hierbei wird eine Aufgabe nur dann als richtig bewertet, wenn eine der Testperson nicht mitgeteilte Anzahl von korrekten Antwortoptionen (und keiner der Distraktoren), ausgewählt wird. Die a-priori Ratewahrscheinlichkeit ist in diesem Fall beispielsweise bei 5 Antwortmöglichkeiten auf $(1/2)^5 = 1/32$, also etwa 3%, reduziert (Kubinger, 2009). Das „x aus 5“-Format weist keinen Unterschied in Bezug auf die Schwierigkeit der Aufgaben im Vergleich zum freien Antwortformat auf (Kubinger & Gottschall, 2007).

Solche Antwortformate sind jedoch nicht für alle Arten von Aufgabenmaterial möglich. In Tests, die Syllogismen nach dem System der klassischen Logik als Aufgabenmaterial verwenden, ist die Anzahl der zur Verfügung stehenden Distraktoren inhaltlich grundsätzlich auf drei beschränkt, da nur vier verschiedene Arten von Aussagen vorkommen, wobei jeweils nur eine der vier möglichen Aussagen als Lösung einer Syllogismen-Aufgabe anzusehen ist. Es ist lediglich eine sprachliche Umformulierung der Aussagen möglich, wodurch dann beispielsweise bei zwei Varianten pro Aussage ein „2 aus 8“-Format in Betracht käme. Für eine Testperson, die über eine ausreichende Fähigkeit verfügt, um zu erkennen, dass jeweils zwei der angebotenen Optionen inhaltlich gleichbedeutend sind, besteht allerdings dann kein Unterschied in Bezug auf die faktische Ratewahrscheinlichkeit

im Vergleich zu einem „1 aus 4“-Format. Die Tatsache, dass eine Person mit mäßiger Fähigkeit außerdem oft zusätzlich eine oder zwei der Antwortmöglichkeiten ausschließen kann, bewirkt in diesem Fall dann eben, dass jeweils gleich zwei bzw. vier gleichbedeutende Antwortmöglichkeiten ausgeschlossen können. So kann also keine wirkliche Reduktion der faktischen Ratewahrscheinlichkeit erreicht werden.

Auch die Verwendung eines *Fill-in*-Formats wie oben beschrieben bietet keinen Vorteil im Vergleich zum „1 aus 4“ Format, da die a-priori Ratewahrscheinlichkeit dabei aufgrund der Kombination von nur zweimal zwei Auswahlmöglichkeiten ebenfalls 1/4 entspricht.

Es besteht allerdings noch eine andere Möglichkeit zur Reduktion der Ratewahrscheinlichkeit, die auch in diesem Fall anwendbar ist, nämlich die des MC-Antwortformats mit sequentieller Vorgabe der Antwortmöglichkeiten. Hierbei werden der Testperson die Antwortoptionen nicht wie beim „klassischen“ MC-Format gleichzeitig zur Auswahl vorgegeben, sondern einzeln nacheinander. Die Testperson muss für jede angebotene Antwortmöglichkeit entscheiden, ob diese richtig oder falsch ist. Erst danach wird die nächste Antwortmöglichkeit zur Entscheidung vorgegeben (Kubinger, 2009).

Dieses Format lässt sich nur bei computergestützten Tests (oder bei Einzeltestung) anwenden, da nur so die Vorgabe der Antwortoptionen einzeln nacheinander und ohne die Möglichkeit des Vor- oder Zurückblätterns erfolgen kann. Bei diesem Format ist die a-priori Ratewahrscheinlichkeit beispielsweise bei 5 Antwortmöglichkeiten abhängig von der Position der Lösung bis auf $(1/2)^5 = 1/32$ bzw. $(1/2)^4 = 1/16$ reduziert, wenn die Lösung an letzter bzw. vorletzter Position zur Entscheidung angeboten wird. Bei Präsentation der korrekten Antwortoption an der dritten Stelle beträgt die Ratewahrscheinlichkeit $(1/2)^3 = 1/8$. Wird die Lösung an erster bzw. zweiter Stelle zur Entscheidung vorgelegt, beträgt die Ratewahrscheinlichkeit dagegen 1/2 bzw. 1/4. Wenn alle Antwortpositionen gleich oft vorkommen, ergibt sich eine Ratewahrscheinlichkeit von 19,4% bei 5 möglichen Antworten (bzw. 23,4% bei 4 möglichen Antworten), also praktisch in gleicher Höhe wie beim „klassischen“ MC-Format. Daher müssen die Lösungspositionen anders verteilt werden, um tatsächlich eine Reduktion der Ratewahrscheinlichkeit über alle Items eines Tests zu erreichen: Wenn die Lösung beispielsweise bei 5

Antwortmöglichkeiten in je 10% der Aufgaben an erster und zweiter Stelle, in 20% der Items an 3. und in je 30% an vierter und fünfter Stelle angeboten wird, reduziert sich diese auf 12,8%⁷.

Gleichzeitig wird verhindert, dass Testpersonen die Strategie, alle Antwortoptionen zu vergleichen und dann die „passendste“ auszuwählen, anwenden. Andererseits ergibt sich hieraus auch eine Einschränkung: Die Lösung muss eindeutig sein und darf sich nicht erst gerade aus dem Vergleich der verschiedenen Antwortmöglichkeiten ergeben (Kubinger, 2009).

Abgesehen von diesen eher praktischen, formalen Methoden zur Reduktion der Ratewahrscheinlichkeit ist auch eine Berücksichtigung im Rahmen der Auswertung unter Verwendung von Modellen der Item-Response-Theorie, die Rateparameter pro Item modellieren, insbesondere des drei-parametrischen logistischen Modells von Birnbaum (3-PL-Modell) bzw. des *Difficulty-plus-Guessing* PL-Modells möglich (Kubinger, 2009).

Wenn das dichotome logistische Modell von Rasch, welches keine Rateparameter enthält, dagegen für einen Test im MC-Format gilt, kann davon ausgegangen werden, dass die Rateeffekte zumindest für alle Items gleich groß und im idealen Fall fast gleich null sind (Kubinger, 2009).

⁷ Siehe dazu 4.1.3.

III. Empirischer Teil

Im empirischen Teil dieser Arbeit wird zu Beginn über die Konstruktion des neuen Syllogistentests berichtet, anschließend werden die Ziele der empirischen Untersuchung samt Hypothesen vorgestellt, dann folgt eine Beschreibung der Planung und Durchführung der Untersuchung und zuletzt die Darstellung der Ergebnisse.

4. Konstruktion der Syllogismen 2010

Testtheoretische Grundlage für die Entwicklung der Items stellt die *Item-Response*-Theorie mit dem Rasch-Modell dar (vgl. 2.2).

Ausgehend von den 19 gültigen Formen von Syllogismen (vgl. 1.2.3), wurden 55 Items erstellt, wobei die Begriffe so gewählt wurden, dass sie in einem inhaltlichen Zusammenhang stehen (soweit es sich nicht um „erfundene“ Begriffe handelt).

4.1 Itemkonstruktion

Die Basis der Itemkonstruktion bildeten die 19 gültigen Syllogismenformen⁸. Diese verschiedenen Formen von Syllogismen unterscheiden sich durch die Art der enthaltenen Urteile und die Position des Mittelbegriffs (Figur des Syllogismus) deutlich in ihrer Schwierigkeit.

Außerdem werden verschiedene sprachliche Formulierungen der Quantoren verwendet, um die Aufgaben abwechslungsreicher zu gestalten und eine Wiederholung derselben Verknüpfungswörter in jedem Item zu vermeiden⁹.

Einen möglichen Einfluss auf die Schwierigkeit der Aufgaben hat auch die Wirklichkeitsentsprechung der Aussagen in den Prämissen und der Konklusion. Da lediglich die Fähigkeit zum logischen Schlussfolgern gemessen werden und Faktenwissen keinen Einfluss auf das Lösen der Aufgaben haben soll, wurden die durch die Autorin entworfenen Aufgaben so konstruiert, dass mindestens eine

⁸ Diese sind im Anhang unter A. dargestellt.

⁹ Eine Auflistung der „Standard“-Quantoren und Varianten findet sich im Anhang unter C.

Prämisse sowie die Konklusion den tatsächlichen Gegebenheiten (der materiellen Wahrheit) widerspricht bzw. dass kein Realitätsbezug gegeben war, indem sinnlose, frei erfundene „Worte“ als Begriffe eingesetzt wurden. So sollte verhindert werden, dass Testpersonen im Fall von der materiellen Wahrheit entsprechenden Beziehungen einfach auf ihr Faktenwissen zu den Beziehungen zwischen den verschiedenen Begriffen zugreifen und so zur Lösung gelangen. Gleichzeitig wird dadurch die Wahrscheinlichkeit verringert, dass Testpersonen vorschnell nicht korrekte Konklusionen akzeptieren bzw. korrekte Konklusionen ablehnen, nur weil diese mit ihrem Erfahrungswissen übereinstimmen bzw. diesem widersprechen.

4.1.1 Variation der Schwierigkeit

Die Figur eines Syllogismus beeinflusst deutlich die Schwierigkeit von Syllogismenaufgaben.

Nach Dickstein (1978) sollten Syllogismen der ersten Figur am einfachsten zu lösen sein, Syllogismen der zweiten und dritten Figur schwieriger und Syllogismen der vierten Figur schließlich am schwierigsten. Nach der Theorie der Mentalen Modelle dagegen sollten Syllogismen der zweiten und dritten Figur schwieriger sein als die der ersten und vierten Figur (vgl. 1.2.4).

In Anlehnung an die Literatur wurden folgende weitere Eigenschaften der Syllogismen variiert, die die Schwierigkeit der Aufgaben beeinflussen können.

4.1.1.1 Bekanntheit der Begriffe

Die Bekanntheit der in Syllogismen enthaltenen Begriffe kann dazu führen, dass Syllogismen-Aufgaben leichter zu lösen sind (Heinrich, 1973).

Es wurden folgende Abstufungen der Bekanntheit der in den Syllogismen enthaltenen Begriffe verwendet: allgemein bekannte Begriffe, kaum bekannte Begriffe und erfundene „Wörter“. Hoffmann verwendete außerdem, wie bereits Heinrich (1973) und Srp (1993), einzelne Buchstaben als abstrakte Symbole.

4.1.1.2 Konjunktiv

Die Aussagen wurden in Anlehnung an Srp (1993) bzw. Heinrich (1973) teilweise im Konjunktiv formuliert. Ursprünglich war diese Variation von Heinrich dazu bestimmt gewesen, eine Verunsicherung der Testpersonen zu vermeiden, welche auftreten könnte, wenn diese mit materiell unwahren Aussagen konfrontiert werden, also als potentiell erleichternder Faktor.

Allerdings erhöht die Verwendung des Konjunktivs die sprachliche Komplexität der Aussagen, so dass diese Variation sich möglicherweise stattdessen erschwerend auswirkt.

Hierfür sprechen auch die Itemparameter im Test *Syllogismen* von Srp (1994): Die Items mit den Nummern 62 bis 67, 86 bis 88 und 92 bis 100, welche im Konjunktiv formuliert sind, weisen fast durchgängig negative Leichtigkeitsparameter auf, sind also im Vergleich zu den andern Aufgaben schwieriger.

Der Konjunktiv wird daher hier als potentiell erschwerender Faktor eingesetzt.

4.1.1.3 Variation in der Formulierung der Quantoren

Die Verwendung von unterschiedlich formulierten Quantoren dient vor allem der Vermeidung des Eindrucks von Eintönigkeit im Aufgabenmaterial, der auftreten könnte, wenn in allen Items immer dieselben vier Verknüpfungen in den Prämissen wie in den Antwortmöglichkeiten auftauchen. Möglicherweise hat diese Variation aber auch eine erschwerende Wirkung, da vor der Leistung des Schlussfolgerns aus den Prämissen auf die Konklusion die Bedeutung der unterschiedlichen Quantoren und deren Übereinstimmung mit den „Standard“-Quantoren erkannt werden muss.

4.1.2 Itemgenerierung

Anhand der oben angeführten Kriterien wurde versucht, systematisch die Schwierigkeit zu variieren. Es wurden 55 Items erstellt, wobei alle 19 gültigen Syllogismenformen (vgl. im Anhang unter A.) als Grundlage für die Itemkonstruktion verwendet wurden.

Hoffmann (in Vorbereitung) verwendete dagegen nur Syllogismen der ersten Figur.

Zu jedem Item wurden als Antwortmöglichkeiten neben der Lösung vier Distraktoren formuliert; drei der Distraktoren ergeben sich durch die Verwendung der drei nicht in der Lösung enthaltenen Quantoren zur Verknüpfung von Subjekt- und Prädikatbegriff. Zusätzlich wurde eine in einem der Distraktoren enthaltene Aussage umformuliert, um einen zusätzlichen vierten Distraktor zu erhalten, beispielsweise „Kein A ist ein B“ zu „Alle A sind keine B“.

4.1.3 Antwortformat

Als Antwortformat wurde aus den oben (unter 3.) dargelegten Gründen ein MC-Format mit sequentieller Vorgabe der AW-Möglichkeiten gewählt. Grundsätzlich ist auch eine Vorgabe im „klassischen“ MC-Format („1 aus 5“) möglich, hierbei ist allerdings mit beachtlichen Rateeffekten zu rechnen. Die relative Lösungshäufigkeit kann bei einer solchen Vorgabe, abhängig von der Stichprobe, sehr hoch sein. Hauser (2011) gab in ihrer Diplomarbeitstestung einer Stichprobe von Studenten die ersten 10 der hier vorgegebenen Items (syll_1 bis syll_10) der *Syllogismen 2010*, also 10 von der Autorin konstruierte Items im klassischen „1 aus 5“-Format vor. Die Lösungshäufigkeit betrug 78,8%. Dies liegt zwar sicher auch am hohen Bildungsniveau der Teilnehmer, das Antwortformat spielt allerdings sicher ebenfalls eine große Rolle. Zu beachten ist die Positionierung der Lösung innerhalb der Antwortmöglichkeiten: Bei Vorgabe im „klassischen“ MC-Format sollten die Lösung gleich häufig auf allen Positionen vorkommen, bei sequentieller Vorgabe der Antwortoptionen ist die Lösung häufiger auf den hinteren Positionen vorzugeben, um eine tatsächliche Verringerung der a-priori Ratewahrscheinlichkeit zu erzielen.

Für die Vorgabe unter Verwendung des MC-Formats mit sequentieller Vorgabe der AW-Möglichkeiten wurde folgende Verteilung in Bezug auf die Lösungspositionen gewählt¹⁰:

1. Stelle	2. Stelle	3. Stelle	4. Stelle	5. Stelle
10%	10%	20%	30%	30%

Tabelle 4.1: Verteilung der Position der Lösung innerhalb der Antwortoptionen

¹⁰ Diese Verteilung wurde bislang so nur für die in der empirischen Untersuchung vorgegebenen Items umgesetzt, in Bezug auf die übrigen Items ist die Lösung etwa gleichverteilt über alle Position. Dies ist je nach Auswahl der Items anzupassen.

Es ergibt sich somit eine a-priori Ratewahrscheinlichkeit von insgesamt $0,1 \cdot 0,5 + 0,1 \cdot 0,5^2 + 0,2 \cdot 0,5^3 + 0,3 \cdot 0,5^4 + 0,3 \cdot 0,5^5 = 0,12825$, also etwa 12,8%.

4.2 Instruktion

Für die Instruktion wurde die von Hansmann (2010) in ihrer Diplomarbeitstestung für die *Syllogismen*-Items verwendete Instruktion als Grundlage verwendet und überarbeitet. Die Formulierungen der Instruktion wurden klarer und eindeutiger gestaltet und ergänzt, außerdem wurde ein Kreisdiagramm zur Veranschaulichung eingefügt. Im „Hilfe“-Text, in welchem die Regeln zum logischen Schlussfolgern bzw. zur Symmetrie der verschiedenen logischen Urteile beschrieben sind, wurde, unter anderem aufgrund der Erfahrung aus der Vorstudie, dass viele Teilnehmer den Quantor „einige“ als strikt partikulär interpretieren, ergänzt, dass eine „Folgerung“ von „Einige A sind B“ auf „Einige A sind nicht B“ nicht zulässig ist. Auf den „Hilfe“-Text wird in der Instruktion verwiesen und er kann jederzeit während der Testung aufgerufen werden.

Screenshots der Instruktion, des Hilfetextes und der ersten Beispielaufgabe finden sich im Anhang unter E.

5. Ziel der Untersuchung und Hypothesen

Ziel der empirischen Untersuchung war die Prüfung der neu erstellten Items der *Syllogismen 2010* und der zwölf verwendeten Items des *FRRT* in Bezug darauf, ob Konformität mit dem dichotomen logistischen Modell von Rasch gegeben ist und gegebenenfalls die Herstellung der a-posteriori Modellgültigkeit durch Ausscheiden von eventuell enthaltenen nicht modellkonformen Aufgaben. Es sollte also eine Kalibrierung des neuen Tests bzw. der hier untersuchten Items anhand des Rasch-Modells erfolgen.

Des Weiteren sollten die Items des neuen Tests *Syllogismen 2010* auf Übereinstimmungsvalidität mit dem *FRRT* geprüft werden, indem die Stärke der Korrelation der Leistung in diesem Test mit der Leistung im *FRRT* als Verfahren zur Messung von Reasoning im sprachlichen Bereich festgestellt wird.

Gleichzeitig sollte im Rahmen der durchgeführten Untersuchung auch die Frage geklärt werden, ob die Verwendung von Papier und Stift bei der Bearbeitung dieser beiden Tests zu einer Verbesserung der Testleistung führt, also ob sich die Leistung von Testpersonen, die diese Hilfsmittel bei der Testung verwenden (dürfen) von denjenigen, die dies nicht tun (dürfen) unterscheidet. Diese Fragestellung wird allerdings nicht hier behandelt, sondern bei Hoffmann (in Vorbereitung).

Aus den beiden erstgenannten Fragestellungen leiten sich die folgenden Hypothesen ab:

$H_0(1)$

Die Items des Tests *Syllogismen 2010* sind (gegebenenfalls nach Ausschluss einzelner Items) konform mit dem dichotomen logistischen Modell von Rasch.

$H_0(2)$

Die Items des *FRRT* sind (gegebenenfalls nach Ausschluss einzelner Items) konform mit dem dichotomen logistischen Modell von Rasch.

$H_0(3)$

Es besteht kein Zusammenhang zwischen der Leistung in *Syllogismen 2010* und *FRRT*.

6. Untersuchungsplanung und Durchführung der Untersuchung

Zum Zweck der Kalibrierung nach dem Rasch-Modell sollten die Items der *Syllogismen 2010* und des *FRRT* etwa 200 bis 250 Testpersonen vorgegeben werden. Um eine sonst sehr oft verwendete rein studentische Stichprobe zu vermeiden, wurde eine Testung von Stellungspflichtigen angestrebt. Auf diese Weise sollte eine Stichprobe mit breiterer Streuung in Bezug auf Bildungsniveau und sozialen Status erhalten werden.

Den Personen der Stichprobe sollten 20 Items der *Syllogismen 2010*, zehn von der Autorin und zehn weitere von Hoffmann (in Vorbereitung) konstruierte Items, sowie zwölf ausgewählte Items des *FRRT* mit breit gestreuten Itemschwierigkeitsparametern vorgegeben werden.

Die Untersuchung wurde gemeinsam mit Frau Hoffmann im März 2011 an einer Stichprobe von 245 Stellungspflichtigen bei der Stellungskommission durchgeführt.

Jedem Teilnehmer wurden per Computerprogramm zuerst die 20 Items der *Syllogismen 2010* und anschließend zwölf *FRRT*-Items vorgegeben. Alle Personen bearbeiteten dieselben Items. Am Ende der Instruktion und vor Beginn der Testaufgaben wurden jeweils zwei Beispielaufgaben mit Rückmeldung, ob die Aufgabe korrekt gelöst wurde, vorgegeben, welche beliebig oft wiederholt werden können.

Die Testung erfolgte in Gruppen von zunächst sieben bis acht Personen in der ersten Phase der Untersuchung und in der zweiten Phase in Gruppen von vier bis 15 Personen. Die Testungen fanden in den Räumen des Heerespsychologischen Dienstes bei der Stellungskommission Wien statt. Während der ersten Phase wurde an den fünf Netbooks, die der Arbeitsbereich Psychologische Diagnostik der Fakultät für Psychologie an der Universität Wien für Diplomanden zur Verfügung stellt, bzw. drei Laptops der Testleiterinnen getestet. Hierbei wurden die Tests über ein von Poinstingl am Arbeitsbereich Psychologische Diagnostik entwickeltes Software-Programm vorgegeben, in welches die Items der *Syllogismen 2010* eingespielt worden waren. Diese Testungen erfolgten morgens im Zeitraum zwischen sieben und acht Uhr. In der zweiten Phase der Untersuchung wurden ebenfalls zuerst die 20 Items der *Syllogismen 2010* und anschließend die zwölf Items des *FRRT* vorgegeben. Die Testung erfolgte über das Testsystem des Heerespsychologischen Dienstes an den dort für die computergestützte Diagnostik im Rahmen der Stellung vorgesehenen Computern. Hierbei wurden die Testungen im Anschluss an die etwa einstündige psychologische Testung im Rahmen der Stellung an den Computern des dortigen Testsystems vorgegeben, die Testpersonen konnten allerdings auf Wunsch eine kurze Pause zwischen der Stellungstestung und der Diplomarbeitstestung machen. Die Testungen fanden vormittags zwischen ca. zehn und elf Uhr und nachmittags zwischen ca. 13.30 und 14.30 Uhr statt. Da die Dauer der vorangehenden Testung im Rahmen der Stellung von der Bearbeitungsgeschwindigkeit der Testpersonen abhängt, also individuell unterschiedlich ist und das zur Verfügung stehende Zeitfenster relativ knapp bemessen war, konnten diejenigen Personen, die relativ lange für die

computergestützte psychologische Stellungstestung benötigten, im Rahmen der vorliegenden Untersuchung nicht getestet werden, so dass die in der Stichprobe enthaltenen Personen wahrscheinlich insgesamt eher schnell bei der Bearbeitung von psychologischen Verfahren sind.

Die Teilnehmer wurden darauf hingewiesen, dass ihre Leistung in der vorliegenden Testung nicht in ihre Bewertung im Rahmen der Stellung mit einfließt.

Da die Testvorgabe ohne Zeitbeschränkung erfolgte, variierte die Testdauer erheblich. Die reine Bearbeitungszeit insgesamt für die Items beider Testungen (exklusive der Zeit zum Lesen der Instruktion und Bearbeiten der zwei Übungsbeispiele) lag zwischen einer Minute (bei einer Person) und 44 Minuten, die durchschnittliche reine Bearbeitungszeit betrug ca. 19 Minuten. Aufgrund der extrem kurzen Bearbeitungszeiten, die einzelne Personen aufwiesen, musste davon ausgegangen werden, dass diese die Aufgaben nicht ernsthaft bearbeitet hatten, so dass sie für die weiteren Analysen nicht berücksichtigt wurden. Die genauen Ausschlusskriterien werden im nächsten Kapitel beschrieben.

Die Teilnehmer wurden im Rahmen des gleichzeitig durchgeführten Experiments randomisiert zur Versuchsgruppe mit bzw. zur Kontrollgruppe ohne Möglichkeit der Verwendung von Papier und Stift als Hilfsmittel zugewiesen. Hierzu wurde die letzte Ziffer der fortlaufenden sogenannten Vorgangsnummer herangezogen, die die Personen im Rahmen der Stellung zur Identifizierung zugeteilt bekommen. War die letzte Ziffer dieser Nummer gerade, so bearbeiteten die Testpersonen die Aufgaben ohne Hilfsmittel (Kontrollgruppe), war diese Ziffer ungerade, so bekamen sie Papier und Bleistift als Hilfsmittel zur Verfügung gestellt (Versuchsgruppe). Außerdem wurde den Personen in der Versuchsgruppe der Hinweis gegeben, dass sie Papier und Stift bei der Bearbeitung der Testaufgaben verwenden dürfen, um sich Notizen zu machen.

7. Stichprobe

Die Stichprobe setzt sich aus insgesamt 245 Stellungspflichtigen zusammen, die Testpersonen sind also alle männliche österreichische Staatsangehörige.

Da jeder männliche Österreicher grundsätzlich verpflichtet ist, an der Stellung teilzunehmen, ist davon auszugehen, dass die Stichprobe mehr oder weniger repräsentativ für männliche österreichische Jugendliche dieses Alters aus dem Einzugsgebiet der Stellungskommission Wien sein sollte. Zu beachten ist jedoch, dass die Aufforderung zur Teilnahme an der Stellung nach Bezirken aufgeteilt erfolgt, so dass die Zusammensetzung der Gruppe derjenigen, die am selben Tag zur Stellung einberufen werden, deutlich in Hinblick auf sozioökonomischen Status, Bildung und Migrationshintergrund, somit auch Muttersprache, schwankt.

Die Einberufung zur Stellung erfolgt in Abhängigkeit vom Geburtsjahr, daher ist die weit überwiegende Mehrheit der Testpersonen 17 bis 18 Jahre alt (17 bis 34 Jahre, durchschnittlich 17,4 Jahre).

In der ersten Phase der Untersuchung wurden die Teilnehmer aufgefordert, ihre Muttersprache einzugeben, in der zweiten Phase wurde nur gefragt, ob die Muttersprache Deutsch oder eine andere Sprache ist. 39,6% der Personen in der Stichprobe gaben an, dass Deutsch ihre Muttersprache ist, die übrigen gaben an, eine andere Muttersprache als Deutsch zu haben.

Zur Erfassung des Bildungsniveaus wurde aufgrund des Alters der Personen in der Stichprobe nach dem aktuellen Besuch einer höheren Schule (bzw. wenn keine Schule mehr besucht wird, dem Abschluss einer solchen) bzw. in Phase 1 nach dem Typ der zurzeit besuchten (bzw. bereits abgeschlossenen) Schule gefragt. 42,9% der Teilnehmer besuchen eine höhere Schule, also eine Schule mit Maturaabschluss (bzw. haben eine solche bereits abgeschlossen).

Aufgrund der extrem kurzen Bearbeitungszeiten in einem der beiden Tests von durchschnittlich unter 10 Sekunden pro Item musste bei einigen Personen davon ausgegangen werden, dass sie den jeweiligen Test nicht ernsthaft bearbeitet hatten, zumal diese Personen auch nur sehr wenige Aufgaben gelöst hatten. Da einige dieser Personen aber einen der beiden Tests offenbar sehr wohl ernsthaft bearbeitet hatten (normale bis lange Bearbeitungszeit und mittlere bis sehr hohe Anzahl gelöster Aufgaben), den anderen jedoch nicht, werden die Personen jeweils nur für die Auswertung desjenigen Tests, in welchem die auffällig kurzen Bearbeitungszeiten auftraten, nicht berücksichtigt. Außerdem wurden Personen, bei welchen die

Testungen aus organisatorischen Gründen (beispielsweise eine unerwartet angeordnete zusätzliche ärztliche Untersuchung) abgebrochen werden musste, in Bezug auf den unvollständig bearbeiteten Test nicht berücksichtigt.

Im Fall der *Syllogismen 2010* waren dies 14 Personen¹¹ mit auffällig kurzen Bearbeitungszeiten sowie 3 Personen¹², bei welchen der Test abgebrochen werden musste, beim *FRRT* 18 Personen¹³ (kurze Bearbeitungszeit) bzw. 8 Personen¹⁴ (Testabbruch während Bearbeitung des *FRRT* oder vor dessen Beginn).

Da die Teilnehmer im Rahmen der verpflichtenden Stellungsuntersuchung akquiriert wurden, handelt es sich nicht um eine typische „Freiwilligenstichprobe“, so dass sich das Problem der Non-Responder (vgl. Kubinger, 2009) hier nicht stellt.

8. Ergebnisse

In diesem Abschnitt werden zuerst die Ergebnisse der Prüfung der Geltung des Rasch-Modells in Bezug auf *Syllogismen 2010* und *FRRT* dargestellt, anschließend die relativen Lösungshäufigkeiten der Items (deskriptive Ergebnisse) und zuletzt die Ergebnisse zur Korrelation zwischen den Items der beiden eingesetzten Verfahren. Die Auswertung zum Vergleich zwischen Versuchs- und Kontrollgruppe im Experiment zum Einfluss von Hilfsmitteln findet sich bei Hoffmann (in Vorbereitung).

8.1 Ergebnisse der Rasch-Modell-Analysen

Die Analysen zur Prüfung der Geltung des dichotomen logistischen Modells nach Rasch wurden mit der frei verfügbaren Statistiksoftware R (Version 2.13.0) unter Verwendung des Programmpakets eRm (*extended Rasch modelling*)¹⁵ durchgeführt.

Es wurden die folgenden vier Teilungskriterien (Splitkriterien) verwendet:

¹¹ VPN 24, 138, 150, 155, 166, 167, 175, 180, 182, 188, 190, 213, 216 und 243

¹² VPN 98, 200 und 224

¹³ VPN 63, 119, 138, 153, 155, 160, 167, 177, 178, 180, 182, 188, 191, 198, 213, 214, 216 und 245

¹⁴ VPN 16, 72, 73, 98, 166, 200, 205, 224

¹⁵ Zur Anwendung des Programmpakets eRm vgl. Poinstingl, Mair, & Hatzinger (2007)

- a. Anzahl der gelösten Aufgaben (kleiner/gleich Median vs. größer Median)
- b. Bildungsniveau (kein Besuch vs. Besuch einer höheren Schule)
- c. Muttersprache (Nicht-Deutsch vs. Deutsch)
- d. Versuchsbedingung (ohne Hilfsmittel vs. mit Hilfsmitteln)

Das Geschlecht als üblicherweise verwendetes Teilungskriterium kann vorliegend nicht angewendet werden, da sich die Stichprobe ausschließlich aus Männern zusammensetzt. Die Anwendung des Alters als Teilungskriterium erscheint infolge der geringen Unterschiede zwischen den Teilnehmern – die überwiegende Mehrheit ist 17 Jahre alt und nur 4 Personen sind älter als 18 Jahre – nicht sinnvoll.

Anhand der vier oben genannten Kriterien wird die Stichprobe jeweils in zwei Teilstichproben unterteilt, um anhand des Likelihood-Ratio-Tests nach Andersen als globalem Modelltest zu prüfen, ob die Schätzungen der Itemparameter statistisch gleich sind (Kubinger, 1989), also die Eigenschaft der „Stichprobenunabhängigkeit“ erfüllt und somit Modellgeltung gegeben ist. Da vier Signifikanztests durchgeführt werden und das untersuchungsbezogene Risiko erster Art durch Kumulierung des Risikos eines Fehlers erster Art hierbei deutlich über demjenigen der einzelnen Signifikanztests liegt, wird das akzeptierte Risiko für einen Fehler erster Art auf $\alpha = 0,01$ festgesetzt (vgl. Kubinger, 2005).

Im Fall eines signifikanten Modelltests wird anhand des Graphischen Modell-„Tests“ (als deskriptive Methode) überprüft, welche Items sich auffällig verhalten. Hierbei werden die innerhalb der jeweiligen zwei Teilstichproben geschätzten Itemparameter in einem Koordinatenkreuz gegeneinander aufgetragen (d. h. pro Item stellt der in einer Teilstichprobe geschätzte Wert die Ordinate, der in der anderen Teilstichprobe geschätzte Wert die Abszisse eines Punktes dar). Wenn Modellgeltung gegeben ist, liegen alle Punkte auf oder in der Nähe der 45°-Geraden; liegen einzelne Items relativ weit entfernt von der Geraden, so sind diese als nicht modellkonform anzusehen und schrittweise zu entfernen, bis das Ergebnis des LRT für alle Teilungskriterien nicht mehr signifikant ist und a-posteriori Modellgeltung gegeben ist (Kubinger, 2005). Hierbei werden auch die sogenannten Konfidenzellipsen

berücksichtigt, welche den jeweiligen Standardschätzfehler pro Item in den beiden Teilstichproben darstellen.

Der z-Test nach Fischer und Scheiblechner wird nur zur zusätzlichen Absicherung bei der Auswahl der auszuscheidenden Items verwendet. Wie oben dargelegt sind dessen Ergebnisse aufgrund der extremen Überhöhung des Risikos für einen Fehler erster Art nur deskriptiv zu werten (vgl. Kubinger, Rasch, & Yanagida, in Druck).

8.1.1 Rasch-Modell-Analysen hinsichtlich der Items der *Syllogismen 2010*

Im Folgenden werden die Ergebnisse der Likelihood-Ratio-Tests mit den Teilungskriterien Anzahl gelöster Aufgaben, Bildungsniveau, Muttersprache und Versuchsbedingung sowie die entsprechenden Graphischen Modell-„Tests“ (Graphische Modellkontrolle) für die 20 Items der *Syllogismen 2010* dargestellt.

8.1.1.1 Erster Berechnungsdurchgang (alle Items)

Im ersten Berechnungsdurchgang wird die Rasch-Modell-Geltung für alle 20 vorgegebenen Items geprüft.

8.1.1.1.1 Teilungskriterium Anzahl der gelösten Aufgaben

Das Ergebnis des LRT nach Andersen mit dem Teilungskriterium Anzahl gelöster Aufgaben (kleiner/gleich Median vs. größer Median) ist signifikant:

χ^2 -Wert	df	p-Wert
53,126	19	0

Tabelle 8.1: LRT *Syllogismen 2010*, Score (alle Items)

Im Graphischen Modell-„Test“ ist zu erkennen, dass Item 8 besonders auffällig ist.

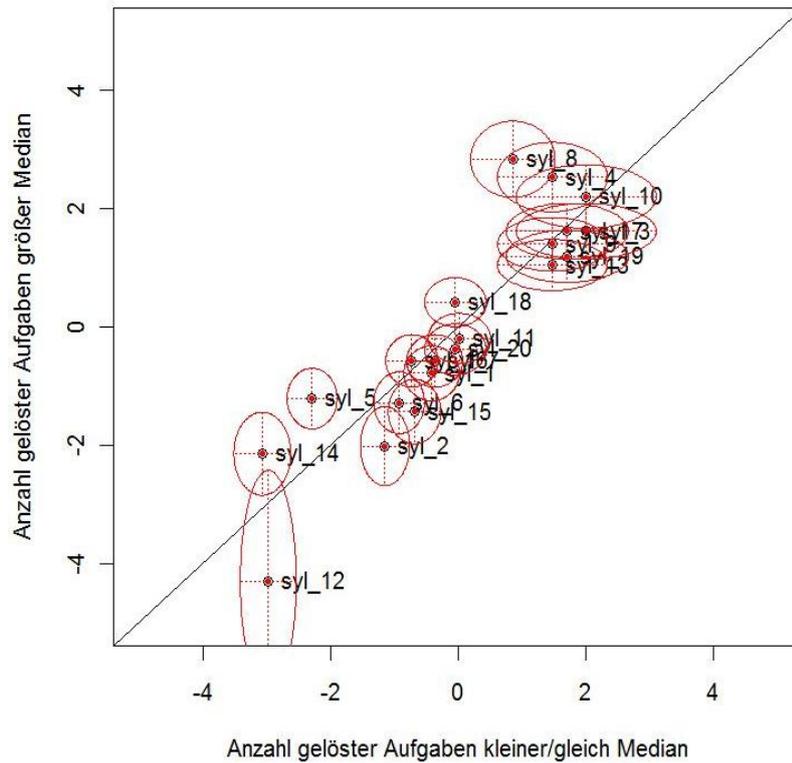


Abb. 8.1: Graphische Modellkontrolle *Syllogismen 2010* (alle Items), Score

Auch im z-Test nach Fischer und Scheiblechner zeigt sich, dass Item 8 besonders auffällig ist:

	z-Wert	p-Wert
syl_1	-1,127	0,260
syl_2	-2,200	0,028
syl_3	-0,626	0,531
syl_4	1,995	0,046
syl_5	3,282	0,001
syl_6	-1,062	0,288
syl_7	-0,637	0,524
syl_8	4,208	0,000
syl_9	-0,150	0,881
syl_10	0,329	0,742
syl_11	-0,629	0,529
syl_12	-1,348	0,178
syl_13	-0,863	0,388
syl_14	2,173	0,030
syl_15	-2,137	0,033
syl_16	0,500	0,617

syl_17	-0,158	0,875
syl_18	1,439	0,150
syl_19	-0,949	0,343
syl_20	-1,006	0,314

Tabelle 8.2: Z-Test nach Fischer und Scheiblechner *Syllogismen 2010*, Score

8.1.1.1.2 Teilungskriterium Bildungsniveau

Das Ergebnis des LRT mit dem Teilungskriterium Bildungsniveau (Besuch einer höheren Schule vs. kein Besuch einer höheren Schule) ist dagegen nicht signifikant auf einem Niveau von $\alpha = 0,01$:

χ^2 -Wert	df	p-Wert
32,375	19	0,028

Tabelle 8.3: LRT *Syllogismen 2010*, Bildungsniveau (alle Items)

Auch das Ergebnis der Graphischen Modellkontrolle ist zufriedenstellend.

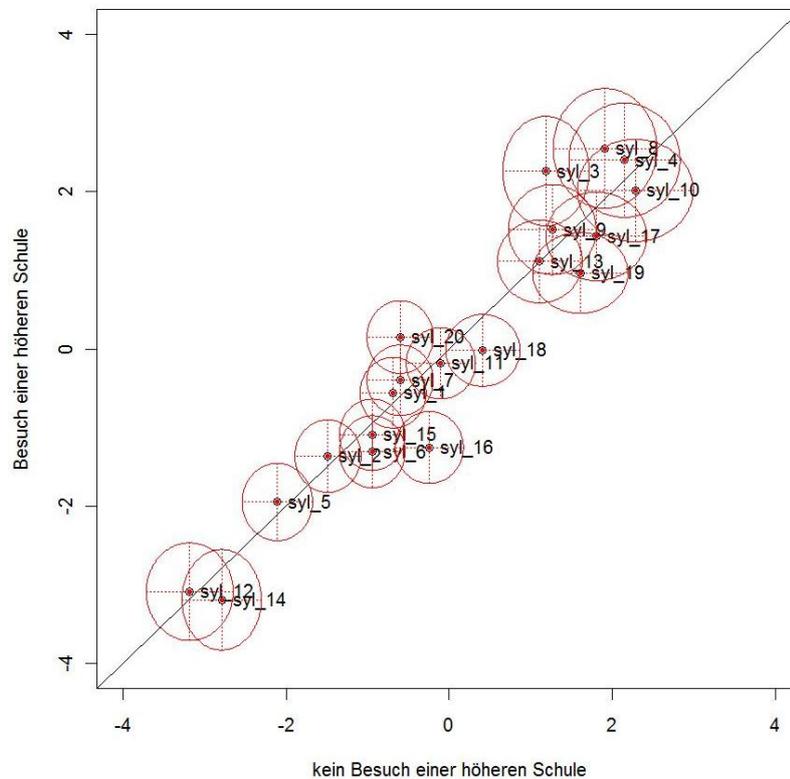


Abb. 8.2: Graphische Modellkontrolle *Syllogismen 2010* (alle Items), Bildungsniveau

8.1.1.1.3 Teilungskriterium Muttersprache

Der LRT mit dem Teilungskriterium Muttersprache (Deutsch vs. nicht Deutsch) ist ebenfalls nicht signifikant:

χ^2 -Wert	df	p-Wert
22.334	19	0.268

Tabelle 8.4: LRT *Syllogismen 2010*, Muttersprache (alle Items)

Auch hier ist das Ergebnis der Graphischen Modellkontrolle zufriedenstellend.

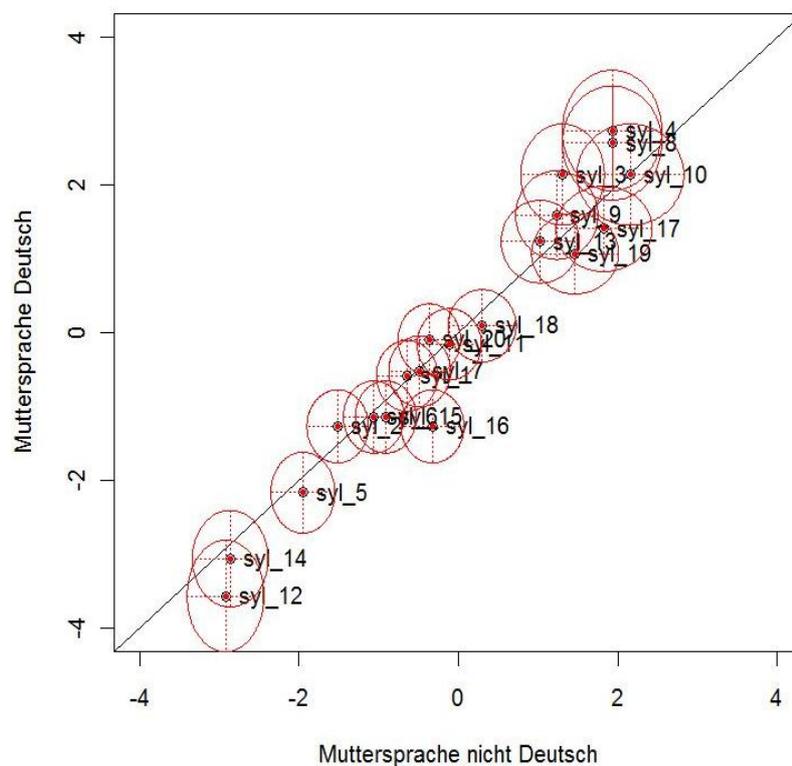


Abb. 8.3: Graphische Modellkontrolle *Syllogismen 2010* (alle Items), Muttersprache

8.1.1.1.4 Teilungskriterium Versuchsbedingung

Auch das Ergebnis für den LRT mit dem Teilungskriterium Versuchsbedingung (mit Hilfsmittel vs. ohne Hilfsmittel) ist nicht signifikant:

χ^2 -Wert	df	p-Wert
21,554	19	0,307

Tabelle 8.5: LRT *Syllogismen 2010*, Versuchsbedingung (alle Items)

Ebenso ist das Ergebnis der Graphischen Modellkontrolle zufriedenstellend.

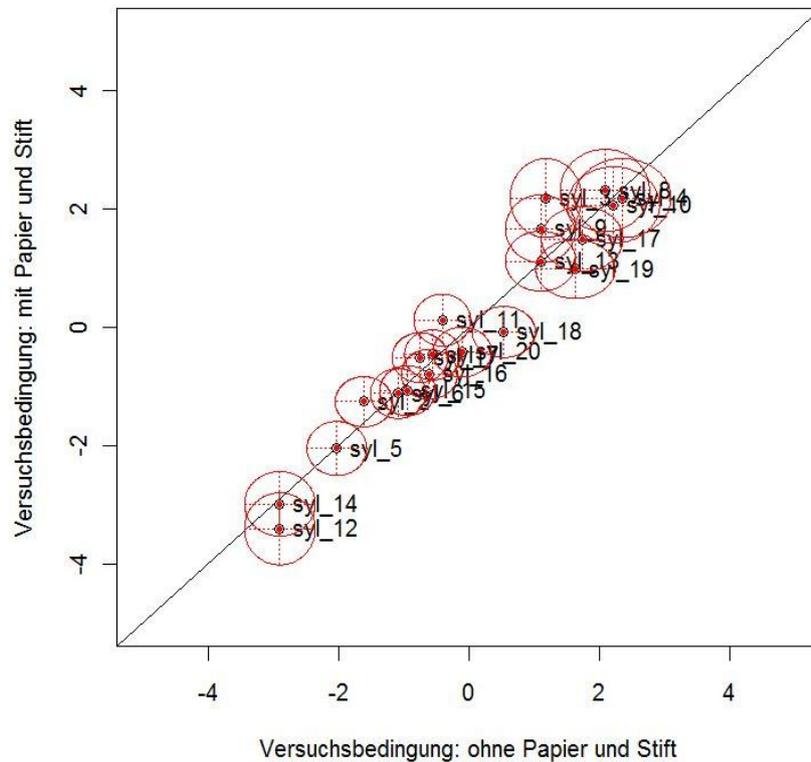


Abb. 8.4: Graphische Modellkontrolle *Syllogismen 2010* (alle Items), Versuchsbedingung

8.1.1.1.5 Zusammenfassung zum ersten Berechnungsdurchgang

Aufgrund des signifikanten Ergebnisses im LRT mit dem Teilungskriterium Anzahl gelöster Aufgaben wird das Item 8 (syl_8), das sich in der Graphischen Modellkontrolle (und auch im z-Test nach Fischer und Scheiblechner) als sehr auffällig erweist, ausgeschieden.

8.1.1.2 Zweiter Berechnungsdurchgang (nach Ausscheiden von Item 8)

Der zweite Berechnungsdurchgang wird ohne das Item 8 durchgeführt.

8.1.1.2.1 Teilungskriterium Anzahl der gelösten Aufgaben

Das Ergebnis der LRT mit dem Teilungskriterium Anzahl gelöster Aufgaben ist weiterhin signifikant.

χ^2 -Wert	df	p-Wert
39,62	18	0,002

Tabelle 8.6: LRT *Syllogismen 2010*, Score (ohne Item 5)

In der Graphischen Modellkontrolle erweist sich vor allem Item 5 als auffällig:

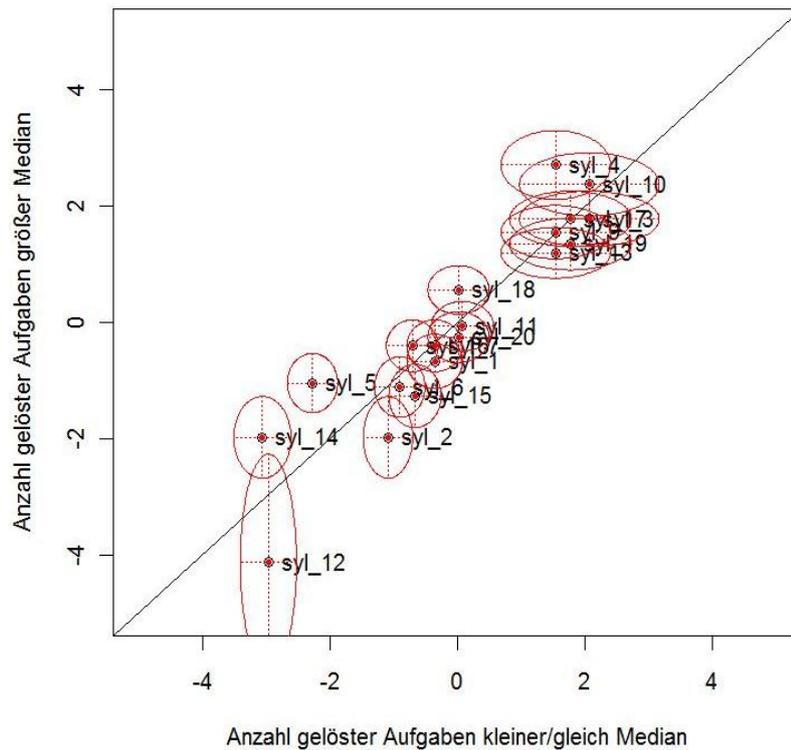


Abb. 8.5: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 8), Score

Auch im z-Test nach Fischer und Scheiblechner zeigt sich, dass Item 5 besonders auffällig ist:

	z-Wert	p-Wert
syl_1	-0,956	0,339
syl_2	-2,166	0,030
syl_3	-0,467	0,640
syl_4	2,197	0,028
syl_5	3,736	0,000
syl_6	-0,621	0,535
syl_7	-0,142	0,887
syl_9	0,037	0,971
syl_10	0,499	0,618
syl_11	-0,428	0,669
syl_12	-1,192	0,233
syl_13	-0,687	0,492
syl_14	2,539	0,011
syl_15	-1,711	0,087
syl_16	0,995	0,320
syl_17	0,020	0,984

syl_18	1,668	0,095
syl_19	-0,784	0,433
syl_20	-0,813	0,416

Tabelle 8.7: Z-Test nach Fischer und Scheiblechner *Syllogismen 2010*, Bildung (ohne Item 8)

8.1.1.2.2 Teilungskriterium Bildungsniveau

Das Ergebnis des LRT mit dem Teilungskriterium Bildungsniveau ist weiterhin nicht signifikant.

χ^2 -Wert	df	p-Wert
31,117	18	0,028

Tabelle 8.8: LRT *Syllogismen 2010*, Bildungsniveau (ohne Item 8)

Auch die Graphische Modellkontrolle liefert weiterhin ein zufriedenstellendes Ergebnis:

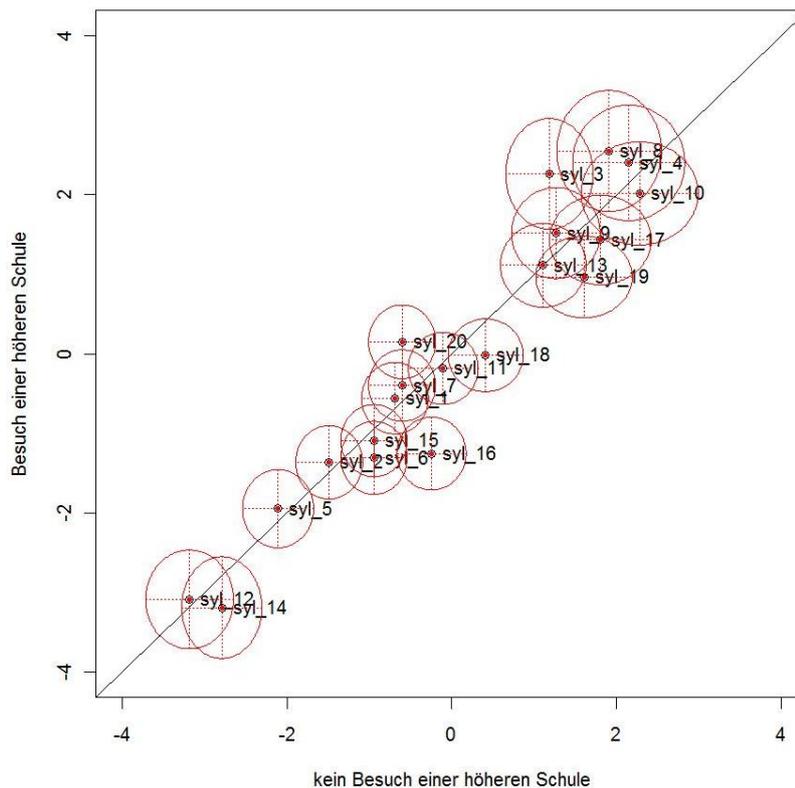


Abb. 8.6: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 8), Bildungsniveau

8.1.1.2.3 Teilungskriterium Muttersprache

Auch das Ergebnis des LRT mit dem Teilungskriterium Muttersprache ist weiterhin nicht signifikant.

χ^2 -Wert	df	p-Wert
21,075	18	0,276

Tabelle 8.9: LRT *Syllogismen 2010*, Muttersprache (ohne Item 8)

Die Graphische Modellkontrolle liefert ebenfalls weiterhin ein zufriedenstellendes Ergebnis:

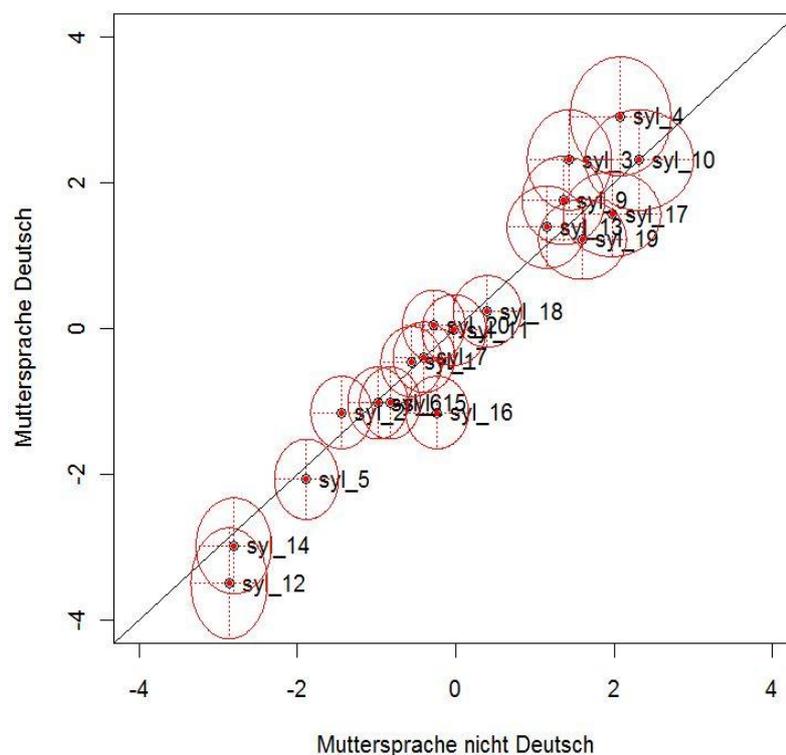


Abb. 8.7: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 8), Muttersprache

8.1.1.2.4 Teilungskriterium Versuchsbedingung

Das Ergebnis des LRT mit dem Teilungskriterium Versuchsbedingung ist ebenfalls weiterhin nicht signifikant.

χ^2 -Wert	df	p-Wert
21,736	18	0,244

Tabelle 8.10: LRT *Syllogismen 2010*, Versuchsbedingung (ohne Item 8)

Auch die Graphische Modellkontrolle liefert weiterhin ein zufriedenstellendes Ergebnis:

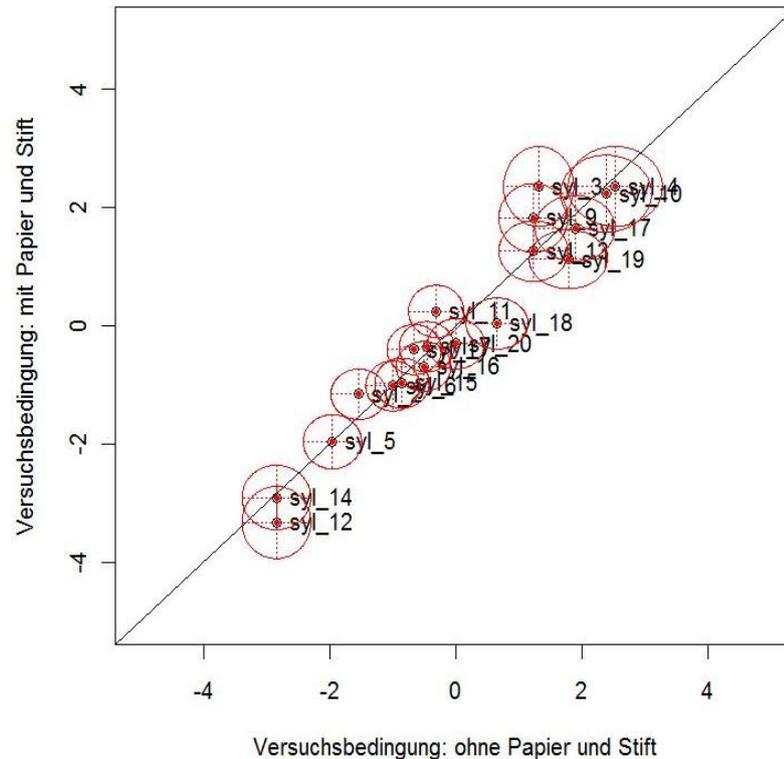


Abb. 8.8: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 8), Versuchsbedingung

8.1.1.2.5 Zusammenfassung zum zweiten Berechnungsdurchgang

Aufgrund des (weiterhin) signifikanten Ergebnisses im LRT mit dem Teilungskriterium Anzahl gelöster Aufgaben wird zusätzlich das Item 5 (syl_5), welches nach dem Graphischen Modell-„Test“ am auffälligsten ist (und auch im z-Test nach Fischer und Scheiblechner den auffälligsten z-Wert aufweist) ausgeschieden.

8.1.1.3 Dritter Berechnungsdurchgang (nach Ausscheiden von Items 5 und 8)

Nach dem zusätzlichen Ausscheiden von Item 5 wird ein dritter Berechnungsdurchgang durchgeführt.

8.1.1.3.1 Teilungskriterium Anzahl der gelösten Aufgaben

Das Ergebnis des LRT mit dem Teilungskriterium Anzahl der gelösten Aufgaben ist nun nicht mehr signifikant.

χ^2 -Wert	df	p-Wert
25,073	17	0,093

Tabelle 8.11: LRT *Syllogismen 2010*, Score (ohne Items 5, 8)

Auch die Graphische Modellkontrolle liefert jetzt ein zufriedenstellendes Ergebnis:

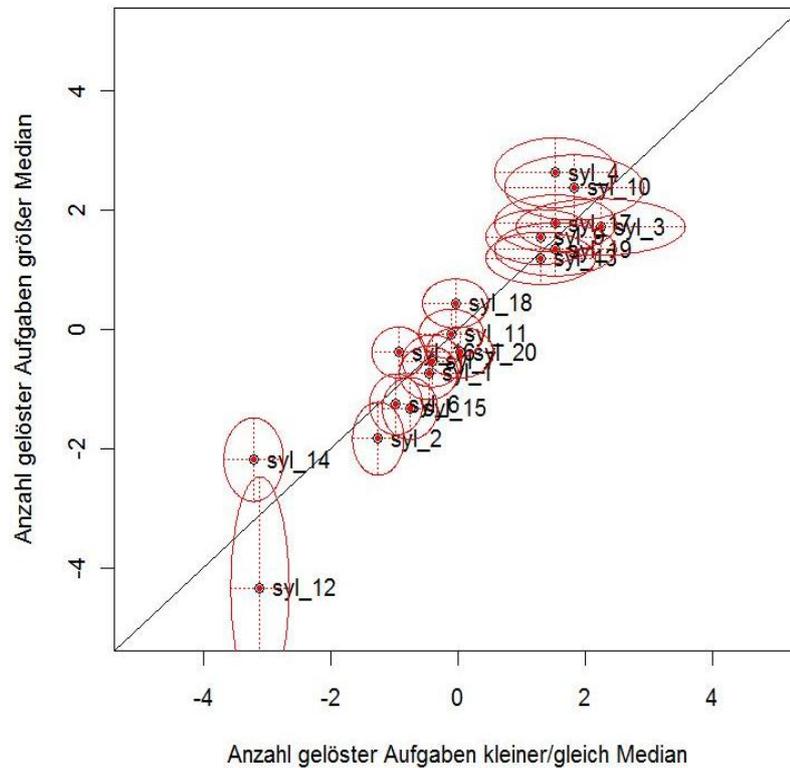


Abb. 8.9: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 5, 8), Score

8.1.1.3.2 Teilungskriterium Bildungsniveau

Das Ergebnis des LRT mit dem Teilungskriterium Bildungsniveau ist weiterhin nicht signifikant (mit einem p-Wert größer als 0,01).

χ^2 -Wert	df	p-Wert
31,288	17	0,018

Tabelle 8.12: LRT *Syllogismen 2010*, Bildungsniveau (ohne Items 5, 8)

Ebenso liefert der die Graphische Modellkontrolle weiterhin ein zufriedenstellendes Ergebnis:

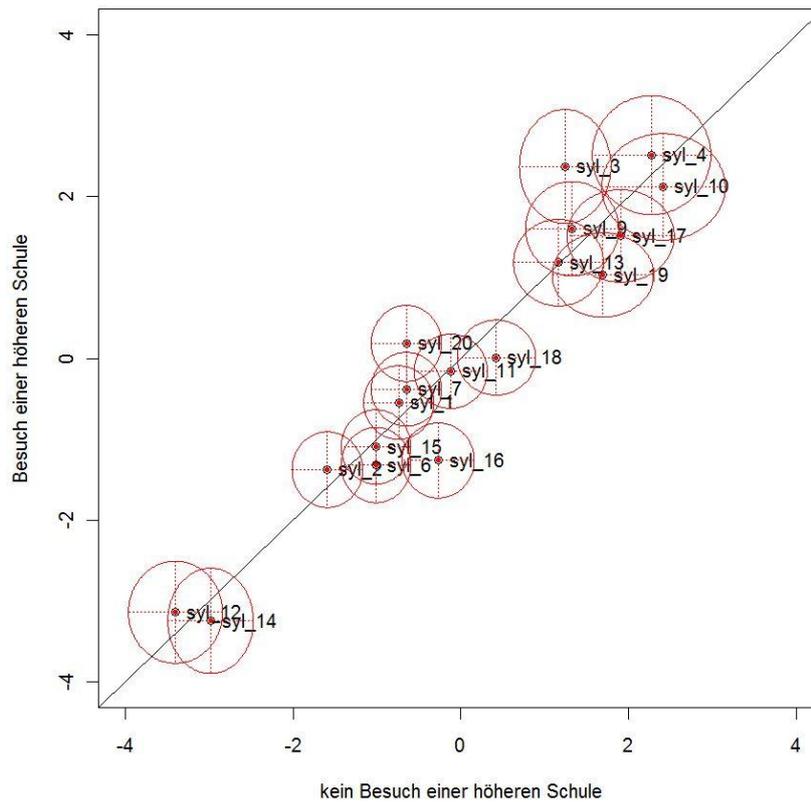


Abb. 8.10: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 5, 8), Bildungsniveau

8.1.1.3.3 Teilungskriterium Muttersprache

Auch das Ergebnis des LRT mit dem Teilungskriterium Muttersprache ist weiterhin nicht signifikant.

χ^2 -Wert	df	p-Wert
21,74	17	0,195

Tabelle 8.13: LRT *Syllogismen 2010*, Muttersprache (ohne Items 5, 8)

Das Ergebnis der Graphischen Modellkontrolle ist weiterhin zufriedenstellend:

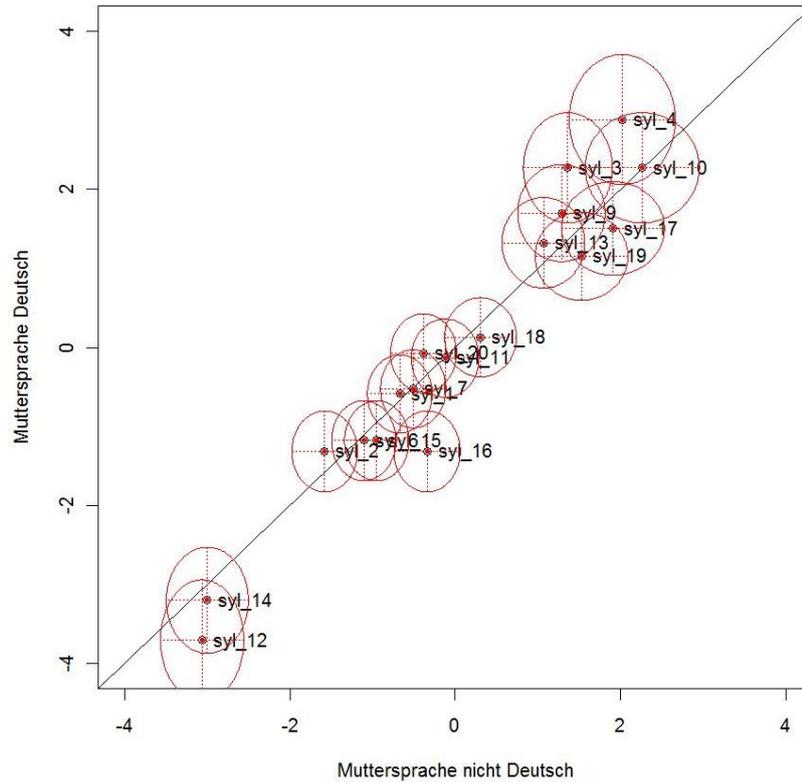


Abb. 8.11: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 5, 8), Muttersprache

8.1.1.3.4 Teilungskriterium Versuchsbedingung

Auch das Ergebnis des LRT mit dem Teilungskriterium Versuchsbedingung ist weiterhin nicht signifikant.

χ^2 -Wert	df	p-Wert
22,721	17	0,159

Tabelle 8.14: LRT *Syllogismen 2010*, Versuchsbedingung (ohne Items 5, 8)

Die Graphische Modellkontrolle liefert ebenfalls ein zufriedenstellendes Ergebnis:

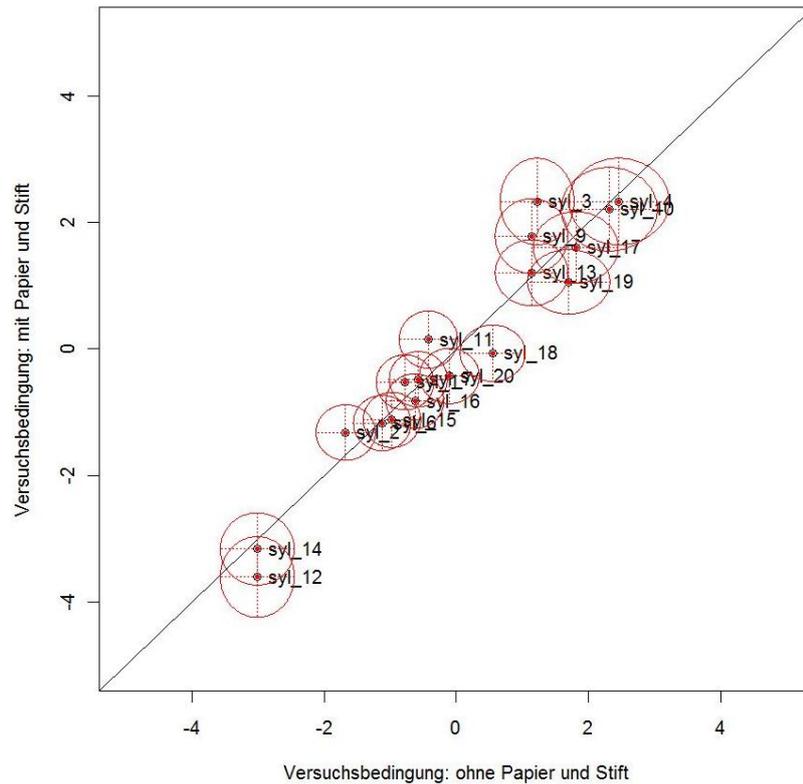


Abb. 8.12: Graphische Modellkontrolle *Syllogismen 2010* (ohne Item 5, 8), Versuchsbedingung

8.1.1.3.5 Zusammenfassung und Interpretation der Ergebnisse zu *Syllogismen 2010*

Nach Ausschluss der Items 5 und 8 ist a-posteriori Modellgültigkeit gegeben. Die Nullhypothese $H_0(1)$ kann somit beibehalten werden. Die verbleibenden 18 Items können als konform mit dem dichotomen logistischen Testmodell von Rasch angesehen werden. Sie messen eindimensional und die Summe der gelösten Aufgaben bildet die Leistung entsprechend der empirisch gegebenen Unterschiede ab.

Inhaltliche Gründe für die fehlende Modellgültigkeit der Items 5 und 8 konnten nicht gefunden werden. Zwar sind in beiden Items die Begriffe erfunden bzw. nicht aus der Alltagssprache entnommen, also eher unbekannt, und ein Syllogismus der dritten Figur wird verwendet, diese beiden Eigenschaften kommen jedoch auch bei anderen der vorgegebenen Items vor, auch in dieser Kombination (in Item 6), so dass es als sehr unwahrscheinlich erscheint, dass die Ursache für die Modellabweichung hierin begründet ist.

Im Fall der a-posteriori Modellgültigkeit nach Ausschluss einer bestimmten Anzahl von Items ist im Sinne einer Art Kreuzvalidierung die Vorgabe des reduzierten Itempools und erneute Modellprüfung anhand einer anderen Stichprobe notwendig. Erst wenn sich dann eine a-priori Modelgeltung zeigt, kann die Konformität der Items mit dem Modell von Rasch als gesichert angesehen werden (Kubinger, 2005).

Die testtheoretische Überprüfung der weiteren konstruierten Items ist in zukünftigen Untersuchungen durchzuführen.

Die Schätzungen der Itemleichtigkeitsparameter stellen sich wie folgt dar:

Item	Schätzwert	Std. Schätzfehler	Untere Grenze KI	Obere Grenze KI
syl_1	0,642	0,157	0,335	0,949
syl_2	1,483	0,160	1,170	1,795
syl_3	-1,741	0,218	-2,168	-1,313
syl_4	-2,384	0,259	-2,891	-1,877
syl_6	1,134	0,157	0,827	1,442
syl_7	0,517	0,157	0,209	0,825
syl_9	-1,461	0,205	-1,862	-1,060
syl_10	-2,249	0,249	-2,737	-1,761
syl_11	0,131	0,161	-0,184	0,446
syl_12	3,273	0,214	2,853	3,693
syl_13	-1,173	0,193	-1,550	-0,795
syl_14	3,063	0,203	2,665	3,461
syl_15	1,036	0,156	0,729	1,343
syl_16	0,716	0,156	0,410	1,023
syl_17	-1,691	0,216	-2,114	-1,269
syl_18	-0,225	0,166	-0,550	0,101
syl_19	-1,333	0,199	-1,723	-0,943
syl_20	0,262	0,159	-0,050	0,574

Tabelle 8.15: Schätzwerte der Itemleichtigkeitsparameter für *Syllogismen 2010* (ohne Items 5,8) mit 95%-Konfidenzintervall

8.1.2 Rasch-Modell-Analysen hinsichtlich der Items des *FRRT*

Für die 12 verwendeten Items des *FRRT* wird ebenfalls eine Prüfung der Geltung des Rasch-Modells mittels (bedingten) Likelihood-Ratio-Tests (LRT) nach Andersen mit den Teilungskriterien Anzahl gelöster Aufgaben, Bildungsniveau, Muttersprache und Versuchsbedingung durchgeführt.

8.1.2.1 Teilungskriterium Anzahl der gelösten Aufgaben

Der LRT mit dem Teilungskriterium Anzahl gelöster Aufgaben (Score) führt zu einem nicht signifikanten Ergebnis¹⁶:

χ^2 -Wert	df	p-Wert
14,204	10	0,164

Tabelle 8.16: LRT *FRRT*, Score

Auch das Ergebnis der Graphischen Modellkontrolle ist zufriedenstellend:

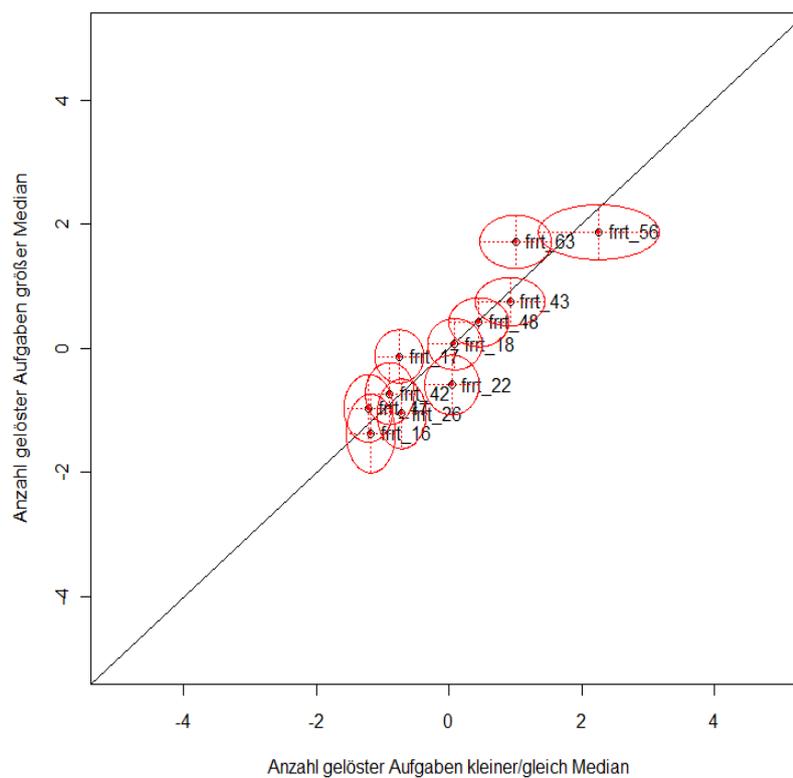


Abb. 8.13: Graphische Modellkontrolle *FRRT*, Score

8.1.2.2 Teilungskriterium Bildungsniveau

Für das Teilungskriterium Bildungsniveau ist der LRT nach Andersen ebenfalls nicht signifikant:

¹⁶ Das erste vorgegebene Item (frrt_2) konnte hier nicht berücksichtigt werden, da es von allen Personen der oberen Leistungsgruppe (Anzahl von gelösten Aufgaben größer als der Median) gelöst worden war.

χ^2 -Wert	df	p-Wert
12,201	11	0,349

Tabelle 8.17: LRT *FRRT*, Bildungsniveau

Auch die Graphische Modellkontrolle liefert ein zufriedenstellendes Ergebnis:

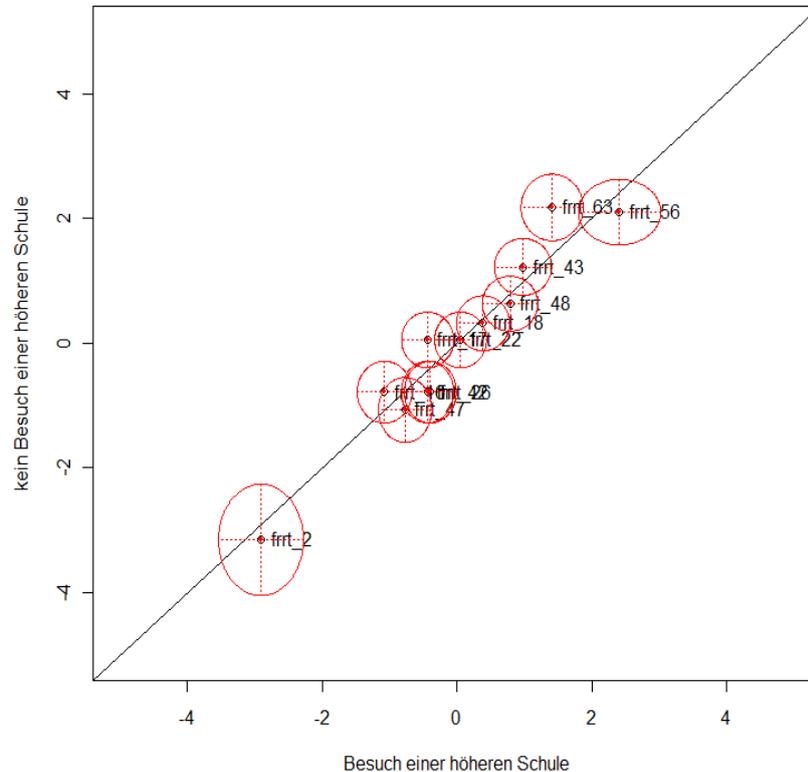


Abb. 8.14: Graphische Modellkontrolle *FRRT*, Bildungsniveau

8.1.2.3 Teilungskriterium Muttersprache

Das Ergebnis des LRT für das Teilungskriterium Muttersprache ist ebenfalls nicht signifikant:

χ^2 -Wert	df	p-Wert
17,046	11	0,107

Tabelle 8.18: LRT *FRRT*, Muttersprache

Auch die Graphische Modellkontrolle liefert ein zufriedenstellendes Ergebnis.

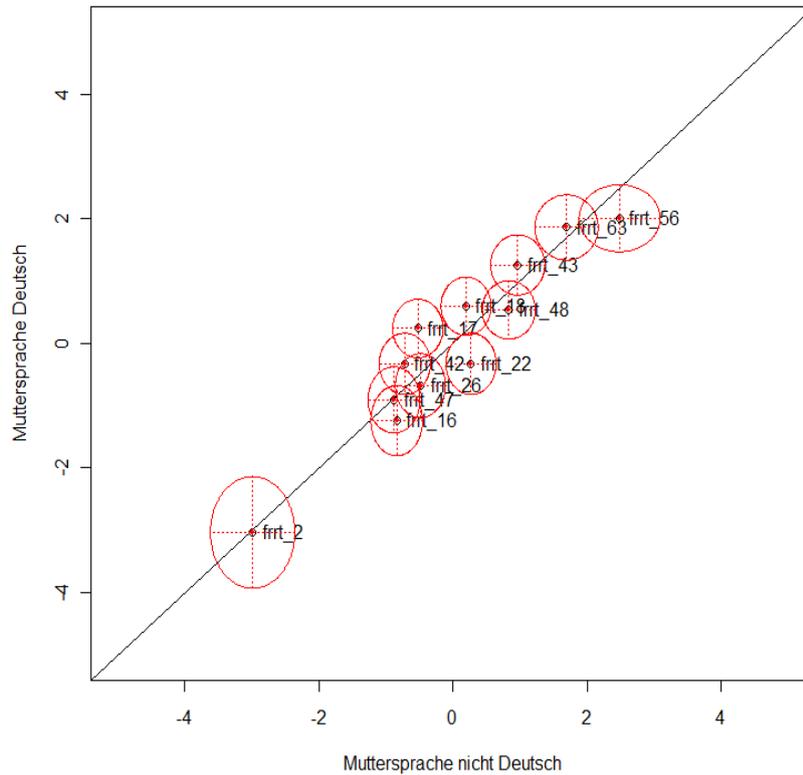


Abb. 8.15: Graphische Modellkontrolle *FRRT*, Muttersprache

8.1.2.4 Teilungskriterium Versuchsbedingung

Ebenso ist das Ergebnis des LRT mit dem Teilungskriterium Versuchsbedingung nicht signifikant.

χ^2 -Wert	df	p-Wert
11,99	11	0,364

Tabelle 8.19: LRT *FRRT*, Versuchsbedingung

Die Graphische Modellkontrolle liefert ebenfalls ein zufriedenstellendes Ergebnis.

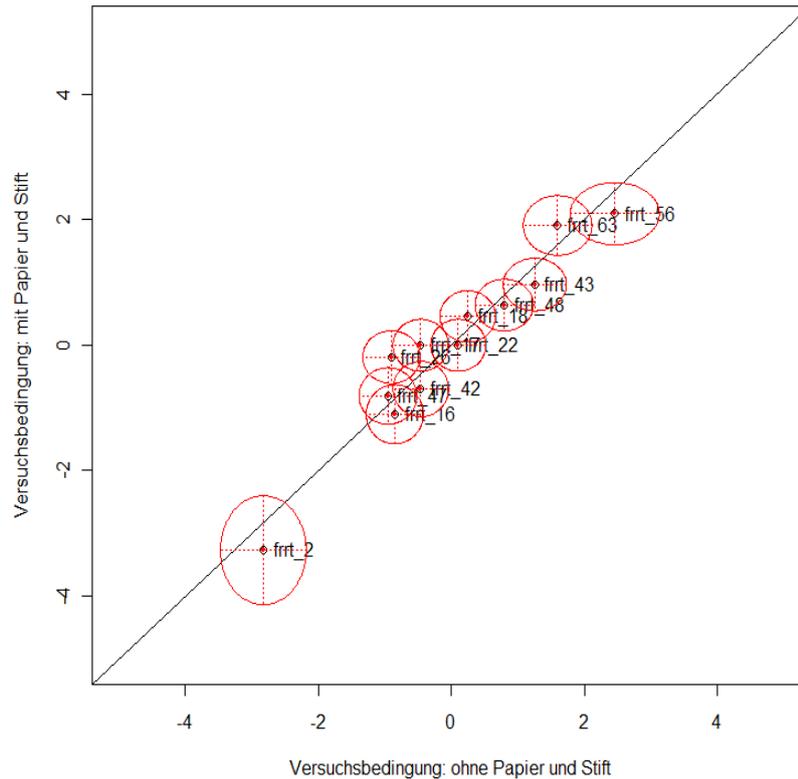


Abb. 8.16: Graphische Modellkontrolle *FRRT*, Versuchsbedingung

8.1.2.5 Zusammenfassung und Interpretation der Ergebnisse zum *FRRT*

Es kann Modellgültigkeit für alle 12 vorgegeben Items des *FRRT* angenommen werden. Die Nullhypothese $H_0(2)$ kann beibehalten werden.

Es ergibt sich somit eine weitere Absicherung der Rasch-Modell-Geltung in Bezug auf die hier verwendeten Items des *FRRT*.

Die Ergebnisse der Schätzungen der Itemleichtigkeitsparameter fallen wie folgt aus:

Item	Schätzwert	Std. Schätzfehler	Untere Grenze KI	Obere Grenze KI
frrt_2	2,946	0,256	2,444	3,448
frrt_47	0,882	0,159	0,571	1,193
frrt_16	0,963	0,160	0,649	1,278
frrt_17	0,234	0,150	-0,060	0,528
frrt_18	-0,350	0,149	-0,642	-0,058
frrt_22	-0,048	0,149	-0,339	0,243
frrt_26	0,549	0,153	0,248	0,849
frrt_42	0,599	0,154	0,297	0,900
frrt_43	-1,085	0,158	-1,396	-0,774
frrt_48	-0,706	0,152	-1,004	-0,408

frtt_56	-2,217	0,200	-2,610	-1,825
frtt_63	-1,766	0,179	-2,117	-1,415

Tabelle 8.20: Schätzung der Itemleichtigkeitsparameter für *FRRT* mit 95%-Konfidenzintervall

8.2 Deskriptive Ergebnisse

Im Folgenden werden die relativen Lösungshäufigkeiten pro Aufgabe für die Items der *Syllogismen 2010* sowie die Items des *FRRT* (ohne Berücksichtigung der jeweils pro Test ausgeschlossenen Personen) dargestellt, außerdem die durchschnittliche Lösungshäufigkeit innerhalb der Stichprobe über alle Aufgaben sowie die Verteilung der Anzahl gelöster Aufgaben in der Stichprobe. Zur Berechnung der deskriptiven Statistiken wurde PASW Statistics 18 verwendet.

8.2.1 Deskriptive Ergebnisse zu *Syllogismen 2010*

In der folgenden Tabelle sind die relativen Lösungshäufigkeiten in der Stichprobe für die Items der *Syllogismen 2010* (ohne Item 5 und 8) dargestellt:

Item	n	Mittelwert
syl_1	228	0,46
syl_2	228	0,61
syl_3	228	0,13
syl_4	228	0,08
syl_6	228	0,55
syl_7	228	0,44
syl_9	228	0,16
syl_10	228	0,09
syl_11	228	0,37
syl_12	228	0,86
syl_13	228	0,19
syl_14	228	0,84
syl_15	228	0,53
syl_16	228	0,47
syl_17	228	0,14
syl_18	228	0,32
syl_19	228	0,17
syl_20	228	0,39

Tabelle 8.21: Relative Lösungshäufigkeit *Syllogismen 2010* (ohne Items 5, 8)

Die durchschnittliche Lösungshäufigkeit über alle modellkonformen Items hinweg beträgt in der Stichprobe (ohne die ausgeschlossenen Personen) ca. 0,378; durchschnittlich lösten die im Rahmen der Auswertung berücksichtigten 228 Teilnehmer 6,81 Aufgaben (0 bis 17 Aufgaben, Standardabweichung: 3,95).

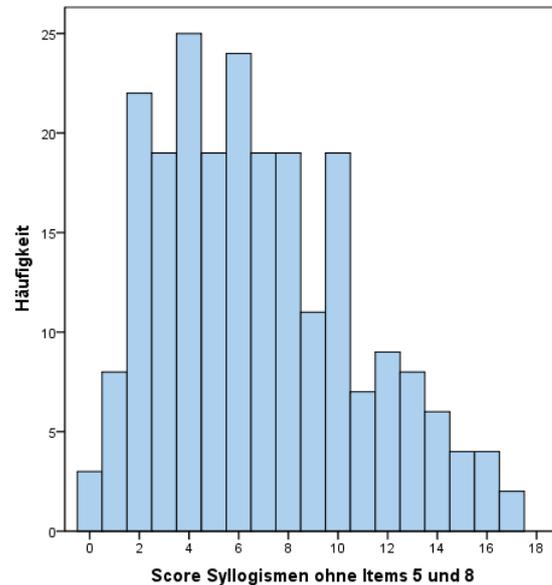


Abb. 8.17: Histogramm Anzahl gelöster Aufgaben *Syllogismen 2010* (ohne Items 5, 8)

Die relative Lösungshäufigkeit für die beiden ausgeschiedenen Items 5 und 8 beträgt 0,71 bzw. 0,09.

8.2.2 Deskriptive Ergebnisse zum *FRRT*

Für den *FRRT* ergaben sich die folgenden relativen Lösungswahrscheinlichkeiten:

Item	n	Mittelwert
frrt_2	219	0,92
frrt_47	219	0,68
frrt_16	219	0,69
frrt_17	219	0,56
frrt_18	219	0,44
frrt_22	219	0,50
frrt_26	219	0,62
frrt_42	219	0,62
frrt_43	219	0,31
frrt_48	219	0,37
frrt_56	219	0,15
frrt_63	219	0,20

Tabelle 8.22: Relative Lösungshäufigkeiten *FRRT*

Die relative Lösungshäufigkeit über alle Items beträgt in der Stichprobe (ohne die ausgeschlossenen Personen) durchschnittlich circa 0,50; es wurden durchschnittlich 6,05 Items gelöst (0 bis 12, Standardabweichung 2,72).

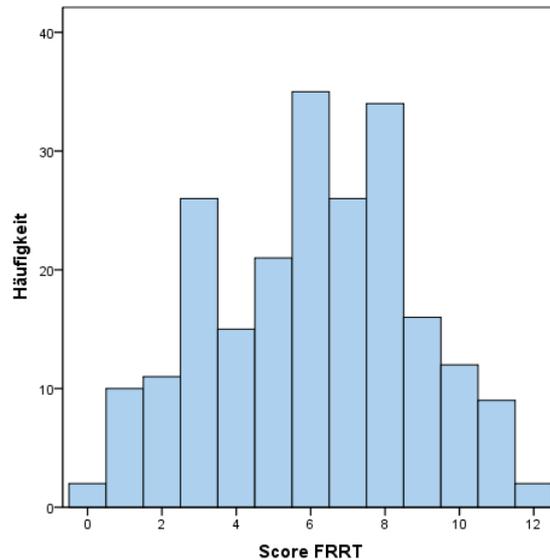


Abb. 8.18: Histogramm Anzahl gelöster Aufgaben *FRRT*

8.3 Korrelation zwischen *Syllogismen 2010* und *FRRT*

Da aufgrund der Geltung des Rasch-Modells (für die nicht ausgeschiedenen Items) Normalverteilung und Intervallskalierung der Testscores (Anzahl gelöster Aufgaben) angenommen werden kann, wird die Berechnung des Zusammenhangs an Hand des Pearson-Korrelationskoeffizienten durchgeführt; diese erfolgt unter Verwendung der Software PASW Statistics 18.

Es besteht ein signifikanter positiver Zusammenhang zwischen der Leistung (Anzahl gelöster Aufgaben) in den Rasch-Modell-konformen Items der *Syllogismen 2010* und im *FRRT*. Die Hypothese $H_0(3)$ kann somit verworfen und die $H_1(3)$ angenommen werden. Der Korrelationskoeffizient nach Pearson beträgt $r = 0,525$, das Bestimmtheitsmaß $B = 0,2756$. Das Ergebnis in einem der beiden Tests erklärt somit ca. 28 % der Varianz der Leistung im jeweils anderen Test.

Korrelationen

		Score_Syll_ohne_5_8	Score_FRRT
Score_Syll_ohne_5_8	Korrelation nach Pearson	1	,525
	Signifikanz (2-seitig)		,000
	N	218	214

Tabelle 8.23: Korrelation zwischen Anzahl gelöster Aufgaben in *Syllogismen 2010* (RM-konforme Items) und *FRRT*¹⁷

Die Übereinstimmungsvalidität kann also als gegeben betrachtet werden, beide Tests messen grundsätzlich dasselbe Konstrukt¹⁸.

9. Diskussion

Es ist im Rahmen dieser Arbeit gelungen, Rasch-Modell-konforme Aufgaben zu konstruieren, denen Syllogismen unterschiedlicher Figuren zugrunde liegen, wie dies bereits bei Srp (1993) der Fall war. Syllogismen eignen sich also offenbar als Material für die Testkonstruktion im Rahmen der Item-Response-Theorie.

Hinsichtlich der Stichprobe, die aus Stellungspflichtigen bei der Stellungskommission Wien besteht, ist zu sagen, dass diese im Gegensatz zu einer studentischen Stichprobe einerseits eine breite Streuung in Bezug auf das Bildungsniveau aufweist und auch viele Personen mit nicht-deutscher Muttersprache enthält, andererseits allerdings nur männliche Personen. Eine Prüfung der Geltung des Rasch-Modells mittels LRT anhand des Teilungskriteriums Geschlecht war daher nicht möglich, so dass bislang nicht sichergestellt werden konnte, dass sich die Parameterschätzungen in nach Geschlecht getrennten Gruppen statistisch nicht unterscheiden.

In Hinblick auf die empirische Untersuchung ist als kritisch zu betrachten, dass in der zweiten Phase der Untersuchung die Motivation einiger Teilnehmer eher gering gewesen zu sein scheint. Es ist auch nicht auszuschließen, dass Ermüdungseffekte

¹⁷ Die Korrelation ist auf dem zuvor festgelegten Niveau für das Risiko 1. Art von $\alpha = 0,05$ signifikant (vgl. dazu Rasch, Kubinger, Schmidtke, & Häusler, 2004)

¹⁸ Bei der Beurteilung der (Übereinstimmungs-)Validität als Korrelation zwischen zwei jeweils messfehler-behafteten Testwerten ist zu beachten, dass die maximal erreichbare Validität in diesem Sinn durch die Reliabilität der beiden Tests beschränkt ist und dadurch die „wahre“ Validität als Korrelation zwischen den wahren Messwerten stets unterschätzt wird (vgl. Rost, 2004).

durch die vorangegangene psychologische Stellungstestung auftraten. Dies ließ sich organisatorisch leider nicht anders lösen. Die 14 bzw. 18 Testpersonen, die auffällig kurze Bearbeitungszeiten von durchschnittlich weniger als 10 Sekunden bei den *Syllogismen 2010* bzw. dem *FRRT* aufwiesen (wobei alle gleichzeitig nur eine sehr geringe Anzahl von Aufgaben gelöst hatten) wurden im Rahmen der Auswertung nicht berücksichtigt, da in diesem Fall davon ausgegangen werden musste, dass die betreffenden Personen den jeweiligen Test nicht ernsthaft bearbeitet haben.

Ein Problem, das sich bei den *Syllogismen* (Srp, 1994) gezeigt hatte, und auch im neu konstruierten Test nicht gelöst werden konnte, ist, dass sehr fähige Testpersonen teilweise zu Recht bemängelten, dass bei einem Syllogismus, der eine universelle Konklusion enthält, eigentlich stets nicht nur diese universelle, sondern gleichzeitig die entsprechende partikuläre Konklusion logisch korrekt gefolgt werden kann. Dies ergibt sich daraus, dass die universelle Konklusion die partikuläre impliziert: Wenn „alle“ A B sind, dann sind immer auch „einige“ A B, da der partikuläre Quantor „einige“ im Sinne von „mindestens ein“ definiert ist. Eine mögliche Lösung für dieses Problem könnte sein, die Testpersonen über die Instruktion anzuweisen, nur die universelle Konklusion als richtig zu betrachten, wenn diese möglich ist. Allerdings ist die Instruktion bereits ohne diesen zusätzlichen Hinweis relativ lang und eher kompliziert, so dass ein solcher Hinweis weniger begabte Testpersonen möglicherweise überfordert bzw. verwirrt. Alle Syllogismenformen, bei denen dieses Problem auftritt, bei der Itemkonstruktion nicht zu berücksichtigen ist keine Alternative: Dann wären nur noch Items mit partikulärer Konklusion enthalten, die Lösung also in jedem Fall eine die Quantoren „einige“ oder „einige nicht“ enthaltende Aussage darstellen.

Ein weiteres noch nicht vollständig gelöstes Problem stellt der unterschiedliche Gebrauch des Quantors „einige“ in der formalen Logik (in der Bedeutung von „mindestens ein“) im Vergleich zur Alltagssprache (in der Bedeutung von einige (wenige), aber nicht alle; also strikt partikulär) dar. Dies kann dazu führen, dass Prämissen, die diesen Quantor enthalten, falsch interpretiert werden (dazu beispielsweise auch schon Neubauer, 1992). Auch in der von der Autorin durchgeführten Voruntersuchung zu den neu konstruierten Items zeigte sich, dass nicht wenige Personen den Quantor „einige“ als strikt partikulär interpretieren und

daher davon ausgehen, dass aus einem I-Urteil („Einige A sind B.“) auf ein O-Urteil („Einige A sind nicht B.“) geschlossen werden könne. Das führt gleichzeitig dazu, dass die Lösung einer Syllogismenaufgabe mit partikulärer Konklusion als uneindeutig angesehen wird: Ist die korrekte Konklusion ein I-Urteil, so wird gleichzeitig die Geltung des entsprechenden O-Urteils angenommen. In der vorliegenden Arbeit wurde versucht, dieses Problem zu lösen, indem eine Erklärung der Bedeutung der partikulären Quantoren über den „Hilfe“-Text gegeben wurde, in welchem ausgeführt ist, dass „einige“ gleichbedeutend ist mit „mindestens ein“ und eben nicht die Möglichkeit ausschließt, dass auch allen Subjekten die zugeschriebene Eigenschaft zukommen (bzw. im Fall des O-Urteils nicht zukommen) kann. Problematisch hierbei ist jedoch, dass möglicherweise nicht alle Testpersonen diese Erklärung, auf welche in der Instruktion verwiesen wird und welche während des Tests jederzeit aufgerufen werden kann, beachtet haben. Um dies zu verhindern wurde allerdings das zweite Beispielitem so konstruiert, dass die Testpersonen nur zur Lösung und damit zu einer positiven Rückmeldung kommen, wenn sie die entsprechende falsche Antwortmöglichkeit (in diesem Fall eine partikuläre negative Aussage) ablehnen, da diese vor der korrekten (hier partikulären affirmativen Aussage) angeboten wird. Da neben den Standard-Quantoren aber auch alternative sprachliche Formulierungen verwendet wurden, könnte es auch zu Verständnisproblemen in Bezug darauf, ob die zum partikulären Quantor gegebenen Erklärungen ebenfalls für die sprachlich anders formulierten Varianten dieses Quantoren gelten, kommen, weil angenommen werden könnte, dass die unterschiedliche Formulierung als Hinweis auf eine andere Bedeutung zu interpretieren sei.

Als mögliche Lösung für dieses Problem erscheint die Verwendung des Quantors „einige, aber nicht alle“ (strikt partikuläres Urteil) in Entsprechung zur Alltagssprache statt des in der formalen Logik üblichen partikulären Quantors. Hierbei stellt sich jedoch das Problem, dass die Mehrzahl der gültigen Syllogismen der klassischen Logik dann nicht mehr funktioniert – ein komplett neues System der Logik wäre erforderlich. Ein neues System der Quantorenlogik mit eindeutigen Urteilen, in welchem auch das strikt partikuläre Urteil enthalten ist, hat Menne (1954) entwickelt. Dieses hat große theoretische Vorteile, allerdings ist eine verständliche sprachliche Umsetzung der Quantoren nach Menne im Rahmen von Testaufgaben wohl kaum

möglich (Neubauer, 1992). Denkbar wäre jedoch, diese Quantoren in den Items eines Tests einzusetzen, sie dabei aber nicht sprachlich zu formulieren, sondern durch Symbole darzustellen, deren Bedeutung den Testpersonen in der Instruktion und in einem während der Testung jederzeit aufrufbaren „Hilfe“-Text zu beschreiben. Hierdurch könnte das Problem gelöst werden, gleichzeitig würde jedoch der Reiz der sprachlichen Formulierung des Itemmaterials, wodurch sich ein Syllogismen-Test von anderen Tests zur Messung der Reasoningfähigkeit, die figurales Material verwenden, unterscheidet, und wodurch die Aufgaben wahrscheinlich von vielen Testpersonen als ansprechender wahrgenommen werden, größtenteils verloren gehen.

Zur Verringerung der Ratewahrscheinlichkeit wurde jeweils eine der drei nicht korrekten Konklusionen (welche Aussagen mit den drei nicht in der Lösung vorkommenden Quantoren enthalten) in einer anderen sprachlichen Formulierung als zusätzliche Antwortmöglichkeit angeboten, um insgesamt 5 Antwortmöglichkeiten zu erhalten. Dies ist wenig problematisch bei der sequentiellen Darbietung der Antwortmöglichkeiten, wie sie in der vorliegenden Arbeit erfolgte. Bei einer Vorgabe im „klassischen“ Multiple-Choice-Antwortformat ist dies jedoch kontraproduktiv, da sich den Testpersonen hierdurch die Möglichkeit bietet, mithilfe einer Ausschluss technik die faktische Ratewahrscheinlichkeit durch das Verwerfen derjenigen Antwortmöglichkeit die zweimal, nur in unterschiedlichen Formulierungen angeboten wird, noch über diejenige, die bei einem gewöhnlichen „1 aus 4“-Antwortformat gegeben ist, zu erhöhen. Es stehen dann praktisch nur noch drei tatsächlich in Frage kommende Möglichkeiten zur Auswahl, von denen Testpersonen mit mäßiger Fähigkeit oft nochmals eine ausschließen können (beispielsweise eine verneinende Aussage, wenn in den Prämissen keine Verneinung enthalten war), so dass letztendlich eine Wahrscheinlichkeit für das „Erraten“, also das zufällige Wählen der richtigen Antwortmöglichkeit, von $1/2$ gegeben ist. Stattdessen sollten daher bei einer Vorgabe im klassischen MC-Format besser nur vier Antwortmöglichkeiten oder statt der zusätzlichen fünften Antwortmöglichkeit eine Antwortmöglichkeit in der Art von „Keine der angegebenen Schlussfolgerungen ist richtig“ angeboten werden. Bei Syllogismenaufgaben erscheint diese Antwortmöglichkeit auch nicht von Vorneherein als falsch, wie dies bei anderen Tests mit der Antwortoption „Keine der anderen angegebenen Antworten ist richtig“ der Fall sein mag. In einigen (aus psychologisch-

diagnostischer Sicht wohl wenig sinnvollen) Syllogismen-Tests, wie sie beispielsweise im Internet zu finden sind, aber auch bei Syllogismenaufgaben wie sie teilweise in der kognitionspsychologischen Forschung verwendet werden, ist diese Antwortmöglichkeit bei einem Teil der Aufgaben tatsächlich korrekt. Denkbar wäre auch die Aufnahme von ungültigen Syllogismenformen in den Test, bei denen eben diese Antwortmöglichkeit die Lösung darstellt. Im Fall der Syllogismen erscheint dies nicht als „Irreführung“ der Testpersonen, die in der Regel nicht erwarten, dass ihnen in einem Test unlösbare Aufgaben vorgegeben werden, da es ja dann ausdrücklich auch darum geht, zu beurteilen, ob sich aus den Prämissen überhaupt ein gültiger Schluss ableiten lässt. Andererseits kann nicht ausgeschlossen werden, dass in Bezug auf die Wahl dieser Antwortmöglichkeit andere Faktoren als die zu messende Fähigkeit einen Einfluss haben, insbesondere bestimmte Persönlichkeitseigenschaften (vgl. Kubinger & Wolfsbauer, 2010).

Jedenfalls ist es im Rahmen dieser Arbeit gelungen, Items zu konstruieren, die sich in dieser ersten testtheoretischen Überprüfung ganz überwiegend als Rasch-Modellkonform erwiesen haben und eine zufriedenstellende Übereinstimmungsvalidität mit dem *FRRT* zeigen. Das verwendete Material ist ansprechend und die Items sind sprachlich abwechslungsreich gestaltet.

Wenn weitere Untersuchungen zur Geltung des Rasch-Modells durchgeführt worden sind, besteht auch die Möglichkeit des Adaptiven Testens in der Form des *tailored testing* als äußerst ökonomischer Form des Testens (vgl. Kubinger, 1996).

10. Zusammenfassung

Im Rahmen der vorliegenden Diplomarbeit wurden insgesamt 55 Items für den neu entwickelten Test *Syllogismen 2010* zur Messung der verbalen Reasoning-Fähigkeit entworfen. Ausgehend von den 19 schlüssigen Syllogismen in der klassischen Logik wurden Items erstellt, wobei die sprachliche Formulierung der Quantoren variiert wurde, um die Aufgaben abwechslungsreicher zu gestalten. Im Vergleich zum Test *Syllogismen* (Srp, 1994), in Anlehnung an welchen die Testkonstruktion erfolgte, wirken die Aufgaben dadurch weniger gleichförmig. Durch die unterschiedliche Art und Kombination der enthaltenen Quantoren sowie die verschiedenen Figuren der Syllogismen (Stellung des Mittelbegriffs) ist bereits durch die Verwendung aller 19

Typen von schlüssigen Syllogismen eine starke Streuung der Schwierigkeit der Aufgaben gegeben. Zur weiteren Variation der Schwierigkeit wurde die Bekanntheit der verwendeten Begriffe variiert und die Formulierung der Aussagen im Konjunktiv eingesetzt. Weitere Items für den Test *Syllogismen 2010* konstruierte Hoffmann (in Vorbereitung) nach ähnlichen Prinzipien im Rahmen ihrer Diplomarbeit.

In der empirischen Untersuchung zur Kalibrierung nach dem dichotomen logistischen Modell von Rasch, welche gemeinsam mit Hoffmann durchgeführt wurde, wurden bei der Stellungskommission Wien insgesamt 245 Personen zuerst 20 Items des neu konstruierten Tests *Syllogismen 2010* und im Anschluss daran 12 Items des *FRRT* (Poinstingl, Kubinger, Skoda & Schechtner, in Vorbereitung) vorgegeben. Zur Prüfung der Frage, ob die (Möglichkeit der) Verwendung von Papier und Stift während der Bearbeitung der verbalen Reasoningtests zu einer Verbesserung der Testleistung führt, wurde außerdem ein Experiment durchgeführt; die Ergebnisse hierzu finden sich bei Hoffmann (in Vorbereitung).

Die testtheoretische Analyse der Items ergab, dass 18 der untersuchten Items der *Syllogismen 2010* mit dem Rasch-Modell konform sind. Von den ursprünglich 20 Items mussten 2 Aufgaben ausgeschieden werden, um eine a-posteriori Modellgeltung zu erreichen. Eine inhaltliche Begründung für die fehlende Modellgeltung in Bezug auf diese Items konnte bislang nicht gefunden werden.

Es wurde also ein Test zur Messung der Fähigkeit des schlussfolgernden Denkens im sprachlichen Bereich entwickelt, dessen Items (soweit sie hier untersucht wurden), dem Modell von Rasch entsprechen und somit das Gütekriterium der Skalierung erfüllen sowie eindimensional messen. Zur Absicherung der diesbezüglichen Ergebnisse ist die Vorgabe des um die nicht modellkonformen Items reduzierten Itempools und die erneute Prüfung der Modellgeltung anhand einer anderen Stichprobe erforderlich, also eine „Art Kreuzvalidierung“ durchzuführen. Erst wenn bei dieser a-priori Modellgeltung festgestellt wird, kann die Konformität der Items mit dem Modell von Rasch als gesichert angesehen werden (Kubinger, 2005). In Bezug auf die bislang noch nicht vorgegebenen Items ist in zukünftigen Untersuchungen in derselben Weise vorzugehen.

Die 12 verwendeten Items des *FRRT* entsprechen dem Rasch-Modell; für diesen Test war Modellgeltung von Beginn an für alle Items gegeben. Aus der vorliegenden Untersuchung ergibt sich somit eine weitere Absicherung der Rasch-Modell-Konformität in Bezug auf die verwendeten Items dieses noch unveröffentlichten Tests.

Die Berechnung der Korrelation zwischen den Rasch-Modell-konformen Items der *Syllogismen 2010* und dem *FRRT* ergab einen Korrelationskoeffizienten nach Pearson von $r = 0,525$. Es kann somit die sogenannte Übereinstimmungsvalidität (vgl. Kubinger, 2009) angenommen und davon ausgegangen werden, dass die beiden Verfahren grundsätzlich dasselbe Konstrukt messen.

11. Abbildungsverzeichnis

Abb. 8.1: Graphische Modellkontrolle <i>Syllogismen 2010</i> (alle Items), Score	34
Abb. 8.2: Graphische Modellkontrolle <i>Syllogismen 2010</i> (alle Items), Bildungsniveau	35
Abb. 8.3: Graphische Modellkontrolle <i>Syllogismen 2010</i> (alle Items), Muttersprache	36
Abb. 8.4: Graphische Modellkontrolle <i>Syllogismen 2010</i> (alle Items), Versuchsbedingung.....	37
Abb. 8.5: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 8), Score	38
Abb. 8.6: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 8), Bildungsniveau	39
Abb. 8.7: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 8), Muttersprache	40
Abb. 8.8: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 8), Versuchsbedingung.....	41
Abb. 8.9: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 5, 8), Score.....	42
Abb. 8.10: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 5, 8), Bildungsniveau	43
Abb. 8.11: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 5, 8), Muttersprache	44
Abb. 8.12: Graphische Modellkontrolle <i>Syllogismen 2010</i> (ohne Item 5, 8), Versuchsbedingung.....	45
Abb. 8.13: Graphische Modellkontrolle <i>FRRT</i> , Score	47
Abb. 8.14: Graphische Modellkontrolle <i>FRRT</i> , Bildungsniveau.....	48
Abb. 8.15: Graphische Modellkontrolle <i>FRRT</i> , Muttersprache	49
Abb. 8.16: Graphische Modellkontrolle <i>FRRT</i> , Versuchsbedingung.....	50
Abb. 8.17: Histogramm Anzahl gelöster Aufgaben <i>Syllogismen 2010</i> (ohne Items 5, 8).....	52
Abb. 8.18: Histogramm Anzahl gelöster Aufgaben <i>FRRT</i>	53

12. Tabellenverzeichnis

Tabelle 1.1: Arten von Urteilen	4
Tabelle 1.2: Figuren des Syllogismus (Standardform)	5
Tabelle 4.1: Verteilung der Position der Lösung innerhalb der Antwortoptionen.....	25
Tabelle 8.1: LRT <i>Syllogismen 2010</i> , Score (alle Items).....	33
Tabelle 8.2: Z-Test nach Fischer und Scheiblechner <i>Syllogismen 2010</i> , Score.....	35
Tabelle 8.3: LRT <i>Syllogismen 2010</i> , Bildungsniveau (alle Items)	35
Tabelle 8.4: LRT <i>Syllogismen 2010</i> , Muttersprache (alle Items)	36
Tabelle 8.5: LRT <i>Syllogismen 2010</i> , Versuchsbedingung (alle Items).....	36
Tabelle 8.6: LRT <i>Syllogismen 2010</i> , Score (ohne Item 5).....	37
Tabelle 8.7: Z-Test nach Fischer und Scheiblechner <i>Syllogismen 2010</i> , Bildung (ohne Item 8).....	39
Tabelle 8.8: LRT <i>Syllogismen 2010</i> , Bildungsniveau (ohne Item 8)	39
Tabelle 8.9: LRT <i>Syllogismen 2010</i> , Muttersprache (ohne Item 8)	40
Tabelle 8.10: LRT <i>Syllogismen 2010</i> , Versuchsbedingung (ohne Item 8).....	40
Tabelle 8.11: LRT <i>Syllogismen 2010</i> , Score (ohne Items 5, 8)	42
Tabelle 8.12: LRT <i>Syllogismen 2010</i> , Bildungsniveau (ohne Items 5, 8).....	42
Tabelle 8.13: LRT <i>Syllogismen 2010</i> , Muttersprache (ohne Items 5, 8).....	43
Tabelle 8.14: LRT <i>Syllogismen 2010</i> , Versuchsbedingung (ohne Items 5, 8)	44
Tabelle 8.15: Schätzwerte der Itemleichtigkeitsparameter für <i>Syllogismen 2010</i> (ohne Items 5,8) mit 95%-Konfidenzintervall	46
Tabelle 8.16: LRT <i>FRRT</i> , Score.....	47
Tabelle 8.17: LRT <i>FRRT</i> , Bildungsniveau	48
Tabelle 8.18: LRT <i>FRRT</i> , Muttersprache	48
Tabelle 8.19: LRT <i>FRRT</i> , Versuchsbedingung.....	49
Tabelle 8.20: Schätzung der Itemleichtigkeitsparameter für <i>FRRT</i> mit 95%-Konfidenzintervall	51
Tabelle 8.21: Relative Lösungshäufigkeit <i>Syllogismen 2010</i> (ohne Items 5, 8).....	51
Tabelle 8.22: Relative Lösungshäufigkeiten <i>FRRT</i>	52
Tabelle 8.23: Korrelation zwischen Anzahl gelöster Aufgaben in <i>Syllogismen 2010</i> (RM-konforme Items) und <i>FRRT</i>	54

13. Literaturverzeichnis

- Andersen, E. B. (1973). A goodness of fit test for the Rasch Model. *Psychometrika*, 38(1), 123-140.
- Ball, L. J., & Quayle, J. D. (2009). Phonological and visual distinctiveness effects in syllogistic reasoning: implications for mental models theory. *Memory & Cognition*, 37(6), 759-768.
- Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Cognition*, 6(1), 76-83.
- Formann, A. K., & Piswanger, K. (1979). *Wiener Matrizen-Test. WMT*. Weinheim: Beltz-Test.
- Freytag-Löringhoff, B. (1961). *Ihr System und ihr Verhältnis zur Logik*. Stuttgart: Kohlhammer.
- Glas, C. A. W., & Verhelst, N. D. (1995). Testing the Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models* (pp. 69-95). New York: Springer.
- Hansmann, B. C. (2010). *About the psychometric quality of various multiple choice response formats in the context of cultural differences between Austria and the United States of America*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Hauser, M. (2011). *Extremgruppenvalidierung von LAMBDA-neu*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Heinrich, I. G. (1973). *Veränderung des logischen Schlussprozesses bei semantischer Variation seiner Elemente. Ein Versuch mit Syllogismen*. Unveröffentlichte Dissertation, Universität Wien.
- Hoffmann, I. (in Vorbereitung). *Konstruktion des Verfahrens Syllogismen 2010 und Auswirkung der Verwendung von Hilfsmitteln auf die Leistung in verbalen Reasoning-Tests*.
- Hornke, L. F., Etzel, S., & Rettig, K. (1997). *Adaptiver Matrizentest (AMT)*. Mödling: Schuhfried.
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in cognitive sciences*, 5(10), 434-442.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 2010.
- Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 19-83). Weinheim: Beltz.

- Kubinger, K. D. (1996). Methoden der Psychologischen Diagnostik. In E. Erdfelder, R. Mausfeld, T. Meiser, & G. Rudinger (Hrsg.), *Handbuch Quantitative Methoden* (S. 567-576). Weinheim: Beltz.
- Kubinger, K. D. (2003). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz.
- Kubinger, K. D. (2005). Psychological Test Calibration Using the Rasch Model — Some Critical Suggestions on Traditional Approaches. *International Journal of Testing*, 5(4), 377-394.
- Kubinger, K. D. (2009). *Psychologische Diagnostik*. Göttingen: Hogrefe.
- Kubinger, K. D., & Gottschall, C. H. (2007). Item difficulty of multiple choice tests dependant on different item response formats – An experiment in fundamental research on psychological assessment. *Psychology Science*, 49(4), 361-374.
- Kubinger, K. D., Holocher-Ertl, S., Reif, M., Hohensinn, C., & Frebort, M. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. *International Journal of Selection and Assessment*, 18(1), 111-115.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (in Druck). *Statistik in der Psychologie*. Göttingen: Hogrefe.
- Kubinger, K. D., & Wolfsbauer, C. (2010). On the Risk of Certain Psychotechnological Response Options in Multiple-Choice Tests. *European Journal of Psychological Assessment*, 26(4), 302-308.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2001). *Intelligenz-Struktur-Test 2000 R. Intelligenz-Struktur-Test zweitausend*. Göttingen: Hogrefe.
- Menne, A. (1954). *Logik und Existenz*. Meisenheim/Glan: Hain.
- Morley, N. J., Evans, J. S., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 57(4), 666-92.
- Neubauer, G. (1992). *Logisches Denken bei Syllogismen*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Poinstingl, H. (2009). The Linear Logistic Test Model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychology Science Quarterly*, 51(2), 123 - 134.
- Poinstingl, H., Kubinger, K. D., Skoda, S., & Schechtner, C. (in Vorbereitung). *Family Relation Reasoning Test (FRRT)*.

- Poinstingl, H., Mair, P., & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling) Anwendung des Rasch-Modells (1-PL Modell)*. Lengerich: Pabst Science Publishers.
- Rasch, D., Kubinger, K. D., Schmidtke, J., & Häusler, J. (2004). The misuse of asterisks in hypothesis testing. *Psychology Science, 46*(2), 227-242.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Huber.
- Salmon, W. C. (1983). *Logik*. Stuttgart: Reclam.
- Skoda, S. (2005). *Der Verwandtschaften-Reasoning-Test. Konzept und erste empirische Erprobung*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Srp, G. (1993). *Syllogismen als Aufgaben für einen computerisierten adaptiven Test*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Srp, G. (1994). *Syllogismen*. Frankfurt am Main: Swets Test Service.
- Stupple, E. J. N., & Ball, L. J. (2007). Figural Effects in a Syllogistic Evaluation Paradigm. *Experimental Psychology, 54*(2), 120-127.
- Wolff, M. (2009). *Abhandlung über die Prinzipien der Logik*. Frankfurt am Main: Klostermann.
- Wolff, M. (2010). Vollkommene Syllogismen und reine Vernunftschlüsse: Aristoteles und Kant. *Journal for General Philosophy of Science, 40*(2), 341-355.

14. Anhang

A. 19 gültige Formen von Syllogismen

Nummer	Urteile	Figur	Form des Syllogismus
1	AAA	1	Alle M sind P. Alle S sind M. Alle S sind P.
2	AAI	3	Alle M sind P. Alle M sind S. Einige S sind P.
3	AII	3	Alle M sind P. Einige M sind S. Einige S sind P.
4	AII	1	Alle M sind P. Einige S sind M. Einige S sind P.
5	IAI	3	Einige M sind P. Alle M sind S. Einige S sind P.
6	IAI	4	Einige P sind M. Alle M sind S. Einige S sind P.
7	AEE	2	Alle P sind M. Alle S sind nicht M. Alle S sind nicht P.
8	AEE	4	Alle P sind M. Alle M sind nicht S. Alle S sind nicht P.
9	EAE	1	Alle M sind nicht P. Alle S sind M. Alle S sind nicht P.
10	EAE	2	Alle P sind nicht M. Alle S sind M. Alle S sind nicht P.
11	EAO	3	Alle M sind nicht P. Alle M sind S. Einige S sind nicht P.
12	EAO	4	Alle P sind nicht M. Alle M sind S. Einige S sind nicht P.
13	EIO	3	Alle M sind nicht P. Einige M sind S. Einige S sind nicht P.
14	EIO	4	Alle P sind nicht M. Einige M sind S. Einige S sind nicht P.

Nummer	Urteile	Figur	Form des Syllogismus
15	EIO	1	Alle M sind nicht P. Einige S sind M. Einige S sind nicht P.
16	EIO	2	Alle P sind nicht M. Einige S sind M. Einige S sind nicht P.
17	OA0	3	Einige M sind nicht P. Alle M sind S. Einige S sind nicht P.
18	AO0	2	Alle P sind M. Einige S sind nicht M. Einige S sind nicht P.
19	AAI	4	Alle P sind M. Alle M sind S. Einige S sind P.

B. 20 Items der *Syllogismen 2010* für die empirische Untersuchung

Items von Sarah Treiber (syl_1 bis syl_10)

syl_1

Alle Juristen sind Rechtsanwälte.
Alle Richter sind Juristen.

Kein Richter ist ein Rechtsanwalt.
Einige Richter sind Rechtsanwälte.
Einige Richter sind keine Rechtsanwälte.
Alle Richter sind Rechtsanwälte.
Alle Richter sind keine Rechtsanwälte.

syl_2

Jede Katze ist ein Säugetier.
Einige Igel sind Katzen.

Jeder Igel ist ein Säugetier.
Einige Igel sind Säugetiere.
Kein Igel ist ein Säugetier.
Einige Igel sind keine Säugetiere.
Alle Igel sind Säugetiere.

syl_3

Es gibt Karbunkel, die Furunkel sind.
Jedes Furunkel ist ein Abszess.

Jeder Abszess ist ein Karbunkel.
Es gibt Abszesse, die keine Karbunkel sind.
Kein Abszess ist ein Karbunkel.
Alle Abszesse sind Karbunkel.
Es gibt Abszesse, die Karbunkel sind.

syl_4

Für alle Klavierstücke gilt, dass sie Kunstwerke sind.
Für alle Klavierstücke gilt, dass sie Melodien sind.

Für alle Melodien gilt, dass sie keine Kunstwerke sind.
Für alle Melodien gilt, dass sie Kunstwerke sind.
Für keine Melodie gilt, dass sie ein Kunstwerk ist.
Für einige Melodien gilt, dass sie Kunstwerke sind.
Für einige Melodien gilt, dass sie keine Kunstwerke sind.

syl_5

Jedes Aliquod ist ein Substitut.
Mindestens ein Aliquod ist ein Aliud.

Mindestens ein Aliud ist ein Substitut.
Mindestens ein Aliud ist kein Substitut.
Kein Aliud ist ein Substitut.
Jedes Aliud ist ein Substitut.
Jedes Aliud ist kein Substitut.

syl_6

Wenn einige Rotaugen Salmoniden wären
und alle Rotaugen Weißfische wären, dann...
wären alle Weißfische Salmoniden.
wären alle Weißfische keine Salmoniden.
wären einige Weißfische Salmoniden.
wären einige Weißfische keine Salmoniden.
wären keine Weißfische Salmoniden.

syl_7

Jede Barke ist eine Barkasse.
Jede Barkasse ist keine Bilge.

Einige Bilgen sind keine Barken.
Jede Bilge ist eine Barke.
Einige Bilgen sind Barken.
Alle Bilgen sind Barken.
Keine Bilge ist eine Barke.

syl_8

Kein Quopp ist ein Wabb.
Jeder Quopp ist ein Fif.

Kein Fif ist ein Wabb.
Alle Fifs sind Wabbs.
Einige Fifs sind keine Wabbs.
Jeder Fif ist kein Wabb.
Einige Fifs sind Wabbs.

syl_9

Kein Wobbler ist ein Spinner.
Einige Twister sind Spinner.

Kein Twister ist ein Wobbler.
Jeder Twister ist ein Wobbler.
Einige Twister sind Wobbler.
Alle Twister sind keine Wobbler.
Einige Twister sind keine Wobbler.

syl_10

Wenn einige Sandkörner keine Kieselsteine wären
und alle Sandkörner Felsbrocken wären, dann...
wäre kein Felsbrocken ein Kieselstein.
wären alle Felsbrocken Kieselsteine.
wären einige Felsbrocken Kieselsteine.
wären einige Felsbrocken keine Kieselsteine.
wären alle Felsbrocken keine Kieselsteine.

Items von Ines Hoffmann (syl_11 bis syl_20)

syl_11

Keine Schildkröte hat Kiemen.
Alle im Meer lebenden Panzertiere sind Schildkröten.

Kein im Meer lebendes Panzertier hat keine Kiemen.
Alle im Meer lebenden Panzertiere haben Kiemen.
Manche im Meer lebenden Panzertiere haben Kiemen.
Alle im Meer lebenden Panzertiere haben keine Kiemen.
Kein im Meer lebendes Panzertier hat Kiemen.

syl_12

Alle Ärzte sind Doktoren.
Alle Radiologen sind Ärzte.

Manche Radiologen sind keine Doktoren.
Alle Radiologen sind Doktoren.
Keine Radiologen sind Doktoren.
Manche Radiologen sind Doktoren.
Alle Radiologen sind keine Doktoren.

syl_13

Kein Proton ist ein Lepton.
Ein paar Hadronen sind Protonen.

Kein Hadron ist ein Lepton.
Alle Hadronen sind keine Leptonen.
Ein paar Hadronen sind Leptonen.
Ein paar Hadronen sind keine Leptonen.
Alle Hadronen sind Leptonen.

syl_14

Alle Zahlen haben Vorzeichen.
Einige mathematische Objekte sind Zahlen.

Einige mathematische Objekte haben Vorzeichen.
Einige mathematische Objekte haben keine Vorzeichen.
Alle mathematischen Objekte haben Vorzeichen.
Alle mathematischen Objekte haben keine Vorzeichen.
Kein mathematisches Objekt hat Vorzeichen.

syl_15

Kein Chamäleon stammt aus Amerika.
Alle Stummelschwanzchamäleons sind Chamäleons.

Manche Stummelschwanzchamäleons stammen nicht aus Amerika.
Kein Stummelschwanzchamäleon stammt nicht aus Amerika.
Manche Stummelschwanzchamäleons stammen aus Amerika.
Alle Stummelschwanzchamäleons stammen aus Amerika.
Kein Stummelschwanzchamäleon stammt aus Amerika.

syl_16

Kein F ist ein K.
Jedes H ist ein F.

Kein H ist kein K.
Jedes H ist ein K.
Kein H ist ein K.
Ein paar H sind keine K.
Ein paar H sind K.

syl_17

Keine Erdbeere ist eine Frucht.
Manche Früchte sind Beeren.

Keine Beere ist keine Erdbeere.
Alle Beeren sind keine Erdbeeren.
Manche Beeren sind Erdbeeren.
Manche Beeren sind keine Erdbeeren.
Alle Beeren sind Erdbeeren.

syl_18

Kein Mandignan hat Flügel.
Ein paar Gadner sind Mandignans.

Kein Gadner hat keine Flügel.
Ein paar Gadner haben Flügel.
Ein paar Gadner haben keine Flügel.
Alle Gadner haben Flügel.
Alle Gadner haben keine Flügel.

syl_19

Jeder Nasder ist glänzend.
Einige Polgen sind nicht glänzend.

Jeder Polge ist kein Nasder.
Einige Polgen sind Nasder.
Jeder Polge ist ein Nasder.
Kein Polgen ist ein Nasder.
Einige Polgen sind keine Nasder.

syl_20

Kein einziges Monster mag Läuse.
Manche Spinnen sind Monster.

Keine Spinne mag Läuse.
Manche Spinnen mögen Läuse.
Manche Spinnen mögen keine Läuse.
Alle Spinnen mögen Läuse.
Alle Spinnen mögen keine Läuse.

C. Standard-Quantoren und Varianten

Universell bejahend (A)	Universell verneinend (E)	Partikulär bejahend (I)	Partikulär verneinend (E)
Alle A sind B.	Kein A ist B.	Einige A sind B.	Einige A sind nicht B.
Jedes A ist ein B.	Alle A sind nicht B.		
Für alle A gilt, dass sie B sind.	Für alle A gilt, dass sie keine B sind.	Mindestens ein A ist B.	Mindestens ein A ist nicht/kein B.
Wenn etwas ein A ist, dann ist es auch ein B.	Wenn etwas ein A ist, dann ist es kein B.	Es gibt mindestens ein A für das gilt, dass es B ist.	Es gibt mindestens ein A für das gilt, dass es kein B ist.
	Es gibt kein A für das gilt, dass es ein B ist.	Manche A sind B.	Manche A sind keine B.

D. 55 Items der *Syllogismen 2010* (durch die Autorin erstellt)

1.

Alle Juristen sind Rechtsanwälte.

Alle Richter sind Juristen.

Kein Richter ist ein Rechtsanwalt.

Einige Richter sind Rechtsanwälte.

Alle Richter sind Rechtsanwälte.

Einige Richter sind keine Rechtsanwälte.

Alle Richter sind keine Rechtsanwälte.

2.

Jede Libelle ist ein Insekt.

Jede Ameise ist eine Libelle.

Einige Ameisen sind keine Insekten.

Jede Ameise ist ein Insekt.

Keine Ameise ist ein Insekt.

Einige Ameisen sind Insekten.

Alle Ameisen sind keine Insekten.

3.

Wenn alle Kitesurfer Wellenreiter wären,
und alle Windsurfer Kitesurfer wären, dann wären...

alle Windsurfer Wellenreiter.

einige Windsurfer keine Wellenreiter.

alle Windsurfer keine Wellenreiter.

kein Windsurfer ist ein Wellenreiter.

einige Windsurfer Wellenreiter.

4.

Für jedes Nachtschattengewächs gilt, dass es eine Pflanze ist.

Für alle Tomaten gilt, dass sie Nachtschattengewächse sind.

Für einige Tomaten gilt, dass sie keine Pflanzen sind.

Für jede Tomate gilt, dass sie keine Pflanze ist.

Für keine Tomate gilt, dass sie keine Pflanzen ist.

Für jede Tomate gilt, dass sie eine Pflanze ist.

Für einige Tomaten gilt, dass sie Pflanzen sind.

5.

Alle Meisen sind Spechte.

Alle Meisen sind Vögel.

Alle Vögel sind Spechte.

Kein Vogel ist ein Specht.

Einige Vögel sind keine Spechte.

Alle Vögel sind keine Spechte.

Einige Vögel sind Spechte.

6.

Alle Menschen sind Vegetarier.
Alle Menschen sind Lebewesen.

Einige Lebewesen sind Vegetarier.

Kein Lebewesen ist ein Vegetarier.
Alle Lebewesen sind Vegetarier.
Einige Lebewesen sind keine Vegetarier.
Alle Lebewesen sind keine Vegetarier.

7.

Für alle Klavierstücke gilt, dass sie Kunstwerke sind.
Für alle Klavierstücke gilt, dass sie Melodien sind.

Für alle Melodien gilt, dass sie keine Kunstwerke sind.
Für einige Melodien gilt, dass sie Kunstwerke sind.
Für alle Melodien gilt, dass sie Kunstwerke sind.
Für keine Melodie gilt, dass sie ein Kunstwerk ist.
Für einige Melodien gilt, dass sie keine Kunstwerke sind.

8.

Alle Maschinen sind Geräte.
Es gibt mindestens eine Maschine, die eine Pflanze ist.

Keine Pflanze ist ein Gerät.
Es gibt mindestens eine Pflanze, die kein Gerät ist.
Es gibt mindestens eine Pflanze, die ein Gerät ist.
Jede Pflanze ist kein Gerät.
Jede Pflanze ist ein Gerät.

9.

Wenn alle Brillen konkav wären,
und einige Brillen rund wären, dann wären...

alle runden Dinge konkav.
kein rundes Ding konkav.
einige runde Dinge konkav.
einige runde Dinge nicht konkav.
alle runden Dinge nicht konkav.

10.

Alle Bakterien sind Krankheitserreger.
Einige Bakterien sind Viren.

Kein Virus ist ein Krankheitserreger.
Alle Viren sind Krankheitserreger.
Einige Viren sind keine Krankheitserreger.
Alle Viren sind keine Krankheitserreger.
Einige Viren sind Krankheitserreger.

11.

Jedes Aliquod ist ein Substitut.
Mindestens ein Aliquod ist ein Aliud.

Mindestens ein Aliud ist ein Substitut.

Mindestens ein Aliud ist kein Substitut.
Kein Aliud ist ein Substitut.
Jedes Aliud ist ein Substitut.
Jedes Aliud ist kein Substitut.

12.

Jede Katze ist ein Säugetier.
Einige Igel sind Katzen.

Jeder Igel ist ein Säugetier.

Einige Igel sind Säugetiere.

Kein Igel ist ein Säugetier.
Einige Igel sind keine Säugetiere.
Alle Igel sind Säugetiere.

13.

Für alle Würfel gilt, dass sie Quader sind.
Es gibt Pyramiden, für die gilt, dass sie Würfel sind.

Für alle Pyramiden gilt, dass sie Quader sind.

Für alle Pyramiden gilt, dass sie keine Quader sind.

Es gibt Pyramiden, für die gilt, dass sie Quader sind.

Es gibt Pyramiden, für die gilt, dass sie keine Quader sind.
Es gibt keine Pyramiden, für die gilt, dass sie Quader sind.

14.

Wenn alle Monozyten Granulozyten wären,
und einige Lymphozyten Monozyten wären, dann wären...

alle Lymphozyten Granulozyten.

einige Lymphozyten keine Granulozyten.

alle Lymphozyten keine Granulozyten.

einige Lymphozyten Granulozyten.

keine Lymphozyten Granulozyten.

15.

Wenn jedes Meeting eine Konferenz wäre,
und einige Geschäftsessen Meetings wären, dann...

wären alle Geschäftsessen keine Konferenzen.

wären einige Geschäftsessen keine Konferenzen.

wären alle Geschäftsessen Konferenzen.

wäre kein Geschäftsessen eine Konferenz.

wären einige Geschäftsessen Konferenzen.

16.

Einige Monolithen sind Reliquien.
Alle Monolithen sind antike Gegenstände.

Einige antike Gegenstände sind Reliquien.

Alle antiken Gegenstände keine Reliquien.
Einige antike Gegenstände sind keine Reliquien.
Alle antiken Gegenstände sind Reliquien.
Kein antiker Gegenstand ist eine Reliquie.

17.

Für einige Ringelnattern gilt, dass sie Kreuzottern sind.
Jede Ringelnatter ist eine Sandviper.

Jede Sandviper ist eine Kreuzotter.
Keine Sandviper ist eine Kreuzotter.
Einige Sandvipern sind keine Kreuzottern.

Einige Sandvipern sind Kreuzottern.

Alle Sandvipern sind Kreuzottern.

18.

Wenn einige Rotaugen Salmoniden wären,
und alle Rotaugen Weißfische wären, dann...

wären alle Weißfische Salmoniden.
wären alle Weißfische keine Salmoniden.
wären einige Weißfische Salmoniden.
wären einige Weißfische keine Salmoniden.
wären keine Weißfische Salmoniden.

19.

Einige Magnolien sind Chrysanthemen.
Alle Chrysanthemen sind Lupinen.

Einige Lupinen sind keine Magnolien.

Einige Lupinen sind Magnolien.

Alle Lupinen sind Magnolien.
Alle Lupinen sind keine Magnolien.
Jede Lupine ist eine Magnolie.

20.

Es gibt Karbunkel, die Furunkel sind.
Jedes Furunkel ist ein Abszess.

Jeder Abszess ist ein Karbunkel.
Es gibt Abszesse, die keine Karbunkel sind.
Kein Abszess ist ein Karbunkel.
Alle Abszesse sind Karbunkel.
Es gibt Abszesse, die Karbunkel sind.

21.

Wenn einige Filarien Protozoen wären,
und alle Protozoen Erythrozyten wären, dann...

wären einige Erythrozyten Filarien.

wären alle Erythrozyten keine Filarien.

wären einige Erythrozyten keine Filarien.

wären alle Erythrozyten Filarien.

wären keine Erythrozyten Filarien.

22.

Wenn jedes Gsels eine Ebira wäre
und kein Guatsle eine Ebira wäre, dann wäre...

jedes Guatsle ein Gsels.

kein Guatsle ein Gsels.

mindestens ein Guatsle ein Gsels.

manche Guatsle kein Gsels.

manche Guatsle ein Gsels.

23.

Alle Eosinophile sind Basophile.

Alle Neutrophile sind keine Basophile.

Einige Neutrophile sind Eosinophile.

Alle Neutrophile sind keine Eosinophile.

Alle Neutrophile sind Eosinophile.

Einige Neutrophile sind keine Eosinophile.

Jeder Neutrophile ist ein Eosinophiler.

24.

Für alle Kreise gilt, dass sie Rechtecke sind.

Für alle Ovale gilt, dass sie keine Rechtecke sind.

Für einige Ovale gilt, dass sie Kreise sind.

Für alle Ovale gilt, dass sie Kreise sind.

Für einige Ovale gilt, dass sie keine Kreise sind.

Für alle Ovale gilt, dass sie keine Kreise sind.

Für mindestens ein Oval gilt, dass es ein Kreis ist.

25.

Alle Kirschen sind Trauben.

Keine Traube ist eine Beere.

Einige Beeren sind keine Kirschen.

Alle Beeren sind Kirschen.

Einige Beeren sind Kirschen.

Jede Beere ist eine Kirsche.

Keine Beere ist eine Kirsche.

26.

Jede Barke ist eine Barkasse.
Jede Barkasse ist keine Bilge.

Einige Bilgen sind keine Barken.
Jede Bilge ist eine Barke.

Keine Bilge ist eine Barke.

Einige Bilgen sind Barken.
Alle Bilgen sind Barken.

27.

Jede Hütte ist ein Gebäude.
Kein Gebäude ist eine Hecke.

Keine Hecke ist eine Hütte.

Alle Hecken sind Hütten.
Einige Hecken sind keine Hütten.
Einige Hecken sind Hütten.
Jede Hecke ist eine Hütte.

28.

Alle Brücken sind keine Tiefbauten.
Alle Hochbauten sind Brücken.

Alle Hochbauten sind Tiefbauten.

Alle Hochbauten sind keine Tiefbauten.

Einige Hochbauten sind Tiefbauten.
Einige Hochbauten sind keine Tiefbauten.
Manche Hochbauten sind Tiefbauten.

29.

Wenn kein Arbeitnehmer ein Freiberufler wäre,
und alle Arbeitgeber Arbeitnehmer wären, dann wären....

einige Arbeitgeber Freiberufler.
einige Arbeitgeber keine Freiberufler.
alle Arbeitgeber Freiberufler.
alle Arbeitgeber keine Freiberufler.
manche Arbeitgeber Freiberufler.

30.

Für jede Pflanze gilt, dass sie kein Gegenstand ist.
Für jeden Baum gilt, dass er eine Pflanze ist.

Für einige Bäume gilt, dass sie keine Gegenstände sind.
Für jeden Baum gilt, dass er ein Gegenstand ist.
Für einige Bäume gilt, dass sie Gegenstände sind.
Für alle Bäume gilt, dass sie Gegenstände sind.
Für jeden Baum gilt, dass er kein Gegenstand ist.

31.

Wenn etwas eine Guave ist, dann ist es kein Gemüse.

Wenn etwas eine Gurke ist, dann ist es ein Gemüse.

Keine Gurke ist eine Guave.

Einige Gurken sind keine Guaven.

Jede Gurke ist eine Guave.

Einige Gurken sind Guaven.

Alle Gurken sind Guaven.

32.

Wenn alle Menschen keine Lebewesen wären,
und alle Tiere Lebewesen wären, dann wären....

einige Tiere Menschen.

einige Tiere keine Menschen.

alle Tiere Menschen.

alle Tiere keine Menschen.

manche Tiere Menschen.

33.

Alle Söhne sind keine alten Menschen.

Alle Großväter sind alte Menschen.

Einige Großväter sind Söhne.

Alle Großväter sind Söhne.

Einige Großväter sind keine Söhne.

Jeder Großvater ist ein Sohn.

Alle Großväter sind keine Söhne.

34.

Alle Planeten sind keine Scheiben.

Alle Planeten sind Sterne.

Kein Stern ist eine Scheibe.

Jeder Stern ist eine Scheibe.

Einige Sterne sind keine Scheiben.

Einige Sterne sind Scheiben.

Alle Sterne sind keine Scheiben.

35.

Kein Quopp ist ein Wabb.

Jeder Quopp ist ein Fif.

Kein Fif ist ein Wabb.

Einige Fifs sind keine Wabbs.

Alle Fifs sind Wabbs.

Einige Fifs sind Wabbs.

Jeder Fif ist kein Wabb.

36.

Wenn für jeden Tisch gelten würde, dass er kein Stuhl ist,
und für jeden Tisch, dass er ein Schemel ist, dann...

wäre kein Schemel ein Stuhl.

wären alle Schemel Stühle.

wären einige Schemel Stühle.

wären alle Schemel keine Stühle.

wären einige Schemel keine Stühle.

37.

Kein Kartenspiel ist ein Glücksspiel.

Jedes Glücksspiel ist ein Würfelspiel.

Kein Würfelspiel ist ein Kartenspiel.

Einige Würfelspiele sind Kartenspiele.

Alle Würfelspiele sind Kartenspiele.

Einige Würfelspiele sind keine Kartenspiele.

Alle Würfelspiele sind keine Kartenspiele.

38.

Wenn keine Dalbe ein Dolle wäre,
und alle Dollen Duchten wären, dann...

wären alle Duchten Dalben.

wäre keine Ducht eine Dalbe.

wären einige Duchten keine Dalben.

wären einige Duchten Dalben.

wären alle Duchten keine Dalben.

39.

Wenn etwas ein Laa ist, dann ist es kein Gebi.

Einige Laa sind Folp.

Einige Folps sind keine Gebi.

Alle Folps sind Gebi.

Einige Folps sind Gebi.

Kein Folp ist ein Gebi.

Jedes Folp ist ein Gebi.

40.

Kein Faustpfand ist ein Sicherungsmittel.

Einige Faustpfänder sind Hypotheken.

Keine Hypothek ist ein Sicherungsmittel.

Einige Hypotheken sind keine Sicherungsmittel.

Alle Hypotheken sind Sicherungsmittel.

Einige Hypotheken sind Sicherungsmittel.

Alle Hypotheken sind keine Sicherungsmittel.

41.

Wenn kein Champignon eine Wildpflanze wären,
und einige Wildpflanzen Pilze wären, dann....

wären alle Pilze Champignons.

wären alle Pilze keine Champignons.

wären einige Pilze Champignons.

wären einige Pilze keine Champignons.

wäre kein Pilz ein Champignon.

42.

Keine Galosche ist eine Kartusche.

Einige Kartuschen sind Gefäße.

Kein Gefäß ist eine Galosche.

Alle Gefäße sind Galoschen.

Einige Gefäße sind keine Galoschen.

Einige Gefäße sind Galoschen.

Alle Gefäße sind keine Galoschen.

43.

Für alle Rosen gilt ausnahmslos, dass sie keine Kreuzblütler sind.

Für einige Margeriten gilt, dass sie Rosen sind.

Für alle Margeriten gilt, dass sie keine Kreuzblütler sind.

Für einige Margeriten gilt, dass sie keine Kreuzblütler sind.

Für einige Margeriten gilt, dass sie Kreuzblütler sind.

Für alle Margeriten gilt, dass sie Kreuzblütler sind.

Für keine Margerite gilt, dass sie ein Kreuzblütler ist.

44.

Es gibt keine Zeitschriften, die Bücher sind.

Einige Monographien sind Zeitschriften.

Einige Monographien sind keine Bücher.

Keine Monographie ist ein Buch.

Jede Monographie ist ein Buch.

Einige Monographien sind Bücher.

Alle Monographien sind Bücher.

45.

Kein Wobbler ist ein Spinner.

Einige Twister sind Spinner.

Kein Twister ist ein Wobbler.

Jeder Twister ist ein Wobbler.

Einige Twister sind Wobbler.

Alle Twister sind keine Wobbler.

Einige Twister sind keine Wobbler.

46.

Wenn kein Blumenstrauß ein Geschenk wäre,
und einige Pralinenschachteln Geschenke wären, dann...

wäre keine Pralinenschachtel ein Blumenstrauß.

wären einige Pralinenschachteln keine Blumensträuße.

wären alle Pralinenschachteln Blumensträuße.

wären einige Pralinenschachteln Blumensträuße.

wären alle Pralinenschachteln keine Blumensträuße.

47.

Einige Matrosen sind keine Kapitäne.

Alle Matrosen sind Seemänner.

Jeder Seemann ist Kapitän.

Kein Seemann ist Kapitän.

Einige Seemänner sind Kapitäne.

Einige Seemänner sind keine Kapitäne.

Alle Seemänner sind keine Kapitäne.

48.

Es gibt Mäuse, für die gilt, dass sie keine Katzen sind.

Wenn etwas eine Maus ist, dann ist es ein Hund.

Es gibt Hunde, die keine Katzen sind.

Wenn etwas ein Hund ist, dann ist es keine Katze.

Es gibt Hunde, die Katzen sind.

Wenn etwas ein Hund ist, dann ist es eine Katze.

Wenn etwas kein Hund ist, dann ist es eine Katze.

49.

Wenn einige Sandkörner keine Kieselsteine wären,

und alle Sandkörner Felsbrocken wären, dann...

wären einige Felsbrocken keine Kieselsteine.

wäre kein Felsbrocken ein Kieselstein.

wären alle Felsbrocken Kieselsteine.

wären einige Felsbrocken Kieselsteine.

wären alle Felsbrocken keine Kieselsteine.

50.

Alle Kegel sind Quader.

Einige Kugeln sind keine Quader.

Alle Kugeln sind Kegel.

Einige Kugeln sind keine Kegel.

Alle Kugeln sind keine Kegel.

Einige Kugeln sind Kegel.

Keine Kugel ist ein Kegel.

51.

Jedes Flipp ist ein Flarr.
Es gibt Hupfs die keine Flarrs sind.

Jedes Hupf ist ein Flipp.
Es gibt Hupfs, die keine Flipps sind.
Kein Hupf ist ein Flipp.
Es gibt Hupfs, die Flipps sind.
Alle Hupfs sind keine Flipps.

52.

Wenn alle Münzen Goldstücke wären,
und einige Geldscheine keine Goldstücke wären, dann...

wären alle Geldscheine Münzen.
wären alle Geldscheine keine Münzen.
wären einige Geldscheine Münzen.
wären einige Geldscheine keine Münzen.
wäre kein Geldschein eine Münze.

53.

Alle Piraten sind Freibeuter.
Alle Freibeuter sind Seeräuber.

Alle Seeräuber sind Piraten.
Kein Seeräuber ist ein Pirat.
Einige Seeräuber sind Piraten.
Einige Seeräuber sind keine Piraten.
Alle Seeräuber sind keine Piraten.

54.

Wenn etwas eine Amsel ist, dann ist es eine Drossel.
Wenn etwas eine Drossel ist, dann ist es ein Star.

Alle Stare sind Amseln.
Kein Star ist eine Amsel.
Einige Stare sind keine Amseln.
Alle Stare sind keine Amseln.
Einige Stare sind Amseln.

55.

Wenn jedes Bild ein Kunstwerk wäre,
und alle Kunstwerke Sammlerstücke wären, dann...

wären einige Sammlerstücke Bilder.
wären alle Sammlerstücke keine Bilder.
wären einige Sammlerstücke keine Bilder.
wären alle Sammlerstücke Bilder
wäre kein Sammlerstück ein Bild.

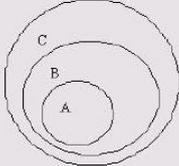
E. Screenshots zum Test *Syllogismen 2010* - Instruktion, Hilfetext und Beispielaufgabe 1

Syllogismen 2010
Testleiter

Syllogismen 2010

Syllogismen 2010 ist ein Computertest, in dem bei jeder Aufgabe zwei Aussagen (Prämissen) angeboten werden. Mit Hilfe dieser beiden Aussagen sollen Sie eine zwingende logische Schlussfolgerung ziehen.

Beispiel
Die Prämissen: **Alle A sind B. Alle B sind C.**
führen zu der Lösung: **Alle A sind C.**
Da alle A B sind und alle B C sind, müssen alle A auch C sein.



[zu den Beispielen](#)

Syllogismen 2010
Testleiter

Es ist möglich, dass Sie zu Lösungen kommen, die Ihnen absurd oder inhaltlich falsch vorkommen. Dies beruht darauf, dass man auch aus absurden Aussagen logisch richtige Schlüsse ableiten kann. Ob ein logischer Schluss richtig ist, hängt nicht von der „Wahrheit“ der Aussage in der realen Welt ab.

Die Testung

In jeder Aufgabe werden zwei Aussagen dargeboten und jeweils eine mögliche Schlussfolgerung. Sie müssen entscheiden, ob diese Schlussfolgerung auf Grund der beiden vorgegebenen Aussagen logisch richtig ist. Wenn Sie der Meinung sind, dass die gezeigte Antwortmöglichkeit richtig ist, dann wählen Sie die WAHR-Option und klicken auf den [WEITER]-Button. Der Computer geht dann zur nächsten Aufgabe über.

Wenn Sie der Meinung sind, dass die angebotene Aussage falsch ist, dann wählen Sie die FALSCH-Option und drücken anschließend den [WEITER]-Button. Der Computer bietet Ihnen dann weitere Antwortmöglichkeiten an. Haben Sie alle angebotenen Antwortmöglichkeiten zu einer Aufgabe als falsch abgelehnt, so geht der Test automatisch zur nächsten Aufgabe über.

[zu den Beispielen](#)

Syllogismen 2010
Testleiter

vorgegebenen Aussagen logisch richtig ist. Wenn Sie der Meinung sind, dass die gezeigte Antwortmöglichkeit richtig ist, dann wählen Sie die WAHR-Option und klicken auf den [WEITER]-Button. Der Computer geht dann zur nächsten Aufgabe über.

Wenn Sie der Meinung sind, dass die angebotene Aussage falsch ist, dann wählen Sie die FALSCH-Option und drücken anschließend den [WEITER]-Button. Der Computer bietet Ihnen dann weitere Antwortmöglichkeiten an. Haben Sie alle angebotenen Antwortmöglichkeiten zu einer Aufgabe als falsch abgelehnt, so geht der Test automatisch zur nächsten Aufgabe über.

Wichtige Informationen
Es gibt keine Zeitbegrenzung für die Bearbeitung der Aufgaben.
Nach einem Klick auf den [WEITER]-Button kann eine Antwort nicht mehr geändert werden.
Während der Testung können Sie die Hilfe aufrufen, indem Sie auf das Fragezeichen klicken.

Viel Erfolg!

[zu den Beispielen](#)

Hilfe

1. Wenn „Alle A sind B“, so gilt nicht immer „Alle B sind A“ (z.B.: Alle Dichter sind Menschen, aber nicht alle Menschen sind Dichter.)
2. Wenn „Einige A sind B“, dann gilt auch „Einige B sind A“ (z.B.: Einige Architekten sind Techniker, daher müssen auch einige Techniker Architekten sein.)
3. Wenn „Alle A sind keine B“ gilt, dann gilt auch „Alle B sind keine A“ (z.B.: Alle Fische sind keine Reptilien, daher sind auch alle Reptilien keine Fische.)
4. Wenn „Einige A sind keine B“ gilt, gilt nicht automatisch „Einige B sind keine A.“ (z.B.: Einige Hunde sind keine Pudel, dagegen können Pudel auch Hunde sein.)
5. Wenn gilt „Einige A sind B“, dann gilt nicht immer „Einige A sind keine B.“ („Einige“ bedeutet in der Logik „mindestens ein“, es können also theoretisch auch alle A B sein.)

[zurück](#)

Beispiel 1 von 2

Die Prämissen lauten

Jeder Baum ist eine Pflanze

Jede Birke ist ein Baum

Schlussfolgerung

(1) Einige Birken sind keine Pflanzen

- richtig
- falsch

weiter



Curriculum Vitae

Adresse Neustiftgasse 53/15, 1070 Wien
E-Mail treibersarah@gmail.com
Geburtsdaten 11. Juni 1981, Engen (Deutschland)
Nationalität deutsch
Akademischer Grad Diplomjuristin

Schule

06/2000 Salier Gymnasium Waiblingen, Deutschland
Abitur, Note: 1.4

Studium

10/2000 - 09/2003 Universität Konstanz, Deutschland
Rechtswissenschaft

10/2003 - 02/2005 Eberhard-Karls-Universität, Tübingen, Deutschland
Rechtswissenschaft
1. Staatsexamen 02/2005, 11.27 Punkte
Vertiefungsfächer: Kriminologie, Jugendstrafrecht, Strafvollzug

seit 03/2007 Universität Wien
Diplomstudium Psychologie
1. Diplomprüfung 08/2008 mit Auszeichnung bestanden
Vertiefungsfächer: Spezielle Psychologische Diagnostik, Angewandte Sozialpsychologie

02/2010 - 06/2010 Instituto Superior das Ciencias do Trabalho e da Empresa (ISCTE), Lisboa, Portugal
ERASMUS-Aufenthalt im Rahmen des Diplomstudiums Psychologie

Berufserfahrung

02/2002 - 06/2002 **Universität Konstanz**, Deutschland
Studentische Hilfskraft am Lehrstuhl für Öffentliches Baurecht

04/2005 - 04/2007 **Juristisches Referendariat**, Landgericht Mannheim, Deutschland
2. Staatsexamen 04/2007, 9.02 Punkte

04/2007 - 08/2008 **Rechtsanwalt Mag. Haas**, Wien
Wissenschaftliche Mitarbeiterin

02/2009 - 04/2009 **Universität Wien**
Freie Dienstnehmerin Zulassungs- und Studienbeitragserslassstelle

11/2009 - 02/2010 **Universität Wien**
Praktikum Arbeitsbereich Allgemeine Psychologie

seit 10/2010 **Medizinische Universität Wien**
Projektmitarbeit Institut für Medizinische Psychologie

seit 11/2010 **Universität Wien**
Projektmitarbeit Fakultät für Psychologie, Arbeitsbereich Psychologische Diagnostik