



universität
wien

Diplomarbeit

Titel der Arbeit

Erfassung wesentlicher kognitiver Operationen bei der
Lösung von Aufgaben zum logisch-schlussfolgernden
Denken im Kindesalter unter Berücksichtigung der
Antwortformate Multiple-Choice und freies Antwortformat

Verfasser

Johann Ecker

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juli 2011

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Mag. Dr. Stefana Holocher-Ertl

Abstract

Die vorliegende Arbeit widmet sich der Erforschung von Unterschieden in der Bearbeitung von Aufgaben zum logischen Schlussfolgern in den Antwortformaten Multiple-Choice vs. freie Antwort im Altersbereich von 6- bis 9-jährigen Kindern anhand einer Testung von insgesamt 80 bis 164 Testpersonen aus zwei Wiener Schulen. Aufgrund von bisherigen Forschungsergebnissen wird hypothetisiert, dass Kinder bei der Bearbeitung von Reasoning-Aufgaben im MC-Design schlechtere Leistungen erzielen als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat, wobei zusätzlich der Einfluss des Alters sowie der Arbeitsstile „Impulsiv vs. Reflexiv“ auf die Aufgabenbearbeitung überprüft wurde. Zur Erfassung des logischen Schlussfolgerns wurden im Rahmen der vorliegenden Arbeit drei verschiedene Verfahren eingesetzt, welche von den Testpersonen sowohl im freien Antwortformat als auch im MC-Format bearbeitet wurden. Die Hypothesenprüfung erfolgte in mehreren Schritten, wobei T-Tests für gepaarte Stichproben sowie Rasch-Modell-Analysen durchgeführt wurden. Zusammenfassend lässt sich sagen, dass zwar vereinzelte Trends dahingehend gefunden wurden, dass Kinder aufgrund von unausgereiften Kontrollstrategien und eines Arbeitsgedächtnisses, welches sich noch in der Entwicklung befindet, geringere Testleistungen bei der Bearbeitung von Aufgaben zum logischen Schlussfolgern im MC-Format erzielen. Dieser Trend scheint jedoch nicht auf alle Verfahren zu generalisieren sein.

The following study is devoted to the investigation of differences in the processing of reasoning tasks considering multiple-choice formats vs. free item response of 6 - to 9-year-old children based on a total sample of 80-164 subjects from two Viennese schools. Previous research indicates the hypothesis that, given the same tasks, children perform worse solving reasoning tasks in the MC design than by solving tasks with a free item response format. In addition, the influence of age and the working styles „Impulsive vs. Reflexive“ on the task processing was checked. To acquire the reasoning ability in the present work, three different types of tests were carried out by the subjects in both, the free item response format as well

as in the MC-format. A tiered approach was used for hypothesis testing, where paired-sample T-Tests and Rasch Model analysis were performed. In summary, isolated trends were found in such a way that children achieve, due to immature control strategies and a working memory which is still developing, lower test scores in the processing of reasoning tasks in the MC-format. However, it does not appear that this trend can be generalised for all tests.

Inhaltsverzeichnis

1. Einleitung	7
2. Theoretischer Teil	8
2.1. Logisch-schlussfolgerndes Denken	9
2.2. Die Bedeutung des Arbeitsgedächtnisses	12
2.3. Der Einfluss von Kontrollstrategien	15
2.4. Antwortformate.....	18
2.4.1. Das freie Antwortformat	19
2.4.1.1. Kontrollstrategien und das freie Antwortformat	21
2.4.2. Das Multiple-Choice-Antwortformat	22
2.4.2.1. Kontrollstrategien und das Multiple-Choice-Antwortformat	23
2.5. Die Relevanz der Erforschung genannter Konstrukte im Kindesalter	24
2.6. Arbeitsstil	27
2.6.1. Der Arbeitsstil und das Multiple-Choice-Antwortformat	29
2.6.2. Der Arbeitsstil und das freie Antwortformat	30
3. Empirischer Teil.....	31
3.1. Fragestellung	32
3.1.1. Hypothesen	33
3.1.1.1. H 1.....	33
3.1.1.2. H2.....	35
3.1.1.3. H3.....	38
3.2. Verwendete Verfahren	41
3.2.1. Das Prüfsystem für Schul- und Bildungsberatung (PSB)	42
3.2.1.1. Das Prüfsystem für Schul- und Bildungsberatung im freien Antwortformat	44
3.2.1.2. Das Prüfsystem für Schul- und Bildungsberatung im Multiple-Choice- Antwortformat	47
3.2.2. Matrizentest	49

3.2.2.1.	Der Matrizentest im freien Antwortformat.....	51
3.2.2.2.	Der Matrizentest im Multiple-Choice-Antwortformat.....	54
3.2.1.	Der Untertest „Formale Folgerichtigkeit“ aus dem AID 3.....	58
3.2.1.1.	Der Untertest „Formale Folgerichtigkeit“ aus dem AID 3 im freien Antwortformat.....	61
3.2.1.2.	Der Untertest „Formale Folgerichtigkeit“ aus dem AID 3 im Multiple- Choice-Antwortformat.....	63
3.2.2.	Die Erfassung des Arbeitsstils.....	66
3.3.	Planung und Durchführung	68
3.4.	Stichprobe.....	70
3.5.	Ergebnisse.....	75
3.5.1.	Generelle Unterschiede in der Bearbeitung des Multiple-Choice-Designs und des freien Antwortformats.....	78
3.5.1.1.	T-Test für gepaarte Stichproben.....	79
3.5.1.2.	Rasch-Modell-Analysen	84
3.5.2.	Unterschiede in der Bearbeitung des Multiple-Choice-Designs und des freien Antwortformats unter Einbezug des Alters	92
3.5.2.1.	Rasch-Modell-Analysen	93
3.5.3.	Unterschiede in der Bearbeitung des Multiple- Choice-Designs und des freien Antwortformats unter Einbezug des Arbeitsstils.....	100
3.5.3.1.	T-Tests für gepaarte Stichproben	100
3.5.3.2.	Rasch-Modell-Analysen	110
4.	Diskussion	112
5.	Zusammenfassung	119
6.	Literaturverzeichnis	122
7.	Anhang	125
	Abbildungsverzeichnis.....	125
	Tabellenverzeichnis	127

1. Einleitung

Zum gegenwärtigen Zeitpunkt ist es Standard in der psychologischen Diagnostik, bei einem Großteil der Fragestellungen neben spezifischen Merkmalen auch die generelle kognitive Leistungsfähigkeit des Klienten zu erheben. Dabei wird fast immer auch ein Verfahren eingesetzt, das die Fähigkeit zum logisch-schlussfolgernden Denken misst; das ist die Art und Weise, wie Menschen von etwas Gegebenem auf etwas Neues kommen oder schließen. Dieser Prozess ist mit einer Reihe von kognitiven Abläufen und impliziten Regeln verbunden. Wichtig dabei ist, dass diese Prozesse in den meisten Fällen nicht in einem geordneten Denkablauf, sondern vielmehr unbewusst stattfinden.

Wenig untersucht ist in diesem Zusammenhang die Frage, welche kognitiven Operationen speziell ausschlaggebend sind, um in einem solchen Verfahren zum logisch-schlussfolgernden Denken möglichst viele Aufgaben zu lösen. Während die Bedeutung des Arbeitsgedächtnisses bereits gut nachgewiesen ist (Baddeley & Hitch, 1974; Baddeley, 1986; Kyllonen & Christal, 1990; Süß et al., 2004; Kane, 2004; Unsworth & Engle, 2004), deuten jüngere Ergebnisse darauf hin, dass hier auch Kontrollprozesse und der Arbeitsstil eine wichtige Rolle spielen (Baddeley, 1994, 1996; Embretson, 1995, 1998; Unsworth & Engle, 2004; Kane et al., 2004).

Viele aktuelle Verfahren, welche die Fähigkeit zum logisch-schlussfolgernden Denken erfassen, sind mit einem Multiple-Choice-Antwortformat konzipiert, wobei eine Vielzahl an unterschiedlichen Konzeptionen zu beobachten ist. Unter diesen ist das Design „1 aus 6“ (sechs Antwortmöglichkeiten, von denen eine die richtige ist) am häufigsten anzutreffen. Dem gegenüber sind Tests, welche dieselbe Fähigkeit mit freiem Antwortformat erfassen, deutlich in der Unterzahl. Dabei zeigt sich, dass auffallend selten kritisch reflektiert wurde, inwieweit sich das gewählte Antwortdesign auf die Erhebung der interessierenden Fähigkeit auswirkt. Dies scheint gerade im Bereich der Diagnostik im Kindesalter von besonderer Bedeutung zu sein, da vor allem Kinder jüngeren Alters oftmals noch wenig ausgereifte Kontrollstrategien aufweisen, die sich für die Lösung von MC-Aufgaben als wichtig herausgestellt haben (Baddeley, 1994, 1996; Ziegler, 1994; Embretson, 1995, 1998; Goswami, 2001; Unsworth & Engle, 2004; Kane et al., 2004). Auch in der alltäglichen

Beobachtung in der diagnostischen Arbeit fällt auf, dass nicht alle Kinder adäquat mit Multiple-Choice-Antwortformaten umgehen können, da sie in der Aufgabenbearbeitung mit diesem Antwortformat oftmals ein vorschnelles Antwortverhalten zeigen.

Diese Arbeit befasst sich mit der Frage, welche kognitiven Operationen bei Kindern im Alter von 6-9 J. wesentlich bei der Lösung von Aufgaben zum logisch-schlussfolgernden Denken beteiligt sind, wobei das Hauptaugenmerk auf dem Einflussfaktor der Antwortformate (Multiple-Choice versus freie Antwort) liegt.

Die hier untersuchte Thematik ist sowohl im Hinblick auf die zukünftige Testkonstruktion relevant als auch allgemein für den Bereich der angewandten Kinder- und Jugenddiagnostik.

2. Theoretischer Teil

Die Fähigkeit zum logisch-schlussfolgernden Denken, oft auch kurz *reasoning ability* oder nur *reasoning* genannt, stellt derzeit in der Psychologie ein anerkanntes und oft gebrauchtes Konstrukt dar. Dabei ist allgemeiner Konsens, dass diese Fähigkeit einen zentralen Aspekt in der kognitiven Leistungsfähigkeit eines Menschen bildet. Für die Bedeutung der Reasoning Ability spricht zum einen die enorme Anzahl an verbreiteten eigenständigen Verfahren, die alleinig die Fähigkeit zum logischen Schlussfolgern erfassen (im deutschsprachigen Raum z.B.: SPM, Raven, Court, Raven, 1988; APM, Raven, Raven, Court, 1998; CFT 1, Cattell, Weiß, Osterland, 1997; BOMAT, Hossiep, Turck, Hasella, 1999;...). Viele dieser Tests beanspruchen durch die Erfassung dieses Merkmals sogar die Erhebung der generellen Intelligenz einer Person. Dies spricht ebenso deutlich für die Bedeutung dieses Konstrukts wie die Tatsache, dass viele einschlägige Intelligenztestbatterien zumindest eine Skala beinhalten, welche Reasoning erfasst.

Dabei stellt sich die Frage, welche zugrunde liegenden Fähigkeiten eines Menschen mit Verfahren zum logisch-schlussfolgernden Denken erfasst werden und damit im Sinne des Konstrukts als bedeutend angesehen werden können. Im Folgenden wird aufgezeigt, dass

die Bedeutung des Arbeitsgedächtnisses bereits nachgewiesen und umfassend untersucht werden konnte, während andere Faktoren eher vernachlässigt wurden. So deuten neuere Forschungsergebnisse auf einen deutlichen Einfluss von generellen Kontrollstrategien im Zusammenhang mit der Erfassung des logisch-schlussfolgernden Denkens hin. Dies scheint besonders eine Relevanz für die weitere Forschung im Kindesalter nahezulegen, da hier zu vermuten ist, dass vor allem junge Kinder aufgrund ihres Alters und ihres noch in der Entwicklung befindlichen Arbeitsstils schwächer ausgeprägte Kontrollstrategien besitzen und dadurch ihre tatsächliche Fähigkeit zum logisch-schlussfolgernden Denken, abhängig vom jeweiligen Verfahren, nur schwer erfasst werden kann.

Unter diesen Gesichtspunkten wird die besondere Bedeutung des Testdesigns, speziell des Antwortformats, bewusst. Die nachfolgenden Erläuterungen verdeutlichen die Vermutung, dass gerade bei jüngeren Kindern das Design des Antwortformats, vor allem in Hinblick auf unausgereifte Kontrollstrategien, eine entscheidende Rolle bei der Erfassung der Fähigkeit zum logisch-schlussfolgernden Denken spielen kann.

2.1. Logisch-schlussfolgerndes Denken

Im Rahmen der untersuchten Thematik der vorliegenden Arbeit stellt das Konstrukt des logisch-schlussfolgernden Denkens einen zentralen Aspekt dar. Aus diesem Grund soll im Folgenden zunächst näher auf die Begriffsbestimmung sowie die Definition dieses Konstrukts eingegangen werden, sowie deren weitere Verwendung in der vorliegenden Arbeit determiniert werden.

Allgemein betrachtet kann schlussfolgerndes Denken als die Art und Weise beschrieben werden, wie Menschen von einem gegebenen Sachverhalt auf einen neuen kommen oder schließen (Beckmann & Guthke, 1999). In der Literatur über das Konstrukt des logisch-schlussfolgernden Denkens wird oftmals zwischen verschiedenen Arten von Schlüssen unterschieden, wobei im Zusammenhang mit der vorliegenden Arbeit besonders das induktive sowie das deduktive Schließen hervorzuheben ist. Dabei zeichnet sich der

induktive Schluss dadurch aus, dass in dessen Ergebnis eine allgemeine Regel oder Aussage enthalten ist (Beckmann & Guthke, 1999; Holyoak & Morrison, 2005). Ein Beispiel hierfür wäre:

Eine Gabel sowie ein Teller fallen.

Gabel und Teller sind Gegenstände.

Also: Alle Gegenstände fallen.

Anhand dieses Beispiels lässt sich zeigen, dass ein induktiver Schluss vor allem dadurch gekennzeichnet ist, dass in ihm eine Verallgemeinerung enthalten ist, sprich dass von Teilfakten auf eine allgemeine Regel geschlossen wird. Dadurch, dass nur eine begrenzte Anzahl an Informationen (nicht alle) zur Verfügung steht, ist ein induktiver Schluss immer mit einer gewissen Unsicherheit verbunden (Beckmann & Guthke, 1999).

Ein deduktiver Schluss zeichnet sich demgegenüber durch eine Eindeutigkeit in der Schlussfolgerung aus. Das heißt, dass bei dieser Art des Schlusses keine weiteren Informationen zu den gegebenen Prämissen angenommen werden. Deduktives Schließen ist laut Beckmann und Guthke (1999, S. 5) unter anderem notwendig, „... um Pläne zu formulieren, [...] Regeln und generelle Prinzipien anzuwenden, Abwägen von Beweisen und Werten von Daten...“.

Die Autoren schreiben weiters, dass zum gegenwärtigen Stand der psychologischen Diagnostik, logisch-schlussfolgerndes Denken oftmals mit dem induktiven Schließen gleichgesetzt wird. Mit dem Verweis auf weiterführende Literatur geben Beckmann und Guthke (1999, S. 9) eine Auflistung von vier Aufgabentypen, die allgemein dem induktiven Schließen zugeordnet werden können:

- Klassifikationsaufgaben
- Vervollständigung von Reihen
- Analogien
- Matrizen-Probleme (z.B. vom Raven-Test-Typus)

Besonders Matrizenaufgaben mit figuralen Aufgabenmaterial, wie sie unter anderem auch im Rahmen der vorliegenden Arbeit zur Anwendung kamen, werden nach Meinung der Autoren im Allgemeinen als prototypisch für die Erfassung des induktiven Denkens sowie für die generelle Erfassung der allgemeinen Intelligenz angesehen.

Im Hinblick auf diese Aufgabenklassifikation nach Beckmann und Guthke (1999) ist jedoch zu ergänzen, dass die Autoren davon sprechen, dass sich die verschiedenen Typen von Aufgaben in der Erfassung der interessierenden Fähigkeit nicht diagnostisch relevant unterscheiden. Die Unterteilung der Itemtypen ist dabei rein kategorischer Natur und nicht mit inhaltlichen Unterschieden gleichzusetzen. Demnach sollte die Erfassung der Reasoning Ability nicht wesentlich innerhalb der angesprochenen Aufgabenkategorien differieren, was in weiterer Folge bedeutend für die Vergleichbarkeit der gewonnenen Daten der vorliegenden Arbeit durch die verschiedenen Verfahren sein wird.

In der Literatur finden sich zahlreiche Verweise auf Intelligenzmodelle, welche direkten Bezug auf die Bedeutung des induktiven Schließens und damit der Reasoning-Fähigkeit nehmen (Beckmann & Guthke, 1999; Holyoak & Morrison, 2005). Beckmann und Guthke (1999) geben einen guten Überblick über bedeutende Intelligenzmodelle im Zusammenhang mit dem logischen Schlussfolgern, welche nachfolgend auszugsweise erläutert werden sollen. So berichtet bereits Spearman in seiner frühen Theorie des g-Faktors von maßgeblichen Korrelationen des logischen Schlussfolgerns auf die generelle Intelligenz (nach Beckmann & Guthke, 1999). In der Theorie der Primärfaktoren nach Thurstone findet sich sogar ein Faktor „reasoning“ der nach oben definierten Kriterien sowohl deduktives als auch induktives Schließen beinhaltet und als Kernfaktor der Intelligenz angesehen wird (nach Beckmann & Guthke, 1999). In faktorenanalytischen Nachfolgeuntersuchungen zu Thurstones Modell postuliert Guilford eine Aufteilung der Reasoning-Fähigkeit in sieben Subfaktoren, die jedoch hauptsächlich materialspezifischen Teilungskriterien zugrunde liegen, was bei der Erfassung des logisch-schlussfolgernden Denkens jedoch vernachlässigt werden kann (nach Beckmann & Guthke, 1999). Im Berliner Intelligenz Struktur Modell (BIS) von Jäger wird der Faktor „Schlussfolgerndes Denken“, ebenfalls getrennt nach dem Aufgabenmaterial, in drei Ebenen erfasst – numerisch, figural und verbal, wobei hier selbige

Kritik wie auch schon bei der Unterteilung nach Guilford zutreffend ist. In der Theorie der fluiden sowie kristallinen Intelligenz nach Catell wird vor allem die fluide Intelligenz durch Reasoning-Aufgaben, meist mit figural-abstraktem Itemmaterial, erfasst (nach Beckmann & Guthke, 1999).

Beckmann und Guthke (1999) kritisieren jedoch die strikte Trennung von Induktion und Deduktion sowie deren Gleichsetzung mit der Reasoning-Fähigkeit aus dem Grund, dass die oben genannten Aufgabentypen in gewissem Sinne beide Arten des Schließens beinhalten. Dabei begründen sie, dass durch diagnostische Verfahren oftmals die Induktionsfähigkeit vermittelt über die Deduktionsfähigkeit erfasst wird: „Aus der richtigen Anwendung der Regel (Deduktion) schließen wir, daß diese Regel erfolgreich gefunden wurde (Induktion)“ (Beckmann & Guthke, 1999, S. 27). Holyoak und Morrison (2005) berichten ebenfalls über theoretische sowie praktische Probleme bei der strikten Trennung der beiden Konstrukte bei der Erfassung des logisch-schlussfolgernden Denkens. In Anlehnung an die vorherrschende Literatur wird in der vorliegenden Arbeit ebenfalls davon ausgegangen, dass sowohl induktives sowie deduktives Schließen als Teilaspekte der Reasoning-Fähigkeit anzusehen sind und durch die hier verwendeten Aufgabentypen kombiniert erfasst werden.

2.2. Die Bedeutung des Arbeitsgedächtnisses

In der vorherrschenden Literatur bezüglich wesentlicher kognitiver Operationen im Zusammenhang mit Aufgaben zur Erhebung des logisch-schlussfolgernden Denkens finden sich zahlreiche Studien über die Bedeutung des Arbeitsgedächtnisses. Dabei zeigt sich, dass Literatur im deutschsprachigen Raum in diesem Forschungsbereich kaum vertreten ist. Ein Großteil der relevanten Studien und Ergebnisse, die in diese Arbeit einfließen, stammt aus dem englischsprachigen Raum. Dies ist ein erstes Indiz für die Notwendigkeit von Forschung im deutschsprachigen Raum in diesem Bereich.

Baddeley und Hitch (1974) sowie Baddeley (1986) haben in einer Reihe von Vorläuferstudien versucht, die Bedeutung des Arbeitsgedächtnisses für die generelle Leistungsfähigkeit einer

Person zu erfassen. Dabei konnte sie aufzeigen, dass der Arbeitsspeicher des Menschen maßgeblich an der Bearbeitung von sprachlichen Aufgaben verschiedenster Art beteiligt ist (Baddeley & Hitch, 1974; Baddeley, 1986). Er kommt dabei zu dem Schluss, dass im Bereich der verbalen Reasoning Ability die Kapazität des Arbeitsgedächtnisses als Hauptdeterminationsfaktor für dessen Leistungsfähigkeit angesehen werden kann.

Kyllonen und Christal (1990) konnten in einer Nachfolgestudie anhand mehrerer Untersuchungen mit einer Stichprobengröße von $N = 713$ bis $N = 412$ die Ergebnisse von Baddeley und Hitch (1974) sowie Baddeley (1986) bestätigen. Dabei berichten die Autoren von Korrelationen im Bereich von 0,80 bis 0,90, die den Zusammenhang zwischen dem Arbeitsgedächtnis und der Reasoning Ability belegen. Aufgrund der Höhe der Korrelationen in ihrer Studie kommen Kyllonen und Christal (1990) ebenso wie Baddeley (1986) sowie Baddeley und Hitch (1974) zu dem Schluss, dass die Kapazität des Arbeitsgedächtnisses als alleinige Determinante für die Fähigkeit zum logisch-schlussfolgernden Denken angesehen werden kann.

Die Theorie der Autoren stützt sich dabei auf Erkenntnisse der Grundlagenforschung sowie auf das von Baddeley und Hitch (1974) sowie Baddeley (1986, 1994) vorgestellte Modell des Arbeitsgedächtnisses. Das Modell besagt, dass bei jedem Menschen das Arbeitsgedächtnis einer Limitation unterworfen ist. Demnach ist die Kapazität des Arbeitsgedächtnisses, und damit verbunden auch dessen Leistungsfähigkeit, beschränkt. Letztlich besteht das Arbeitsgedächtnis nach Baddeley aus zwei Speichern und einem Kontrollsystem aus exekutiven Funktionen. Je nachdem, wie viele Ressourcen ein Mensch zur Bearbeitung von Aufgaben zur Verfügung hat (je größer also die Kapazität des Arbeitsgedächtnisses), desto größer ist im Sinne der Theorie dessen Fähigkeit zum logisch-schlussfolgernden Denken. Je mehr Spielraum das Arbeitsgedächtnis bietet, desto mehr Regeln und Bedingungen können auf einmal bearbeitet und berücksichtigt werden, und damit auch komplexere Aufgaben bearbeitet werden.

Ähnliche Ergebnisse, welche diese Annahme bestätigen, finden sich in zahlreichen Folgestudien aus der jüngeren Zeit, was den Schluss zulässt, dass die Bedeutung des Arbeitsgedächtnisses zumindest für sprachliches Material als nachgewiesen erscheint (Süß, Oberauer, Wittmann, Wilhelm, Schulze, 2004; Kane, 2004; Unsworth & Engle, 2004).

Studien, die sich mit der Generalisierbarkeit dieser Ergebnisse für sprachfreies Material befassen, kommen zu geteilten Ergebnissen (Carpenter, Just, Shell; 1990; Embretson, 1995, 1998; Unsworth & Engle, 2004).

Carpenter et al. haben in einer Studie aus dem Jahr 1990 untersucht, welche Merkmale sich zwischen Personen mit hoher Fähigkeit zum logisch-schlussfolgernden Denken und mit niedriger Fähigkeit in diesem Bereich unterscheiden. Als Aufgaben, um eben diese Fähigkeit zu erfassen, wurden dabei unter anderem die *Advanced Progressive Matrices* von Raven (APM; Raven, Raven, Court, 1998) eingesetzt. Anhand der Daten von prototypischen Testpersonen mit entweder sehr hohem oder niedrigem Testscore in den APM haben die Autoren eine Computersimulation entwickelt, die ähnlich dem Median der jeweiligen Gruppe agieren soll. Die Ausgangsgedanken der Autoren waren dabei, dass die Belastung des Arbeitsgedächtnisses durch die Anzahl an item-spezifischen Regeln determiniert wird und somit erfasst werden kann. Im Sinne von Baddeley (1986) wird also angenommen, dass umso mehr Kapazität des Arbeitsgedächtnisses verlangt wird, je mehr Regeln zur Lösung einer Aufgabe beachtet werden müssen. Dabei kamen die Autoren zu dem Schluss, dass ein Großteil der Unterschiede zwischen Personen mit hohem und niedrigem Testscore durch einen Unterschied der Kapazität des Arbeitsgedächtnisses erklärt werden kann. Die Ergebnisse von Carpenter et al. (1990) sprechen also für eine Generalisierbarkeit der Annahmen von Baddeley (1986) und anderen Forschern (Baddeley & Hitch, 1974; Kyllonen & Christal, 1990; Süß et al., 2004; Kane, 2004; Unsworth & Engle, 2004) hinsichtlich der Bedeutung des Arbeitsgedächtnisses auch auf die Bearbeitung von Aufgaben mit sprachfreiem Inhalt.

Weitere Studien im Bereich von Reasoning bezüglich sprachfreiem als auch sprachlichem Material nennen jedoch Ergebnisse, die an der alleinigen Determination des Arbeitsgedächtnisses für logisch-schlussfolgerndes Denken zweifeln lassen (Baddeley, 1994, 1996; Embretson, 1995, 1998; Unsworth & Engle, 2004; Kane et al., 2004). Nachfolgend soll die Bedeutung von Kontrollstrategien sowie generellen exekutiven Funktionen für die Reasoning Ability erläutert werden.

2.3. Der Einfluss von Kontrollstrategien

Wie bereits weiter oben beschrieben, sprechen viele Forschungsergebnisse dafür, dass die Kapazität des Arbeitsgedächtnisses eine maßgebliche Determinante für die Fähigkeit zum logisch-schlussfolgernden Denken ist (Baddeley & Hitch, 1974; Baddeley, 1986; Kyllonen & Christal, 1990; Süß et al., 2004; Kane, 2004; Unsworth & Engle, 2004). Jedoch weisen zahlreiche differenzierte Studien in diesem Bereich darauf hin, dass neben dem Arbeitsgedächtnis noch eine wesentliche andere Determinante die Reasoning-Fähigkeit beeinflusst (Baddeley, 1994, 1996; Embretson, 1995, 1998; Unsworth & Engle, 2004; Kane et al., 2004). Die Autoren kommen dabei übergreifend zu dem Schluss, dass generelle Kontrollstrategien einen maßgeblichen Einfluss bei der Bearbeitung von Aufgaben zum schlussfolgernden Denken haben.

Bereits 1996 berichtet Baddeley in einer weiterführenden Studie zu den Ergebnissen aus seinem Artikel von 1986 von Überlegungen über die Wichtigkeit von regulierenden exekutiven Funktionen bei der Leistungsfähigkeit im Reasoning-Bereich. Er untersuchte in dieser Studie an einer Reihe von Experimenten die Beschaffenheit und die Bedeutung dieser Funktionen im Zusammenhang mit dem Arbeitsgedächtnis. Wenngleich Baddeley in seinem Artikel zu keinem eindeutigen Schluss über die Struktur der exekutiven Funktionen gelangt, so scheinen für ihn alle Indizien dafür zu sprechen, dass diese Funktionen als generelle Kontrollstrategien einen bedeutenden Einfluss auf die Bearbeitung von Reasoning-Aufgaben haben, der sogar mit jenem des Arbeitsgedächtnisses gleichgesetzt werden könne.

Embretson hat 1995 in einer Studie gezielt den Einfluss der Kapazität des Arbeitsgedächtnisses bei Aufgaben zum logisch-schlussfolgernden Denken erfasst und dem Einfluss von generellen Kontrollprozessen gegenübergestellt. In ihrer Arbeit stützt sich die Autorin stark auf die Studie von Carpenter et al. aus dem Jahr 1990. Zur Erfassung der Reasoning Ability setzte Embretson in dieser Untersuchung als Ausgangspunkt unter anderem ebenfalls die APM von Raven, Raven und Court (1998) ein, wobei auf der Grundlage von dessen Modell weitere Items entwickelt wurden. Ebenso geht Embretson von der Annahme aus, dass die Belastung des Arbeitsgedächtnisses durch die Anzahl an

benötigten Itemregeln determiniert und erfasst werden kann. Des Weiteren nimmt sie an, dass höhere Regeln die Anwendung von niedrigeren voraussetzen, wobei sie sich dabei auf eine unveröffentlichte Studie von Embretson aus dem Jahr 1993 stützt, wonach die Beanspruchung des Arbeitsgedächtnisses der beste alleinige Prädiktor für die Leistung in den APM sind.

Mittels Auswertung der gewonnenen Daten an N= 577 Vp per MLTM (Multilevel Logistic Test Model), einer veränderten Form des LLTM (Linear Logistic Test Model), kommt Embretson (1995) zu dem Schluss, dass generelle Kontrollprozesse eine maßgebliche Determinante für die Fähigkeit zum logisch-schlussfolgernden Denken darstellen. In allen Versuchsbedingungen zeigte sich ein teilweise deutlich stärkerer Einfluss der exekutiven Kontrollfunktion gegenüber der Kapazität des Arbeitsgedächtnisses auf die Reasoning Ability. Die dabei gefunden Ergebnisse können insofern als übereinstimmend mit jenen von Carpenter et al. (1990) sowie Kyllonen und Christal (1990) gesehen werden, als auch hier die Kapazität des Arbeitsgedächtnisses als relevanter Faktor bei der Lösung von Aufgaben zum logischen Schlussfolgern identifiziert wurde. Jedoch widersprechen die Ergebnisse von Embretson (1995) bezüglich der wesentlich stärkeren Bedingung von generellen Kontrollstrategien auf die Erfassung der Reasoning Ability den Annahmen der obigen Autoren, nach denen die Kapazität des Arbeitsgedächtnisses als alleiniger Determinationsfaktor angesehen werden kann. Als Ursache für die beschriebene Diskrepanz der Ergebnisse nennt Embretson (1995), dass sie in ihrer Studie ein anderes Auswertungsverfahren eingesetzt habe, welches sich deutlich besser zur Differenzierung der beiden Fähigkeiten eigne. Das von ihr angewandte MLTM ist speziell darauf ausgelegt, zwischen dem Einfluss zweier Faktoren zu differenzieren, wodurch es ihr deutlich besser gelingt, den Unterschied zwischen den beiden Fähigkeitskonzepten zu erfassen. Carpenter et al. (1990) sowie Kyllonen und Christal (1990) stützen ihre Ergebnisse dagegen hauptsächlich auf die Prüfung von Strukturgleichungsmodellen sowie Korrelationsberechnungen. Dadurch lässt sich zwar der Zusammenhang zwischen zwei Fähigkeiten bestimmen (in diesem Fall die Kapazität des Arbeitsgedächtnisses und die Reasoning Ability), jedoch ist damit keine Differenzierung zwischen dem Einfluss mehrerer Faktoren möglich. So könnten die Ergebnisse von Carpenter et al. in ihrer Studie aus dem Jahr 1990 vielmehr durch den Einfluss genereller Kontrollprozesse bestimmt sein als durch die Kapazität des

Arbeitsgedächtnisses selber. Da Carpenter et al. jedoch nicht zwischen den beiden Fähigkeitskonzepten differenzieren, lässt sich dies nicht hinreichend klären. Diese Argumentation führt allenfalls zu einer vorsichtigen Interpretation der Ergebnisse von Carpenter et al. (1990) sowie Kyllonen und Christal (1990).

Die Annahmen von Embretson (1995) werden durch einige andere Untersuchungen unterstützt (Unsworth & Engle, 2004; Kane et al., 2004). So kommen beispielsweise Unsworth und Engle (2004) in einer Studie, die speziell den Zusammenhang zwischen der Kapazität des Arbeitsgedächtnisses und den Leistungen bei der Bearbeitung von Aufgaben zum logisch-schlussfolgernden Denken im figuralen Bereich erfasst, zu dem Schluss, dass die Spanne des Arbeitsgedächtnisses als alleiniger Faktor nicht genügend Erklärung für die Reasoning Ability bietet. Die Autoren erfassten dabei die Reasoning-Fähigkeit u.a. ebenfalls mit den Matrizentests von Raven. Unsworth und Engle (2004) gelangen nach einer inhaltlichen Analyse ihrer Ergebnisse zu der Überlegung, dass im Einklang mit Embretson (1995) die generellen Kontrollmechanismen eine entscheidende Rolle bei der Leistungsfähigkeit im Reasoning-Bereich spielen. Auch Kane et al. (2004) kommen in ihrer Arbeit zu Ergebnissen, die eine regulierende und bedeutende Funktion der exekutiven Funktionen nahe legen.

Die hier beschriebenen Ergebnisse aus der einschlägigen Literatur scheinen zunächst im Widerspruch zu früheren Ergebnissen zu stehen, bei denen die Kapazität des Arbeitsgedächtnisses als alleiniger determinierender Faktor für die Leistungsfähigkeit im Reasoning-Bereich (im sprachlichen Bereich) identifiziert wurde. Bei einer differenzierten Betrachtung der Methodik sowie der dahinterstehenden Überlegungen wird jedoch erkennbar, dass die Ergebnisse von Embretson (1995), Unsworth und Engle (2004), Keane et al. (2004) sowie Baddeley (1996) vielmehr als Ergänzung früherer Forschung anzusehen sind. Jedenfalls scheint mit dem Blick auf jüngere Forschungsergebnisse sehr viel für eine starke Beteiligung von generellen Kontrollstrategien im Bereich der Leistungsfähigkeit des logisch-schlussfolgernden Denkens zu sprechen. Trotz einiger Ergebnisse, die sogar auf eine größere Bedeutung der regulierenden exekutiven Funktionen gegenüber der Spanne des

Arbeitsgedächtnisses hindeuten, sind genaue Belege für dessen relativen Einfluss bis heute nicht aufzufinden.

2.4. Antwortformate

Wie in den Ausführungen unter Punkt 2.2 sowie 2.3 bereits erläutert werden konnte, spricht neben der Kapazität des Arbeitsgedächtnisses vieles für eine maßgebliche Beteiligung der generellen Kontrollstrategien an der Leistungsfähigkeit im Reasoning-Bereich. Betrachtet man die einschlägige Literatur unter diesem Gesichtspunkt, so fällt auf, dass besonders häufig die Matrizen tests von Raven (SPM, Raven, Court, Raven, 1988; APM, Raven, Raven, Court, 1998) für die Erfassung dieser Fähigkeit zum Einsatz kommen — zum einen vermutlich aus pragmatischen Gründen der Zugänglichkeit sowie Bekanntheit, zum anderen, da die APM und die SPM einen sehr gut untersuchten sowie prototypischen Test im Bereich des nicht sprachlichen logischen Schlussfolgerns darstellen (Carpenter et al., 1990; Beckmann & Guthke, 1999; Holyoak & Morrison, 2005).

Wenig bis gar nicht reflektiert scheint in diesem Zusammenhang die Bedeutung des Antwortformats. Im Bereich der Reasoning Ability gibt es eine Vielzahl an sprachlichen sowie sprachfreien Verfahren zur Erfassung dieser Fähigkeit, wobei im Zuge der Recherchen der vorliegenden Arbeit auffallend war, dass das Antwortformat Multiple-Choice in verschiedensten Ausführungen bei Weitem am häufigsten anzutreffen war.

Anhand der obigen Überlegungen über die Bedeutung genereller Kontrollstrategien soll im Folgenden aufgezeigt werden, dass eine kritische Reflexion des Antwortformats in diesem Kontext notwendig ist. Zwar lassen sich einige Artikel aus der jüngeren Zeit finden, die sich mit generellen Unterschieden verschiedener Antwortformate bei der Fähigkeitserfassung im psychologisch-diagnostischen Kontext befassen (Kubinger & Gottschall, 2007; Kubinger, 2009), jedoch scheint es derzeit keine publizierte Forschungsarbeit zu geben, die sich speziell dem Einfluss des Antwortformats auf die Bearbeitung von Aufgaben zum logischen Schlussfolgern widmet.

2.4.1. Das freie Antwortformat

Wie bereits in der Einleitung zu Punkt 2.4 erläutert wurde, ist das freie Antwortformat in Tests zur Erfassung des logisch-schlussfolgernden Denkens nur wenig verbreitet.

Ein Vertreter aus dem Bereich der Intelligenzdiagnostik, in dem sich gleich mehrere Untertests finden, die in verschiedenster Art und Weise die Reasoning Ability mit freiem Antwortformat erfassen, ist im Bereich des Kindesalters die Intelligenztestbatterie *AID 2* (Adaptives Intelligenz Diagnostikum – aktuelle Version 2.2, Kubinger, K.D., 2009). Als ein Beispiel für die Erfassung des schlussfolgernden Denkens im mathematischen Bereich sei hier der Untertest „Angewandtes Rechnen“ sowie für sprachliche Inhalte der Untertest „Funktionen Abstrahieren“ genannt. Bei Letztgenanntem werden dem Kind zwei Wörter vorgesprochen, wie z.B. „Korb“ und „Tasche“. Anschließend wird gefragt, was das Gemeinsame an den beiden Wörtern sei, beziehungsweise was man mit beiden Gegenständen machen könne. Laut Manual des AID 2 und Validitätsstudien wird damit die Fähigkeit erfasst, mit logisch-schlussfolgerndem Denken Funktionen zu abstrahieren und diese sprachlich auszudrücken. Zwar gelten bei diesem Untertest nur bestimmte, a priori festgelegte Antworten als richtig; da jedoch dem Kind keine Antwortmöglichkeiten angeboten werden und es frei antworten muss, kann dieses Design als eines der weniger Vertreter des freien Antwortformats gesehen werden. Als Pendant zu den beschriebenen Untertests aus dem AID 2 wäre als weiterer namhafter Vertreter der *HAWIK-IV* (Hamburg-Wechsler-Intelligenztest für Kinder IV, Petermann & Petermann, 2008) mit dem Untertest „Gemeinsamkeiten finden“ zu nennen. Abstraktes logisches Schlussfolgern wird im AID 2 unter anderem durch die Untertests „Antizipieren und Kombinieren“ sowie „Analysieren und Synthetisieren“ erfasst. Dabei muss die Testperson Puzzleaufgaben lösen bzw. ein vorgegebenes Muster mit Bausteinen nachlegen. Die charakteristischen Antwortmöglichkeiten von Multiple-Choice-Formaten, aus denen die richtige auszuwählen ist, fehlen; somit wären diese Aufgaben eher dem freien Antwortdesign zuzuordnen. Klassische Reasoning-Aufgaben (Matrizenaufgaben) mit freiem Antwortformat sind im Kindesbereich keine zu finden.

Die geringe Verbreitung des freien Antwortformats im Bereich der Erfassung des logisch-schlussfolgernden Denkens scheint zunächst erstaunlich, zumal aktuelle Forschungsberichte Ergebnisse präsentieren, die aufzeigen, dass dieses Design die Fähigkeit einer Person am unverfälschtesten erfasst (Kubinger & Gottschall, 2007; Kubinger, 2009). Kubinger zeigt in einem Artikel aus dem Jahr 2009 in einem Vergleich mehrerer Antwortdesigns, dass das oft verwendete MC-Antwortformat „1 aus 6“ zu einer signifikanten Vereinfachung einer Aufgabe im Vergleich zum freien Antwortformat führt, während Designs mit mehreren unterschiedlichen richtigen Antworten „x aus 5“ (bei diesem Design kann, je nach Konzeption, entweder keine, eine, zwei bis alle fünf der gegebenen Antwortmöglichkeiten richtig sein) zu einer Erhöhung der ursprünglichen Aufgabenschwierigkeit führen. Dabei ist zu hypothesieren, dass das freie Antwortformat per Konzeption die zu erfassende Fähigkeit am unverfälschtesten erfasst, sofern eine ausreichend gute Itemkonstruktion gegeben ist. Bei einem freien Antwortdesign kann das Verständnis der Aufgabe als hauptsächlicher Faktor für dessen Lösung angesehen werden. Nur bei diesem Antwortdesign sind jegliche Effekte durch die gegebenen Antwortmöglichkeiten, welche die Lösungen der Testperson beeinflussen könnten, auszuschließen, was einen klaren Vorteil dieses Designs gegenüber Testkonzeptionen mit vorgegebenen Antwortmöglichkeiten darstellt.

Dass Testdesigns mit freiem Antwortformat trotz dieses Vorteils in der psychologisch-diagnostischen Praxis vor allem im Bezug auf die Erfassung des logisch-schlussfolgernden Denkens bei sprachfreiem Material im Verhältnis zum Multiple-Choice-Design nur vermindert vorkommen, ist vermutlich vor allem darauf zurückzuführen, dass mit dem freiem Antwortformat eine erhöhte Schwierigkeit der Test-Erstellung sowie der Verrechnung einhergeht. Bei der Möglichkeit, freie Antworten zu geben, stößt man schnell auf die schwierige Frage der Grenzen der zulässigen Antworten und damit auf das Problem der Interpretation der gegebenen Antworten. Bei uneindeutigen Antworten der Versuchsperson ist leicht vorstellbar, dass verschiedene Testleiter a posteriori zu einem anderen Schluss kommen können. Um das Problem der verminderten Verrechnungsfairness bei diesem Antwortformat zu umgehen, bedarf es zum einen einer äußerst sorgfältigen Konstruktion der Testaufgaben, zum anderen scheinen dafür einige Testinhalte geeigneter als andere. Während bei den Beispielen aus dem AID 2 (Kubinger, 2009) sowie dem HAWIK-IV (Petermann & Petermann, 2008) gut aufgezeigt werden konnte, wie ein freies

Antwortformat im Bereich des logischen Schlussfolgerns bei sprachlichem Inhalt im Kindesalter realisiert werden kann, scheint dies vor allem im Bezug auf abstraktes, sprachfreies Material besonders schwer umsetzbar zu sein, da in diesem Bereich kaum Vertreter dieses Designs zu finden sind.

2.4.1.1. Kontrollstrategien und das freie Antwortformat

Ausgehend von den Überlegungen über die Bedeutung von regulierenden Kontrollprozessen bei der Bearbeitung von Aufgaben zum logisch-schlussfolgernden Denken ergibt sich in weiterer Folge der Schluss, dass dabei die Gestaltung des Antwortformats eine wichtige Determinante spielt, die in diesem Zusammenhang noch nicht ausreichend untersucht worden ist. Unter diesem Gesichtspunkt zeigt sich ein klarer Vorteil des freien Antwortdesigns gegenüber Multiple-Choice-Antwortformaten. Wie bereits am Ende von Punkt 2.4.1 beschrieben, kann davon ausgegangen werden, dass das freie Antwortformat per Konzeption die Fähigkeit einer Person am unverfälschtesten erfassen kann (Kubinger, 2009). Dies scheint im Kindesalter unter der Betrachtung der hier aufgestellten Hypothesen von besonderer Bedeutung. Unter der Annahme, dass die Fähigkeit im logischen Schlussfolgern vor allem bei jüngeren Kindern durch gering ausgeprägte Kontrollstrategien bei der Aufgabenbearbeitung von Multiple-Choice-Designs nicht realistisch geschätzt werden kann, gewinnen Testdesigns mit freiem Antwortformat stark an Bedeutung. Bei dieser Art von Antwortgestaltung ist davon auszugehen, dass die Ausprägung der generellen Kontrollstrategien eine untergeordnete Rolle spielt. Im Vorfeld wird davon ausgegangen, dass durch die Präsentation einer Aufgabe im freien Antwortformat die Testperson die Logik des Items entweder versteht und dieses dadurch lösen kann, oder aber die Aufgabe nicht versteht und somit auch nicht lösen kann. Für die Bearbeitung einer Aufgabe mit freiem Antwortdesign sind daher Kontrollprozesse weit weniger involviert als bei der Bearbeitung

von Multiple-Choice-Aufgaben, was in den Punkten 2.4.2 sowie 2.4.2.1 näher erläutert werden soll.

2.4.2. Das Multiple-Choice-Antwortformat

Wie bereits weiter oben erwähnt, ist das Multiple-Choice-Antwortformat im Bereich der Erfassung von logisch-schlussfolgerndem Denken das vorherrschende Design. Unter den bekanntesten Vertretern sind hier die Matrizentests von Raven (SPM, Raven, Court, Raven, 1988; APM, Raven, Raven, Court, 1998) zu nennen, wobei darin je nach Test und Schwierigkeit der Aufgabe die Designs „1 aus 6“ sowie „1 aus 8“ zur Anwendung kommen.

Die Vorteile eines MC-Antwortformats im Vergleich zu einem freien Antwortformat sind zum einen in einer deutlich besseren Verrechnungssicherheit zu sehen, was bereits unter Punkt 2.4.1 dargelegt werden konnte. Die Testperson muss bei dieser Art des Antwortdesigns in den meisten Fällen die eine richtige Antwort (oder mehrere) markieren. Der Testleiter wertet dann lediglich jene Antworten als korrekt, welche auch entsprechend angekreuzt wurden. Dabei bietet sich wenig Spielraum bei der Auswertung eines Tests, was dem Gütekriterium der Verrechnungssicherheit entgegenkommt.

Studien bezüglich der Vergleichbarkeit verschiedener Antwortdesigns kommen jedoch zu Schlüssen, die eine kritische Betrachtung der Vorherrschaft des klassischen MC-Designs „1 aus 6“ nahe legen. Kubinger präsentiert in einem Artikel aus dem Jahr 2009 Ergebnisse, die eine Vereinfachung der Aufgabenschwierigkeit durch das MC-Format „1 aus 6“ gegenüber dem freien Antwortformat aufzeigen. Die Senkung des Schwierigkeitsgrades, und damit auch die Erhöhung der Lösungswahrscheinlichkeit, sind nach Annahmen des Autors primär auf die Möglichkeit des Ratens, sprich den Rateeffekt, zurückzuführen. Der Rateeffekt beschreibt im Grunde die Möglichkeit einer Testperson, eine Aufgabe unter Umständen auch ohne das Vorhandensein der dafür nötigen Fähigkeiten durch Raten zu lösen. Bei dem klassischen Design „1 aus 6“ liegt die Möglichkeit, ein Item durch bloßes Raten zu lösen, bei einer Wahrscheinlichkeit von $1/6$. Dies gewinnt besondere Bedeutung, wenn man bedenkt, dass

eine Testperson in den meisten Fällen nicht völlig ahnungslos ist, sondern oftmals eine gewisse Ahnung oder Vermutung über die Lösung hat, wodurch sich die Auswahl an möglichen Antworten schnell von sechs auf eine deutlich geringere Zahl verringern kann, was sich wiederum in einer erhöhten Wahrscheinlichkeit niederschlägt, die Aufgabe durch Raten lösen zu können.

Kubinger (2009) präsentiert Ergebnisse für das MC-Format „x aus 5“, die für eine deutliche Erhöhung des Schwierigkeitsgrades dieses Designs gegenüber freien Antwortformaten sprechen. Dies kann, genauso wie die Erhöhung der Lösungswahrscheinlichkeiten durch das Format „1 aus 6“, keinesfalls im Sinne eines Testleiters zu sein, da dadurch die Fähigkeit, die erfasst werden soll, je nach Design systematisch über- oder unterschätzt wird.

2.4.2.1. Kontrollstrategien und das Multiple-Choice-Antwortformat

Wie in Punkt 2.4.2 erläutert wurde, lassen aktuelle Forschungsergebnisse den Schluss zu, dass MC-Antwortformate je nach Design zu einer systematischen Verzerrung der zu erfassenden Fähigkeit führen (Kubinger & Gottschall, 2007; Kubinger, 2009). Einschränkend kann hier jedoch gesagt werden, dass diese Trends zwar hinreichend für das Erwachsenenalter untersucht scheinen, jedoch noch kaum im Kindesalter erforscht worden sind. In diesem Zusammenhang und in Anbetracht der Überlegung aus 2.3 sowie 2.4 scheint dabei die Ausprägung von generellen Kontrollstrategien eine bedeutende Rolle zu spielen. Wie in Punkt 2.5 näher erläutert werden soll, sprechen vorherrschende Theorien sowie aktuelle Forschungsergebnisse dafür, dass Kinder erst im späteren Verlauf ihrer Entwicklung über differenziert ausgeprägte Kontrollstrategien verfügen (Ziegler, 1994; Goswami, 2001). Gerade im Vergleich der beiden Antwortformate liegt die Vermutung nahe, dass für das Lösen eines Items im MC-Format deutlich mehr Kontrollstrategien benötigt werden als für das Lösen dieses Items im freien Antwortformat. Während bei einem Aufgabendesign mit freiem Antwortformat per Konzeption das Verständnis des Items als hauptsächlicher Faktor

für dessen Lösung angesehen werden kann, so ist bei der Aufgabenbearbeitung bei einem MC-Design mit der Auswahl an verschiedenen Antwortmöglichkeiten eine zusätzliche bedeutende Komponente involviert. Die bearbeitende Person muss dabei zusätzlich in der Lage sein, alle gegebenen Antwortmöglichkeiten zu analysieren und zu differenzieren, um anschließend die Lösung (oder auch mehrere Lösungen) auszuwählen. Dabei muss also einerseits die Logik der Aufgabe verstanden werden und andererseits jede einzelne Antwortmöglichkeit auf eine eventuelle Passung untersucht und bearbeitet werden. Bei dieser Art der Bearbeitung einer Aufgabe kommen also weit mehr Kontrollstrategien zum Einsatz als bei der Lösung von Items im freien Antwortformat.

Das legt in weiterer Folge den Schluss nahe, dass Personen mit gering ausgeprägten exekutiven Funktionen größere Schwierigkeiten bei einer Aufgabenbearbeitung durch ein MC-Format haben müssten, als wenn sie selbige Aufgabe im freien Antwortformat bearbeiten würden. In der einschlägigen Literatur lassen sich zwar Forschungsergebnisse finden, die für eine bedeutende Rolle von generellen Kontrollstrategien bei der Erfassung der Reasoning Ability sprechen, jedoch scheint der weitere Schritt dieser Überlegungen — der Einfluss der exekutiven Funktion auf die Bearbeitung verschiedener Antwortformate — noch kaum untersucht (Baddeley, 1994, 1996; Embretson, 1995, 1998; Unsworth & Engle, 2004; Kane et al., 2004).

2.5. Die Relevanz der Erforschung genannter Konstrukte im Kindesalter

Auf Basis der vorherrschenden Literatur konnte in den vorigen Abschnitten die bedeutende Rolle von generellen Kontrollstrategien im Zusammenhang mit der Bearbeitung von Aufgaben zum logisch-schlussfolgernden Denken aufgezeigt werden. Zahlreiche Studien liefern Ergebnisse, die neben der Kapazität des Arbeitsgedächtnisses als bedeutendste Determinante für die Reasoning Ability auch eine ebenso große Beteiligung von generellen exekutiven Funktionen als regulierenden Faktor zeigen (Baddeley, 1994, 1996; Embretson,

1995, 1998; Unsworth & Engle, 2004; Kane et al., 2004). Relevante Studien in diesem Bereich beschränken sich jedoch ausschließlich auf die Erforschung genannter Faktoren im Erwachsenenbereich, während die Generalisierbarkeit dieser Ergebnisse auf das Kindesalter noch unzureichend erforscht ist.

Goswami beschreibt 2001 in einem Werk zahlreiche Theorien, die größtenteils auf dem Entwicklungsmodell von Piaget aufbauen. Diese verweisen auf die zentrale Bedeutung der Verarbeitungskomponenten des Gedächtnisses. Als besonders relevant beschreibt Goswami (2001) jedoch den Ansatz von Fischer aus dem Jahr 1980. Dieser legt besonderes Augenmerk auf die Entwicklung von Kontrolle und den Aufbau von Fertigkeiten, wobei er unter Fertigkeiten die Fähigkeit versteht, in einem bestimmten Kontext eine Reihe von Handlungen auszuführen. Gemeinsam ist in den Theorien im Sinne Piagets (nach Goswami 2001) die Annahme, dass die Verarbeitungskapazität eine zentrale Rolle bei der Entwicklung und, damit verbunden, auch der kognitiven Leistungsfähigkeit von Kindern spielt. Diesen Annahmen zufolge ergibt sich eine Obergrenze der kognitiven Leistungsfähigkeit, bedingt durch die individuelle Verarbeitungskapazität. Mit zunehmendem Alter verschiebt sich diese Obergrenze weiter nach oben, wodurch eine Steigerung der kognitiven Leistungsfähigkeit mit höherem Alter postuliert wird.

Aktuelle Forschungsergebnisse postulieren eine verminderte Schwierigkeit von Aufgaben im MC-Format „1 aus 6“ gegenüber freien Antwortformaten im Erwachsenenalter (Kubinger & Gottschall, 2007; Kubinger, 2009). Im Bezug auf etablierte Entwicklungstheorien kommt jedoch die Annahme auf, dass sich die unterschiedlichen Lösungswahrscheinlichkeiten verschiedenster Antwortformate, wie sie Kubinger (2009) beschreibt, nicht im gleichen Maße auf das Kindesalter generalisieren lassen. Die beiden identifizierten Hauptfaktoren für das Zustandekommen eines hohen Testergebnisses bei Tests zum figuralen logischen Schlussfolgern, die Kapazität des Arbeitsgedächtnisses und die Ausprägung von generellen Kontrollstrategien sind bei Kindern noch nicht im vollen Maße ausgeprägt und befinden sich noch in der Entwicklung (Ziegler, 1994; Goswami, 2001). Während die Limitierung des Arbeitsgedächtnisses zum großen Teil Leistungsunterschiede zwischen verschiedenen Altersstufen bedingt, ist jedoch im Bezug auf vermindert ausgeprägte Kontrollstrategien eine

Auswirkung hinsichtlich der Bearbeitung des Itempools je nach Antwortdesign zu erwarten. Wie bereits in Punkt 2.4.2.1 erläutert wurde, benötigt die Bearbeitung einer Aufgabe im MC-Design eine weitaus größere Beteiligung von exekutiven Funktionen als das Lösen eines Items im freien Antwortformat. Daraus lässt sich die Schlussfolgerung ableiten, dass bei mangelnden Kontrollstrategien die Bearbeitung einer Aufgabe im MC-Design schwerer fällt als bei Kontrollstrategien, die in vollem Maße ausgeprägt sind. Daraus folgt, dass vor allem jüngere Kinder durch gering ausgeprägte Kontrollstrategien, die sich noch in der Entwicklung befinden, größere Probleme mit der Bearbeitung von MC-Designs haben sollten als Erwachsene. Im Kindesalter ist also mit zu hypothesieren, das sich entgegengesetzt zu den Beobachtungen von Kubinger (2009) im Erwachsenenalter verhält: dass Kinder durch die Antwortmöglichkeiten in einem „1 aus 6“-Design keine erhöhte, sondern vielmehr eine verringerte Lösungswahrscheinlichkeit zeigen. Entsprechend den Ergebnissen von Kubinger aus dem Jahr 2009 ist bei dem Format „x aus 5“ im Kindesalter eine Erhöhung der Schwierigkeit zu erwarten, die jedoch noch deutlich über das beobachtete Maß im Erwachsenenalter hinausgehen dürfte. Dies würde in weitere Folge bedeuten, dass aufgrund unausgereifter Kontrollstrategien bei Kindern die Fähigkeit zum logischen Schlussfolgern durch MC-Designs nicht adäquat geschätzt werden kann, sondern einer Verzerrung unterliegt.

Diese Arbeit hier soll nun eine sinnvolle Ergänzung der dargestellten Forschung im Bereich der Kinderdiagnostik liefern und Anregung für zukünftige Projekte geben.

2.6. Arbeitsstil

Bei der Intention, wesentliche Merkmale zu erfassen, die einen Einfluss auf die Bearbeitung von Reasoning-Aufgaben haben, kommt man nicht umhin, die Bedeutung des Arbeitsstils zu erwähnen. In der vorherrschenden Literatur finden sich zahlreiche Studien, die sich speziell dem Einfluss der Arbeitsstile Reflexiv vs. Impulsiv auf die Bearbeitung verschiedenster Aufgabentypen und Tests widmen und dabei zu teils unterschiedlichen Ergebnissen kommen.

Kagan, Pearson und Welch berichten bereits im Jahr 1966 in einer frühen Studie an einer Stichprobe von N= 155 Kindern der ersten Schulstufe über deutliche Unterschiede in der Leistung von diversen Reasoning-Tests, abhängig vom individuellen Arbeitsstil. In dieser Studie haben Kinder mit einem impulsiven Arbeitsstil signifikant schnellere Antworten gegeben als Kinder mit reflexivem Arbeitsstil, hatten jedoch auch eine deutlich höhere Fehlerquote. Diese Ergebnisse konnten die Autoren in ihrer Studie sowohl für sprachliches als auch für sprachfreies sowie abstraktes Itemmaterial nachweisen. Vergleichbare Ergebnisse bei einer Stichprobe von älteren Testpersonen berichtet Schweizer in einer aktuelleren Studie aus dem Jahr 2001. Der Autor hat bei einer Stichprobe von 108 Studierenden zum Teil substantielle negative Korrelationen zwischen der Ausprägung an Impulsivität sowie Extraversion und den Leistungen beim logisch-schlussfolgernden Denken im abstrakten figuralen Bereich festgestellt. Dabei kommt er zu dem Schluss, dass sich eine hohe Ausprägung an Impulsivität negativ auf die Leistung zum logischen Schlussfolgern auswirkt.

Andere Arbeiten zu dieser Thematik kommen jedoch zu Ergebnissen, die nicht im Einklang mit den oben beschriebenen stehen (Zelniker & Jeffrey, 1976; Phillips & Rabbitt, 1995). Dabei scheinen diese Arbeiten auf den ersten Blick einen entgegengesetzten Trend zu den oben beschriebenen Forschungen darzustellen (Kagan, Pearson, Welch, 1966; Schweizer 2001). Allerdings zeigt sich bei einer differenzierten Betrachtung, dass die Forschung von Phillips und Rabbitt (1995) vielmehr eine Einschränkung beziehungsweise eine Ergänzung der Thematik liefert. Zelniker und Jeffrey berichten bereits im Jahr 1976 in einer Studie zum Einfluss der Arbeitsstile Reflexiv vs. Impulsiv bei der Bearbeitung von Aufgaben zum

logischen Schlussfolgern bei Kindern im Alter von etwa 8-12 J. über Ergebnisse, die sich mit jenen von Phillips und Rabbitt (1995) decken. Die Autoren konnten bei der Vorgabe von sprachlichem sowie sprachfreiem Material nachweisen, dass Kinder mit einem reflexiven Arbeitsstil im Vergleich zu impulsiven Kindern deutlich bessere Ergebnisse bei Aufgaben zeigen, die eine differenzierte und genaue Analyse des Inhaltes verlangen. Bei Aufgabenmaterial, das eher eine globale Analyse verlangt, zeigt sich hingegen, dass impulsiv arbeitende Kinder bessere Leistungen erzielen. Den Ergebnissen von Zelniker und Jeffrey (1976) folgend, ist also keine globale Aussage über die Vorteilhaftigkeit eines der beiden Arbeitsstile bei der Bearbeitung von Reasoning-Aufgaben zulässig. Vielmehr muss diese Aussage in Anbetracht des jeweiligen Testdesigns und Itemmaterials getroffen werden. In Hinblick auf diese Überlegungen scheinen die Unstimmigkeiten in der einschlägigen Literatur gut erklärbar: Kagan, Pearson und Welch (1966) ebenso wie Schweizer (2001) haben in ihren Studien den Einfluss des Arbeitsstils nicht hinsichtlich des zu bearbeitenden Itemmaterials differenziert, und somit die Stärken des impulsiven Arbeitsstils nicht entsprechend erfasst.

Wie hier gezeigt werden konnte, ist der Bereich der Arbeitsstile im Zusammenhang mit der Testbearbeitung ein Bereich, der vor allem in früheren Jahren vielfach untersucht wurde. Während sich einige Studien dem Bereich des Kindesalters widmen, konzentriert sich ein Großteil der Arbeiten auf erwachsene Personen, wobei aufgrund des Faktors der Entwicklung des individuellen Arbeitsstils gerade hier vermehrte Forschung wünschenswert wäre. Des Weiteren fällt auf, dass in diesem Forschungsbereich kaum deutschsprachige Untersuchungen zu finden sind, wobei hier alleine aus Gründen der Generalisierbarkeit Ergebnisse internationaler Studien wichtig wären. In weiterer Folge soll aufgezeigt werden, dass je nach gegebenem Antwortformat mit unterschiedlichen Ergebnissen je nach individuellem Arbeitsstil zu rechnen ist.

2.6.1. Der Arbeitsstil und das Multiple-Choice-Antwortformat

Zahlreiche Forscher in diesem Bereich klassifizieren Personen mit einem reflexiven Arbeitsstil bei der Aufgabenbearbeitung dahingehend, dass diese einen gegebenen Sachverhalt genau analysieren und daraufhin eine differenzierte Antwort geben (Zelniker & Jeffrey, 1976; Schweizer, 2001). Dem gegenüber wird ein impulsiver Arbeitsstil von Schweizer (2001, S. 1) folgendermaßen beschrieben:

„Persons high in impulsivity prefer the immediate over the well-reasoned response.“

Damit ist gemeint, dass der Arbeitsstil von impulsiven Personen durch schnelle, oft nicht umfassend reflektierte Antworten gekennzeichnet ist.

Diese Klassifizierungen führen in weiterer Folge zu der Überlegung, dass die Wahl des Antwortformats Auswirkungen auf die Itembearbeitung je nach Arbeitsstil haben könnte. Durch das Multiple-Choice-Antwortdesign bekommen die Testpersonen mehrere Antwortmöglichkeiten angeboten und sollen dann, je nach Konzeption, die richtige Antwort (oder mehrere Antworten) kennzeichnen. Hier ist bei Personen mit einem reflexiven Arbeitsstil zu erwarten, dass diese die verschiedenen dargebotenen Antwortmöglichkeiten genau bearbeiten, analysieren und somit zu einem sicheren Ergebnis kommen. Bei einem impulsiven Arbeitsstil ist dies jedoch nicht zu vermuten. Vielmehr ist wahrscheinlich, dass Personen mit einem solchen Arbeitsstil die erste Antwort kennzeichnen, die sie für richtig halten — unter Umständen, ohne die weiteren Antwortmöglichkeiten genau bearbeitet zu haben. Durch eine solche Aufgabenbearbeitung ist die Fehleranfälligkeit um einiges höher zu erwarten als durch ein sorgfältiges und genaues Vorgehen im Sinne eines reflexiven Arbeitsstils. Es ist zu hypothesieren, dass Personen mit einem impulsiven Arbeitsverhalten bei Aufgaben im MC-Design systematisch schlechtere Leistungen erzielen als Personen mit einem reflexiven Stil. Dabei kann allerdings nicht automatisch rückgeschlossen werden, dass diese Personen auch eine geringere Fähigkeit beispielsweise im Bereich des logischen Schlussfolgerns haben, da nicht genau differenziert werden kann, ob die daraus

resultierenden Testergebnisse nur Ausdruck der Fähigkeit sind oder auch maßgeblich durch den individuellen Arbeitsstil beeinflusst werden. Demnach ist also zu vermuten, dass durch Multiple-Choice-Testdesigns die zu erfassende Fähigkeit durch eine maßgebliche Beteiligung des Arbeitsstils verzerrt und somit ungenau erfasst wird.

2.6.2. Der Arbeitsstil und das freie Antwortformat

Wie bereits in den Punkten 2.4.1. sowie 2.4.2. erläutert wurde, ist zu hypothetisieren, dass das freie Antwortformat per Konzeption die Fähigkeit einer Person am unverfälschtesten erfasst. Dies bezieht sich in weiterer Folge nicht nur auf verschieden ausgeprägte Kontrollstrategien, sondern auch auf unterschiedliche Arbeitsstile. Dieses Antwortdesign ist derart konzipiert, dass die bearbeitende Person die Logik der Aufgabe durchschauen und die korrekte Antwort aus sich heraus produzieren muss. Dabei sollte es keinen Unterschied machen, über welchen Arbeitsstil eine Person verfügt, d.h. dass es keine Differenzen in der Aufgabenbearbeitung zwischen impulsiven und reflexiven Personen gibt. Anders als bei der Bearbeitung von Items im MC-Format ist hier zu hypothetisieren, dass impulsive sowie reflexive Personen eine annähernd gleiche Leistung erzielen, abhängig von ihrer individuellen Fähigkeit.

Einschränkend muss hier jedoch der Einfluss des Arbeitsstils bei der Aufgabenbearbeitung hinsichtlich des Itemmaterials im Sinne von Zelniker und Jeffrey (1976) sowie Phillips und Rabbitt (1995) betrachtet werden. So könnte sich der Arbeitsstil, wie oben beschrieben, auch bei unterschiedlichen Antwortformaten nur bei bestimmten Aufgabendesigns deutlich zeigen.

3. Empirischer Teil

In der Einleitung der vorliegenden Arbeit sowie im theoretischen Teil konnte bereits ausführlich dargelegt werden, dass eine weiterführende differenzierte Forschung im Bereich der Erfassung von kognitiven Operationen bei der Bearbeitung von Aufgaben zum logischen Schlussfolgern im Kindesalter äußerst wichtig ist. Im Bereich dieser Thematik findet sich ausschließlich Literatur aus dem englischsprachigen Raum, die zu einem großen Teil als veraltet angesehen werden muss und zudem hauptsächlich Personen im Erwachsenenalter einbezieht (Baddeley & Hitch, 1974; Baddeley, 1986, 1996; Embretson, 1995, 1998; Kyllonen & Christal, 1990; Süß et al., 2004; Unsworth & Engle, 2004; Kane et al., 2004). Gerade bei Kindern ist jedoch aufgrund von zahlreichen sich noch in der Entwicklung befindlichen kognitiven Fähigkeiten sowie Arbeitsstilfacetten zu vermuten, dass bei der Bearbeitung von Multiple-Choice-Aufgaben Wirkmechanismen zum Tragen kommen, die sich entgegengesetzt zu Phänomenen in diesem Bereich verhalten, die bei erwachsenen Personen gefunden wurden (Goswami, 2001; Kubinger & Gottschall, 2007; Kubinger, 2009).

Unter diesen Annahmen ist in weiterer Folge eine maßgebliche Beteiligung der Arbeitsstile Reflexiv vs. Impulsiv bei der Aufgabenbearbeitung verschiedener Antwortformate zu vermuten. Die Auswirkung des individuellen Arbeitsverhaltens war vor allem in den 70er-Jahren sowie auch im weiteren Verlauf immer wieder Interessensgegenstand zahlreicher Studien. Dabei wurden Ergebnisse dahingehend gefunden, dass Kinder mit einem impulsiven Arbeitsstil tendenziell eine schnelle Antwort einer wohlüberlegten Antwort vorziehen (Kagan, Pearson, Welch, 1966; Zelniker und Jeffrey, 1976). Auch in diesem Bereich sind zu einem großen Teil nur Forschungsergebnisse aus dem englischsprachigen Raum zu finden, wobei eine gezielte Untersuchung des Konstrukts „Impulsivität vs. Reflexivität“ im Zusammenhang mit der Bearbeitung verschiedenster Antwortformate bisher nicht gegeben ist.

Diese Arbeit versucht einen Beitrag zur Erforschung von wesentlichen kognitiven Operationen beim Bearbeiten von Reasoning-Aufgaben im Kindesalter im deutschsprachigen Raum zu liefern, wobei hier zusätzlich auf die wenig untersuchte Beteiligung der

Antwortformate Multiple-Choice und freie Antwort sowie der Arbeitsstile Impulsiv vs. Reflexiv eingegangen wird.

Nachfolgend sollen die untersuchten Hypothesen, die Planung und Durchführung der Studie sowie das verwendete Arbeitsmaterial näher beschrieben werden. Abschließend folgt in diesem Kapitel noch eine differenzierte Darstellung der gewonnenen Ergebnisse.

3.1. Fragestellung

Wie bereits einleitend erwähnt, ist das Ziel der vorliegenden Arbeit, wesentliche kognitive Operationen bei Kindern im Alter von 6-9 J. bei der Aufgabenbearbeitung von Reasoning-Tasks bei verschiedenen Antwortformaten zu erheben. Dabei wird theoriegeleitet hypothetisiert, dass Kinder bei der Bearbeitung von Reasoning-Aufgaben im MC-Design schlechtere Leistungen erzielen als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat (Baddeley, 1994, 1996; Embretson, 1995, 1998; Goswami 2001; Unsworth & Engle, 2004; Kane et al., 2004). In weiteren Schritten soll zusätzlich der Einfluss des Alters sowie der Arbeitsstile Impulsiv sowie Reflexiv auf die Aufgabenbearbeitung untersucht werden.

Aus den genannten Überlegungen sowie der Zielsetzung der vorliegenden Arbeit lassen sich folgende Forschungshypothesen ableiten:

3.1.1. Hypothesen

3.1.1.1. H 1

Die Hypothese H 1 stellt die Hauptfragestellung der vorliegenden Arbeit dar und widmet sich generell der Frage, ob es Unterschiede in den Testergebnissen bei der Bearbeitung von Aufgaben zum logischen Schlussfolgern, je nach der Gestaltung des Antwortformats, gibt. Im Sinne der vorherrschenden Literatur über die Bedeutung von Kontrollstrategien beim Lösen von Reasoning-Tasks sowie im Hinblick auf gängige Entwicklungstheorien wird dabei einseitig von der Annahme ausgegangen, dass Kinder bei der Bearbeitung von Multiple-Choice-Aufgaben systematisch schlechtere Leistungen erzielen als bei freien Antwortformaten (Baddeley, 1994, 1996; Embretson, 1995, 1998; Goswami 2001; Unsworth & Engle, 2004; Kane et al., 2004). Aus diesen Überlegungen leiten sich folgende übergeordnete statistische Hypothesen ab:

- **H 1.0:** Es gibt keine signifikanten Unterschiede der Testergebnisse zwischen der Bearbeitung der Verfahren im freien Antwortformat und im Multiple-Choice-Format.
- **H 1.1:** In der Bearbeitung des Multiple-Choice-Formats der Verfahren werden signifikant schlechtere Leistungen erzielt als bei der Bearbeitung selbiger Tests im freien Antwortformat.

Als erweiterte Fragestellung des Hauptuntersuchungsgegenstands liegt es nahe, die erwarteten Effekte differenziert nach den einzelnen Testverfahren zu untersuchen. Aufgrund der verschiedenen Testkonzeptionen, der Itemgestaltung sowie des unterschiedlichen theoretischen Hintergrunds ist nicht auszuschließen, dass Unterschiede zwischen den beiden Antwortformaten in einem der Tests anders aufzufinden sind als in einem anderen, sodass

sich als erweiterte Fragestellung der H 1 folgende Nebenhypothesen im statistischen Sinne ableiten lassen:

- **H 1.a.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung des „Matrizentests“ im freien Antwortformat und im Multiple-Choice-Format.
- **H 1.a.1:** In der Bearbeitung des Multiple-Choice-Formats des „Matrizentests“ werden signifikant schlechtere Leistungen erzielt als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat.
- **H 1.b.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung des Untertests aus dem PSB im freien Antwortformat und im Multiple-Choice-Format.
- **H 1.b.1:** In der Bearbeitung des Multiple-Choice-Formats des Untertests aus dem PSB werden signifikant schlechtere Leistungen erzielt als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat.
- **H 1.c.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung des Untertests „Formale Folgerichtigkeit“ im freien Antwortformat und im Multiple-Choice-Format.
- **H 1.c.1:** In der Bearbeitung des Multiple-Choice-Formats des Untertests „Formale Folgerichtigkeit“ werden signifikant schlechtere Leistungen erzielt als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat.

3.1.1.2. H2

Im theoretischen Teil der vorliegenden Arbeit konnte bereits ausführlich erläutert werden, dass Kinder erst im Laufe ihrer Entwicklung differenzierte Kontrollmechanismen entwickeln (Ziegler, 1994; Goswami, 2001). Im Hinblick darauf, sowie unter Berücksichtigung der Forschungsergebnisse über die Wichtigkeit von generellen exekutiven Funktionen beim Lösen von Aufgaben zum logischen Schlussfolgern ist zu hypothetisieren, dass besonders junge Kinder Probleme bei der Bearbeitung von Items im Multiple-Choice-Design haben (Baddeley, 1994, 1996; Embretson, 1995, 1998; Goswami 2001; Unsworth & Engle, 2004;). Die zweite Forschungshypothese der vorliegenden Arbeit widmet sich also speziell der Frage, inwieweit sich ein Einfluss des Alters der Testpersonen auf die Bearbeitung der beiden untersuchten Antwortformate zeigt. Obige Überlegung führt zu der Annahme, dass mit steigendem Alter Unterschiede in der Aufgabenbearbeitung beider Antwortdesigns abnehmen müssten, woraus sich die folgenden statistischen Hypothesen ergeben:

- **H 2.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung der Verfahren im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.
- **H 2.1:** Es gibt einen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung der Verfahren im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.

Wenn man den Einfluss des Alters getrennt nach den beiden Antwortformaten betrachtet, so sollten sich vor allem im Multiple-Choice-Design Unterschiede in der Testleistung zwischen älteren und jüngeren Kindern zeigen. Bei der Bearbeitung der Items im freien Antwortformat hingegen sollte kaum ein Einfluss des Alters zu beobachten sein, da dabei eine wesentlich geringere Beteiligung der Kontrollstrategien vermutet wird. Auch hier macht

es Sinn, die Ergebnisse getrennt nach den einzelnen Verfahren zu betrachten, um etwaige Unterschiede der Effekte besser erfassen zu können.

- **H 2.a.1.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des „Matrizentests“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.
- **H 2.a.1.1:** Es gibt einen signifikanten Unterschied der Testergebnisse des „Matrizentests“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.
- **H 2.a.2.0:** In der Bearbeitung des freien Antwortformats des „Matrizentests“ gibt es keinen signifikanten Unterschied zwischen jüngeren Kindern und älteren Kindern.
- **H 2.a.2.1:** In der Bearbeitung des freien Antwortformats des „Matrizentests“ gibt es einen signifikanten Unterschied zwischen jüngeren Kindern und älteren Kindern.
- **H 2.b.1.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des Untertests aus dem PSB zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.
- **H 2.b.1.1:** Es gibt einen signifikanten Unterschied der Testergebnisse des Untertests aus dem PSB zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.

- **H 2.b.2.0:** In der Bearbeitung des freien Antwortformats des Untertests aus dem PSB gibt es keinen signifikanten Unterschied zwischen jüngeren Kindern und älteren Kindern.
 - **H 2.b.2.1:** In der Bearbeitung des freien Antwortformats des Untertests aus dem PSB gibt es einen signifikanten Unterschied zwischen jüngeren Kindern und älteren Kindern.
-

- **H 2.c.1.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des Untertests „Formale Folgerichtigkeit“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.
- **H 2.c.1.1:** Es gibt einen signifikanten Unterschied der Testergebnisse des Untertests „Formale Folgerichtigkeit“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Alter.
- **H 2.c.2.0:** In der Bearbeitung des freien Antwortformats des Untertests „Formale Folgerichtigkeit“ gibt es keinen signifikanten Unterschied zwischen jüngeren Kindern und älteren Kindern.
- **H 2.c.2.1:** In der Bearbeitung des freien Antwortformats des Untertests „Formale Folgerichtigkeit“ gibt es einen signifikanten Unterschied zwischen jüngeren Kindern und älteren Kindern.

3.1.1.3. H3

Ein weiterer zentraler Aspekt bei der Erfassung von wesentlichen kognitiven Operationen bei der Bearbeitung von Aufgaben zum logisch-schlussfolgernden Denken im Zusammenhang mit den Antwortformaten Multiple-Choice vs. freie Antwort ist die Untersuchung, inwiefern dies vom Arbeitsstil beeinflusst wird. Aufgrund der Ausführungen in Punkt 2.5 ist zu erwarten, dass sich besonders ein impulsives Arbeitsverhalten in Kombination mit einem MC-Antwortdesign negativ auf die Testleistung auswirkt, während sich dieser Effekt bei der Bearbeitung von Items im freien Antwortformat kaum zeigen sollte. Bei reflexiven Kindern ist im Gegensatz dazu aufgrund des genauen und differenzierten Arbeitsverhaltens kaum ein Unterschied zwischen den Testergebnissen je nach Antwortformat zu erwarten, so dass sich primär folgende statistische Hypothesen ergeben:

- **H 3.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung der Verfahren im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach den Arbeitsstilen Impulsiv vs. Reflexiv.
- **H 3.1:** Es gibt einen signifikanten Unterschied der Testergebnisse zwischen der Bearbeitung der Verfahren im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach den Arbeitsstilen Impulsiv vs. Reflexiv.

Betrachtet man die Fragestellung der Forschungshypothese H3, differenziert nach den einzelnen Tests, so ergeben sich ergänzend folgende Hypothesen:

- **H 3.a.1.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des „Matrizentests“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Impulsiv.
 - **H 3.a.1.1:** In der Bearbeitung des Multiple-Choice-Formats des „Matrizentests“ erzielen Kinder mit einem impulsiven Arbeitsstil signifikant schlechtere Leistungen als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat.
 - **H 3.a.2.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des „Matrizentests“ zwischen der Bearbeitung der Verfahren im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Reflexiv.
 - **H 3.a.2.1:** Es gibt einen signifikanten Unterschied der Testergebnisse des „Matrizentests“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Reflexiv.
-
- **H 3.b.1.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des Untertests aus dem PSB zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Impulsiv.
 - **H 3.b.1.1:** In der Bearbeitung des Multiple-Choice-Formats des Untertests aus dem PSB erzielen Kinder mit einem impulsiven Arbeitsstil signifikant schlechtere Leistungen als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat.

- **H 3.b.2.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des Untertests aus dem PSB zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Reflexiv.
 - **H 3.b.2.1:** Es gibt einen signifikanten Unterschied der Testergebnisse des Untertests aus dem PSB zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Reflexiv.
-

- **H 3.c.1.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des Untertests „Formale Folgerichtigkeit“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Impulsiv.
 - **H 3.c.1.1:** In der Bearbeitung des Multiple-Choice-Formats des Untertests „Formale Folgerichtigkeit“ erzielen Kinder mit einem impulsiven Arbeitsstil signifikant schlechtere Leistungen als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat.
 - **H 3.c.2.0:** Es gibt keinen signifikanten Unterschied der Testergebnisse des Untertests „Formale Folgerichtigkeit“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Reflexiv.
 - **H 3.c.2.1:** Es gibt einen signifikanten Unterschied der Testergebnisse des Untertests „Formale Folgerichtigkeit“ zwischen der Bearbeitung des Verfahrens im freien Antwortformat und im Multiple-Choice-Format, betrachtet nach dem Arbeitsstil: Reflexiv.
-

3.2. Verwendete Verfahren

In der vorliegenden Arbeit wurden drei verschiedene Verfahren zur Erfassung des logisch-schlussfolgernden Denkens eingesetzt, wobei Gemeinsamkeiten dahingehend bestehen, dass sämtliche Tests ausschließlich abstraktes, sprachfreies Itemmaterial verwenden; lediglich für die Testinstruktion ist deutsches Sprachverständnis notwendig.

Im Sinne von generalisierbaren Aussagen wurden hier bewusst drei verschiedene Tests eingesetzt, die allesamt dieselbe Fähigkeit erfassen; so kann ausgeschlossen werden, dass beobachtete Effekte auf Eigenheiten einer bestimmten Testkonstruktion zurückgeführt werden können. Aus diesem Grund wurden auch bewusst Tests ausgewählt, die einen unterschiedlichen Grad an Etablierung aufwiesen; so handelt es sich bei dem Prüfsystem für Schul- und Bildungsberatung (PSB; Horn, 1969) um ein älteres, jedoch oft eingesetztes sowie mehrfach überprüftes Instrument im schulpsychologischen Bereich, während die beiden anderen Verfahren nach dem neuesten Stand der Forschung entwickelt wurden beziehungsweise sich teilweise noch in der Erprobung befinden. Im Hinblick auf das Forschungsinteresse der vorliegenden Arbeit war es nötig, Verfahren auszuwählen, die zum einen prototypische Vertreter des Multiple-Choice-Antwortdesigns sind und sich zum anderen auch zu Testversionen mit einem freien Antwortformat umformen lassen. Zwei der drei hier verwendeten Instrumente liegen in ihrer ursprünglichen Form entweder im MC-Design oder per Konzeption mit freiem Antwortformat vor und wurden im Zuge der vorliegenden Diplomarbeit in eine entsprechende Form des jeweils anderen Antwortformats transferiert, sodass jeder Test in beiden Antwortformaten vorliegt. Dabei ist wichtig zu erwähnen, dass die Aufgaben an sich in beiden Testversionen gleich sind und sich nur das Antwortformat unterscheidet. Bei der Auswahl beziehungsweise der Gestaltung der MC-Versionen wurde entschieden, das oft verwendete Antwortdesign „1 aus 6“ für alle Verfahren zu verwenden.

Zusätzlich zu den beschriebenen Verfahren zur Erfassung des logischen Schlussfolgerns wurde zudem der Arbeitsstil aller Kinder während der Testung erfasst. Die Erfassung per Verhaltensbeobachtung geschah dabei durch a priori instruierte sowie geschulte TestleiterInnen.

Im Folgenden sollen die verwendeten Verfahren sowie der Prozess der Adaption in das entsprechend andere Antwortformat zum besseren Verständnis näher beschrieben werden.

3.2.1. Das Prüfsystem für Schul- und Bildungsberatung (PSB)

Das Prüfsystem für Schul- und Bildungsberatung, kurz auch PSB genannt, ist ein etabliertes Instrument zur Erfassung verschiedenster schulrelevanter Faktoren, wurde von Wolfgang Horn entwickelt und erstmals im Jahr 1969 publiziert. Zusätzlich zu dieser ersten Form gibt es zudem eine aktuellere Neuauflage aus dem Jahr 2004. Neben einer neuen Normierung sowie der Überarbeitung zahlreicher Items, zeichnet sich die Neuauflage des PSB vor allem dadurch aus, dass getrennt nach Schulstufen nun zwei Testformen vorliegen. Dies wurde nach Angaben der Autoren unter anderem aus dem Grund etabliert, dass die Aufgaben in der Testversion aus dem Jahr 1969 vor allem für ältere Testpersonen sehr leicht waren und dadurch im oberen Fähigkeitsbereich kaum differenzierte Erhebungen möglich waren.

Im Rahmen dieser Studie wurde ausschließlich ein Untertest aus der älteren Testversion des PSB zur Erfassung des logisch-schlussfolgernden Denkens bei der Bearbeitung von Aufgaben mit abstraktem visuellem Itemmaterial verwendet (Untertest 3). Die Aufgaben aus dem Untertest 4 in der neuen Version des PSB 6-13 von Horn et al. aus dem Jahr 2004 entsprechen nur hinsichtlich ihrer Konzeption und des Grundgedankens den Items des äquivalenten Untertests aus der älteren Testversion. Nicht nur, dass die Regeln zur Lösung der Aufgaben teilweise komplexer sind, es sind zudem noch die Symbole selbst weitaus komplexer und differenzierter ausgestaltet, wenn auch einige davon aus der alten Testversion aus dem Jahr 1969 übernommen wurden. Während in der alten, hier verwendeten Testversion hauptsächlich geometrische Figuren und einfache Symbole wie Kreise und Quadrate verwendet wurden (siehe Abbildung 1 sowie Abbildung 2), wurden in der neuen Versionen dieses Untertests auch Häuser, Männchen und schematische Gesichter verwendet. Da die ursprüngliche Untertestform des PSB sowohl für die Testversion im freien

Antwortformat als auch im Multiple-Choice-Design adaptiert und bearbeitet werden muss, eignet sich die ältere Version dieses Verfahrens, aus Gründen der leichteren Bearbeitung, besser zum Zwecke dieser Untersuchung. Da sich der hier untersuchte Altersbereich auf 6- bis 9-Jährige Kinder beschränkt, ist zudem kaum mit der Gefahr eines Deckeneffekts zu rechnen. Zum Zweck der hier untersuchten Fragestellungen werden nur die Testwerte der Testpersonen im Hinblick auf beide Testformen und nicht mit einer Population verglichen, wodurch keinerlei Normen benötigt werden und die Testversion aus dem Jahr 1969 keine relevanten Nachteile mit sich bringt.

In der ursprünglichen Version des dritten Untertests aus dem PSB bearbeiten die Testpersonen verschiedene Aufgabenreihen mit diversen Symbolen und Figuren, wobei jenes Symbol identifiziert und durchgestrichen werden muss, das nicht in die Reihe passt. Abbildung 1 und 2 zeigen beispielhafte Aufgabenreihen dieses Untertests aus der originalen Testversion, hier jedoch zur besseren Veranschaulichung deutlich vergrößert.



Abbildung 1: Anfängliche Aufgabenreihe aus dem UT 3 des PSB; Horn, 1969

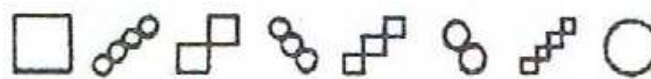


Abbildung 2: Letzte Aufgabenreihe aus dem UT 3 des PSB; Horn, 1969

In den beiden obigen Abbildungen lässt sich gut das Aufgabenprinzip dieses Untertests erkennen. Abbildung 1 zeigt dabei eine Aufgabenreihe, welche nur aus einem Symbol

besteht, wobei die Identifikation der nicht passenden Figur noch sehr leicht fällt. Abbildung 2 zeigt im Gegensatz die letzte Aufgabe dieses Untertests aus dem PSB, welche sich durch eine deutlich komplexere Gestaltung auszeichnet.

Im Sinne der untersuchten Fragestellung ist es nötig, dass jedes Verfahren in einem Multiple-Choice-Design sowie im freien Antwortformat vorliegt. Die ursprüngliche Version des Untertests aus dem PSB entspricht in ihrer Form weder einem klassischen MC-Design, noch ist sie als prototypisch für das Format freie Antwort zu werten. Im Rahmen der vorliegenden Diplomarbeit wurden im Sinne einer möglichst guten Vergleichbarkeit der Verfahren sowie im Hinblick auf die Generalisierbarkeit der hier getroffenen Aussagen auf Basis dieses Untertests aus dem PSB Testversionen mit einem klassischen Multiple-Choice-Format sowie mit einem freien Antwortdesign entwickelt. Nachfolgend soll die Konzeption beider Formate näher beschrieben werden.

Wenn in weiterer Folge der vorliegenden Arbeit die Abkürzung PSB verwendet wird, dann ist damit der hier beschriebene Untertest 3 aus dem Prüfsystem für Schul- und Bildungsberatung gemeint. Im Sinne eines erleichterten Leseflusses wurde beschlossen, nicht immer die ganze Formulierung auszuschreiben, sondern vereinfachend auf die Abkürzung zurückzugreifen.

3.2.1.1. Das Prüfsystem für Schul- und Bildungsberatung im freien Antwortformat

Die Adaption des ursprünglichen Untertests des PSB in ein freies Antwortformat erfolgte zunächst durch das Nachbauen der einzelnen Items. Dies war nötig, da die originale Version von Horn aus dem Jahr 1969 nur eine sehr kleine Aufgabenreihe beinhaltet, was zwar für die Konzeption des Durchstreichens falscher Antworten legitim ist, sich jedoch für die angestrebte Bearbeitung im freien Antwortformat sowie im Multiple-Choice-Design nicht eignet. Zum anderen war die Darstellungsqualität in der Originalfassung nicht mehr zeitgemäß. Wie auch in der Testversion des Matrizentests im freien Antwortformat war hier

der Grundgedanke, dass die Kinder bei der Aufgabenbearbeitung nicht mehr das falsche Symbol in der Reihe durchstreichen sollen, sondern stattdessen in einem freien Feld das Symbol einzeichnen sollen, welches in die Reihe passt und diese logisch ergänzt. Dabei wurde darauf geachtet, die Konzeption der einzelnen Items nicht zu verändern und das freie Feld in der Symbolreihe an die Stelle zu setzen, an der sich ursprünglich die falsche Figur befand. Die Aufgabenreihen des PSB wurden dabei maßgeblich vergrößert, sodass ein Einzeichnen der Lösungen problemlos möglich ist. Von den ursprünglichen Aufgaben mussten einige im Zuge des Adaptionprozesses ausgeschlossen werden, da nicht alle Items im Hinblick auf eine sinnvolle Bearbeitung in beide Antwortformate übertragen werden konnten.

Die freie Testversion des PSB in dieser Studie beinhaltet ebenso wie die Testform im Multiple-Choice-Design jeweils 35 Items, wobei die erste Aufgabe als Beispiel und zur Verdeutlichung des Arbeitsablaufs dient. Die Aufgaben am Anfang des Tests sind dabei wie im Original noch sehr leicht und beinhalten nur wenige Symbole und einfache Regeln, wobei diese im Laufe des Verfahrens immer komplexer werden.

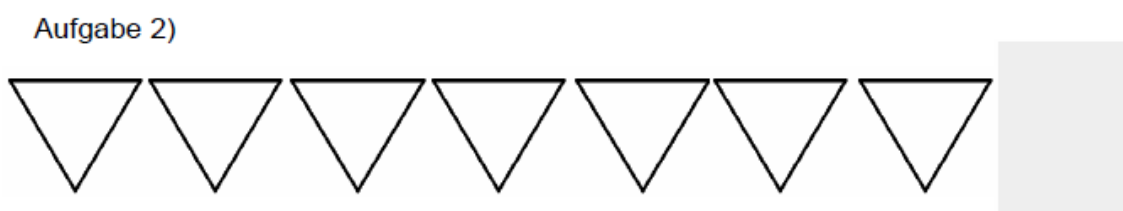


Abbildung 3: Aufgabe 2 aus der adaptierten Testversion des PSB mit freiem Antwortformat

Abbildung 3 zeigt die zweite Aufgabe der adaptierten Testversion des Untertests 3 aus dem PSB mit einem freien Antwortformat. Das gezeigte Item ist das erste, das die Kinder nach der Übungsaufgabe bearbeiten. Dabei lässt sich gut erkennen, dass die ursprüngliche Konzeption

der Aufgabe aus der originalen Testversion beibehalten wurde sowie das freie Feld, in welches die Lösung eingezeichnet werden soll, an der Stelle platziert wurde, an der sich ursprünglich das falsche Symbol befand (siehe Abbildung 1 zum Vergleich). Abbildung 4 zeigt Aufgabe 35 aus der freien Testversion des PSB-Untertests und stellt die letzte zu bearbeitende Aufgabe dar. Das Item in dieser Abbildung stellt das adaptierte Pendant zu der originalen Aufgabenreihe in Abbildung 2 dar. Auch hier wurde bei der Bearbeitung besonders darauf geachtet, die ursprüngliche Konzeption so wenig wie möglich zu verändern.

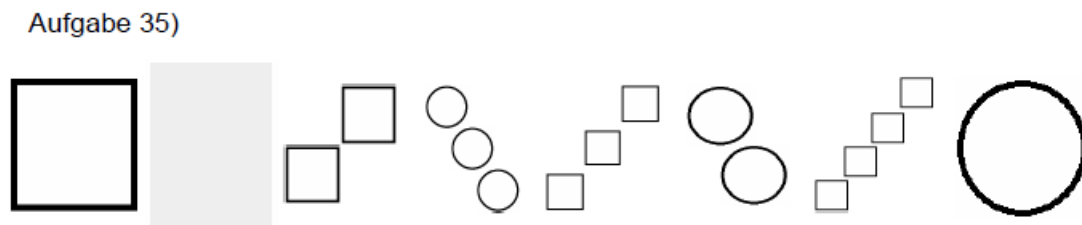


Abbildung 4: Aufgabe 35 aus der adaptierten Testversion des PSB mit freiem Antwortformat

3.2.1.2. Das Prüfsystem für Schul- und Bildungsberatung im Multiple-Choice- Antwortformat

Die Adaption der ursprünglichen Aufgabenreihen aus dem Untertest 3 des PSB in ein Testformat mit einem Multiple-Choice-Design erfolgte in den gleichen Schritten wie bei der Anfertigung der Items mit freiem Antwortformat. Zum Zwecke der Vergleichbarkeit der Ergebnisse aus beiden Antwortformaten wurde streng darauf geachtet, die Aufgabenreihen in jeder Testversion in Bezug auf Form und Größe gleich zu gestalten. In dieser Testversion sollen die Kinder bei der Bearbeitung jedoch nicht die richtige Antwort in das graue Kästchen frei zeichnen, sondern stattdessen die Lösung aus den gegebenen Antwortmöglichkeiten auswählen und einkreisen. Die Antwortmöglichkeiten wurden hier, wie auch in den anderen Multiple-Choice-Testversionen, in einem „1 aus 6“-Design präsentiert. Bei der Konstruktion der Antworten wurde speziell darauf geachtet, diese möglichst logisch und sinnvoll zu gestalten, sodass vorab nicht bereits Antwortmöglichkeiten ohne eine differenzierte Analyse ausgeschlossen werden können. Dabei wurden nur Symbole in der Form und Größe verwendet, wie sie auch in den Aufgabenreihen vorkommen.

Abbildung 5 zeigt die zweite Aufgabe aus der Multiple-Choice-Version des abstrakten Reasoning-Tests des PSB. Dabei ist gut zu sehen, dass die eigentliche Aufgabenreihe die gleiche ist wie in der freien Testversion (Abbildung 3).

Aufgabe 2)

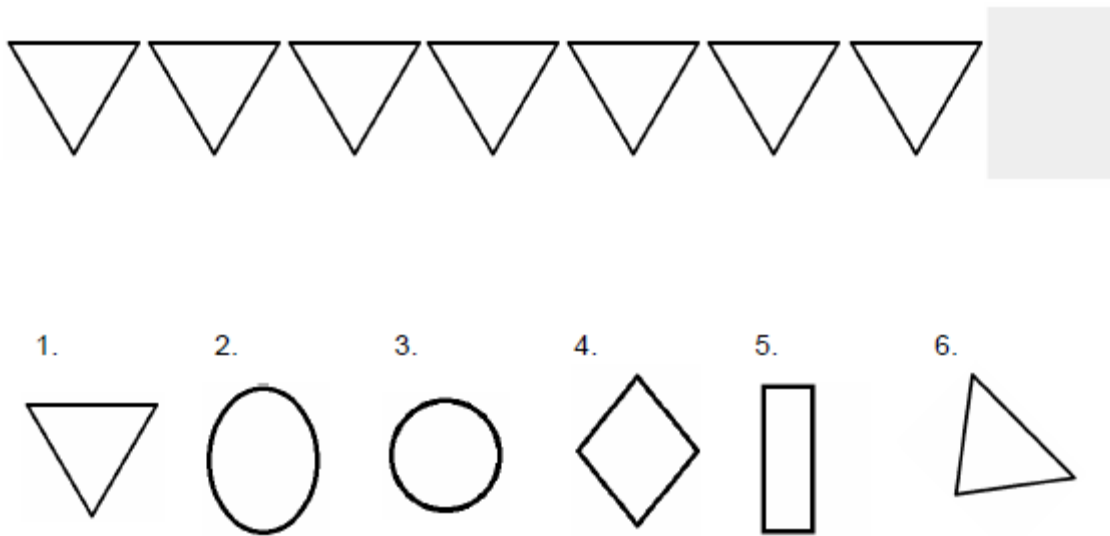


Abbildung 5: Aufgabe 2 aus der adaptierten Testversion des PSB mit Multiple-Choice-Antwortformat

In Abbildung 6 ist die adaptierte Version von Aufgabe 35 aus der Testversion mit freiem Antwortformat (Abbildung 4) in das Multiple-Choice-Design zu sehen. Dieses Beispiel verdeutlicht das Bestreben, bei der Konstruktion der Antwortmöglichkeiten jene Figuren zu verwenden, welche auch für das jeweilige Item sinnvoll sind. Hier wurden fast ausschließlich Symbole derselben Aufgabenreihe verwendet, wodurch es sehr schwer fällt, a priori eine Antwort auszugrenzen und dadurch die Ratewahrscheinlichkeit zu erhöhen.

Aufgabe 35)

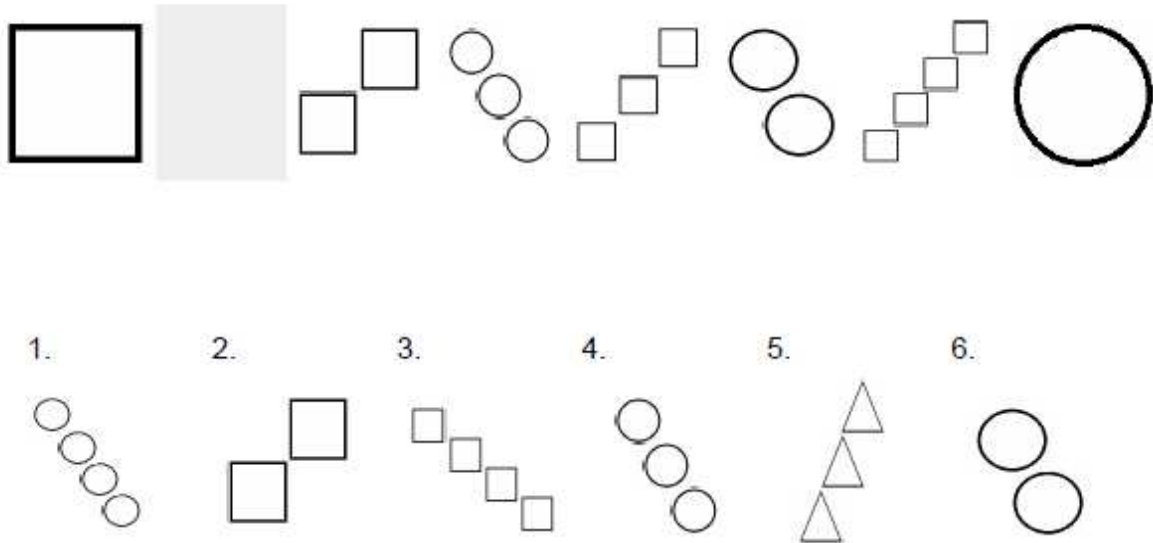


Abbildung 6: Aufgabe 35 aus der adaptierten Testversion des PSB mit Multiple-Choice-Antwortformat

3.2.2. Matrizenest

Bei dem als Matrizenest betitelten Verfahren handelt es sich um ein erst kürzlich entwickeltes Instrument zur Erfassung des logisch-schlussfolgernden Denkens bei der Bearbeitung von abstraktem, sprachfreiem Itemmaterial. Diese Bezeichnung findet hier aus praktischen Gründen statt und stellt einen vorläufigen Arbeitstitel dar, da sich dieser Test nach wie vor noch in der Erprobung und Entwicklung befindet und der endgültige Name noch nicht feststeht. Der Test wurde im Herbst 2010 am Institut für psychologische Diagnostik der Universität Wien unter anderem im Hinblick darauf konstruiert, ein Verfahren zu schaffen, das die Möglichkeit bietet, die abstrakte Reasoning Ability mit einem freien Antwortformat im Kindesalter zu erfassen. Im Zuge der vorliegenden Diplomarbeit wurde viel Recherche im Bereich der Diagnostik des logischen Schlussfolgerns betrieben, wobei die Auswahl an gefundenen Instrumenten, die sich für die Erfassung der Reasoning-Fähigkeit sowohl im Multiple-Choice-Design als auch im freien Antwortformat eignen, äußerst

ernüchternd erscheint. Wie bereits mehrfach erwähnt, sind Verfahren mit freien Antwortformaten, vor allem bei der Verwendung von abstrakten Iteminhalten, kaum vorhanden, wobei in Punkt 2.3 die Vorteile dieses Antwortformats bereits dargelegt wurden. Dies spricht, zusammen mit der Notwendigkeit der Verwendung möglichst vieler Verfahren, um die Generalisierbarkeit der gefunden Ergebnisse zu erhöhen, für die Entscheidung, diesen Test in die vorliegende Arbeit einzubinden.

Der Matrizentest ist bei dieser Studie das einzige angewandte Verfahren, das von Grund auf in beiden Antwortformaten, Multiple-Choice sowie freie Antwort, konzipiert wurde und daher nicht adaptiert werden musste. Die Vorteile eines solchen Vorgehens ergeben sich schnell im Hinblick auf die Itemkonstruktion, da somit sichergestellt ist, dass die verwendeten Aufgaben sich auch tatsächlich für die Bearbeitung in beiden Formaten eignen. Dies kann bei den anderen beiden Instrumenten nicht immer sichergestellt werden, sodass bei diesen vorab einige Items ausgeschlossen werden mussten.

Der Matrizentest ist speziell für das Kindesalter konzipiert und ist ab einem Alter von etwa fünf Jahren angedacht, wobei sich der endgültige Anwendungsbereich erst im Verlauf der Erprobung ergeben wird. Die Testversionen des Multiple-Choice-Antwortformats sowie des freien Antwortformats umfassen jeweils 25 Aufgaben, wobei die Schwierigkeit mit jedem Item immer weiter ansteigt. Neben der Testinstruktion bearbeiten alle Testpersonen zwecks besseren Verständnisses zunächst zwei Beispielaufgaben, anhand derer sichergestellt werden kann, dass die Kinder die Aufgabenstellung auch tatsächlich verstanden haben. Im Sinne des vorläufigen Arbeitstitels dieses Verfahrens wird die Fähigkeit zum logisch-schlussfolgernden Denken hier über das Verständnis sowie das logische Fortsetzen von Matrizenaufgaben erfasst. Neben der Komplexität der Itemregeln erhöht sich im Verlauf des Tests auch die Anzahl an verwendeten Elementen und, damit verbunden, auch die Anzahl an Feldern der Matrix. Die Aufgaben bestehen zunächst noch aus zwei und später aus vier Elementen, dargestellt in einer 2 x 2-Matrix, wobei die Aufgaben am Ende des Tests aus einer 3 x 3-Matrix bestehen. Im Sinne klassischer Matrizenaufgaben ist das rechte untere Feld der Matrix leer und soll von der Testperson ergänzt werden. Ein wesentliches Charakteristikum dieses Tests besteht darin, dass sämtliche Inhalte in vier verschiedenen

Farben dargestellt werden: Schwarz, Gelb, Grün und Grau. Dabei ist je nach Item unterschiedlich, wie viele der vier Farben dabei zum Einsatz kommen, wobei die Schwankungsbreite von einer Farbe bis zur Beinhaltung aller Farben reicht. Die Figuren der jeweiligen Elemente einer Aufgabe setzen sich ausschließlich aus geometrischen Symbolen wie Dreiecken, Quadraten, Strichen etc. zusammen. Zum Lösen einer Aufgabe müssen demnach die richtigen Symbole in der richtigen Farbe ausgewählt werden.

In den beiden nachfolgenden Unterpunkten sollen zum besseren Verständnis exemplarische Aufgabenbeispiele der jeweiligen Antwortformate gegeben werden, sowie auf die Charakteristika beider Testversionen eingegangen werden.

3.2.2.1. Der Matrizentest im freien Antwortformat

Besonders hervorzuheben bei dem Matrizentest im Design mit freien Antworten ist die innovative Lösung beim Kennzeichnen der richtigen Antworten durch die bearbeitenden Personen. Diese Testversion ist dadurch gekennzeichnet, dass die Testpersonen in das jeweils freie Feld der Aufgabenmatrix die korrekte Antwort einzeichnen müssen. Dabei stehen den Kinder vier verschiedene Filzstifte in den Farben Schwarz, Gelb, Grün und Grau zur Verfügung, sodass alle verwendeten Farben in dem Test auch für die Beantwortung zur Auswahl stehen. Bei der Bearbeitung der Items müssen die Kinder also speziell darauf achten, neben dem Finden der richtigen Symbole diese auch in der korrekten Farbe einzuzeichnen.

Ein Kritikpunkt an diesem Vorgehen könnte sicherlich sein, dass dadurch vermehrt Fähigkeiten der visuomotorischen Koordination sowie der Farbdifferenzierung notwendig sind, als dies bei der Bearbeitung eines Multiple-Choice-Designs der Fall wäre. Es wäre durchaus denkbar, dass besonders jüngere Kinder oder generell solche, deren visuomotorische Fähigkeiten noch nicht gut ausgeprägt sind, vermehrt Schwierigkeiten beim

Einzeichnen der Antworten haben. Diesem Aspekt kann jedoch durch eine gezielte und dies berücksichtigende Itemkonstruktion weitestgehend entgegengewirkt werden. In diesem Matrizentest wurden aus diesem Grund nur einfache geometrische Symbole verwendet, die leicht zu zeichnen sind. Außerdem wurde bei der Konzeption der Aufgaben darauf geachtet, dass die Antworten eindeutig sind, sprich: dass keine ähnlich aussehenden Antworten möglich sind. Dadurch ist es nicht notwendig, dass die bearbeitenden Personen die Antwort schön oder genau zeichnen; es genügt, wenn der Testleiter die intendierte Antwort erkennt. Zudem besteht bei der Vorgabe dieses Verfahrens bei einer Einzeltestung immer die Möglichkeit, gezielt nachzufragen, sollte eine Antwort nicht klar zu identifizieren sein. Unter Berücksichtigung dieser Punkte sollten Kinder mit einer gering ausgeprägten Fähigkeit im Bereich der visuomotorischen Koordination kaum benachteiligt gegenüber anderen Kindern sein.

Zur Veranschaulichung der Testkonzeption seien hier prototypische Beispiele des Matrizentests im freien Antwortformat dargeboten. Zunächst die erste Aufgabe am Anfang des Tests, ein Item aus der Mitte sowie ein schwierigeres Beispiel vom Ende.

Beispielaufgaben:

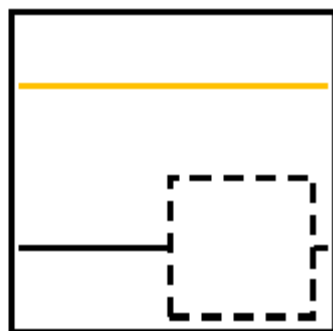


Abbildung 7: Beispielaufgabe aus dem Matrizentest (freies Antwortformat; Seite 2)

In der Übungsaufgabe in Abbildung 7 wäre als korrekte Antwort der untere schwarze Strich, unterbrochen durch das strichlierte Kästchen, durch einen Strich mit dem schwarzen Filzstift zu vervollständigen.

12.

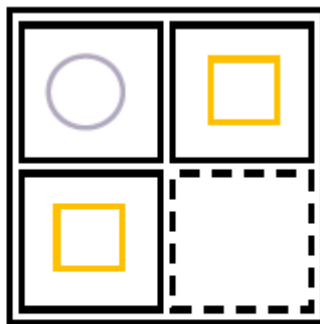


Abbildung 8: Aufgabe 12 aus dem Matrizentest (freies Antwortformat; Seite 6)

Abbildung 8 zeigt eine exemplarische Aufgabe aus der Mitte des Matrizentests in einer 2 x 2-Matrix. Die zu beachtende Regel ist hier, dass abwechselnd ein grauer Kreis und ein gelbes Quadrat dargeboten werden; dabei ist es egal, ob die Regel zeilen- oder spaltenweise angewandt wird. Die richtige Antwort stellt hier (wie oben links in der Matrix) ein grauer Kreis dar.

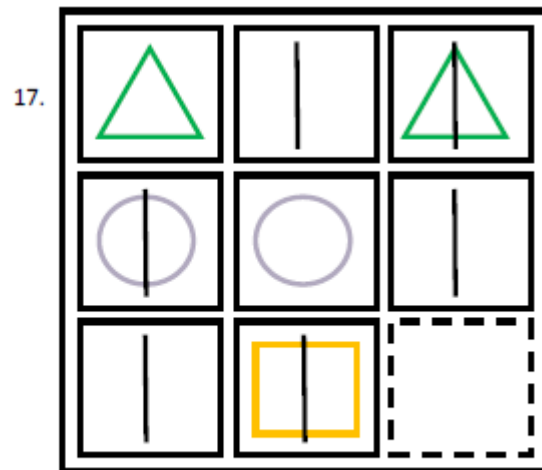


Abbildung 9: Aufgabe 17 aus dem Matrizentest (freies Antwortformat; Seite 9)

Als Beispiel einer 3 x 3-Matrix zeigt Abbildung 9 ein Item relativ weit hinten aus dem Matrizentest. Hier sieht man gut die Zunahme an Komplexität durch einen vermehrten Gebrauch von Symbolen und Farben sowie durch komplexere Itemregeln.

3.2.2.2. Der Matrizentest im Multiple-Choice-Antwortformat

Wie bereits einleitend in diesem Punkt erwähnt, wurde bei der Auswahl der Testverfahren, beziehungsweise bei deren Adaption, streng darauf geachtet, dass die Aufgaben zwischen den beiden Testversionen (MC-Format und freie Antwort) möglichst gleich sind und nur das Antwortformat differiert. An den folgenden Beispielen aus der MC-Testversion des Matrizentests lässt sich gut erkennen, dass die Items der jeweiligen Version genau gleich sind.

Neben einer leicht veränderten Testinstruktion (angepasst an das Antwortformat) zeichnet sich diese Testversion dadurch aus, dass die Antworten der Testpersonen durch ein Multiple-Choice-Design im Format „1 aus 6“ erfasst werden. Dabei soll das Kind die Lösung durch das Ankreuzen der richtigen Antwort kennzeichnen und nicht, wie im freien Design, die Lösung direkt in die Matrix zeichnen. Dies bietet einerseits einen deutlichen Vorteil im Sinne der Verrechnungsfairness, da genau geregelt ist, wie die gegebene Antwort vom Testleiter zu werten ist, andererseits sind genannte Kritikpunkte betreffend der visuomotorischen Koordination bei der Beschreibung der Testversion mit freier Antwort hier obsolet.

Zu Veranschaulichung der Testkonzeption seien hier prototypische Beispiele des Matrizentests im Multiple-Choice-Antwortdesign dargeboten. Es werden hier im Sinne der besseren Vergleichbarkeit die selben Items gezeigt wie in der Testversion mit freiem Antwortformat.

Beispielaufgaben:

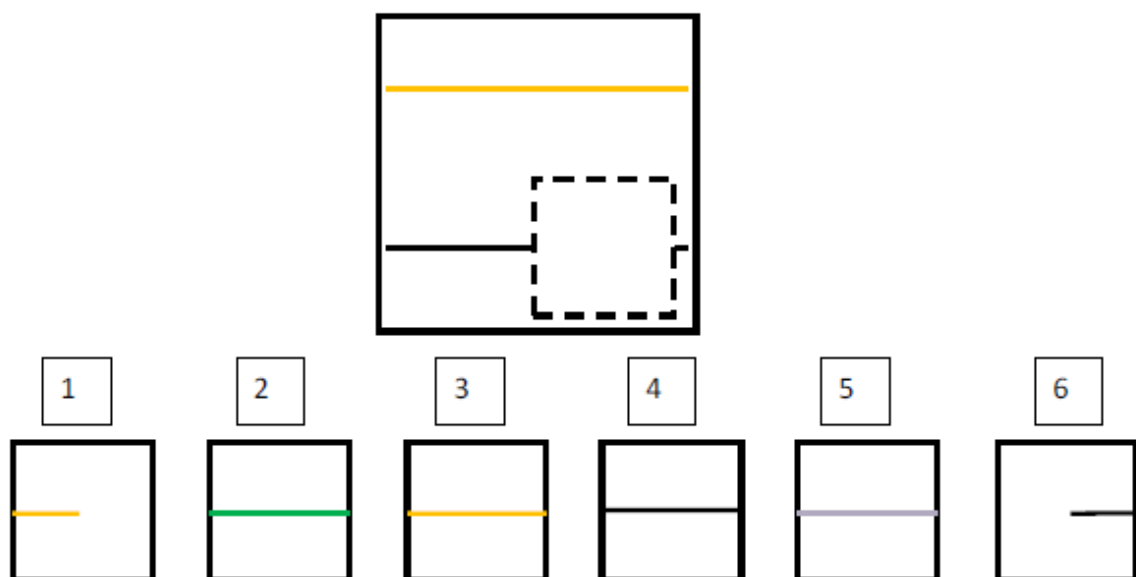


Abbildung 10: Beispielitem aus dem Matrizentest (MC-Version; Seite 2)

Abbildung 10 zeigt die Beispielaufgabe der Multiple-Choice-Testversion des Matrizentests. Dabei soll das Kind in dem Kästchen, in dem die Zahl steht, ein Kreuz bei der Lösung machen. In diesem Fall ist die richtige Antwort an vierter Stelle. Neben der Differenzierung der Farben besteht hier zusätzlich eine Schwierigkeit dahingehend, die richtige Länge des Striches auszuwählen. Bei genauer Betrachtung des Items fällt auf, dass der schwarze Strich nach dem Kästchen noch ein Stück weitergeht, womit Antwort 6 als Lösung hier ausscheidet.

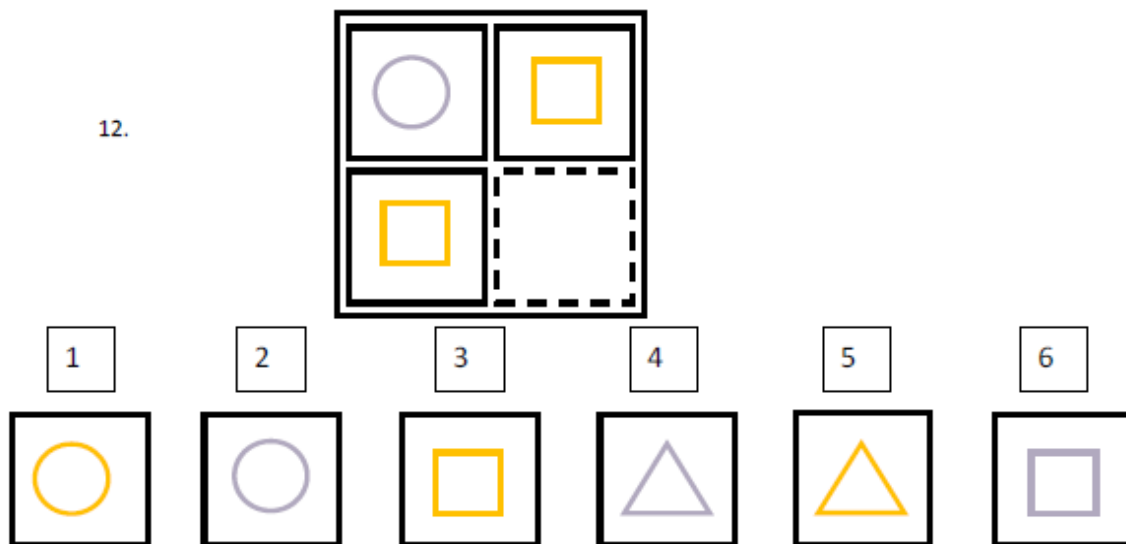


Abbildung 11: Aufgabe 12 aus dem Matrizentest (MC-Version; Seite 6)

Bei der Multiple-Choice-Version von Item 12 des Matrizentests (Abbildung 11) lässt sich gut erkennen, dass bei der Wahl der Antwortalternativen darauf geachtet wurde, diese so zu konzipieren, dass keine Antwort auf den ersten Blick ausgeschlossen werden kann. In diesem Sinne werden hier nur Möglichkeiten geboten, die allesamt in den Farben Grau und Gelb sind, wie auch in der Aufgabenmatrix.

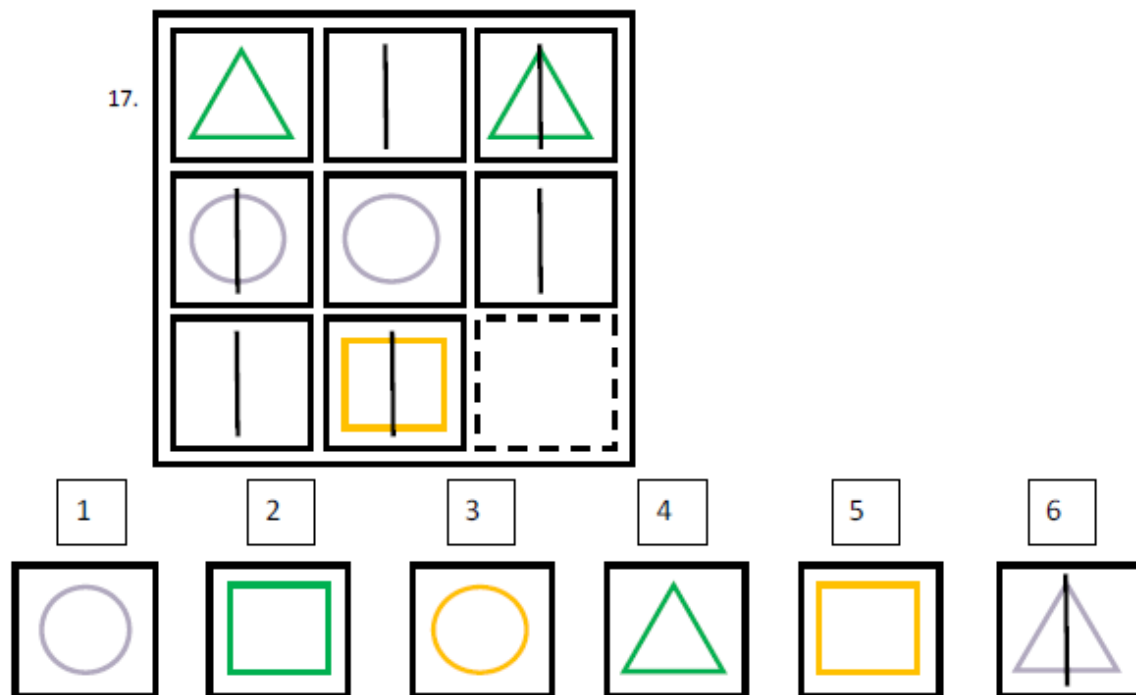


Abbildung 12: Aufgabe 17 aus dem Matrizen-test (MC-Version; Seite 9)

Abbildung 12 veranschaulicht die Konstruktion von Item 17 aus der Multiple-Choice-Testversion des Matrizen-tests. Auch hier wurde darauf geachtet, die Antwortmöglichkeiten allesamt plausibel zu gestalten.

3.2.1. Der Untertest „Formale Folgerichtigkeit“ aus dem AID 3

Bei dem Verfahren „Formale Folgerichtigkeit“ handelt es sich um einen Untertest zur Erfassung des logisch-schlussfolgernden Denkens aus der, sich noch in der Entwicklung befindlichen, Intelligenztestbatterie AID 3 (Adaptives Intelligenz Diagnostikum; aktuelle Version: AID 2.2, Kubinger, K.D., 2009). Dieser Untertest wurde von Hagenmüller (2011) im Rahmen ihrer Diplomarbeit entwickelt, wobei im Bezug auf Einzelheiten bei der Testkonstruktion auf diese Arbeit verwiesen wird. Validitätsstudien des Tests „Formale Folgerichtigkeit“ sind gerade in Arbeit (Heuschober, in Arbeit), wobei erste Ergebnisse für eine insgesamt gute Validität des Verfahrens sprechen.

Das Verfahren deckt prinzipiell einen Altersbereich von 6 bis 15 Jahren ab, wobei gestaffelte Testversionen nach bestimmten Altersgruppen vorliegen; so gibt es eine Version für 6- bis 7-jährige Kinder, eine weitere für 8- bis 11-jährige usw. Für jede Altersgruppe liegen darüber hinaus jeweils zwei Testversionen vor, wobei alle Items der finalen Version der Rasch-Modell-Homogenität entsprechen (Hagenmüller, 2011). Jede Testversion beinhaltet dabei 20 Aufgaben. Wie die beiden anderen, bereits weiter oben beschriebenen und verwendeten Verfahren der vorliegenden Arbeit, so erfasst auch der Untertest „Formale Folgerichtigkeit“ die Reasoning Ability durch die Bearbeitung von Aufgaben mit abstraktem, sprachfreiem Material. Das gesamte Itemmaterial besteht in diesem Instrument aus einfachen geometrischen Figuren (Dreieck, Quadrat, Kreis, Rechteck) in zwei verschiedenen Größen und zwei Farben (Grün und Gelb).

Als Besonderheit des Untertests „Formale Folgerichtigkeit“ muss erwähnt werden, dass dieses Verfahren in seiner ursprünglichen Konzeption ein Antwortdesign hat, das weder eindeutig als freies Antwortformat noch als Multiple-Choice-Format im klassischen Sinne zu klassifizieren ist. Die bearbeitenden Kinder bekommen bei der Testung die Aufgaben einzeln in einem Testheft vorgelegt, wobei die Aufgaben jeweils aus Reihen mit unterschiedlichen Symbolen bestehen. Die Kinder sollen dabei die Reihen logisch fortsetzen. Neben dem

Testheft werden den Kindern noch insgesamt 16 Bausteine in einer bestimmten Anordnung aufgelegt (siehe Abbildung 13). Diese Bausteine stellen alle verwendeten Symbole des Instruments dar, sprich: ein großes, grünes Quadrat und ein kleines, grünes Quadrat sowie Quadrate in beiden Größen in gelb, usw.

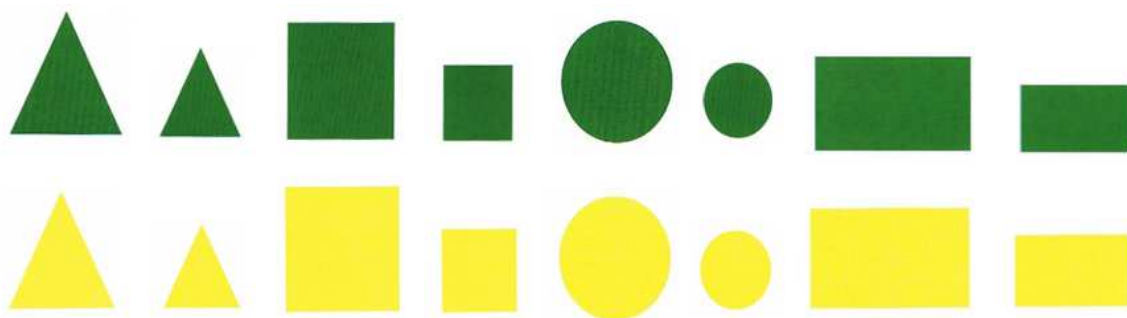


Abbildung 13: Darstellung der Antwortsteine aus dem UT „Formale Folgerichtigkeit“

Abbildung 13 zeigt eine Darstellung der Antwortsteine aus der Sicht der Testperson. Das Kind soll aus diesen Möglichkeiten die jeweils richtige Antwort auswählen und an das Ende der Aufgabenreihe legen, um diese logisch zu ergänzen. Danach wird die gegebene Antwort durch den Testleiter in einem separaten Antwortbogen verzeichnet sowie die Reihe wieder in ihre Ausgangsform gebracht. Zu Beginn des Verfahrens gibt es immer eine Beispielaufgabe, anhand derer sichergestellt werden kann, dass die Testperson die Aufgabeninstruktion verstanden hat. Der Untertest „Formale Folgerichtigkeit“ ist in seiner ursprünglichen Form zur Einzeltestung konzipiert.

Wie bereits weiter oben erwähnt, handelt es sich dabei um eine Spezialform des freien Antwortformats, die auch einigen Kriterien des MC-Formats entspricht. Aufgrund der großen Anzahl an Distraktoren ist davon auszugehen, dass bei der Bearbeitung der Rateeffekt, wie

ihn Kubinger und Gottschall (2007) sowie Kubinger (2009) bei Erwachsenen postulieren, äußerst gering ist. Eine weitere Erklärung dafür, dass dieses Design als freies Antwortformat gesehen werden kann, gibt Hagenmüller (2011) in ihrer Arbeit; die Autorin meint dort, dass aufgrund theoretischer Überlegungen nicht zu erwarten ist, dass Testpersonen andere Symbole als Antwort geben als jene 16, aus denen sich alle Aufgabenreihen zusammensetzen. Da der Test aus den immer gleichen Symbolen besteht, wird also kaum ein Kind z.B. ein Flugzeug als Antwort geben wollen. Eine weitere Besonderheit des Untertests „Formale Folgerichtigkeit“ besteht darin, dass das Kind mit Bausteinen arbeitet, diese in die Hand nehmen, rotieren, austauschen etc. kann. Ein wesentliches Merkmal eines klassischen MC-Designs wäre dagegen ein Ankreuzen der Lösung.

Neben diesen Überlegungen ist jedoch nicht zu leugnen, dass dieses Design auch wesentliche Charakteristika eines Multiple-Choice-Designs erfüllt. Wie der Name schon sagt, zeichnet sich das MC-Antwortformat dadurch aus, dass man seine Antwort aus mehreren Möglichkeiten auswählt, was hier durchaus gegeben ist. Zudem ist denkbar, dass je nach Aufgabe sofort einige der Möglichkeiten ausgeschlossen werden können, wodurch die Wahrscheinlichkeit, die Lösung durch Raten zu finden, steigen würde. Diese Überlegungen werden gerade im Hinblick auf die hier untersuchte Fragestellung besonders zu reflektieren sein. Dennoch sprechen die theoretischen Überlegungen in der Summe dafür, das hier verwendete Design primär als freies Antwortformat zu sehen.

In einigen Abbildungen sowie Tabellen der vorliegenden Arbeit ist die Bezeichnung „Logische Blöcke“ (LB) als Synonym für den Untertest „Formale Folgerichtigkeit“ zu sehen. Dies ist darauf zurückzuführen, dass der Titel dieses Verfahrens im Laufe der Untersuchung und Datenauswertung der vorliegenden Arbeit geändert wurde. Wenn im weiteren Verlauf die Bezeichnung „Logische Blöcke“ zu lesen ist, so ist damit stets das hier beschriebene Verfahren „Formale Folgerichtigkeit“ gemeint.

3.2.1.1. Der Untertest „Formale Folgerichtigkeit“ aus dem AID 3 im freien Antwortformat

In diesem Punkt sollen zum näheren Verständnis exemplarische Beispiele aus der ursprünglichen Testversion des Untertests „Formale Folgerichtigkeit“ von Hagenmüller (2011) im freien Antwortformat gegeben werden. Im Rahmen dieser Forschungsarbeit wurde jeweils nur eine Testversion pro Altersbereich (Testform 1 für 6- bis 7-jährige sowie Testform 3 für 8- bis 9-jährige Kinder) eingesetzt. Dabei gibt es pro Testform sowie altersübergreifend einige Items, die in mehreren Testversionen Verwendung finden, während andere nur in einer Form vorkommen.

5)



Abbildung 14: Aufgabe 5 aus dem UT „Formale Folgerichtigkeit“; Testform 1

Abbildung 14 zeigt exemplarisch eine Aufgabe aus der Testform 1 (für 6- bis 7-jährige Kinder) des Untertests „Formale Folgerichtigkeit“ aus dem AID 3, hier noch in einer einfachen Form.

Zu sehen sind hier nur grüne und kleine Symbole (dies lässt sich leichter im Zusammenhang mit den anderen Items sowie im Vergleich mit den Antwortsteinen erkennen). Um diese Aufgabe richtig zu beantworten, muss das Kind dabei den kleinen, grünen Kreis aus der Reihe der Antwortsteine nehmen und diesen in das Feld mit dem Fragezeichen legen. Im Sinne obiger Kritik bezüglich der Distraktorenzahl wäre dies ein Beispiel, bei dem denkbar wäre, dass vorab alle gelben Steine als mögliche Antwort ausgeschlossen werden können, sodass sich die Anzahl der Antworten unter Umständen schnell auf 8 Steine halbieren lässt.

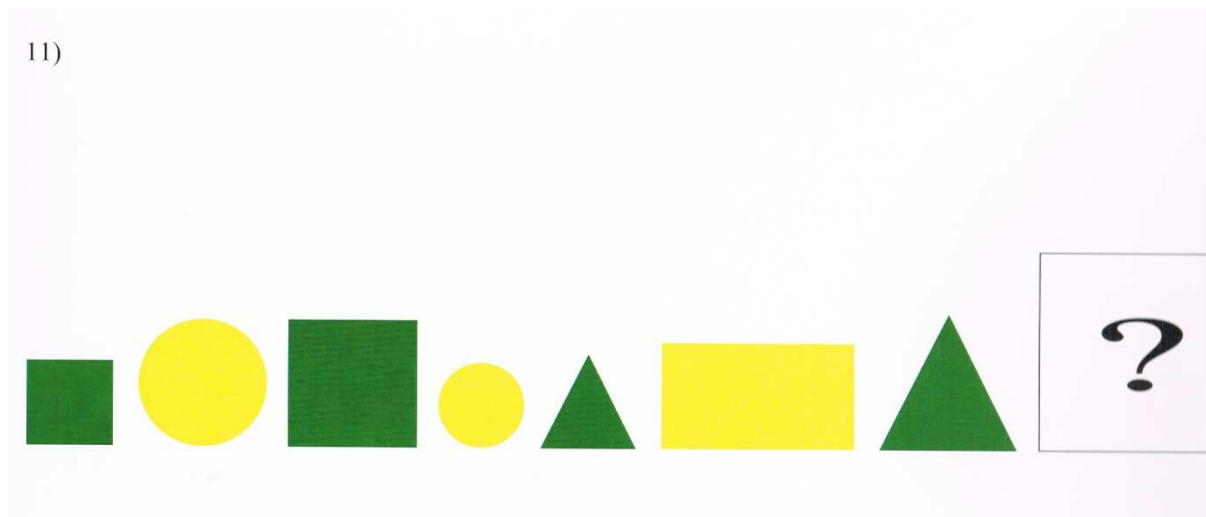


Abbildung 15: Aufgabe 11 aus dem UT „Formale Folgerichtigkeit“; Testform 3

In Abbildung 15 sieht man ein Aufgabenbeispiel aus der Testform 3 für 8- bis 9-jährige, wobei selbiges Item auch in der Testform 1 für 6- bis 7-jährige Testpersonen vorkommt, dort jedoch die 17. Aufgabe darstellt. Diese Aufgabe beinhaltet gegenüber der obigen deutlich mehr Steine sowie komplexere Regeln, die während dem Bearbeiten beachtet werden müssen. Die Lösung wäre hierbei ein kleines, gelbes Rechteck. Hier fällt es nicht leicht, vorschnell Steine auszuschließen, sodass mit der Gesamtzahl der Distraktoren gearbeitet werden muss.

3.2.1.2. Der Untertest „Formale Folgerichtigkeit“ aus dem AID 3 im Multiple-Choice- Antwortformat

Die Multiple-Choice-Version des Untertests „Formale Folgerichtigkeit“ wurde im Hinblick auf die hier untersuchte Fragestellung im Zuge der vorliegenden Diplomarbeit aus der ursprünglichen Version des Verfahrens adaptiert. Um eine möglichst hohe Vergleichbarkeit beider Testversionen zu gewährleisten, wurde streng darauf geachtet, das originale Format so wenig wie möglich zu ändern. Wie u.a. in Abbildung 16 zu sehen ist, wurde jedoch aus ökonomischen Gründen die Entscheidung getroffen, die Multiple-Choice-Version nicht wie ursprünglich farblich zu gestalten, sondern stattdessen das Material in Graustufen darzubieten, sodass aus den Farben **Gelb** und **Grün** schlussendlich **Hell-** sowie **Dunkelgrau** wurden. Da es bei der ursprünglichen Farbwahl jedoch nur um die Unterscheidung zwischen den Symbolen geht und nicht primär um die Farben an sich, sollte die Vorgabe desselben Materials in Graustufen keine Auswirkung auf die Bearbeitung haben. Hier wurde, wie auch in den anderen Multiple-Choice-Testversionen, bewusst im Sinne der Vergleichbarkeit das Antwortdesign „1 aus 6“ gewählt, wobei speziell darauf geachtet wurde, sinnvolle Distraktoren zu wählen, sodass keine Antworten vorschnell ausgeschlossen werden können. Die Instruktion dabei lautet dahingehend, dass die Testpersonen die Zahl der richtigen Antwort einkreisen sollen, was noch einmal speziell an einer Beispielaufgabe verdeutlicht wird.

5)

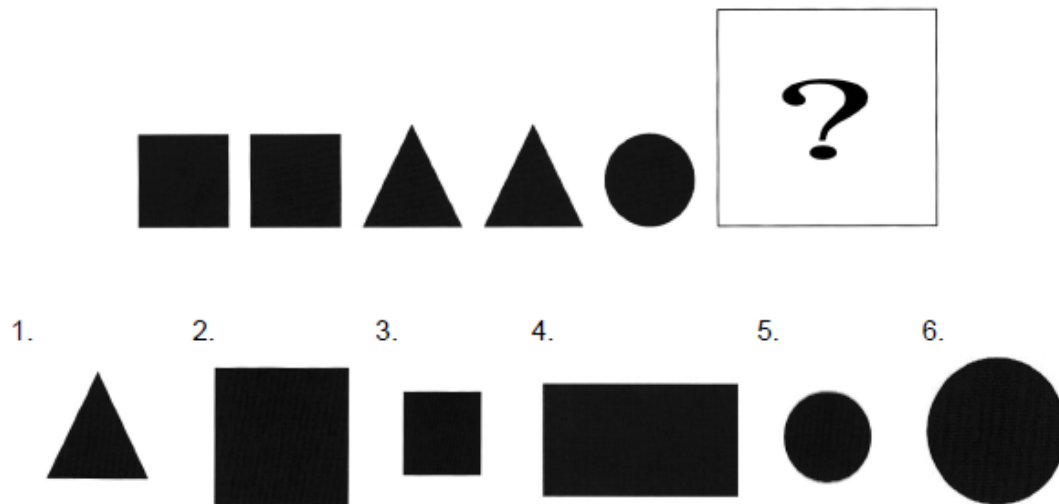


Abbildung 16: Aufgabe 5 aus dem UT „Formale Folgerichtigkeit“; Multiple-Choice-Format

Abbildung 16 zeigt dasselbe Item, das bereits weiter oben in der ursprünglichen Testversion präsentiert wurde. Dabei lässt sich gut die Änderung der grünen Farbe in ein sehr dunkles Grau erkennen, was jedoch keine Auswirkung auf die Bearbeitung haben sollte, da die konzipierten Itemregeln zur Beantwortung nach wie vor dieselben sind. Hier wurde bei der Distraktorenwahl besonders darauf geachtet, dass nur dunkelgraue Symbole verwendet werden, die auch in der Aufgabenreihe vorkommen. Dadurch können keine Antworten vorschnell ausgeschlossen werden.

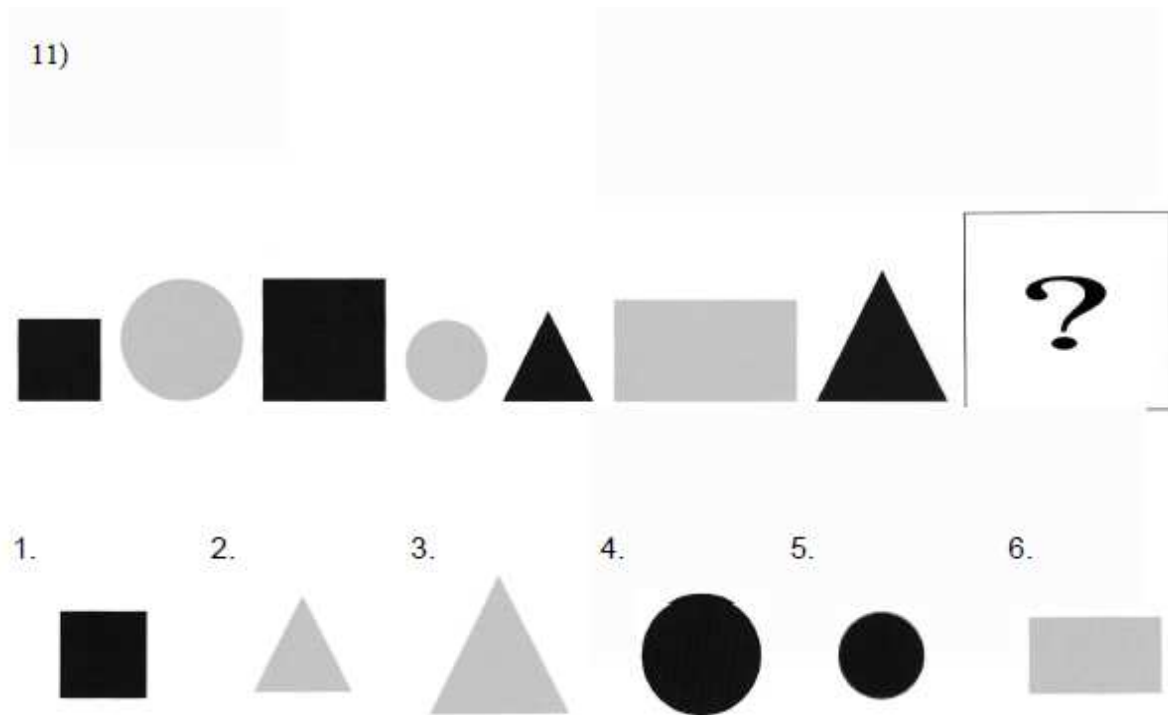


Abbildung 17: Aufgabe 11 aus dem UT „Formale Folgerichtigkeit“; Multiple-Choice-Format

Abbildung 17 zeigt hier das Pendant zu Aufgabe 11 (Abbildung 15) aus dem Untertest „Formale Folgerichtigkeit“ des AID 3 im Multiple-Choice-Design. Dabei lässt sich gut erkennen, dass durch das Umwandeln der ursprünglichen Farben in Graustufen kein Unterschied bei der Aufgabenbearbeitung entsteht. Durch die Auswahl von jeweils drei Symbolen in hell- und dunkelgrau sowie von jeweils zwei großen und vier kleinen Symbolen wurde speziell darauf geachtet, dass bei der Wahl der Antwortmöglichkeiten möglichst ein sinnvolles Gleichgewicht herrscht.

Aus urheberrechtlichen Gründen können die Aufgaben des Matrizentests sowie des Tests „Formale Folgerichtigkeit“ im Anhang der vorliegenden Arbeit nicht angegeben werden. Für nähere Informationen des Tests „Formale Folgerichtigkeit“ sei hier auf die Arbeit von Hagenmüller (2011) sowie für den Matrizentest auf das Institut für psychologische Diagnostik der Universität Wien verwiesen.

3.2.2. Die Erfassung des Arbeitsstils

Die Forschung bezüglich der Hypothese H 3 samt Unterhypothesen erfordert, dass neben der Fähigkeit zum logischen Schlussfolgern auch der Arbeitsstil erfasst wird. Dieser wurde während der Einzeltestung durch vorab geschulte und instruierte Testleiter erfasst. Als Testleiter fungierten ausschließlich Studierende der Fakultät für Psychologie an der Universität Wien, wobei ein Großteil der Personen zudem bereits einschlägige Erfahrungen im Bereich der psychologischen Diagnostik vorweisen konnte.

Im Sinne der untersuchten Fragestellung wurden die getesteten Kinder ausschließlich hinsichtlich der Arbeitsstile Reflexiv vs. Impulsiv klassifiziert, wobei zusätzlich eine neutrale Kategorie eingeführt wurde, sodass sich eine dreikategoriale Abstufung ergibt. Diese wurde aus den Überlegungen in die Klassifikation mit einbezogen, dass viele Kinder weder einem reflexiven noch einem impulsiven Arbeitsverhalten eindeutig zuzuordnen sind; demnach ist in dieser Studie eine Testperson, dessen Arbeitsstil als neutral eingestuft wird, weder als besonders impulsiv noch reflexiv zu sehen. Durch die Möglichkeit einer neutralen Einstufung ergibt sich zwar die Gefahr, dass besonders viele Kinder dieser Kategorie zugeordnet werden, jedoch ist dadurch auch eine wesentlich differenzierte Untersuchung der Fragestellung möglich. Personen, deren Arbeitsverhalten als impulsiv oder reflexiv angesehen werden kann, erfüllen bei einem dreikategorialen System mit einer neutralen Mitte die relevanten Kriterien der beiden Arbeitsstile eher als bei einem Forced-Choice-Design ohne die Möglichkeiten einer neutralen Klassifikation.

Charakteristische Züge der beiden Arbeitsstile bei der Testbearbeitung wurden einerseits aus dem AID 3 abgeleitet sowie im Hinblick auf einschlägige Literatur zur Erfassung des logischen Schlussfolgerns adaptiert und ergänzt (Zelniker & Jeffrey, 1976; Schweizer, 2001). Im Folgenden findet sich eine Beschreibung wesentlicher Merkmale der Arbeitsstile Impulsiv und Reflexiv, wie sie in der vorliegenden Arbeit klassifiziert wurden, sowie ein Auszug einer genaueren Verhaltensbeschreibung während des Matrizentests.

- Facette zum **reflexiven Arbeitsstil**:

Bearbeitet fordernde Aufgaben der Tests zum logischen Schlussfolgern erst, nachdem es sich für eine genaue Analyse des Problems Zeit genommen hat. Achtet dann auf Details und macht wenige Fehler. Tendiert dazu, Antworten zu überdenken und gegebenenfalls zu korrigieren.

- Facette zum **impulsiven Arbeitsstil**:

Bearbeitet fordernde Aufgaben der Tests zum logischen Schlussfolgern direkt, ohne sich vorher Zeit für eine genaue Analyse des Problems zu nehmen. Achtet dann nicht auf Details und macht vermehrt Fehler. Gegebene Antworten werden kaum überdacht. Verweilt nicht lange bei der Bearbeitung einzelner Items.

Beobachtungen während der Einzeltestung mit dem **Matrizentest** zum Arbeitsstil:

Wie ist das Vorbereitungs- und Antwortverhalten bei fordernden Aufgaben?

Reflexiv

Eine längere Betrachtungszeit und vereinzelt (richtige) Antworten deuten auf einen reflexiven Arbeitsstil hin. Wenn das Kind nach dem Zeichnen einer Antwort dazu tendiert, diese noch einmal zu kontrollieren und auszubessern, so deutet dies ebenfalls auf einen reflexiven Arbeitsstil hin.

Impulsiv

Zeichnet sich durch sehr geringe Betrachtungszeiten mit sofortigem Antworten aus. Achtet bei der Bearbeitung nicht auf Details. Eine geringe Tendenz, Antworten zu korrigieren und auszubessern, deutet ebenfalls auf einen impulsiven Arbeitsstil hin.

Abbildung 18: Auszug aus dem Klassifikationsbogen des Arbeitsstils

Abbildung 18 zeigt eine Beschreibung von typischem Verhalten eines reflexiven beziehungsweise impulsiven Arbeitsstils während der Bearbeitung des Matrizentests. Dabei lässt sich, ebenso wie in der Facettenbeschreibung der Arbeitsstile, deutlich erkennen, dass zu einer Klassifikation der Testpersonen das Arbeitsverhalten bei fordernden Aufgaben herangezogen werden muss. Bei zu leichten oder zu schwierigen Aufgaben ist nicht auszuschließen, dass das beobachtete Verhalten durch die Schwierigkeit der Items maßgeblich beeinflusst wird; so ist es zum Beispiel möglich, dass ein reflexiv arbeitendes Kind, entgegen seines üblichen Arbeitsverhaltens, bei besonders leichten Aufgaben ein sehr schnelles Antwortverhalten zeigt, was ursprünglich als Ausdruck von Impulsivität zu werten wäre, hier jedoch durch die Itemschwierigkeit bedingt ist.

3.3. Planung und Durchführung

Um die primäre Fragestellung der vorliegenden Arbeit zu klären, ist es notwendig, dass sämtliche Testpersonen die hier verwendeten Verfahren zur Erfassung des logisch-schlussfolgernden Denkens sowohl im freien Antwortformat als auch im Multiple-Choice-Design bearbeiten. Der Untertest „Formale Folgerichtigkeit“, der Matrizentest sowie der Untertest aus dem PSB wurden allesamt im freien Antwortformat in einer Einzeltestung vorgegeben, während die Testung der Kinder mit den Instrumenten im Multiple-Choice-Format dabei aus ökonomischen Gründen in einem Gruppensetting erfolgte.

Wie bereits in Punkt 2.3 bezüglich der Verfahrensbeschreibung zu entnehmen war, ist vor allem der Untertest „Formale Folgerichtigkeit“ von Hagenmüller (2011) in der ursprünglichen Version dahingehend konzipiert, dass ein Testleiter die Antworten der Kinder auf einem separaten Antwortbogen verzeichnet, was eine Einzeltestung unabdingbar macht. Unter anderem aus diesem Grund wurde die Testung der Kinder zu zwei verschiedenen Terminen durchgeführt. Zu einem Testzeitpunkt wurden alle drei Verfahren im freien Antwortformat vorgegeben, und zu einem zweiten Termin wurden die gleichen Aufgaben von denselben Kindern im Multiple-Choice-Format bearbeitet. Dabei wurden die teilnehmenden

Schulklassen in zwei Hälften geteilt, wobei eine Hälfte die Testversionen zuerst im MC-Design bearbeitet hat und zu einem zweiten Termin im freien Antwortformat, und die zweite Hälfte in der umgekehrten Reihenfolge. So können Reihenfolgeeffekte bezüglich des Antwortdesigns ausgeschlossen werden. Um weiteren Effekten dieser Art vorzubeugen, wurde zudem die Reihenfolge der Testbearbeitung in der Einzel- sowie in der Gruppentestung pro Testperson randomisiert (siehe Anhang), sodass jedes Kind eine zufällige Vorgabe der drei Verfahren erhielt. Zwischen der Bearbeitung der Tests im Multiple-Choice-Design und im freien Antwortformat lag mindestens ein Zeitraum von sechs Wochen, sodass ein Erinnern an das Itemmaterial sowie Übungseffekte minimiert werden konnten. Die Testdauer betrug für beide Testtermine (Einzel- sowie Gruppentestung) jeweils eine Schulstunde mit etwa 50 bis 60 Minuten.

Die Testung im Rahmen der vorliegenden Diplomarbeit fand an einer Wiener Schule statt, die sich vorab bereit erklärt hat, an dieser Studie teilzunehmen. Vor dem Beginn der empirischen Arbeit wurde dieses Forschungsprojekt in einem detaillierten Bericht dem Wiener Stadtschulrat vorgelegt und nach ausführlicher Überprüfung bewilligt. Darüber hinaus wurde allen Eltern sowie Lehrkräften der teilnehmenden Schulklassen eine kurze Beschreibung der Studie vorgelegt sowie das Einverständnis für die Teilnahme der Kinder an der Untersuchung eingehoben. Im Rahmen der vorliegenden Arbeit wurden ausschließlich Kinder getestet, deren Eltern vorab ihr Einverständnis gegeben haben.

In einer kooperativen Diplomarbeit von Carina Heuschober (in Arbeit) zur weiterführenden Überprüfung der Validität des Untertests „Formale Folgerichtigkeit“ wurden darüber hinaus zusätzliche Daten im Sinne der hier untersuchten Fragestellungen erhoben und für die Auswertung verwendet. Dabei wurden ebenfalls die in Punkt 3.2 beschriebenen Verfahren, der Untertest „Formale Folgerichtigkeit“ sowie der PSB in beiden Testdesigns eingesetzt. Der Matrizentest wurde im Rahmen von Heuschobers Untersuchung nicht vorgegeben, was sich in einer verminderten Stichprobengröße bei der Datenerhebung mit diesem Verfahren niederschlägt. Die Testung erfolgte dabei ebenfalls an einer Wiener Schule mit erhobenem Einverständnis der Eltern für die getesteten Kinder und Genehmigung des Wiener Stadtschulrats. Das praktische Vorgehen bei der Datenerhebung erfolgte dabei in gleicher

Weise wie im Rahmen der durchgeführten Studie der vorliegenden Diplomarbeit, um eine Vergleichbarkeit gewährleisten zu können.

Für die Einzeltestungen standen in den jeweiligen Schulen eigene Räume zur Verfügung, die gänzlich für die Untersuchungen dieser Studie genutzt werden konnten, wobei in diesen meist zwei bis vier Testungen parallel stattgefunden haben. Die Kinder wurden einzeln oder in kleinen Gruppen von Testleitern aus der Klasse abgeholt und zu den Testräumen begleitet sowie nach der Testung wieder zurückgebracht. Die Gruppentestungen wurden von jeweils zwei bis drei Testleitern gemeinsam durchgeführt, um einen reibungslosen Ablauf zu garantieren und auf etwaige Fragen der Kinder besser eingehen zu können. Darüber hinaus befanden sich die jeweiligen Klassenlehrerinnen stets im Raum und konnten bei Bedarf Hilfestellungen geben.

Neben den oben beschriebenen Verfahren zur Erfassung der Reasoning Ability sowie des Arbeitsstils wurde zusätzlich u.a. zum Zweck demographischer Analysen das Alter sowie das Geschlecht der Kinder verzeichnet.

3.4. Stichprobe

Da für die statistische Überprüfung der aufgestellten Hypothesen Untersuchungsdaten aus einer kooperativen Studie verwendet wurden, muss hier die Stichprobe für den Matrizentest separat angeführt werden, da die Anzahl an untersuchten Personen mit diesem Verfahren deutlich geringer ist als bei den beiden anderen Instrumenten. So ergibt sich, dass insgesamt 164 Personen den Untertest „Formale Folgerichtigkeit“ sowie den PSB bearbeitet haben, wobei etwa die Hälfte der Untersuchungsdaten im Rahmen der Diplomarbeit von Heuschober (in Arbeit) gesammelt wurden. Von der Gesamtzahl an $N = 164$ Personen müssen aufgrund von teilweise fehlenden Daten sowie aufgrund von Ausfällen einige Testpersonen

von den Analysen ausgeschlossen werden. Für den Matrizentest ergibt sich eine Gesamtzahl von N= 80 Testpersonen, ausschließlich im Rahmen der vorliegenden Arbeit erhoben.

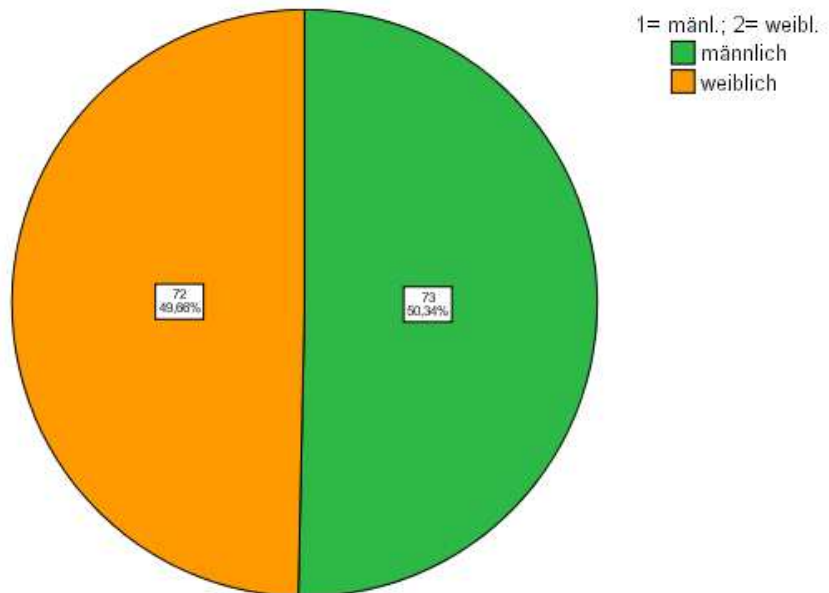


Abbildung 19: Geschlechterverteilung über alle Verfahren

Abbildung 19 zeigt die Verteilung des Geschlechts über alle Tests in dieser Studie einschließlich der Daten aus der Kooperationsstudie von Heuschober (in Arbeit). Dabei lässt sich gut erkennen, dass weibliche und männliche Testpersonen mit einer relativen Prozentzahl von 49,66% (weiblich) zu 50,34% (männlich) nahezu gleich häufig in dieser Studie vertreten sind.

Der Altersbereich der getesteten Kinder liegt in dieser Studie zwischen 6,1 und 9,9 Jahren, wobei das Alter von 8 Jahren mit einer Häufigkeit von 21 Testpersonen dabei am häufigsten vertreten ist (siehe Anhang). Abbildung 20 zeigt in Form eines Balkendiagramms die Verteilung des jeweiligen Alters der Kinder, wobei sich eine Normalverteilung der Daten

optisch gut erkennen lässt. Tabelle 1 zeigt mit einem Wert von 8,04 den Mittelwert sowie die Standardabweichung (SD= 0,84) des Alters.

Tabelle 1: Deskriptive Statistik des Alters über alle Verfahren

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Alter der Kinder	163	6,10	9,90	8,0448	,84347
Gültige Werte (Listenweise)	163				

Histogramm

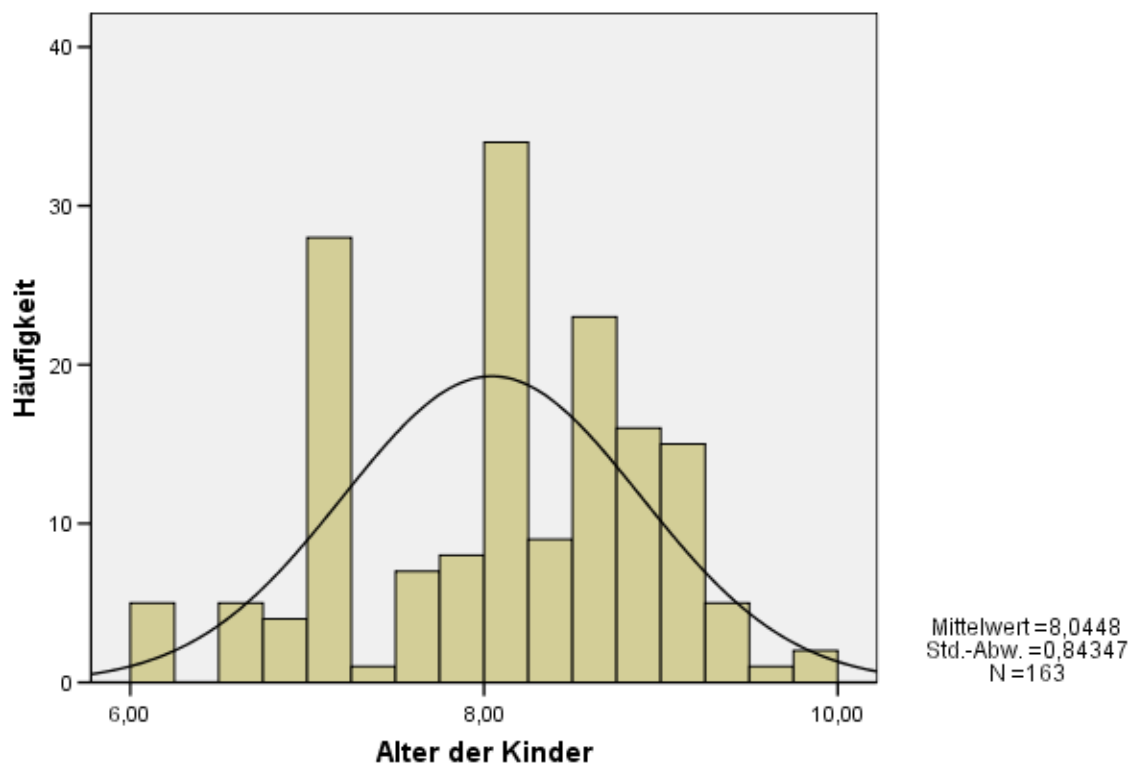


Abbildung 20: Grafische Darstellung der Altersverteilung über alle Verfahren

Abbildung 21 zeigt mit einem relativen Prozentsatz von 48,8% in Schule 1 zu 51,2% in Schule 2 die Verteilung der Testpersonen auf die beiden Wiener Schulen, die an der Testung teilgenommen haben. Dabei repräsentiert Schule 1 jene, an welcher im Rahmen der vorliegenden Arbeit getestet wurde, während die Kinder in Schule 2 im Rahmen der Diplomarbeit von Heuschober (in Arbeit) getestet wurden.

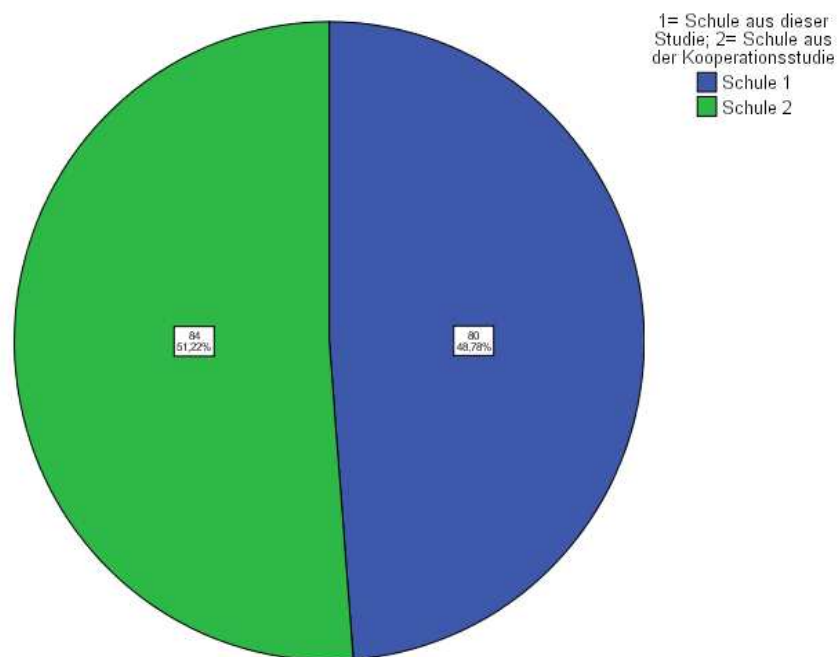


Abbildung 21: Verteilung der Testpersonen auf beide Schulen

Abbildung 22 zeigt die Verteilung der Arbeitsstile Impulsiv vs. Reflexiv in der erhobenen Stichprobe gemeinsam mit der neutralen Kategorie. Dabei lässt sich gut erkennen, dass ein Großteil der Testpersonen erwartungsgemäß mit einem neutralen Arbeitsverhalten klassifiziert wurde; in dieser Kategorie befinden sich mit 97 Kindern 61,39% der Gesamtstichprobe (graues Feld in Abbildung 22). Darüber hinaus zeigt sich, dass mit 30 Testpersonen, die als reflexiv eingestuft wurden (oranges Feld in Abbildung 22) zu 31

Kindern, deren Arbeitsstil als impulsiv beurteilt wurde (blaues Feld in Abbildung 22), nahezu eine Gleichverteilung hinsichtlich der relevanten Kategorien herrscht.

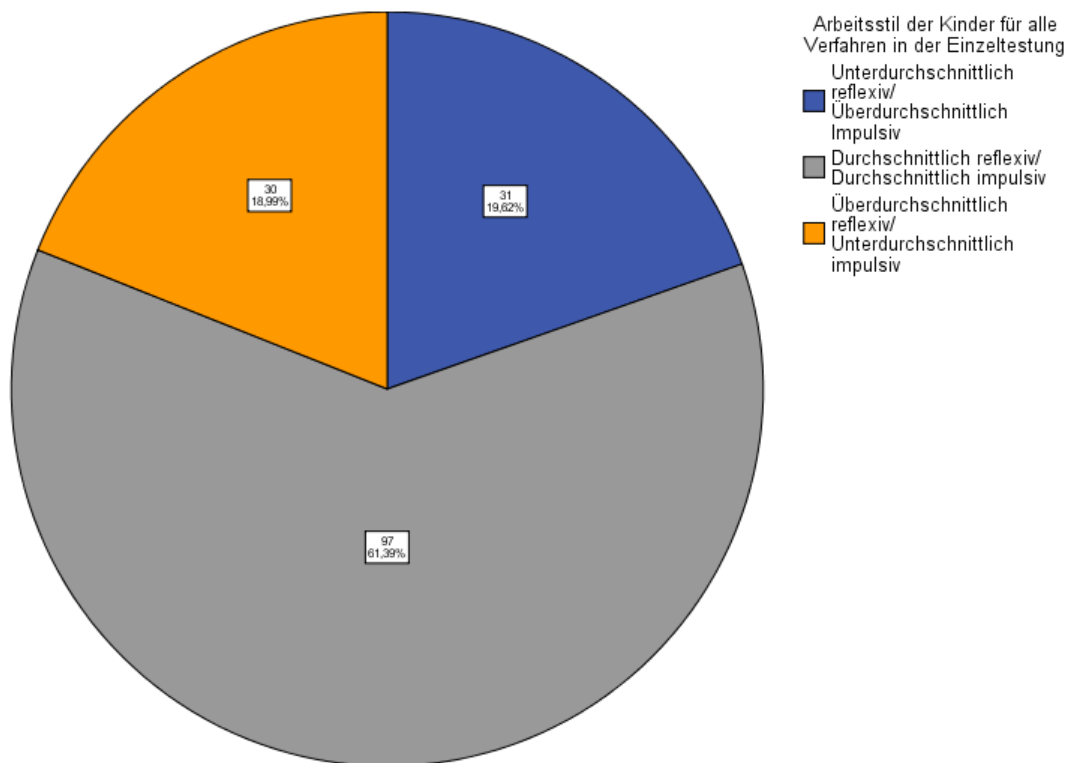


Abbildung 22: Verteilung des Arbeitsstils in der Stichprobe

3.5. Ergebnisse

Dieser Abschnitt widmet sich der Auswertung der gewonnenen Daten im Sinne der vorab aufgestellten Forschungshypothesen. Dabei soll zunächst auf die generelle Methodik der Auswertung eingegangen werden, um anschließend in einzelnen Punkten die spezifischen Ergebnisse genau darzulegen.

Um die Hypothesen aus Punkt 3.1.1 zu untersuchen, wurden neben einer ausführlichen deskriptiven Analyse des Datensatzes zunächst T-Tests für gepaarte Stichproben gerechnet, um etwaige globale Unterschiede in den Testleistungen je nach Antwortdesign zu erfassen. Zum Zwecke einer differenzierten Analyse wurde zusätzlich für jedes der drei Verfahren eine Prüfung auf Gültigkeit des Rasch-Modells durchgeführt. Dabei wurden dieselben Items pro Test, einmal im freien Antwortformat und einmal im Multiple-Choice-Format, als unterschiedliche Aufgaben aufgefasst und, in einem Test kombiniert, der Untersuchung unterzogen. Der Matrizentest beinhaltet in der Testversion mit freien Antworten ebenso wie im Multiple-Choice-Design jeweils 25 Aufgaben; zum Zwecke der Rasch-Modell-Analysen wurden die Antworten auf die Items beider Testversionen zu einem virtuellen Verfahren zusammengefasst, sodass aus den beiden Versionen des Matrizentests eine Version mit allen 50 Items entstand, die gemeinsam der Modellprüfung unterzogen wurden. Selbiges wurde auch für die Untertests des PSB sowie „Formale Folgerichtigkeit“ gemacht. Im Sinne der untersuchten Fragestellung wird dabei von der Annahme ausgegangen, dass die Items, je nach Darbietung in den beiden Antwortformaten, unterschiedlich schwierig sind. In diesem Sinne ist es nur logisch, die selbe Aufgabe, einmal im MC-Format und einmal im freien Antwortformat, als zwei verschiedene aufzufassen.

Die Rasch-Modell-Analysen der Verfahren mit den Items beider Antwortformate eignet sich besonders gut, um Auffälligkeiten einzelner Items in Abhängigkeit des Antwortdesigns zu identifizieren. Wenn Aufgaben in dieser Modellprüfung unauffällig sind, dann deutet dies auf keine Unterschiede bei der Testbearbeitung im Sinne der untersuchten Hypothesen hin. Entsprechen einzelne Items hingegen nicht der Rasch-Modell-Konformität, so wurde in einer weiterführenden inhaltlichen Analyse der Aufgaben nach erklärenden Ursachen für diese

Testergebnisse gesucht. Die Rasch-Modell-Analysen wurden dabei mit folgenden Teilungskriterien durchgeführt:

- **Median der Gesamtestleistung**
- **Alter**, geteilt nach dem Median
- **Geschlecht**
- **Arbeitsstil I:** Dabei wurden die Kategorien Reflexiv und Neutral zusammengefasst und dem Arbeitsstil Impulsiv gegenübergestellt.
- **Arbeitsstil II:** Hierbei wurden nur die beiden interessierenden Arbeitsstile einbezogen (Impulsiv vs. Reflexiv) und die neutrale Kategorie nicht berücksichtigt. Daraus ergibt sich eine relativ kleine Stichprobe.

In Kombination der drei Testverfahren sowie der Teilungskriterien ergibt sich eine Gesamtzahl von 15 Rasch-Modell-Analysen. Tabelle 2 veranschaulicht das Testdesign und gibt einen ersten Eindruck über die gewonnenen Ergebnisse. Jene Felder, die mit *R+* gekennzeichnet sind, beschreiben eine globale Passung des Rasch-Modells für diesen Test mit dem jeweiligen Teilungskriterium, während *R-* (rote Felder) keine Gültigkeit des Modells ausdrückt. Wie bereits einleitend erwähnt, sind besonders jene Rasch-Modell-Prüfungen interessant, bei welchen keine Passung des Modells gefunden wurde, da damit keine Eindimensionalität der Testung gegeben ist und somit eine Auswirkung des Antwortformats indiziert ist.

Tabelle 2: Darstellung des Testdesigns der Rasch-Modell-Analysen

	Teilungs- kriterium	PSB	Formale Folgerichtig- keit	Matrizen
Median	-	RM -	RM +	RM -
	+			
Sex	♂	RM +	RM +	RM -
	♀			
Alter	6 - 8,1	RM -	RM -	RM +
	8,2 – 9,9			
AS I	Reflexiv Neutral	RM +	RM +	RM +
	Impulsiv			
AS II	Reflexiv	RM +	RM -	RM +
	Impulsiv			

Im Anschluss an die Modellprüfung mittels Rasch-Modell-Analyse wurden zudem Wald-Tests zur Identifikation einzelner signifikanter Items, auch bei jenen Analysen, welche global der Gültigkeit entsprechen, durchgeführt. Aufgrund der Fülle an dabei gewonnenen Daten sowie im Hinblick auf deren Relevanz, beschränkt sich die Ergebnisdarstellung ausschließlich auf die nähere Erläuterung besonders wichtiger Ergebnisse, wobei besonderes Augenmerk auf die Analyse jener Rasch-Modell-Prüfungen gelegt wird, deren Ergebnisse in Tabelle 2 mit *R*-gekennzeichnet sind. Die Ergebnisse der näheren Analysen sollen in den nächsten Punkten näher beschrieben werden.

Sämtliche deskriptive Auswertungen, die durchgeführten T-Tests für gepaarte Stichproben samt Prüfung der Voraussetzungen sowie die Analysen der relativen Lösungshäufigkeiten

wurden mit der Statistiksoftware SPSS für Windows in der Version 14.0.1 durchgeführt und berechnet. Sämtliche Rasch-Modell-Analysen der vorliegenden Arbeit sowie weiterführende Modellprüfungen wurden mit der freien Statistiksoftware *R* in der Version 2.12.1 durchgeführt, wobei speziell das für *R* programmierte Zusatzpaket *eRm* in der Version 0.13-0 Anwendung fand (Poinstingl, Mair, Hatzinger, 2007; Mair, Hatzinger, Maier, 2011).

Im Rahmen der vorliegenden Diplomarbeit konnte für den Matrizentest mit $N = 80$ nur eine relativ geringe Anzahl an Testpersonen gewonnen werden. Unter diesen Umständen wurde hier auf eine Darstellung der Interpretation der Rasch-Modell-Analysen dieses Verfahrens verzichtet, da die gewonnenen Ergebnisse unter diesen Gesichtspunkten nur einen geringen Erklärungswert hätten. Im Zuge weiterführender Forschungen wäre hier eine größere Stichprobengröße wünschenswert.

3.5.1. Generelle Unterschiede in der Bearbeitung des Multiple-Choice-Designs und des freien Antwortformats

In einem ersten Schritt bei der Auswertung der gesammelten Daten soll die primäre Forschungshypothese der vorliegenden Arbeit untersucht werden, die sich der Frage widmet, ob es Unterschiede in den Testleistungen bei der Bearbeitung von Aufgaben zur Erfassung des logisch-schlussfolgernden Denkens gibt, je nach Vorgabe der Items in verschiedenen Antwortformaten. Die formulierten Hypothesen der H 1 legen dabei das Hauptaugenmerk auf die Unterschiede zwischen den jeweiligen Multiple-Choice-Versionen und den Testformen im freien Antwortformat eines jeden Verfahrens. Eine Testung auf Unterschiede aller Ergebnisse, zusammengefasst in einem Wert pro Testperson, scheint dabei wenig sinnvoll; zum einen aus dem Grund, dass nur die Hälfte der Testpersonen den Matrizentest bearbeitet hat, zum anderen wegen der Tatsache, dass der Untertest „Formale

Folgerichtigkeit“ aus dem AID 3 aus zwei unterschiedlichen Testversionen (getrennt nach dem Alter) besteht. Aus diesen Überlegungen ergibt sich, dass für die jeweiligen Testversionen eine unterschiedliche Anzahl an Testpersonen entsteht und dadurch ein kombinierter Vergleich der Werte deutlich an Aussagekraft verlieren würde.

3.5.1.1. T-Test für gepaarte Stichproben

Zur globalen Analyse der Ergebnisse wurde zunächst ein T-Test bei gepaarten Stichproben gerechnet, um etwaige Unterschiede in der Testleistung zwischen den beiden Antwortformaten eines jeden Verfahrens zu erfassen. Die Voraussetzung der Normalverteilung der Differenz zwischen den Rohwerten eines jeden Verfahrens im MC-Design sowie im freien Antwortformat wurde einerseits mittels grafischer Modellkontrolle sowie mittels Kolmogorov-Smirnov-Test (K-S-Test) überprüft und kann für alle Instrumente als erfüllt angesehen werden (siehe Anhang).

Tabelle 3 und Tabelle 4 zeigen den SPSS-Output der Testung auf Unterschiede zwischen den Rohwerten des Matrizentests im freien Antwortformat und im Multiple-Choice-Format. Die zweiseitige Signifikanzprüfung zeigt hier mit einem Wert von $p = 0,005$ bei einem Alphaniveau von 5% einen signifikanten Unterschied in der Testleistung zwischen den beiden Testformen des Matrizentests. Im Sinne einer einseitigen Formulierung der H 1.1 ergibt sich hier mit $p = 0,0025$ ebenfalls ein signifikantes Ergebnis bei einer einseitigen Prüfung auf Unterschiede. Das bedeutet, dass generelle Unterschiede der Leistungen im Matrizentest je nach bearbeitetem Antwortformat in dieser Stichprobe anzunehmen sind. Unter dem Gesichtspunkt der vorab aufgestellten Annahme, dass die Testleistung im MC-Design aufgrund von erhöhten Anforderungen an generelle Kontrollstrategien bei der Aufgabenbearbeitung geringer sein müsste als bei der Bearbeitung von selbigen Items im freien Antwortdesign, ist eine Betrachtung der Rohscore-Mittelwerte der jeweiligen Testform in Tabelle 3 überraschend. Dabei zeigt sich, entgegen den Erwartungen, mit einer

Differenz von 0,961 eine leicht erhöhte Testleistung in der Version mit dem Multiple-Choice-Antwortformat. Die hier festgestellte Signifikanz der Unterschiede zwischen den beiden Testformen des Matrizentests wird allerdings durch das Einbeziehen der Standardabweichung relativiert. Die Standardabweichung beträgt hier 2,91 Rohwerte, wohingegen die Mittelwertdifferenz mit 0,961 Rohwertpunkten nicht einmal der Hälfte einer Standardabweichung entspricht und somit aus inhaltlichen Überlegungen als kaum bedeutend angesehen werden muss.

Tabelle 3: SPSS-Output der Statistik bei gepaarten Stichproben für den Matrizentest

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_Matrizen_frei	17,0519	77	3,00830	,34283
	RW_Matrizen_MC	18,0130	77	3,04999	,34758

Tabelle 4: SPSS-Output des T-Tests für gepaarte Stichproben für den Matrizentest

Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz			
					Untere	Obere		
Paaren 1	RW_Matrizen_frei - RW_Matrizen_MC	-,96104	2,91295	,33196	-1,62220	-,29988	-2,895	,005

Die Tabellen 5 und 6 zeigen den SPSS-Output des T-Tests für gepaarte Stichproben für die korrigierte Version des PSB. Die Korrektur der ursprünglichen Testversionen ist darauf zurückzuführen, dass Item 35 aufgrund von einer sehr großen Anzahl an fehlenden Werten in der Multiple-Choice-Version nicht in die Testung miteinbezogen werden konnte und daher in beiden Antwortdesigns aus der Auswertung ausgeschlossen wurde.

Tabelle 5: SPSS-Output der Statistik bei gepaarten Stichproben für den PSB (korrigierte Version)

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	SMEAN(RW_PSB_MC34)	24,2236	164	4,73205	,36951
	SMEAN(RW_PSB_frei34)	24,1835	164	4,06819	,31767

Tabelle 6: SPSS-Output des T-Tests für gepaarte Stichproben für den PSB (korrigierte Version)

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paare n 1	SMEAN(RW_PSB_MC34) – SMEAN(RW_PSB_frei34)	,04006	4,15171	,32419	-,60010	,68022	,124	163	,902

Die Testung auf Unterschiede mittels T-Test kommt bei einem Alphaniveau von 5% mit einem Wert von $p = 0,902$ zu keinem signifikanten Ergebnis bei einer zweiseitigen Prüfung. Im Sinne der einseitig formulierten Hypothese H.1.b.1 ist das Ergebnis einseitig zu betrachten, wobei mit einem Wert von $p = 0,451$ ebenfalls keine signifikanten Unterschiede zwischen den Testversionen des PSB im freien Antwortformat und im Multiple-Choice-Design gefunden wurden. Die Mittelwertdifferenz von 0,040 in Tabelle 6 fällt in Bezug auf die errechnete Standardabweichung von $SD = 4,152$ zudem äußerst gering aus. Aufgrund dieser Analysen kann davon ausgegangen werden, dass die Wahl des Antwortformats nur einen unbedeutenden Einfluss auf die Testleistung des PSB hat.

Um die Fragestellung der Hypothese H 1.c zu beantworten, ist es aufgrund der Testkonzeption der „Formalen Folgerichtigkeit“ hinsichtlich einer Teilung der Testversionen nach dem Alter notwendig, diese auch getrennt nach der Version zu betrachten. Die Tabellen 7 und 8 zeigen den SPSS-Output des T-Tests für gepaarte Stichproben samt Statistik der Testversionen für 6- bis 7-jährige Kinder des UT „Formale Folgerichtigkeit“ aus dem AID 3. Beim Betrachten der Mittelwertdifferenzen zwischen den beiden Testversionen zeigt sich mit einer Differenz von 0,463 Rohwerten ein leicht erhöhter Mittelwert in der Multiple-Choice-Version gegenüber der Konzeption mit freiem Antwortformat, allerdings bei einer Standardabweichung von $SD = 4,479$. Tabelle 8 stellt die zweiseitige Prüfung auf Unterschiede zwischen den Testversionen bei einem üblichen Alphaniveau von 5% dar und zeigt dabei ein nicht signifikantes Ergebnis ($p = 0,451$). Auch im Hinblick auf eine einseitige Prüfung im Sinne der formulierten Hypothese H 1.c.1 zeigt sich mit einem Wert von $p = 0,226$ ein nicht signifikantes Ergebnis, was inhaltlich gesehen auf keine relevanten Unterschiede in der globalen Testleistung, je nach bearbeitetem Antwortformat, hindeutet.

Tabelle 7: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6 bis 7 J.

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_LB_frei	12,2593	54	4,25235	,57867
	RW_LB_MC	12,7222	54	4,40161	,59898

Tabelle 8: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 6 bis 7 J.

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paare n 1	RW_LB_frei - RW_LB_MC	-,46296	4,47936	,60956	-1,68559	,75967	-,759	53	,451

Die Tabellen 9 und 10 zeigen das Pendant zu der Prüfung auf Unterschiede mittels T-Test für gepaarte Stichproben in den Tabellen 7 und 8, hier jedoch für die entsprechenden Testversionen des Altersbereichs 8- bis 11-jähriger Kinder. Dabei zeigt sich bei einer zweiseitigen Prüfung mit einem Wert von $p = 0,267$ kein signifikanter Unterschied zwischen den Testversionen in den beiden Antwortformaten; auch hier wurde ein Konfidenzintervall von 95% verwendet. Eine Umformung des Ergebnisses im Hinblick auf die einseitige Fragestellung ergibt mit einem Wert von $p = 0,136$ ebenfalls ein nicht signifikantes Ergebnis. Genau wie in der Testversion der jüngeren Kinder deutet dies darauf hin, dass Bearbeitungen der Aufgaben in den beiden Antwortformaten nicht maßgeblich voneinander differieren.

Tabelle 9: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_LB_frei	11,8300	100	3,30917	,33092
	RW_LB_MC	12,2400	100	3,82475	,38247

Tabelle 10: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
		Paare n 1	RW_LB_frei - RW_LB_MC	-,41000	3,67381	,36738	-1,13896	,31896	-1,116

3.5.1.2. Rasch-Modell-Analysen

Zum Zwecke einer differenzierteren Analyse der Testergebnisse wurde neben einer Prüfung auf Mittelwertunterschiede zwischen den jeweiligen Testversionen in den beiden Antwortformaten zudem eine Prüfung auf Gültigkeit des Rasch-Modells durchgeführt. Dabei wurden alle Items eines Verfahrens, unabhängig vom Antwortformat, in einer Testversion summiert und gemeinsam einer Analyse unterzogen. Die Ergebnisse dieser Analysen sollen im Folgenden näher erläutert werden.

Tabelle 11 zeigt den R-Output des Andersen Likelihood Ratio-Tests (LR-Test) zur Überprüfung der Gültigkeit des Rasch-Modells für alle Items des PSB kombiniert, mit dem Median der Gesamttestleistung als Teilungskriterium. Inhaltlich kann dies dahingehend beschrieben

werden, dass die Testpersonen, die sich mit ihrer Testleistung unterhalb des Medians befinden, der Einfachheit wegen als Personen mit einer gering ausgeprägten Fähigkeit zum logisch-schlussfolgernden Denken bezeichnet werden können. Personen, deren Testscore über dem Median liegt, können hingegen dahingehend beschrieben werden, dass sie eine hohe oder zumindest höher ausgeprägte Fähigkeit zum logisch-schlussfolgernden Denken haben. Hier wurde, nach dem Ausscheiden von unvollständigen Daten, mit einer Stichprobengröße von $N = 155$ gerechnet, wobei der Median der Testleistung hier bei einem Rohscore von 48 liegt. Dabei musste, ebenso wie bei den Unterschiedsprüfungen mittels T-Test, Item 35 aufgrund von unvollständigen Daten in einer Testversion aus den Analysen ausgeschlossen werden. Die Rasch-Modell-Konformität ist für den gegebenen Datensatz mit dem Median der Testleistung als Teilungskriterium mit $p = 0$ bei einem Alphaniveau von 5% nicht gegeben. Das bedeutet, dass sich die Erfassung des logisch-schlussfolgernden Denkens durch die Gesamtheit der Aufgaben des PSB für die beiden untersuchten Gruppen (geteilt durch den Median der Testleistung) maßgeblich unterscheidet.

Tabelle 11: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Median der Testleistung

N= 155	Median = 48
--------	-------------

Andersen LR-test:	LR-value: 102.812	Chi-square df: 54	p-value: 0
--------------------------	-------------------	-------------------	-------------------

Um eine genauere inhaltliche Analyse der gewonnenen Ergebnisse durchführen zu können, wurde im Anschluss an die Prüfung auf Gültigkeit des Rasch-Modells ein Wald-Test zur Identifikation jener Aufgaben durchgeführt, die nicht dem Rasch-Modell entsprechen. Tabelle 12 zeigt die Ergebnisse des Wald-Tests für den PSB mit dem Median der Testleistung als Teilungskriterium. Jene Items, deren p-Werte die kritische Grenze von 0,05

unterschreiten, sind dabei zur besseren Erkennung hervorgehoben. Dabei ist auffällig, dass nur eine Aufgabe aus der Testversion mit freiem Antwortformat den kritischen p-Wert unterschreitet, während bei der Version im Multiple-Choice-Design sieben Werte signifikant sind.

Beim näheren Betrachten der Aufgaben fällt auf, dass Item 30 aus der Testversion mit freier Antwort mit $p = 0,000$ einen besonders niedrigen Wert aufweist. Eine inhaltliche Analyse dieser Aufgabe legt nahe, dass dieser Wert mit hoher Wahrscheinlichkeit durch eine ungünstige Itemkonstruktion zustande kommt. Dabei ist die Abgrenzung der einzelnen Einheiten kaum oder nur schwer ersichtlich. Das ist vor allem in der Testversion mit freier Antwortkonzeption problematisch, da man bei diesem Antwortformat stärker darauf angewiesen ist, selbstständig die einzelnen Elemente zu erkennen. In der MC-Version hingegen ist das selbstständige Erfassen der einzelnen Elemente weniger gewichtig, da man durch die Antwortmöglichkeiten bereits die Abgrenzung der Elemente vorgegeben hat oder diese zumindest leichter erkennen kann. Diese Aufgabe ist in drei von fünf durchgeführten Rasch-Modell-Analysen mit unterschiedlichen Teilungskriterien signifikant, was die oben getätigten Annahmen bezüglich einer mangelnden Konstruktion bestätigt. Unter diesen Überlegungen kann Item 30 aus weiteren inhaltlichen Analysen ausgeschlossen werden.

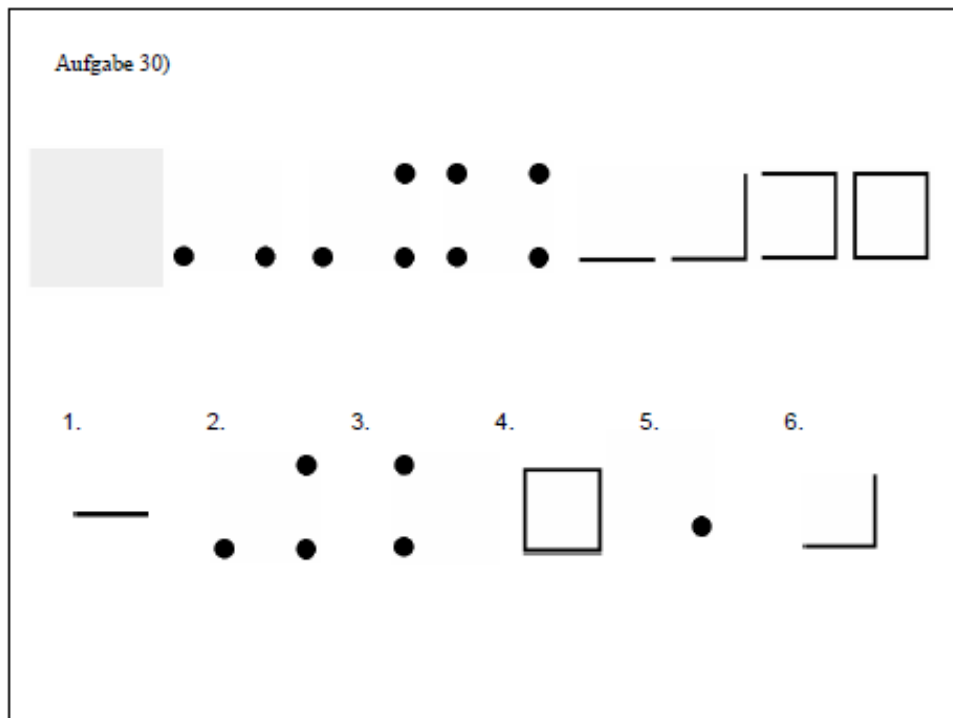


Abbildung 23: Aufgabe 30 aus der adaptierten Testversion des PSB mit Multiple-Choice-Antwortformat

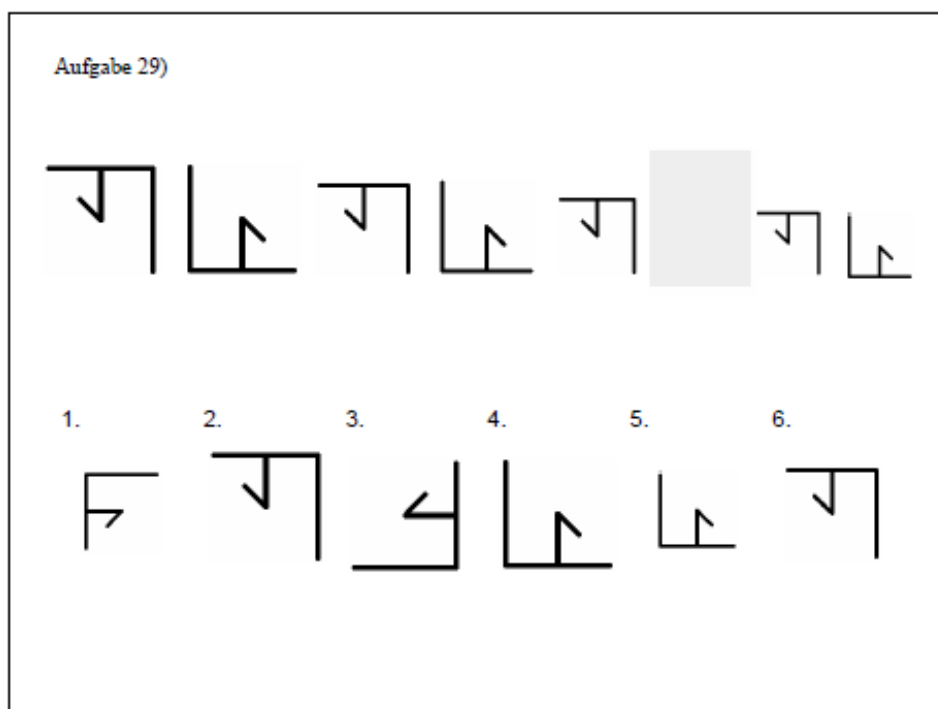


Abbildung 24: Aufgabe 29 aus der adaptierten Testversion des PSB mit Multiple-Choice-Antwortformat

Unter dem Gesichtspunkt der Verrechnungsfairness ist zudem die hier im Wald-Test signifikante Aufgabe 29 kritisch zu betrachten. Das Problem der Verrechnungsfairness bei der Testversion mit freiem Antwortformat liegt vor allem darin, dass sich die Figuren, bezogen auf die Größe, nicht sehr stark unterscheiden. Zwar ist bei einer differenzierten Betrachtung der Figuren ein unterschiedliches Größenverhältnis kaum zu übersehen, jedoch ist es bedeutend schwieriger, die richtige Lösung in der passenden Größe zu zeichnen. Unter Berücksichtigung des Alters der Kinder kommt diesem Argument größere Bedeutung zu. Gerade in diesem Alter ist bei vielen Kindern die nötige Feinmotorik für ein genaues Zeichnen noch nicht sehr gut ausgeprägt. Bei der Auswertung des freien Antwortformats bleibt es schlussendlich der auswertenden Person überlassen, wo sie die Grenze für ein richtig gezeichnetes Item zieht. Bei den MC-Antwortmöglichkeiten gibt es zwei Figuren, die formal und hinsichtlich der Ausrichtung identisch sind. Die richtige Antwort zeichnet sich hier durch eine geringere Größe aus. Um die Aufgabe zu lösen, ist es also nötig, sowohl die richtige Ausrichtung als auch die passende Größe der Figur zu erkennen und die entsprechende Antwort anzukreuzen. Während bei diesem Item in der MC-Version eine genaue und differenzierte Betrachtung der Größenverhältnisse von entscheidender Bedeutung ist, kann unter Umständen bei der Version mit freiem Antwortformat eine gegebene Lösung als richtig interpretiert werden, die von der Testperson falsch angedacht wurde, weil tolerant oder schlicht falsch interpretiert und ausgewertet wurde.

Tabelle 12: R-Output des Wald-Tests für den PSB; Teilungskriterium: Median der Testleistung

Wald test on item level (z-values):		
Items	z-statistic	p-value
beta PSB_frei_2	1.163	0.245
beta PSB_frei_4	1.162	0.245
beta PSB_frei_7	-0.016	0.987
beta PSB_frei_8	-0.777	0.437
beta PSB_frei_11	0.545	0.586
beta PSB_frei_13	-0.323	0.747
beta PSB_frei_16	-1.456	0.145
beta PSB_frei_17	-0.606	0.544
beta PSB_frei_19	-1.456	0.145
beta PSB_frei_20	0.251	0.801
beta PSB_frei_21	-0.321	0.748
beta PSB_frei_22	-0.950	0.342
beta PSB_frei_23	-1.249	0.212
beta PSB_frei_24	-1.694	0.090
beta PSB_frei_25	1.008	0.314
beta PSB_frei_26	-1.027	0.304
beta PSB_frei_27	0.279	0.780
beta PSB_frei_28	0.645	0.519
beta PSB_frei_29	-1.572	0.116
beta PSB_frei_30	3.514	0.000
beta PSB_frei_31	-0.905	0.366
beta PSB_frei_32	-1.125	0.261
beta PSB_frei_33	-1.086	0.278
beta PSB_frei_34	1.323	0.186
beta PSB_MC_2	1.623	0.105
beta PSB_MC_3	0.167	0.867
beta PSB_MC_6	-0.051	0.959
Items	z-statistic	p-value
beta PSB_MC_7	2.247	0.025
beta PSB_MC_8	-0.143	0.887
beta PSB_MC_9	2.248	0.025
beta PSB_MC_10	-1.024	0.306
beta PSB_MC_11	-0.058	0.954
beta PSB_MC_12	1.625	0.104
beta PSB_MC_13	2.883	0.004
beta PSB_MC_14	0.429	0.668
beta PSB_MC_15	-0.837	0.403
beta PSB_MC_16	-2.053	0.040
beta PSB_MC_17	1.304	0.192
beta PSB_MC_18	1.323	0.186
beta PSB_MC_19	0.322	0.747
beta PSB_MC_20	0.269	0.788
beta PSB_MC_21	-0.467	0.641
beta PSB_MC_22	-0.365	0.715
beta PSB_MC_23	-1.261	0.207
beta PSB_MC_24	-3.298	0.001
beta PSB_MC_25	-0.813	0.416
beta PSB_MC_26	0.322	0.747
beta PSB_MC_27	1.292	0.196
beta PSB_MC_28	-1.825	0.068
beta PSB_MC_29	-2.053	0.040
beta PSB_MC_30	-0.520	0.603
beta PSB_MC_31	-2.769	0.006
beta PSB_MC_32	-0.931	0.352
beta PSB_MC_33	-0.068	0.946
beta PSB_MC_34	0.322	0.747

Lässt man diese beiden Items bei der Interpretation außen vor, so fällt auf, dass von den übrigen 6 Aufgaben zumindest 5 in den global beobachteten Lösungen eine Tendenz dahingehend zeigen, dass sie öfter in der Testversion mit freiem Antwortformat gelöst wurden als in der Version im Multiple-Choice-Design. Betrachtet man die

Lösungshäufigkeiten geteilt nach dem Median der Testleistung, so zeigen sich widersprüchliche Ergebnisse zwischen den beiden Vergleichsgruppen. Einige Items zeigen eine Zunahme der Lösungen bei der Gruppe oberhalb des Medians in der Testversion mit freiem Antwortformat, während andere Items bei der gleichen Gruppe eine Zunahme der richtigen Antworten in der MC-Version zeigen.

Die gewonnenen Ergebnisse dieser Analysen scheinen insgesamt für die a priori aufgestellten Annahmen der vorliegenden Arbeit, insbesondere aber für die Hypothese H 1.b.1 bezüglich der Leistungsunterschiede des PSB in den beiden Antwortformaten zu sprechen. So zeigt sich bei der inhaltlichen Analyse der signifikanten Items aus dem Wald-Test ein klarer Trend dahingehend, dass diese öfter im freien Antwortformat gelöst wurden als in der Multiple-Choice-Konzeption, wobei für diese Ergebnisse kein anderer inhaltlicher Grund als die verschiedenen Antwortformate gefunden werden konnte.

Tabelle 13 zeigt den Output der Prüfung auf Rasch-Modell-Konformität des PSB mittels Andersen LR-Test mit dem Geschlecht als Teilungskriterium. Nach dem Entfernen von unvollständigen Daten erhält man hier bei einer Stichprobe von $N = 138$ ein $p = 0,243$ und somit eine Gültigkeit des Rasch-Modells für beide Geschlechter. Somit ist aus inhaltlicher Sicht kein maßgeblicher Unterschied in der Erfassung des logischen Schlussfolgerns in beiden Gruppen, hier definiert durch das Geschlecht, anzunehmen.

Tabelle 13: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Geschlecht

N= 138

Andersen LR-test:	LR-value: 62.987	Chi-square df: 56	p-value: 0.243
--------------------------	------------------	-------------------	-----------------------

Die Ergebnisse des LR-Tests der Rasch-Modell-Analyse für das Verfahren „Formale Folgerichtigkeit“ aus dem AID 3, ebenfalls mit dem Median der Testleistung als Teilungskriterium, sind in Tabelle 14 wiedergegeben. Dabei zeigt sich mit einem $p = 0,287$ eine Gültigkeit des Rasch-Modells für den gegebenen Datensatz für die beiden Gruppen, geteilt durch den Median der Testleistung bei einer Stichprobe von $N = 164$. Anders als bei den gerechneten T-Tests für gepaarte Stichproben ist es hier möglich, eine Prüfung auf Gültigkeit des Rasch-Modells für alle Aufgaben gemeinsam, und nicht separat für die Testversionen des jeweiligen Alters, durchzuführen.

Durch die Gültigkeit des Rasch-Modells ist auch hier anzunehmen, dass die Erfassung des logisch-schlussfolgernden Denkens durch das Verfahren „Formale Folgerichtigkeit“ nicht maßgeblich in den beiden Gruppen (die Stichprobe geteilt durch den Median der Testleistung) differiert. In weiterer Folge ist für diese beiden Gruppen kein Unterschied in der Erfassung der Reasoning-Fähigkeit durch die Antwortformate Multiple-Choice und freie Antwort dieses Verfahrens anzunehmen.

Tabelle 14: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Median der Testleistung

N= 164	Median = 25
--------	-------------

Andersen LR-test:	LR-value: 59.354	Chi-square df: 54	p-value: 0.287
--------------------------	------------------	-------------------	-----------------------

Tabelle 15 zeigt die Ergebnisse der Rasch-Modell-Analyse der „Formalen Folgerichtigkeit“ mittels LR-Test und dem Geschlecht als Teilungskriterium. Dabei zeigt sich mit einem $p = 0,076$ eine Gültigkeit des Rasch-Modells für die Gesamtheit der Items in beiden Antwortformaten. Inhaltlich gesehen, ist somit anzunehmen, dass die Gesamtheit der

Aufgaben aus dem Verfahren „Formale Folgerichtigkeit“ in beiden Antwortformaten Rasch-Modell-konform für beide Geschlechter dieser Stichprobe messen.

Tabelle 15: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Geschlecht

N= 145

Andersen LR-test:	LR-value: 74.018	Chi-square df: 58	p-value: 0.076
-------------------	------------------	-------------------	----------------

3.5.2. Unterschiede in der Bearbeitung des Multiple-Choice-Designs und des freien Antwortformats unter Einbezug des Alters

In einem zweiten Auswertungsschritt sollen die gesammelten Daten hinsichtlich der erweiterten Fragestellung der Forschungshypothese H 2 samt Unterhypothesen näher untersucht werden. Dabei liegt das Hauptaugenmerk der aufgestellten Hypothesen auf der zusätzlichen Erfassung des Einflusses des Alters auf die Bearbeitung von Aufgaben zur Erfassung des logisch-schlussfolgernden Denkens in den Antwortformaten Multiple-Choice sowie freie Antwort. Unter dem Gesichtspunkt der vorherrschenden Literatur über die relative Bedeutung von generellen Kontrollstrategien bei der Bearbeitung solcher Aufgaben sowie im Bezug auf die stetige kognitive Entwicklung von Kindern in diesem Alter wird hier von der Annahme ausgegangen, dass besonders jüngere Kinder erhöhte Schwierigkeiten mit dem Antwortdesign Multiple-Choice haben müssten. Die Ergebnisse der durchgeführten Analysen zur Beantwortung der obigen Fragestellung sollen im weiteren Verlauf näher erläutert werden.

3.5.2.1. Rasch-Modell-Analysen

Um den Einfluss des Alters auf die Bearbeitung der Aufgaben in den beiden Antwortformaten erfassen zu können, wurden Rasch-Modell-Analysen für alle Verfahren mit dem Alter als Teilungskriterium durchgeführt. Dabei wurden wieder die Items im freien Antwortformat und im Multiple-Choice-Design als unterschiedliche Aufgaben aufgefasst und für die Testung in einem Verfahren kombiniert. Die Teilung der erhobenen Daten in zwei Gruppen erfolgt hier durch den Median des Alters. Inhaltlich bedeutet dies, dass sich in der Gruppe der Testpersonen unterhalb des Medians die, relativ an der erhobenen Stichprobe gesehen, jüngeren Kinder befinden, während sich die Gruppe oberhalb des Medians durch ein relativ höheres Alter auszeichnet. Zum Zwecke einer leichteren Veranschaulichung wird im Folgenden von jüngeren vs. älteren Kindern gesprochen, wobei sich diese Klassifikation auf das relative Alter der Testpersonen bezieht und damit die Angehörigkeit zur jeweiligen Gruppe oberhalb oder unterhalb des Medians bezeichnet.

Tabelle 16 zeigt die Ergebnisse der Prüfung auf Rasch-Modell-Konformität des PSB mit dem Alter als Teilungskriterium. Der Median des Alters liegt hier bei 8,1 Jahren; Kinder, deren Alter unterhalb dieses Werts liegt, werden der ersten Gruppe zugeordnet, während Kinder mit einem Alter über dem kritischen Wert in eine zweite Gruppe klassifiziert werden. Die Prüfung mittels LR-Test für die beiden Gruppen fällt hier bei einer Stichprobe von $N = 155$ mit einem Wert von $p = 0$ signifikant aus, sodass keine Rasch-Modell-Gültigkeit für die Gesamtheit der Items angenommen werden kann. Somit scheint die Erfassung des logischen Schlussfolgerns durch die Items des PSB in den beiden Altersgruppen nicht gleich zu erfolgen, sondern vielmehr im Bezug auf gruppenspezifische Faktoren zu differieren.

Tabelle 16: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Median des Alters

N= 155	Median = 8,1 J.
--------	-----------------

Andersen LR-test:	LR-value: 99.87	Chi-square df: 57	p-value: 0
--------------------------	-----------------	-------------------	-------------------

In einem weiteren Auswertungsschritt wurde auch hier ein Wald-Test zur Identifikation einzelner Items, die nicht dem Rasch-Modell entsprechen, durchgeführt. Tabelle 17 zeigt den R-Output des Wald-Tests für den PSB mit dem Teilungskriterium des Alters; dabei sind jene Items, deren p-Werte die kritische Grenze von 0,05 unterschreiten, zur besseren Erkennung hervorgehoben.

Tabelle 17: R-Output des Wald-Tests für den PSB; Teilungskriterium: Median des Alters

Wald test on item level (z-values):

Items	z-statistic	p-value
beta PSB_frei_2	-0.478	0.632
beta PSB_frei_4	-0.478	0.633
beta PSB_frei_6	-0.478	0.633
beta PSB_frei_7	-0.030	0.976
beta PSB_frei_8	1.273	0.203
beta PSB_frei_11	0.640	0.522
beta PSB_frei_13	-0.171	0.864
beta PSB_frei_14	0.037	0.971
beta PSB_frei_15	0.411	0.681
beta PSB_frei_16	0.362	0.717
beta PSB_frei_17	-0.299	0.765
beta PSB_frei_19	1.612	0.107
beta PSB_frei_20	-0.853	0.393
beta PSB_frei_21	-2.649	0.008
beta PSB_frei_22	0.498	0.619
beta PSB_frei_23	0.910	0.363
beta PSB_frei_24	1.366	0.172
beta PSB_frei_25	-0.341	0.733
beta PSB_frei_26	-0.317	0.751
beta PSB_frei_27	-2.808	0.005
beta PSB_frei_28	-1.169	0.243
beta PSB_frei_29	-0.095	0.925
(beta PSB_frei_30)	-2.745	0.006
beta PSB_frei_31	2.324	0.020
beta PSB_frei_32	0.384	0.701
beta PSB_frei_33	-1.095	0.273
beta PSB_frei_34	0.299	0.765
beta PSB_MC_2	-0.662	0.508
beta PSB_MC_3	0.704	0.482
beta PSB_MC_6	0.944	0.345

Items	z-statistic	p-value
beta PSB_MC_7	-1.182	0.237
beta PSB_MC_8	0.462	0.644
beta PSB_MC_9	-1.339	0.180
beta PSB_MC_10	-0.171	0.864
beta PSB_MC_11	0.507	0.612
beta PSB_MC_12	-1.686	0.092
beta PSB_MC_13	-1.339	0.180
beta PSB_MC_14	-1.565	0.118
beta PSB_MC_15	0.711	0.477
beta PSB_MC_16	2.420	0.016
beta PSB_MC_17	-0.635	0.526
beta PSB_MC_18	-0.382	0.702
beta PSB_MC_19	-1.996	0.046
beta PSB_MC_20	1.594	0.111
beta PSB_MC_21	-0.821	0.412
beta PSB_MC_22	2.104	0.035
beta PSB_MC_23	1.281	0.200
beta PSB_MC_24	1.977	0.048
beta PSB_MC_25	1.603	0.109
beta PSB_MC_26	-1.280	0.201
beta PSB_MC_27	-0.124	0.902
beta PSB_MC_28	0.668	0.504
beta PSB_MC_29	-0.291	0.771
beta PSB_MC_30	0.844	0.399
beta PSB_MC_31	1.158	0.247
beta PSB_MC_32	1.788	0.074
beta PSB_MC_33	2.636	0.008
beta PSB_MC_34	1.972	0.049

Aufgrund der hohen Anzahl an Items mit signifikanten p-Werten wird hier von einer detaillierten Besprechung der einzelnen Items abgesehen. Auffällig beim Betrachten der Ergebnisse ist, dass insgesamt drei Aufgaben aus der Testversion mit freiem Antwortformat mit dem Median des Alters als Teilungskriterium nicht dem Rasch-Modell entsprechen (Item

21, 27 und 31), während im Gegenzug sechs Items aus der Version mit Multiple-Choice-Antwortdesign einen signifikanten Wert aufweisen (Item 16, 19, 22, 24, 33 und 34). Auch hier kann Aufgabe 30 von der Interpretation der Ergebnisse aufgrund von schlechter Konstruktion (siehe Punkt 3.5.1.1) ausgeschlossen werden. Bei einem Betrachten der relativen Lösungshäufigkeiten der Aufgaben in den jeweiligen Gruppen, geteilt durch den Median des Alters, fällt weiters auf, dass die Kinder der Altersgruppe über dem Median in fünf der insgesamt neun auffälligen Items eine Zunahme der richtigen Lösungen in der MC-Version gegenüber den Kindern in der Altersgruppe unterhalb des Medians zeigen. Bei den übrigen vier Aufgaben lässt sich kein eindeutiger Trend bei der Analyse der Lösungen in den beiden Gruppen ablesen.

Die formulierten Hypothesen H 2.b beinhalten die Annahme, dass das Alter einen Einfluss auf die Bearbeitung der Aufgaben des PSB in den verschiedenen Antwortformaten hat. Dabei wird theoriegeleitet von der Annahme ausgegangen, dass jüngere Kinder im Vergleich zu älteren schlechtere Testleistungen im MC-Design erzielen. Der hier zu verzeichnende Trend, dass ältere Kinder im Vergleich zu den jüngeren Kindern in vielen der signifikanten Items eine Zunahme an Lösungen in der MC-Version zeigen, steht dabei im Einklang mit den obigen Hypothesen. Kubinger (2009) postuliert, dass die Schwierigkeit eines Items durch das Multiple-Choice-Antwortformat abnimmt, sich also dadurch die Lösungswahrscheinlichkeit erhöht. Im Rahmen der vorliegenden Arbeit wird dagegen von der Annahme ausgegangen, dass bei jüngeren Kindern durch einen unausgereiften Arbeitsstil und mangelnde Kontrollstrategien die Lösungswahrscheinlichkeit durch das MC-Format verringert anstatt erhöht wird. Eine Interpretation der Ergebnisse hinsichtlich dieser Annahmen wäre, dass die getesteten Kinder in der älteren Gruppe bereits über besser ausgeprägte Kontrollstrategien verfügen und deshalb mehr von den Antwortmöglichkeiten des MC-Formats profitieren können, sodass sich Ergebnisse zeigen, die sich jenen von Kubinger (2009) angleichen; sprich: dass die beobachteten Lösungen in der Multiple-Choice-Version höher sind als in der Testversion mit freier Antwort. Bei den Kindern der jüngeren Altersgruppe hingegen sind die Kontrollstrategien noch derart unausgereift, dass sie noch nicht von den Antwortmöglichkeiten profitieren können und sich teilweise sogar ein Trend in die entgegengesetzte Richtung zeigt.

Die Ergebnisse der Überprüfung auf Rasch-Modell-Konformität des Untertests „Formale Folgerichtigkeit“ mit dem Median des Alters als Teilungskriterium sind in Tabelle 18 wiedergegeben. Auch hier liegt der Median des Alters bei 8,1 Jahren bei einer Stichprobengröße von N= 145. Der errechnete p-Wert des Andersen LR-Tests fällt mit $p=0$ signifikant aus, sodass von keiner Gültigkeit des Rasch-Modells für die Gesamtheit der Items unter diesem Teilungskriterium gesprochen werden kann. Auffallend bei der Modellkontrolle mittels LR-Test ist, dass bei dem gegebenen Datensatz 21 von insgesamt 60 Aufgaben nicht geschätzt werden können. Dies ist vermutlich darauf zurückzuführen, dass aufgrund der gemeinsamen Modellprüfung aller Items der Testversionen der „Formalen Folgerichtigkeit“ einige der Aufgaben nicht von allen Testpersonen bearbeitet wurden. Tabelle 19 gibt die Ergebnisse des anschließenden Wald-Tests zur Identifikation einzelner signifikanter Items wieder. Aufgaben mit einem signifikanten p-Wert sind auch hier hervorgehoben.

Tabelle 18: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Median des Alters

N= 145		Median = 8,1 J.	
Andersen LR-test:	LR-value: 114.299	Chi-square df: 38	p-value: 0

Tabelle 19: R-Output des Wald-Tests für den UT „Formale Folgerichtigkeit“; Teilungskriterium: Median des Alters

Wald test on item level (z-values):		
Items	z-statistic	p-value
beta LB_frei_22	-0.942	0.346
beta LB_frei_23	0.661	0.508
beta LB_frei_24	3.252	0.001
beta LB_frei_28	-1.630	0.103
beta LB_frei_34	-0.087	0.930
beta LB_frei_35	1.993	0.046
beta LB_frei_44	-0.190	0.849
beta LB_frei_45	-1.279	0.201
beta LB_frei_46	-0.780	0.435
beta LB_frei_18	-1.709	0.087
beta LB_frei_26	0.696	0.486
beta LB_frei_30	-0.376	0.707
beta LB_frei_32	-2.368	0.018
beta LB_frei_36	-0.829	0.407
beta LB_frei_38	-2.935	0.003
beta LB_frei_40	-0.278	0.781
beta LB_frei_42	0.070	0.944
beta LB_frei_59	0.347	0.729
beta LB_frei_64	-0.950	0.342
beta LB_MC_22	1.714	0.087
beta LB_MC_23	-1.370	0.171
beta LB_MC_24	3.243	0.001
beta LB_MC_28	1.542	0.123
beta LB_MC_34	0.527	0.599
beta LB_MC_35	-0.209	0.835
beta LB_MC_44	0.714	0.475
beta LB_MC_45	0.813	0.416
beta LB_MC_46	-4.091	0.000
beta LB_MC_10	-0.811	0.417
beta LB_MC_18	2.353	0.019
beta LB_MC_26	2.342	0.019
beta LB_MC_30	0.556	0.578
beta LB_MC_32	0.373	0.709
beta LB_MC_36	0.769	0.442
beta LB_MC_38	1.631	0.103
beta LB_MC_40	-2.927	0.003
beta LB_MC_42	-1.267	0.205
beta LB_MC_59	0.204	0.839
beta LB_MC_64	2.355	0.019

Wie in Tabelle 19 zu sehen ist, sind vier Aufgaben aus den Testversionen mit freiem Antwortformat (Item 24, 35, 32 und 38) und insgesamt sechs Items aus den Versionen mit Multiple-Choice-Konzeption (Item 24, 46, 18, 26, 40 und 64) in der Modellprüfung mittels Wald-Test signifikant. Dabei kann jedoch Aufgabe 38 aus den Analysen ausgeschlossen werden, da bereits Hagenmüller (2011) bei der Testerprobung eine Zweideutigkeit der Itemkonzeption nachweisen konnte und dieses Item auch im weiteren Verlauf der Entwicklung ausgeschlossen hat. Aufgabe 18 muss aufgrund von unvollständigen Daten ebenfalls aus weiteren Analysen ausgeschlossen werden. Beim Betrachten der relativen Lösungshäufigkeiten der Items, welche nicht dem Rasch-Modell entsprechen, geteilt nach dem Median des Alters, zeigen sich teils widersprüchliche Ergebnisse, sodass sich nur ein

schwacher Trend innerhalb der Ergebnisse abzeichnet. Bei den Aufgaben 32 sowie 46 sind Lösungshäufigkeiten dahingehend zu beobachten, dass ältere Kinder deutlich weniger Probleme mit der Bearbeitung der Items in der Multiple-Choice-Version gegenüber den jüngeren Kindern haben. Im Gegensatz dazu ist bei den Aufgaben 26 sowie 64 eine Zunahme an Lösungen beider Altersgruppen in der Testversion im MC-Design zu beobachten. Item 64 stellt die letzte Aufgabe der „Formalen Folgerichtigkeit“ in der Version für 8- bis 11-jährige Kinder dar und ist damit per Konzeption sehr schwer zu lösen. Dadurch bedingt, könnte ein Rateverhalten beim Lösen der Aufgabe gefördert werden, was wiederum einer MC-Konzeption zugute kommt und die hier beobachteten Ergebnisse erklären könnte. Die beobachteten Lösungen bei Item 35 zeigt auch eine Zunahme an richtigen Antworten beider Altersgruppen in der MC-Darbietung, wobei hier vor allem jüngere Kinder von diesem Antwortformat zu profitieren scheinen. Bei der Analyse von Aufgabe 40 zeigt sich in beiden Altersgruppen eine Abnahme an Lösungen in der Multiple-Choice-Version gegenüber dem freien Antwortformat, wobei sich dieser Trend bei den älteren Kindern deutlicher abzeichnet. Item 24 ist in Bezug auf die relativen Lösungshäufigkeiten weitestgehend unauffällig.

Die hier gefundenen Ergebnisse sind recht widersprüchlich, sodass sich kein eindeutiger Trend abzeichnet. Während bei den Items 32 und 46 ein hypothesenkonformes Bild der Lösungen zu beobachten ist, ergibt die Analyse der Items 35 sowie 64 ein Bild, welches gegen die Annahmen der vorliegenden Arbeit spricht. Die Aufgaben 24 und 40 ergeben kein eindeutiges Bild.

3.5.3. Unterschiede in der Bearbeitung des Multiple-Choice-Designs und des freien Antwortformats unter Einbezug des Arbeitsstils

Die Forschungsfragen der Hypothese H 3 samt Unterhypothesen beschäftigen sich mit dem relativen Einfluss des Arbeitsstils auf die Bearbeitung von Aufgaben zur Erfassung des logisch-schlussfolgernden Denkens in den Antwortformaten Multiple-Choice sowie freie Antwort. Im Sinne der vorab definierten Fragestellung wird dabei zwischen den Arbeitsstilen Impulsiv vs. Reflexiv unterschieden. Bei der Klassifikation der Testpersonen kommt schlussendlich ein dreikategoriales System zur Anwendung, wobei eine zusätzliche Kategorie im Sinne eines neutralen Verhaltens vorhanden ist. Um den Einfluss des Arbeitsstils auf die Aufgabenbearbeitung beider Antwortformate bestmöglich erfassen zu können, wurde auch bei der Datenauswertung in mehreren Schritten vorgegangen: Zunächst wurde für alle Verfahren ein T-Test für gepaarte Stichproben mit dem Arbeitsstil II als Teilungskriterium (siehe Beschreibung der Teilungskriterien in Punkt 3.5) gerechnet. In einem weiteren Schritt wurden Rasch-Analysen unter Einbezug des Arbeitsverhaltens für die Verfahren durchgeführt, mit einer anschließenden inhaltlichen Analyse derjenigen Items, welche nicht dem Modell entsprechen.

3.5.3.1. T-Tests für gepaarte Stichproben

Im Sinne eines ersten Auswertungsschrittes wurden zunächst T-Tests für gepaarte Stichproben für alle drei Verfahren durchgeführt. Dabei wurden lediglich die Daten jener Testpersonen verwendet, die während der Testung durch die Testleiter als entweder eindeutig impulsiv oder reflexiv im Hinblick auf ihr jeweiliges Arbeitsverhalten klassifiziert wurden. Jene Testpersonen, die in die Kategorie *Neutral* eingeordnet wurden, sind bei diesen Auswertungen ausgeschlossen, um etwaige auftretende Effekte möglichst gut und eindeutig erkennen zu können. Diese Überlegungen bedingen jedoch, dass hier, gemessen

an den gesammelten Daten, mit einer relativ kleinen Stichprobengröße gearbeitet wurde. Die Voraussetzung der Normalverteilung der Differenzen für jede Gruppe wurde einerseits mittels grafischer Modellkontrolle sowie mittels K-S-Test überprüft und kann für alle Instrumente als erfüllt angesehen werden.

Die Tabellen 20 und 21 zeigen die Ergebnisse der Testungen auf Unterschiede zwischen dem Matrizentest im freien Antwortformat sowie im Multiple-Choice-Design einerseits und dem PSB in der korrigierten Fassung für die Gruppe der als impulsiv klassifizierten Kinder in der erhobenen Stichprobe andererseits. Der T-Test für gepaarte Stichproben wurde bei dem Matrizentest für eine Stichprobe von $N = 21$ impulsiv arbeitenden Kindern durchgeführt, wobei mit einem Wert von $p = 0,006$ ein signifikantes Ergebnis gefunden wurde, bei einer zweiseitigen Testung mit einem Konfidenzintervall von 95%. Im Sinne einer einseitig formulierten Forschungshypothese muss das hier gefundene Ergebnis halbiert werden, wobei man einen weiterhin signifikanten Wert von $p = 0,003$ erhält. Betrachtet man unter diesem Gesichtspunkt die Mittelwerte in der jeweiligen Testversion, so fällt auf, dass die impulsiven Kinder in der MC-Version des Matrizentests mit einer Differenz von 1,809 Rohwerten einen höheren Mittelwert erzielt haben als in der Testversion mit freier Antwort. Jedoch muss auch dieses Ergebnis im Hinblick auf die errechnete Standardabweichung von $SD = 2,695$ relativiert werden. Diese Ergebnisse scheinen auf den ersten Blick gegen die hier aufgestellten Hypothesen zu sprechen. Entsprechend dieser, wäre ein Ergebnis dahingehend zu erwarten gewesen, dass impulsiv arbeitende Kinder einen erhöhten Mittelwert in der Testversion mit freier Antwort zeigen, während hier Unterschiede in die entgegengesetzte Richtung zu beobachten sind. Jedoch sind die hier gefunden Werte unter Einbezug der Standardabweichung bestenfalls als Trend zu bewerten.

Die Testung auf Unterschiede zwischen den beiden Testformen des PSB wurde hier bei einer Stichprobe von $N = 31$ Kindern durchgeführt, deren Arbeitsstil ebenfalls als impulsiv eingestuft wurde. Dabei ergibt die zweiseitige Prüfung bei einem Alphaniveau von 5% einen nicht signifikanten Wert von $p = 0,588$ (Tabelle 21). Eine einseitige Signifikanzprüfung zeigt mit einem Wert von $p = 0,290$ ebenfalls keine Unterschiede zwischen den beiden

Testversionen in dieser Gruppe. Ein Vergleich der Rohscore-Mittelwerte beider Testversionen in Tabelle 20 ergibt einen Unterschied von 0,6125 zu Gunsten des Multiple-Choice-Designs bei einer Standardabweichung von $SD = 4,264$. Somit kann in der Gruppe der Kinder mit einem impulsiven Arbeitsstil von keinem Unterschied in der Bearbeitung des PSB in den Antwortformaten Multiple-Choice sowie freie Antwort gesprochen werden.

Tabelle 20: SPSS-Output der Statistik bei gepaarten Stichproben für den Matrizentest und den PSB (korrigierte Version); Arbeitsstil: Impulsiv

Statistik bei gepaarten Stichproben(a)

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_Matrizen_frei	14,8571	21	3,41007	,74414
	RW_Matrizen_MC	16,6667	21	3,15172	,68776
Paaren 2	RW_PSB_frei34	20,6452	31	4,03773	,72520
	RW_PSB_MC34	21,0645	31	5,81341	1,04412

a. Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Unterdurchschnittlich reflexiv/
Überdurchschnittlich Impulsiv

Tabelle 21: SPSS Output des T-Tests für gepaarten Stichproben für den Matrizentest und den PSB (Korrigierte Version); Arbeitsstil: Impulsiv

Test bei gepaarten Stichproben(a)

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paare n 1	RW_Matrizen_frei - RW_Matrizen_MC	-1,80952	2,69479	,58805	-3,03618	-,58287	-3,077	20	,006
Paare n 2	RW_PSB_frei34 - RW_PSB_MC34	-,41935	4,26438	,76591	-1,98354	1,14483	-,548	30	,588

a. Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Unterdurchschnittlich reflexiv/
Überdurchschnittlich Impulsiv

Die Tabellen 22 sowie 23 zeigen den SPSS-Output der Unterschiedsprüfung mittels T-Test für gepaarte Stichproben der Testversionen des UT „Formale Folgerichtigkeit“ für den Altersbereich 6 bis 7 Jahre, ebenfalls für die Gruppe der impulsiv arbeitenden Kinder. Bei einer Stichprobengröße von N= 15 zeigt sich hier ein nicht signifikantes Ergebnis mit einem Wert von $p = 0,760$ bei einer zweiseitigen Prüfung mit 95% Konfidenzintervall. Die einseitige Unterschiedsprüfung fällt dabei mit einem Wert von $p = 0,380$ ebenfalls nicht signifikant aus. Die beobachtete Mittelwertsdifferenz zwischen den beiden Antwortformaten in Tabelle 20 liegt bei 0,333 Rohwertpunkten bei einer errechneten Standardabweichung von $SD = 4,152$. Inhaltlich gesehen, sind diese Ergebnisse dahingehend zu deuten, dass ein impulsiver Arbeitsstil bei Kindern im Alter zwischen sechs und sieben Jahren einen nur unbedeutenden Einfluss auf die Bearbeitung der „Formalen Folgerichtigkeit“ in den jeweiligen Antwortformaten hat.

Tabelle 22: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6 bis 7 J.; Arbeitsstil: Impulsiv

Statistik bei gepaarten Stichproben(a)

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_LB_frei	9,4667	15	4,48596	1,15827
	RW_LB_MC	9,8000	15	4,93095	1,27317

a. Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Unterdurchschnittlich reflexiv/
Überdurchschnittlich Impulsiv

Tabelle 23: SPSS-Output des T-Tests für gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6-7 J.; Arbeitsstil: Impulsiv

Test bei gepaarten Stichproben(a)

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paare n 1	RW_LB_frei - RW_LB_MC	-,33333	4,15188	1,07201	-2,63257	1,96590	-,311	14	,760

- a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Unterdurchschnittlich reflexiv/
Überdurchschnittlich Impulsiv

Die Tabellen 24 und 25 zeigen ebenfalls die Ergebnisse der Prüfung mittels T-Test, allerdings für die Testversionen des UT „Formale Folgerichtigkeit“ für den Altersbereich der 8- bis 11-jährigen Kinder. Dabei zeigt sich bei einer Stichprobe von N= 16 Kindern mit einem impulsiven Arbeitsstil mit einem Wert von $p= 0,479$ kein signifikanter Unterschied zwischen den beiden Antwortformaten bei zweiseitiger Prüfung mit einem Alpha von 0,05. Eine einseitige Prüfung im Sinne der Fragestellung ergibt mit einem Wert von $p= 0,240$ ebenfalls ein nicht signifikantes Ergebnis bei gleichbleibendem Konfidenzintervall. Die beobachtete Mittelwertdifferenz liegt hier bei 0,625 Rohwertpunkten bei einer Standardabweichung von $SD= 3,442$. Somit ist in Anlehnung an obige Ergebnisse kein relevanter Einfluss des impulsiven Arbeitsstils auf die Bearbeitung der MC-Testversion der „Formalen Folgerichtigkeit“ für 8- bis 11-jährige Kinder, sowie jener Version im freien Antwortformat, anzunehmen.

Tabelle 24: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Impulsiv

Statistik bei gepaarten Stichproben(a)

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_LB_frei	9,3125	16	3,28062	,82016
	RW_LB_MC	9,9375	16	4,44925	1,11231

- a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Unterdurchschnittlich reflexiv/
Überdurchschnittlich Impulsiv

Tabelle 25: : SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Impulsiv

Test bei gepaarten Stichproben(a)

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paaren 1	RW_LB_frei - RW_LB_MC	-,62500	3,44238	,86060	-2,45932	1,20932	-,726	15	,479

a. Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Unterdurchschnittlich reflexiv/
Überdurchschnittlich Impulsiv

Im Folgenden sollen die Ergebnisse der Unterschiedsprüfungen mittels T-Tests für gepaarte Stichproben aller drei Verfahren zwischen den beiden Antwortformaten in der Gruppe der als reflexiv eingestuften Kinder näher erläutert werden.

Die Tabellen 26 und 27 zeigen den SPSS-Output der Prüfung auf Unterschiede samt Statistiken zwischen den Testversionen des Matrizentests und des PSB im MC-Design und mit freier Antwort für die Gruppe der reflexiv arbeitenden Kinder. Für die Testversionen des Matrizentests zeigt sich hier bei einer Stichprobe von N= 16 mit einem Wert von $p = 0,924$ ein nicht signifikantes Ergebnis bei einer zweiseitigen Prüfung mit einem Alphaniveau von 5%. Beim Betrachten der Mittelwerte in Tabelle 26 zeigt sich eine kleine Differenz der Rohwerte mit 0,063 zugunsten der Testversion mit freier Antwort bei einer Standardabweichung von $SD = 2,568$. Die Gruppe der reflexiv arbeitenden Kinder zeigt hier keine unterschiedlichen Testergebnisse zwischen der Multiple-Choice-Version des Matrizentests und der Testversion mit freier Antwort. In Anbetracht dieser Ergebnisse scheint somit auch ein reflexiver Arbeitsstil keinen maßgeblichen Einfluss auf die Aufgabenbearbeitung des Matrizentests in den beiden Antwortformaten zu haben.

Das Ergebnis der Unterschiedsprüfung zwischen den Testversionen des PSB in dieser Gruppe fällt bei einer Stichprobengröße von $N = 29$ mit einem Wert von $p = 0,213$ nicht signifikant aus, bei einem üblichen Konfidenzintervall von 95% und zweiseitiger Testung. Hier ist mit einer Differenz von 0,879 ein erhöhter Mittelwert in der MC-Testversion zu verzeichnen, bei einer Standardabweichung von $SD = 3,792$. Ebenso wie bei der Gruppe der impulsiv arbeitenden Kinder, zeigt sich auch in dieser Gruppe kein signifikanter Einfluss des Arbeitsstils auf die Aufgabenbearbeitung des PSB in den beiden Antwortformaten. Die hier gefundenen Kennwerte sprechen für keine maßgebliche Beteiligung des Arbeitsverhaltens von reflexiv arbeitenden Kindern bei dem Lösen der PSB-Aufgaben je nach Antwortformat.

Tabelle 26: SPSS-Output der Statistik bei gepaarten Stichproben für den Matrizentest und den PSB (korrigierte Version); Arbeitsstil: Reflexiv

Statistik bei gepaarten Stichproben(a)

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_Matrizen_frei	18,8750	16	2,60448	,65112
	RW_Matrizen_MC	18,8125	16	3,31097	,82774
Paaren 2	RW_PSB_frei34	24,7241	29	3,96319	,73595
	RW_PSB_MC34	25,6207	29	4,25453	,79005

a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Überdurchschnittlich reflexiv/
Unterdurchschnittlich impulsiv

Tabelle 27: SPSS-Output des T-Tests für gepaarte Stichproben für den Matrizentest und den PSB (korrigierte Version); Arbeitsstil: Reflexiv

Test bei gepaarten Stichproben(a)

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paare n 1	RW_Matrizen_frei - RW_Matrizen_MC	,06250	2,56824	,64206	-1,30602	1,43102	,097	15	,924
Paare n 2	RW_PSB_frei34 - RW_PSB_MC34	-,89655	3,79233	,70422	-2,33908	,54597	-1,273	28	,213

a. Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Überdurchschnittlich reflexiv/
Unterdurchschnittlich impulsiv

Die Ergebnisse der Unterschiedsprüfung zwischen der MC-Version der „Formalen Folgerichtigkeit“ und der Version im freien Antwortformat für den Altersbereich 6 bis 7 Jahre sowie für die Gruppe der reflexiven Kinder sind in den Tabellen 28 und 29 zu sehen. Auffallend ist dabei die sehr kleine Stichprobengröße von N= 5 dieser Gruppe für dieses Verfahren. Der T-Test liefert hier bei einem Alpha von 0,05 sowie zweiseitiger Testung ein nicht signifikantes Ergebnis mit einem Wert von $p = 0,322$. Hier zeigt sich mit einer Differenz von 2,800 ein erhöhter Mittelwert in der Testversion mit freier Antwort bei einer Standardabweichung von $SD = 5,541$. Somit ist bei dieser Altersversion der „Formalen Folgerichtigkeit“ sowohl im Multiple-Choice-Design als auch im freien Antwortformat kein signifikanter Einfluss eines reflexiven Arbeitsverhaltens zu sehen. Aufgrund der geringen Stichprobengröße muss dieses Ergebnis jedoch mit Vorbehalt interpretiert werden.

Tabelle 28: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6-7 J.; Arbeitsstil: Reflexiv

Statistik bei gepaarten Stichproben(a)

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_LB_frei	15,4000	5	3,57771	1,60000
	RW_LB_MC	12,6000	5	4,39318	1,96469

a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Überdurchschnittlich reflexiv/
Unterdurchschnittlich impulsiv

Tabelle 29: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 6-7 J.; Arbeitsstil: Reflexiv

Test bei gepaarten Stichproben(a)

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paare n 1	RW_LB_frei - RW_LB_MC	2,80000	5,54076	2,47790	-4,07976	9,67976	1,130	4	,322

a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Überdurchschnittlich reflexiv/
Unterdurchschnittlich impulsiv

Die folgenden Tabellen 30 sowie 31 zeigen die Ergebnisse des T-Tests für gepaarte Stichproben für die Testversionen der 8- bis 11-jährigen Kinder der „Formalen Folgerichtigkeit“, ebenfalls für den Arbeitsstil: Reflexiv. Hier wurde mit einer Stichprobengröße von N= 23 Testpersonen gerechnet, wobei bei einer zweiseitigen Prüfung mit einem Wert von $p = 0,223$ kein signifikantes Ergebnis gefunden wurde, bei einem Konfidenzintervall von 95%. Bei einem Vergleich der Mittelwerte zeigt sich mit einer Differenz von 0,957 ein leicht erhöhter Wert in der MC-Version gegenüber dem Design mit freier Antwort bei einer errechneten Standardabweichung von $SD = 3,662$. Genau wie in der Testversion für 6- bis 7-jährige Kinder scheint auch hier der reflexive Arbeitsstil keinen maßgeblichen Einfluss auf die Aufgabenbearbeitung der „Formalen Folgerichtigkeit“ in dem

jeweiligen Antwortformat zu haben. So zeigen die hier gefundenen Ergebnisse keinen Unterschied in der Testleistung zwischen dem MC-Design und der Konzeption mit freier Antwort dieses Verfahrens in dieser Gruppe.

Tabelle 30: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Reflexiv

Statistik bei gepaarten Stichproben(a)

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	RW_LB_frei	12,3913	23	2,51792	,52502
	RW_LB_MC	13,3478	23	3,57528	,74550

a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Überdurchschnittlich reflexiv/
Unterdurchschnittlich impulsiv

Tabelle 31: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Reflexiv

Test bei gepaarten Stichproben(a)

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paaren 1	RW_LB_frei - RW_LB_MC	-,95652	3,66157	,76349	-2,53990	,62686	-1,253	22	,223

a Arbeitsstil der Kinder für alle Verfahren in der Einzeltestung = Überdurchschnittlich reflexiv/
Unterdurchschnittlich impulsiv

3.5.3.2. Rasch-Modell-Analysen

In einem erweiterten Auswertungsschritt zur Beantwortung der Fragestellungen unter der Hypothese H 3 und den damit verbundenen weiteren Hypothesen wurden die gesammelten Daten aller Aufgaben der jeweiligen Verfahren Rasch-Modell-Analysen mit dem Arbeitsstil als Teilungskriterium unterzogen. Dabei wurde der Arbeitsstil I, wie in der Einführung zu Punkt 3.5 beschrieben, zur Teilung der Stichprobe verwendet, wobei die Kategorien „Reflexiv“ und „Neutral“ zusammengefasst und dem Arbeitsstil „Impulsiv“ gegenübergestellt wurden. Die Zusammenlegung gerade dieser Gruppen geschieht vor dem Hintergrund der aufgestellten Hypothesen, dass sich die Testleistungen gerade von impulsiv arbeitenden Kindern zwischen den Antwortformaten unterscheiden müssten, während bei reflexiv eingestuften Kindern kaum ein Unterschied zu beobachten sein sollte. Diese Methode hat den Vorteil, dass alle erhobenen Daten für die Auswertung verwendet werden können und so mit einer relativ großen Stichprobenanzahl gearbeitet werden kann. Dies bringt jedoch den Nachteil mit sich, dass mögliche Effekte durch die Kombination mit der neutralen Kategorie abgeschwächt werden können und dadurch auch schwieriger zu erkennen sind. Die Teilung der Gesamtstichprobe nach den Gruppen Impulsiv vs. Reflexiv würde dem oben definierten Kriterium des Arbeitsstils II entsprechen. Dadurch wäre es zwar einfacher, eventuelle Effekte zu entdecken, allerdings mit dem Nachteil einer drastischen Reduzierung der Stichprobe, was sich gerade bei Analysen mittels Rasch-Modell als problematisch erweist. Aus diesem Grund, und weil dieses Teilungskriterium bereits ausführlich durch diverse T-Tests untersucht wurde, wird hier auf eine Prüfung der Rasch-Modell-Konformität mit dem Arbeitsstil II als Teilungskriterium verzichtet. Zum Zwecke weiterführender inhaltlicher Analysen wurden im Anschluss an die jeweiligen Rasch-Modell-Analysen Wald-Tests mit selbigem Teilungskriterium durchgeführt.

Tabelle 32 zeigt die Ergebnisse des Andersen LR-Tests auf Prüfung der Rasch-Modell-Konformität des PSB mit dem Arbeitsstil I als Teilungskriterium. Dabei wurden, wie oben beschrieben, die Kinder mit einem impulsiven Arbeitsstil jenen mit einem reflexiven oder neutralen Stil gegenübergestellt. Hier zeigt sich bei einer Stichprobe von $N = 155$ ein nicht signifikanter p-Wert von $p = 0,067$ sodass eine Gültigkeit des Rasch-Modells für die

Gesamtheit der Items über beide Gruppen gegeben ist. Inhaltlich bedeutet dies, dass die Mehrzahl der Aufgaben sowohl im MC-Design als auch im freien Antwortformat dieses Verfahrens nicht maßgeblich durch spezielle Charakteristika der verglichenen Gruppen (geteilt durch den Arbeitsstil I) beeinflusst wird.

Tabelle 32: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Arbeitsstil I

N= 155

Andersen LR-test:	LR-value: 63.446	Chi-square df: 48	p-value: 0.067
--------------------------	------------------	-------------------	-----------------------

Tabelle 33 zeigt den R-Output der Prüfung auf Gültigkeit des Rasch-Modells für die Gesamtheit der Aufgaben aus dem Untertest „Formale Folgerichtigkeit“ aus dem AID 3. Die Stichprobengröße beträgt hier N= 158 Testpersonen. Der errechnete p-Wert des LR-Tests ist hier mit $p = 0.056$ nicht signifikant, sodass von einer Gültigkeit des Rasch-Modells ausgegangen wird. Somit sind eventuelle Unterschiede zwischen den beiden Antwortformaten für die verglichenen Gruppen nicht so groß, als dass für die Gesamtheit der Items die Modellgültigkeit verworfen werden müsste. Im Sinne der untersuchten Fragestellung ist davon auszugehen, dass keine maßgeblichen Unterschiede in der Aufgabenbearbeitung zwischen den beiden Antwortformaten in dieser Gruppe bestehen.

Tabelle 33: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Arbeitsstil I

N= 158

Andersen LR-test:	LR-value: 77.183	Chi-square df: 59	p-value: 0.056
-------------------	------------------	-------------------	----------------

4. Diskussion

Das zentrale Anliegen der vorliegenden Arbeit ist es, die Forschung über die Beteiligung von wesentlichen kognitiven Operationen bei der Bearbeitung von Aufgaben zum logisch-schlussfolgernden Denken im Alter von 6- bis 11-jährigen Kindern zu ergänzen und fortzuführen, wobei in einem weiteren Schritt die Auswirkungen des gewählten Antwortformats unter diesen Gesichtspunkten untersucht werden sollen.

Ausgehend von der vorherrschenden Literatur über den relativen Einfluss der Kapazität des Arbeitsgedächtnisses und von generellen Kontrollstrategien sowie im Hinblick auf entwicklungspsychologische Aspekte (Baddeley, 1996; Ziegler, 1994; Embretson, 1995, 1998; Goswami, 2001; Kane et al., 2004; Kyllonen & Christal, 1990; Unsworth & Engle, 2004) wurde bei der Formulierung der Forschungshypothese H 1 von der Annahme ausgegangen, dass sich bei Kindern die Art des gewählten Antwortformats dahingehend auswirkt, dass diese bei der Aufgabenbearbeitung im Multiple-Choice-Design signifikant schlechtere Leistungen erzielen als bei dem Lösen selbiger Aufgaben im freien Antwortformat. Die Ergebnisse des T-Tests für gepaarte Stichproben unter Punkt 3.5.1.1 zeigen sowohl für den PSB als auch für beide Altersversionen der „Formalen Folgerichtigkeit“ keine signifikanten Unterschiede in den Testergebnissen zwischen den beiden Antwortformaten. Bei der Unterschiedsprüfung der Leistungen in den Testversionen des Matrizentests zeigt sich zunächst ein signifikanter Unterschied zwischen den Antwortformaten, welcher jedoch unter Einbezug der

Standardabweichung relativiert werden muss. Dabei ist in der Multiple-Choice-Testversion, entgegen den Erwartungen, ein erhöhter Mittelwert gegenüber der Version mit freiem Antwortformat zu beobachten, was als ein Trend zu klassifizieren ist, der gegen die hier getroffenen Annahmen spricht. Die Prüfungen auf Gültigkeit des Rasch-Modells der Items beider Antwortformate mit dem Median der Testleistung sowie dem Geschlecht als Teilungskriterium zeigt für die Aufgaben der „Formalen Folgerichtigkeit“ eine Passung des Modells nach beiden Kriterien. Somit kann für diese Gruppen angenommen werden, dass sich die Aufgabenbearbeitung in diesem Test in den beiden Antwortformaten nicht maßgeblich voneinander unterscheidet, sodass die Hypothese H 1.c.0 beibehalten werden muss. Bei selbiger entsprechender Prüfung der Aufgaben des PSB zeigt sich infolge einer Teilung der Stichprobe durch das Geschlecht ebenfalls eine Gültigkeit des Rasch-Modells für die Gesamtheit der Items, wohingegen die Modellprüfung mittels Median der Testleistung als Teilungskriterium signifikant ausfiel. Bei einer inhaltlichen Analyse der Items, welche nicht dem Rasch-Modell entsprechen, zeigt sich ein hypothesenkonformer Trend dahingehend, dass bei einem Großteil der Aufgaben vermehrt Lösungen in der Testversion mit freiem Antwortformat gefunden wurden, wobei keine andere Erklärung dieser Ergebnisse als die Eigenheiten der Antwortformate plausibel erscheint.

Die hier aufgestellten und untersuchten Annahmen bezüglich der erhöhten Schwierigkeit von Aufgaben im Multiple-Choice-Design im Kindesalter wurden in einem weiteren Schritt durch die Analyse des relativen Einflusses des Alters in der Stichprobe untersucht. So zeigt das Ergebnis der Rasch-Modell-Analysen des PSB mit dem Alter als Teilungskriterium samt inhaltlicher Analyse einen deutlich hypothesenkonformen Trend. Der älteren Kinder zeigt sich bei einem Großteil der Aufgaben, welche nicht dem Modell entsprechen, eine Zunahme an Lösungen in der MC-Version gegenüber den jüngeren. Somit kann hier die Nullhypothese verworfen und die Hypothese H 2.b.1.1 bezüglich einer verminderten Testleistung jüngerer Kinder im Vergleich zu älteren im MC-Format des PSB angenommen werden. Die Prüfung auf Gültigkeit des Rasch-Modells für die Aufgaben beider Antwortformate der „Formalen Folgerichtigkeit“ ist mit dem Median des Alters als Teilungskriterium ebenfalls signifikant ausgefallen. Bei der inhaltlichen Analyse der nicht modellkonformen Items ergibt sich jedoch kein so klares Bild wie bei den Ergebnissen des

PSB. In diesem Verfahren ist bei zwei der signifikanten Items ein Bild der Lösungen zu beobachten, das mit den Annahmen der a priori aufgestellten Hypothesen konform ist (Item 32 und 46) und sich mit den Ergebnissen des PSB deckt, nämlich dass jüngere Kinder im freien Antwortformat bessere Testleistungen erzielen als im MC-Antwortformat. Drei andere Items zeigen einen entgegengesetzten Trend (Item 26, 35 und 64), nämlich dass sie für jüngere Kinder im MC-Format leichter zu lösen sind als im freien Antwortformat. Dabei ist auffallend, dass zwei davon (nämlich Aufgaben 35 und 64) in beiden Formaten generell selten gelöst wurden und sich durch eine hohe Schwierigkeit auszeichnen. Die hohe Grundschwierigkeit könnte dabei eine vermehrte Lösung der Items in der MC-Version für jüngere Kinder erklären; wenn diese Aufgabe derart schwierig ist, dass kaum ein Kind das Prinzip der Konzeption durchschaut, so wird öfter versucht, die Antwort zu erraten, als dies sonst der Fall wäre, wodurch eine erhöhte Lösungswahrscheinlichkeit des MC-Designs verstärkt wird. Die Ergebnisse sprechen also insgesamt dafür, dass jüngere Kinder aufgrund fehlender Kontrollstrategien bei MC-Antwortformaten benachteiligt sind und diese für sie schwieriger sind als für ältere Kinder mit besser entwickelten Kontrollstrategien. Dies aber nur dann, wenn die Aufgaben hinsichtlich der Schwierigkeit leicht sind bzw. annähernd der Fähigkeit des Kindes entsprechen. Nur bei viel zu schwierigen Aufgaben, bei denen die Lösungswahrscheinlichkeit von vornherein sehr gering ist, profitieren auch jüngere Kinder von dem MC-Antwortformat und kommen teilweise durch Raten doch noch zu einer richtigen Lösung, wodurch diese Aufgaben im Vergleich zum freien Antwortformat im Multiple-Choice-Design leichter zu lösen sind.

Eine Interpretation der gefundenen Ergebnisse der Rasch-Modell-Analyse des PSB mit dem Alter als Teilungskriterium im Einklang mit den Hypothesen der vorliegenden Arbeit wäre, dass die getesteten Kinder in der älteren Gruppe bereits über besser ausgeprägte Kontrollstrategien verfügen als die, relativ gesehen, jüngeren Kinder, und deshalb mehr von den Antwortmöglichkeiten des MC-Formats profitieren können, sodass sich Ergebnisse zeigen, die sich jenen des Erwachsenenalters angleichen. Bei Kindern der jüngeren Altersgruppe sind die Kontrollstrategien hingegen noch unausgereifter, sodass diese Kinder noch nicht von den Antwortmöglichkeiten des Multiple-Choice-Designs profitieren können.

Dass in dem UT „Formale Folgerichtigkeit“ kein eindeutiger Trend gefunden wurde, könnte auf die Konzeption des freien Antwortformats zurückzuführen sein. Wie bereits in der

Verfahrensbeschreibung dargelegt, zeichnet sich das ursprüngliche Design des Antwortformats von Hagenmüller (2011) durch diverse Charakteristika aus, welche einem MC-Format zuzuschreiben wären, allen voran die Wahlmöglichkeit der Antwort. Zwar wird aufgrund der hohen Anzahl an Antwortmöglichkeiten sowie aus inhaltlichen Überlegungen davon ausgegangen, dass dieses Design als freies Antwortformat zu sehen ist, jedoch wäre hier denkbar, dass sich die MC-Charakteristika stärker auf die Aufgabenbearbeitung auswirken als ursprünglich angenommen. Hier wären weiterführende Forschungen bezüglich der Eigenheiten dieser Antwortkonzeption, unter Umständen an einer größeren Stichprobengröße, wünschenswert.

Der Einfluss der Arbeitsstile Impulsiv vs. Reflexiv auf die Bearbeitung von Reasoning-Aufgaben wurde in einem weiteren Schritt in die Auswertung mit einbezogen. Ausgehend von bisherigen Forschungsergebnissen (Kagan, Pearson & Welch, 1969; Zelniker & Jeffrey, 1976; Schweizer, 2001) ist dabei eine Annahme der vorliegenden Arbeit, dass Kinder mit einem impulsiven Arbeitsstil im Vergleich zu reflexiv arbeitenden Kindern schlechtere Leistungen bei der Bearbeitung von Multiple-Choice-Aufgaben erzielen. Beim Bearbeiten eines freien Antwortformats hingegen sollte sich kaum ein Unterschied in den Ergebnissen zeigen. Eine erste Überprüfung dieser Hypothese mittels T-Tests für gepaarte Stichproben für die Gruppen der impulsiv sowie reflexiv eingestuften Kinder (Teilung nach dem Arbeitsstil II) zeigt sowohl für den PSB sowie beide Verfahren der „Formalen Folgerichtigkeit“ (6 bis 7 J. und 8 bis 11 J.) keine Bestätigung dieser Annahmen. In der Gruppe der reflexiv arbeitenden Kinder wie auch der impulsiv arbeitenden wurden keine signifikanten Unterschiede zwischen den Testergebnissen der MC-Version und der Konzeption mit freier Antwort gefunden, sodass die Hypothesen H 3.b.1.1 und H 3.c.1.1 zurückgewiesen sowie die Hypothesen H 3.b.2.0 und H 3.c.2.0 angenommen werden müssen. Bei den Unterschiedsprüfungen in den Testversionen des Matrizentests, mit dem Arbeitsstil II als Teilungskriterium, wurde für die Gruppe der impulsiv arbeitenden Kinder ein signifikantes Ergebnis gefunden. Bei einem Vergleich der Mittelwerte der jeweiligen Testversionen zeigt sich entgegen den Erwartungen ein erhöhter Wert in der Version mit einem Multiple-Choice-Antwortformat. In Anbetracht der Standardabweichung muss der inhaltliche Gehalt auch dieser Ergebnisse relativiert werden, wobei ein Trend entgegen den hier getroffenen Annahmen bleibt. In der Gruppe

der reflexiv arbeitenden Kinder zeigt sich für den Matrizentest, konform den Erwartungen, kein signifikanter Unterschied zwischen den Testversionen in den beiden Antwortformaten, sodass hier die Hypothesen H 3.a.1.0 sowie H 3.a.2.0 beibehalten werden.

Für die Überprüfung der Rasch-Modell-Konformität der Aufgaben des PSB sowie der „Formalen Folgerichtigkeit“ wurde auf den Arbeitsstil I als Teilungskriterium zum Zweck einer größeren Stichprobe zurückgegriffen. Die Prüfung auf Gültigkeit des Rasch-Modells fiel dabei für beide Verfahren nicht signifikant aus, sodass für die untersuchten Gruppen kein Unterschied in der Aufgabenbearbeitung angenommen muss. In Anbetracht dieser Ergebnisse ergibt sich der Schluss, dass im Rahmen der vorliegenden Arbeit entgegen den Erwartungen kein Einfluss des Arbeitsstils auf die Aufgabenbearbeitung der beiden Antwortformate nachgewiesen werden konnte, wenngleich für den Matrizentest ein Trend zu beobachten ist, der sich entgegen den Hypothesen verhält. Eine mögliche Erklärung für diese Ergebnisse könnte in der Auswahl der verwendeten Verfahren liegen. Wie bereits weiter oben erwähnt, wäre auch hier denkbar, dass das freie Antwortformat der „Formalen Folgerichtigkeit“ mit diversen Charakteristika eines MC-Designs die gefundenen Ergebnisse verfälscht. In der vorliegenden Literatur wird darauf hingewiesen, dass sich der Einfluss des Arbeitsstils besonders bei der Bearbeitung von Aufgaben zeigt, welche eine genaue Differenzierung des Inhaltes verlangen (Zelniker & Jeffrey, 1976; Phillips & Rabbitt, 1995). Gerade der UT „Formale Folgerichtigkeit“ sowie ein Großteil der Items aus dem PSB beinhalten hauptsächlich große und einfache Symbole, wodurch sich der erwartete Effekt u.U. schwächer zeigen könnte.

Auffallend ist, dass nur in den Unterschiedsprüfungen der beiden Antwortformate des Matrizentests Ergebnisse gefunden wurden, die einen Trend entgegen den hier aufgestellten Hypothesen zeigen. Bei näherer Betrachtung ergibt sich die Überlegung, dass die spezielle Gestaltung des freien Antwortformats dieses Verfahrens einen entscheidenden Unterschied gegenüber den beiden anderen Tests enthält: In der Testversion mit freier Antwort müssen die Kinder zwischen vier verschiedenen Farben differenzieren und arbeiten demnach auch mit vier verschiedenen Stiften, wobei sie jeweils den oder die richtigen auswählen und zusätzlich noch die richtige Antwort entsprechend den Farben einzeichnen müssen. Für die

Bearbeitung von Aufgaben des Matrizentest in dieser Version ist also denkbar, dass dafür deutlich mehr kognitive Schritte durchlaufen werden müssen als bei den freien Testversionen des PSB sowie der „Formalen Folgerichtigkeit“. Dadurch bedingt wäre auch denkbar, dass dafür mehr Kontrolloperationen notwendig sind als in den anderen Verfahren oder u.U. auch als für die Bearbeitung selbiger Items im MC-Format, wodurch sich die gefundenen Ergebnisse erklären ließen.

Kritisch zu betrachten ist an der vorliegenden Arbeit die relativ geringe Anzahl an Testpersonen für diese Untersuchung, gerade im Hinblick auf die verwendeten Auswertungsmethoden. So konnte für die Bearbeitung des Matrizentests nur eine Gesamtzahl von $N = 80$ Testpersonen gefunden werden, und für die „Formale Folgerichtigkeit“ sowie den PSB $N = 164$ Personen, wobei je nach Teilungskriterium noch mit zusätzlichen Drop-Outs gerechnet werden muss. Dies bedingt die Entscheidung gegen eine inhaltliche Analyse der Prüfung auf Rasch-Modell-Konformität des Matrizentests. In Anbetracht der gewählten Auswertung mittels Rasch-Modell sowie im Hinblick auf weiterführende Forschung wäre eine größere Stichprobengröße für alle Verfahren, allen voran den Matrizentest, wünschenswert, was jedoch den Rahmen der vorliegenden Arbeit überschritten hätte.

Die Stichprobengröße stellt ein besonderes Problem bei der Teilung der Untersuchungsdaten nach dem Arbeitsstil dar. Das Vorhandensein einer neutralen Kategorie bei der Beurteilung des Arbeitsverhaltens bedingt, dass, relativ gesehen, wenige Testpersonen als impulsiv oder reflexiv bewertet wurden. Bei einer Teilung nach dem Arbeitsstil II wurde diese geringe Anzahl an Testpersonen in den Extremkategorien in weiterer Folge zusätzlich durch die diversen Verfahren getrennt, was die Anzahl an Personen in den vergleichenden Gruppen noch verringerte. Eine Teilung der Stichprobe nach dem Arbeitsstil I schafft hier zwar Abhilfe; jedoch ist unklar, inwieweit sich das Zusammenfügen der neutralen mit der reflexiven Gruppe verzerrend auf die Ergebnisse auswirkt. Im Rahmen zukünftiger Studien zu dieser Thematik wäre auch unter diesem Gesichtspunkt eine größere Stichprobenanzahl empfehlenswert, um den Einfluss der Arbeitsstile differenzierter und genauer erfassen zu können.

Im Hinblick auf die Generalisierbarkeit der gewonnenen Ergebnisse ist kritisch zu betrachten, dass sämtliche Testpersonen ausschließlich an zwei Wiener Schulen akquiriert wurden. Dabei kann nicht sichergestellt werden, dass die gefundenen Effekte auch auf andere Kinder desselben Alters zutreffen. Hier wäre eine Forschung mit Datenmaterial aus dem gesamten deutschsprachigen Raum wünschenswert, um etwaige regionale Effekte ausschließen zu können und die Repräsentativität der Ergebnisse zu erhöhen.

Ein weiterer kritisch zu betrachtender Punkt der vorliegenden Arbeit ist die Wahl der Instrumente zur Erfassung des logisch-schlussfolgernden Denkens. Wie bereits mehrfach erwähnt, wäre eine genauere Erforschung der Auswirkungen der Charakteristika des spezifischen Antwortformats der „Formalen Folgerichtigkeit“ wünschenswert. Die Wahl des Matrizentests für diese Untersuchung bringt zwar den Vorteil, dass damit ein Verfahren zum Einsatz kam, das speziell in den beiden untersuchten Antwortformaten konzipiert wurde. Jedoch wurde somit auch ein Instrument eingesetzt, das noch kaum erprobt und überprüft ist und somit die mit diesem Verfahren gewonnenen Ergebnisse nur mit Vorbehalt zu interpretieren sind. Des Weiteren scheint unzureichend geklärt, inwieweit die Charakteristika des freien Antwortformats dieses Verfahrens die hier gefundenen Ergebnisse bedingen können. Auffallend an den erhobenen Daten ist, dass gerade für den PSB, der das einzige etablierte und vielfach erprobte Verfahren in dieser Untersuchung darstellt, hypothesenkonforme Ergebnisse gefunden wurden.

Gerade im Hinblick auf die Erfassung des Einflusses des Arbeitsstils wäre im Rahmen zukünftiger Forschung eine Auswahl an Verfahren wünschenswert, welches Itemmaterial enthält, das eine genauere Analyse des Inhaltes erfordert (siehe auch Zelniker & Jeffrey, 1976; Phillips & Rabbitt, 1995;).

Zusammenfassend lässt sich sagen, dass die vorab aufgestellten Hypothesen der vorliegenden Arbeit zu einem großen Teil nicht bestätigt werden konnten. Lediglich für den PSB zeigen sich mit dem Median der Testleistung sowie dem Alter als Teilungskriterium hypothesenkonforme Ergebnisse, die jedoch aufgrund der Stichprobengröße nur als Trend zu klassifizieren sind. Für den Matrizentest sowie die Testversionen der „Formalen Folgerichtigkeit“ zeigen sich größtenteils unauffällige und teilweise widersprüchliche

Ergebnisse. Insgesamt lässt sich sagen, dass zwar vereinzelte Trends dahingehend gefunden wurden, dass Kinder aufgrund von unausgereiften Kontrollstrategien und eines Arbeitsgedächtnisses, welches sich noch in der Entwicklung befindet, geringere Testleistungen bei der Bearbeitung von Aufgaben zum logischen Schlussfolgern im MC-Format erzielen. Dieser Trend scheint jedoch nicht auf alle Verfahren zu generalisieren sein. Weiterführende Studien zu dieser Thematik, mit einer größeren Stichprobe und eventuell anderen Verfahren, könnten hier genaueren Aufschluss über die Bedeutung des verwendeten Antwortformats bei der Bearbeitung von Aufgaben zur Erfassung des logisch-schlussfolgernden Denkens geben.

5. Zusammenfassung

Die bisherige Forschung im Bereich der kognitiven Determinanten für die Leistungsfähigkeit bei Aufgaben zum logisch-schlussfolgernden Denken spricht sowohl für eine maßgebliche Beteiligung der Kapazität des Arbeitsgedächtnisses als auch für die Bedeutung von generellen Kontrollstrategien (Baddeley & Hitch, 1974; Baddeley, 1986, 1996; Kyllonen & Christal, 1990; Embretson, 1995, 1998; Süß et al., 2004; Kane et al., 2004; Unsworth & Engle, 2004). Im Zusammenhang mit der Bearbeitung von Reasoning-Tasks konnte des Weiteren ein deutlicher Einfluss des individuellen Arbeitsstils auf die Testleistung nachgewiesen werden, wobei die Itemkonzeption bei der Stärke des gefundenen Effekts ebenfalls eine Rolle spielt (Kagan, Pearson & Welch, 1969; Zelniker & Jeffrey, 1976; Phillips & Rabbitt, 1995; Schweizer, 2001). Auffallend an der Forschung in diesem Themenbereich ist, dass kaum Studien im deutschsprachigen Raum durchgeführt wurden sowie bisher keine Ergebnisse über die relative Beteiligung genannter Konstrukte im Kindesalter vorliegen. Gerade im Hinblick auf die stetige kognitive Entwicklung von Kindern scheint jedoch eine Forschung in diesem Altersbereich besonders wichtig.

Das zentrale Ziel der vorliegenden Arbeit bestand darin, etwaige Unterschiede in der Bearbeitung von Aufgaben zum logischen Schlussfolgern in den Antwortformaten Multiple-Choice vs. freie Antwort im Altersbereich von 6- bis 9-jährigen Kindern zu erfassen. Dabei

wurde theoriegeleitet hypothetisiert, dass Kinder bei der Bearbeitung von Reasoning-Aufgaben im MC-Design schlechtere Leistungen erzielen als bei der Bearbeitung selbiger Aufgaben im freien Antwortformat. In weiteren Schritten wurde zusätzlich der Einfluss des Alters sowie der Arbeitsstile Impulsiv vs. Reflexiv auf die Aufgabenbearbeitung untersucht. Als Verfahren zur Erfassung des logischen Schlussfolgerns wurden im Rahmen der vorliegenden Arbeit ein Untertest aus dem PSB (Horn, 1969), ein Matrizentest sowie der Untertest „Formale Folgerichtigkeit“ aus dem AID 3 (Hagenmüller, 2011) eingesetzt. Jede Testperson bearbeitete dabei alle drei Verfahren sowohl im Multiple-Choice-Design als auch im freien Antwortformat, wobei einige Testversionen im Rahmen der vorliegenden Arbeit konzipiert und adaptiert wurden.

Zu diesem Zweck wurden, in Kooperation mit einer anderen Diplomarbeit (Heuschober, in Arbeit), an zwei Wiener Schulen Kinder im Altersbereich von 6 bis 9 Jahren akquiriert. Für die Bearbeitung des PSB sowie der „Formalen Folgerichtigkeit“ konnten insgesamt 164 und für den Matrizentest 80 Personen gewonnen werden. In einer randomisierten Vorgabe wurden alle Verfahren von den Kindern in einer einstündigen Einzeltestung an den jeweiligen Schulen im freien Antwortformat bearbeitet, während zu einem weiteren Termin die selben Aufgaben im MC-Format in einer Gruppentestung bearbeitet wurden. Dabei wurde der Arbeitsstil der Kinder von a priori instruierten sowie geschulten Testleitern erfasst.

Die Hypothesenprüfung erfolgte in mehreren Schritten, wobei zunächst zur globalen Unterschiedsprüfung T-Tests für gepaarte Stichproben gerechnet wurden. In weiteren Auswertungsschritten wurden die Items der Verfahren in beiden Antwortformaten kombiniert und zusammen Rasch-Modell-Analysen mit dem Median der Testleistung, dem Geschlecht, dem Alter sowie dem Arbeitsstil als Teilungskriterium unterzogen.

Zusammenfassend lässt sich sagen, dass ein Großteil der a priori getroffenen Hypothesen in dieser Studie nicht bestätigt werden konnte. Lediglich für den PSB mit dem Median der Testleistung sowie dem Alter als Teilungskriterium zeigt sich ein hypothesenkonformer Trend, während in den beiden anderen Verfahren keine Bestätigung dieser Ergebnisse gefunden werden konnte. Als mögliche Erklärung dieser unerwarteten Ergebnisse ist zum einen die Wahl der Erhebungsinstrumente zu nennen. So weist das Design des freien Antwortformats der „Formalen Folgerichtigkeit“ maßgebliche Charakteristika eines MC-

Designs auf, was u.U. einen größeren Einfluss auf die Bearbeitung hat als ursprünglich angenommen. Des Weiteren wurde der Matrizentest zwar eigens in den beiden Antwortformaten entwickelt, bis dato aber noch ungenügend erprobt und validiert. Im Hinblick auf die gewählten Auswertungsmethoden wäre eine größere Stichprobenanzahl wünschenswert gewesen, sodass etwaige Effekte, vor allem im Bezug auf den Einfluss des Arbeitsstils, differenzierter und genauer zu erfassen gewesen wären. Für die Erfassung des Einflusses des Arbeitsstils wäre im Bereich zukünftiger Forschungen eine Auswahl an Verfahren wünschenswert, welches Itemmaterial enthält, das von den Testpersonen eine genauere Analyse des Inhaltes erfordert, um mögliche Effekte besser erfassen zu können.

6. Literaturverzeichnis

- Baddeley, A.D. (1986) . *Working memory*. Oxford: Oxford University Press.
- Baddeley, A., (1996). Exploring the Central Executive. *The quarterly journal of experimental psychology*, 49A (1), 5-28. (1996)
- Baddeley, A.D., & Hitch, G.J. (1974) . Working memory. In Bower, G., H. (Ed.), *The psychology of learning and motivation*, Vol. 8 (pp. 47- 89). New York: Academic Press.
- Baddeley, A., D., & Hitch, G. (1994). Developments in the Concept of Working Memory. *Neuropsychology*, Vol. 8, No. 4, 485-493. (APA 1994)
- Beckmann, J., F., & Guthke, J. (1999). *Psychodiagnostik des schlußfolgernden Denkens: Handbuch zur Adaptiven Computergestützten Intelligenz-Lerntestbatterie für Schlußfolgerndes Denken (ACIL)*. Göttingen: Hogrefe.
- Carpenter, P., A., Just, M., A., Shell, P. (1990). What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test. *Psychological Review*, Vol. 97, No. 3, 404-431. (APA 1990)
- Cattell, R., B., Weiß, R., H., & Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1)*. Göttingen: Hogrefe.
- Embretson, S., E. (1995). The Role of Working Memory Capacity and General Control Processes in Intelligence. *Intelligence*, 20, 169- 189.
- Embretson, S., E. (1998). A Cognitive Design System Approach to Generating Valid Tests: Application to Abstract Reasoning. *Psychological Methods*, Vol. 3, No. 3, 380-396. (APA 1998)
- Goswami, U. (2001). *So denken Kinder - Einführung in die Psychologie der kognitiven Entwicklung*. Bern: Verlag Hans Huber.
- Hagenmüller, B. (2011). Entwicklung des Untertests „Formale Folgerichtigkeit“ zur Erfassung von Reasoning in der Intelligenz-Testbatterie AID 3. Unveröffentlichte Diplomarbeit, Universität Wien.
- Heuschober, C. (In Arbeit). Konstruktvalidierung des Untertests „Formale Folgerichtigkeit“ des AID 3. Unveröffentlichte Diplomarbeit, Universität Wien.
- Holyoak, K., J., & Morrison R., G. (2005). *The Cambridge Handbook of Thinking and Reasoning*. New York: Cambridge University Press.
- Horn, W. (1996). *Prüfsystem für Schul- und Bildungsberatung (P-S-B)*. Göttingen: Hogrefe.
- Horn, W., Lukesch, H., Mayrhofer, S., Kormann, A. (2004). *Prüfsystem für Schul- und Bildungsberatung für 6. Bis 13. Klassen (PSB-R 6-13)*. Göttingen: Hogrefe.

- Hossiep, R., Turck, D., & Hasella, M (1999). *BOMAT - advanced - Bochumer Matrizentest*. Göttingen: Hogrefe.
- Kagan, J., Pearson, L., Welch, L. (1966). Conceptual Impulsivity and Inductive Reasoning. *Child Development*, Vol. 37, No. 3, pp. 583-594, (1966)
- Kane, M., J., Hambrick, D., Z., Tuholski, S., W., Wilhelm, O., Payne, W., Engle, W. E. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology*, Vol. 133, No. 2, 189–217. (APA 2004)
- Kubinger, K.D. (2009a). *Adaptives Intelligenz Diagnostikum – Version 2.2 (AID 2) samt AID 2-Turkisch*. Göttingen: Beltz.
- Kubinger, K., D. (2009b). Applications of the Linear Logistic Test Model in Psychometric Research. *Educational and Psychological Measurement*, 69: 232 (2009)
- 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*, Volume 49, (4), p. 361-374. (2007)
- Kyllonen, P., C., & Christal, R., E. (1990). Reasoning Ability Is (Little More Than) Working-Memory Capacity?!. *Intelligence*, 14, 389–433. (1990)
- Mair, P., Hatzinger, R., Maier, M. (2011). Package „eRm“ Extended Rasch Modeling [online]. URL: cran.r-project.org/web/packages/eRm/eRm.pdf [23.07.2011].
- Petermann, F. & Petermann, U. (2008). *Hamburg-Wechsler-Intelligenztest für Kinder- IV (HAWIK-IV)*. Göttingen: Hogrefe.
- Phillips, L., H., & Rabbitt, P., M., A. (1995). Impulsivity and Speed-Accuracy Strategies in Intelligence Test Performance. *Intelligence*, 21, 13–29. (1995)
- Poinstingl, H., Mair, P., & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling)*. Lengerich: Pabst.
- Raven, J. C., Court, J., & Raven, J. (1988). *Standard Progressive Matrices* (2. Aufl). Weinheim: Beltz Test Gesellschaft.
- Raven J.C., Raven, J., & Court, J., H. (1998). *Raven's Progressive Matrices und Vocabulary Scales: Teil 4 Advanced Progressive Matrices*. Frankfurt: Swets Test Services
- Schweizer, K. (2001). Does impulsivity influence performance in reasoning?. *Personality and Individual Differences*, 33, 1031–1043. (2002)
- Süß, H.-M., Oberauer, K., Wittmann, W., W., Wilhelm, O., Schulze, R. (1996). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, 30, 261–288. (2002)
- Unsworth, N., & Engle, R., W. (2004). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, 33, 67–81. (2005)

- Zelniker, T., & Jeffrey, W., E. (1976). Reflective and Impulsive Children: Strategies of Information Processing Underlying Differences in Problem Solving. *Monographs of the Society for Research in Child Development*, Vol. 41, No. 5, pp. 1-59. (1976)
- Ziegler, A. (1994). *Die Entwicklung schlußfolgernden Denkens*. Frankfurt am Main: Europäischer Verlag der Wissenschaften.

7. Anhang

Abbildungsverzeichnis

Abbildung 1: Anfängliche Aufgabenreihe aus dem UT 3 des PSB; Horn, 1969.....	43
Abbildung 2: Letzte Aufgabenreihe aus dem UT 3 des PSB; Horn, 1969.....	43
Abbildung 3: Aufgabe 2 aus der adaptierten Testversion des PSB mit freiem Antwortformat	45
Abbildung 4: Aufgabe 35 aus der adaptierten Testversion des PSB mit freiem Antwortformat	46
Abbildung 5: Aufgabe 2 aus der adaptierten Testversion des PSB mit Multiple-Choice- Antwortformat	48
Abbildung 6: Aufgabe 35 aus der adaptierten Testversion des PSB mit Multiple-Choice- Antwortformat	49
Abbildung 7: Beispielaufgabe aus dem Matrizentest (freies Antwortformat; Seite 2)	52
Abbildung 8: Aufgabe 12 aus dem Matrizentest (freies Antwortformat; Seite 6)	53
Abbildung 9: Aufgabe 17 aus dem Matrizentest (freies Antwortformat; Seite 9)	54
Abbildung 10: Beispielitem aus dem Matrizentest (MC-Version; Seite 2)	55
Abbildung 11: Aufgabe 12 aus dem Matrizentest (MC-Version; Seite 6).....	56
Abbildung 12: Aufgabe 17 aus dem Matrizentest (MC-Version; Seite 9).....	57
Abbildung 13: Darstellung der Antwortsteine aus dem UT „Formale Folgerichtigkeit“	59
Abbildung 14: Aufgabe 5 aus dem UT „Formale Folgerichtigkeit“; Testform 1	61
Abbildung 15: Aufgabe 11 aus dem UT „Formale Folgerichtigkeit“; Testform 3	62
Abbildung 16: Aufgabe 5 aus dem UT „Formale Folgerichtigkeit“; Multiple-Choice- Format	64
Abbildung 17: Aufgabe 11 aus dem UT „Formale Folgerichtigkeit“; Multiple-Choice- Format	65
Abbildung 18: Auszug aus dem Klassifikationsbogen des Arbeitsstils.....	67

Abbildung 19: Geschlechterverteilung über alle Verfahren.....	71
Abbildung 20: Grafische Darstellung der Altersverteilung über alle Verfahren	72
Abbildung 21: Verteilung der Testpersonen auf beide Schulen	73
Abbildung 22: Verteilung des Arbeitsstils in der Stichprobe.....	74
Abbildung 23: Aufgabe 30 aus der adaptierten Testversion des PSB mit Multiple-Choice-Antwortformat	87
Abbildung 24: Aufgabe 29 aus der adaptierten Testversion des PSB mit Multiple-Choice-Antwortformat	87

Tabellenverzeichnis

Tabelle 1: Deskriptive Statistik des Alters über alle Verfahren.....	72
Tabelle 2: Darstellung des Testdesigns der Rasch-Modell-Analysen.....	77
Tabelle 3: SPSS-Output der Statistik bei gepaarten Stichproben für den Matrizentest.....	80
Tabelle 4: SPSS-Output des T-Tests für gepaarte Stichproben für den Matrizentest.....	80
Tabelle 5: SPSS-Output der Statistik bei gepaarten Stichproben für den PSB (korrigierte Version).....	81
Tabelle 6: SPSS-Output des T-Tests für gepaarte Stichproben für den PSB (korrigierte Version).....	81
Tabelle 7: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6 bis 7 J.....	83
Tabelle 8: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 6 bis 7 J.....	83
Tabelle 9: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.....	84
Tabelle 10: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.....	84
Tabelle 11: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Median der Testleistung.....	85
Tabelle 12: R-Output des Wald-Tests für den PSB; Teilungskriterium: Median der Testleistung.....	89
Tabelle 13: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Geschlecht.....	90
Tabelle 14: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Median der Testleistung.....	91
Tabelle 15: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Geschlecht.....	92
Tabelle 16: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Median des Alters.....	94

Tabelle 17: R-Output des Wald-Tests für den PSB; Teilungskriterium: Median des Alters	95
Tabelle 18: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Median des Alters	97
Tabelle 19: R-Output des Wald-Tests für den UT „Formale Folgerichtigkeit“; Teilungskriterium: Median des Alters.....	98
Tabelle 20: SPSS-Output der Statistik bei gepaarten Stichproben für den Matrizentest und den PSB (korrigierte Version); Arbeitsstil: Impulsiv.....	102
Tabelle 21: SPSS Output des T-Tests für gepaarten Stichproben für den Matrizentest und den PSB (Korrigierte Version); Arbeitsstil: Impulsiv.....	102
Tabelle 22: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6 bis 7 J.; Arbeitsstil: Impulsiv	103
Tabelle 23: SPSS-Output des T-Tests für gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6-7 J.; Arbeitsstil: Impulsiv.....	103
Tabelle 24: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Impulsiv	104
Tabelle 25: : SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Impulsiv	105
Tabelle 26: SPSS-Output der Statistik bei gepaarten Stichproben für den Matrizentest und den PSB (korrigierte Version); Arbeitsstil: Reflexiv	106
Tabelle 27: SPSS-Output des T-Tests für gepaarte Stichproben für den Matrizentest und den PSB (korrigierte Version); Arbeitsstil: Reflexiv	107
Tabelle 28: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 6-7 J.; Arbeitsstil: Reflexiv	108
Tabelle 29: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 6-7 J.; Arbeitsstil: Reflexiv	108
Tabelle 30: SPSS-Output der Statistik bei gepaarten Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Reflexiv	109
Tabelle 31: SPSS-Output des T-Tests für gepaarte Stichproben für den UT „Formale Folgerichtigkeit“; Version 8 bis 11 J.; Arbeitsstil: Reflexiv	109
Tabelle 32: R-Output des LR-Tests der Rasch-Modell-Analyse des PSB; Teilungskriterium: Arbeitsstil I.....	111
Tabelle 33: R-Output des LR-Tests der Rasch-Modell-Analyse des UT „Formale Folgerichtigkeit“; Teilungskriterium: Arbeitsstil I	112

Tabelle: Relative Verteilung des Alters über alle Verfahren

Alter der Kinder

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	6,10	4	2,4	2,5	2,5
	6,11	1	,6	,6	3,1
	6,50	1	,6	,6	3,7
	6,60	2	1,2	1,2	4,9
	6,70	2	1,2	1,2	6,1
	6,80	1	,6	,6	6,7
	6,90	3	1,8	1,8	8,6
	7,00	17	10,4	10,4	19,0
	7,10	7	4,3	4,3	23,3
	7,11	1	,6	,6	23,9
	7,20	3	1,8	1,8	25,8
	7,30	1	,6	,6	26,4
	7,50	1	,6	,6	27,0
	7,60	3	1,8	1,8	28,8
	7,70	3	1,8	1,8	30,7
	7,80	3	1,8	1,8	32,5
	7,90	5	3,0	3,1	35,6
	8,00	21	12,8	12,9	48,5
	8,10	6	3,7	3,7	52,1
	8,11	4	2,4	2,5	54,6
	8,12	2	1,2	1,2	55,8
	8,20	1	,6	,6	56,4
	8,30	4	2,4	2,5	58,9
	8,40	5	3,0	3,1	62,0
	8,50	13	7,9	8,0	69,9
	8,60	8	4,9	4,9	74,8
	8,70	2	1,2	1,2	76,1
	8,80	6	3,7	3,7	79,8
	8,90	10	6,1	6,1	85,9
	9,00	8	4,9	4,9	90,8
	9,10	3	1,8	1,8	92,6
	9,20	4	2,4	2,5	95,1
	9,30	4	2,4	2,5	97,5
	9,40	1	,6	,6	98,2
	9,60	1	,6	,6	98,8
	9,80	1	,6	,6	99,4
	9,90	1	,6	,6	100,0
	Gesamt	163	99,4	100,0	
Fehlend	System	1	,6		
Gesamt		164	100,0		

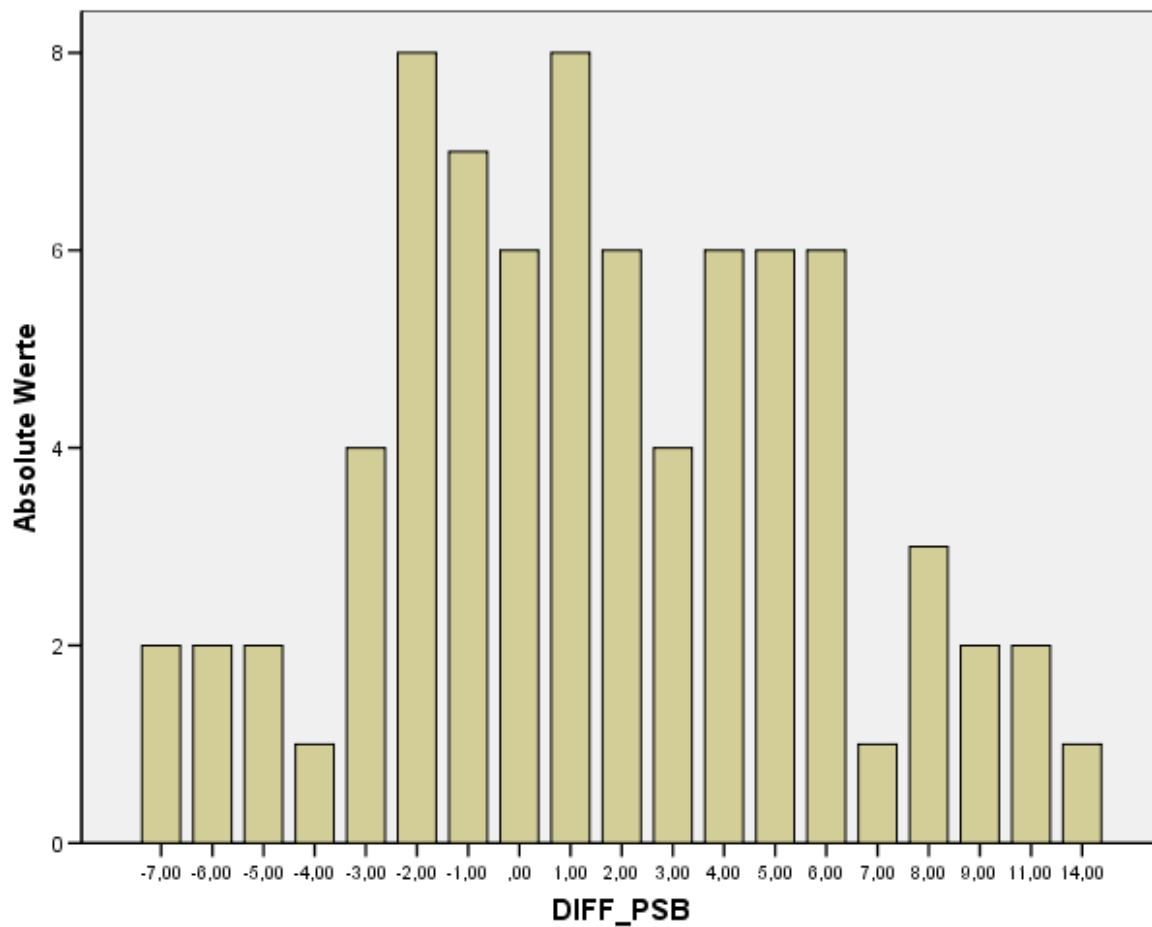


Abbildung: Grafische Modellkontrolle der Normalverteilung der Differenzen zwischen der MC-Version und freier Antwort des PSB (korrigierte Version)

Tabelle: SPSS Output des KS-Tests zur Überprüfung der Normalverteilung der Differenzen zwischen der MC-Version und freier Antwort des PSB (korrigierte Version)

Kolmogorov-Smirnov-Anpassungstest

		DIFF_PSB
N		77
Parameter der Normalverteilung(a,b)	Mittelwert	1,7143
	Standardabweichung	4,37654
Extremste Differenzen	Absolut	,084
	Positiv	,084
	Negativ	-,055
Kolmogorov-Smirnov-Z		,740
Asymptotische Signifikanz (2-seitig)		,644

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

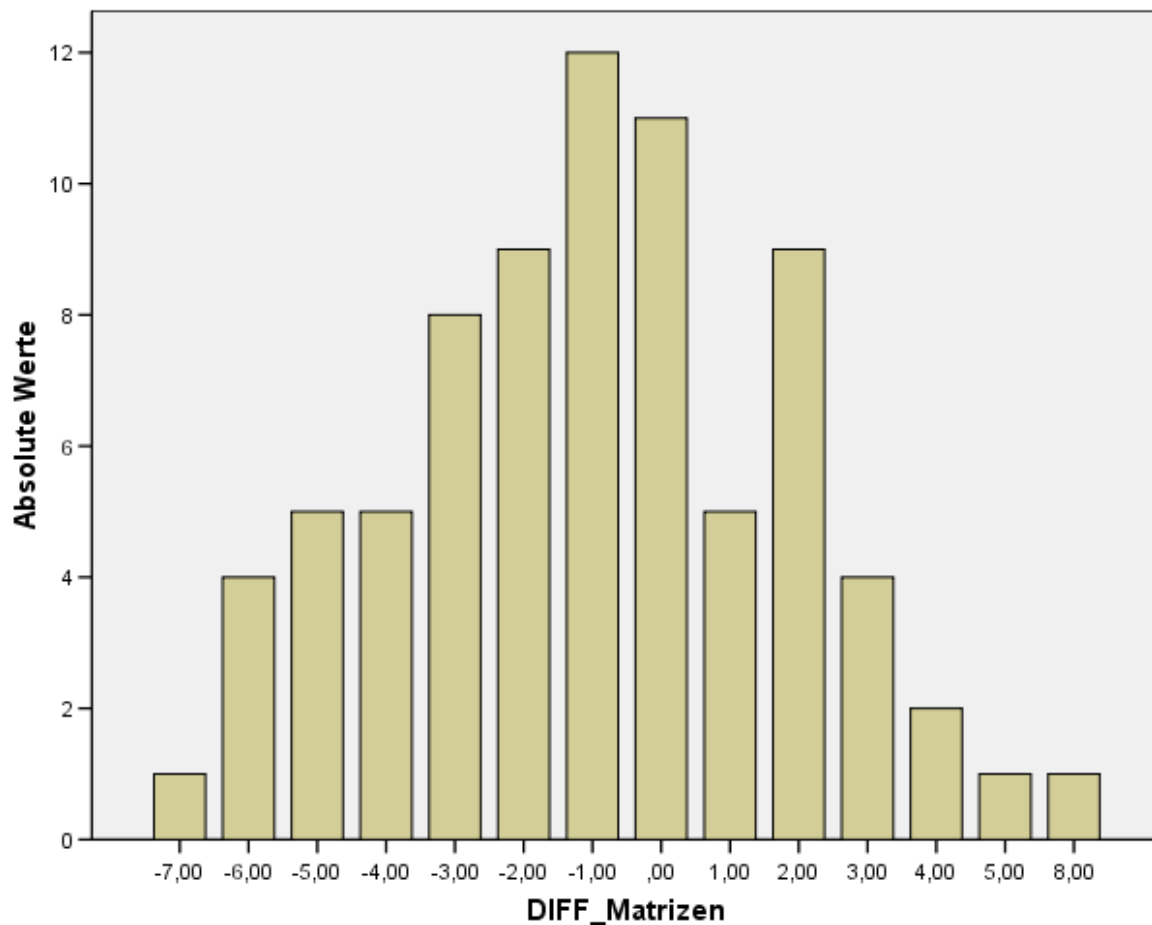


Abbildung: Grafische Modellkontrolle der Normalverteilung der Differenzen zwischen der MC-Version und freier Antwort des Matrizentests

Tabelle: SPSS Output des KS-Tests zur Überprüfung der Normalverteilung der Differenzen zwischen der MC-Version und freier Antwort des Matrizentests

Kolmogorov-Smirnov-Anpassungstest

		DIFF_Matrizen
N		77
Parameter der Normalverteilung(a,b)	Mittelwert	-,9610
	Standardabweichung	2,91295
Extremste Differenzen	Absolut	,085
	Positiv	,085
	Negativ	-,079
Kolmogorov-Smirnov-Z		,746
Asymptotische Signifikanz (2-seitig)		,634

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

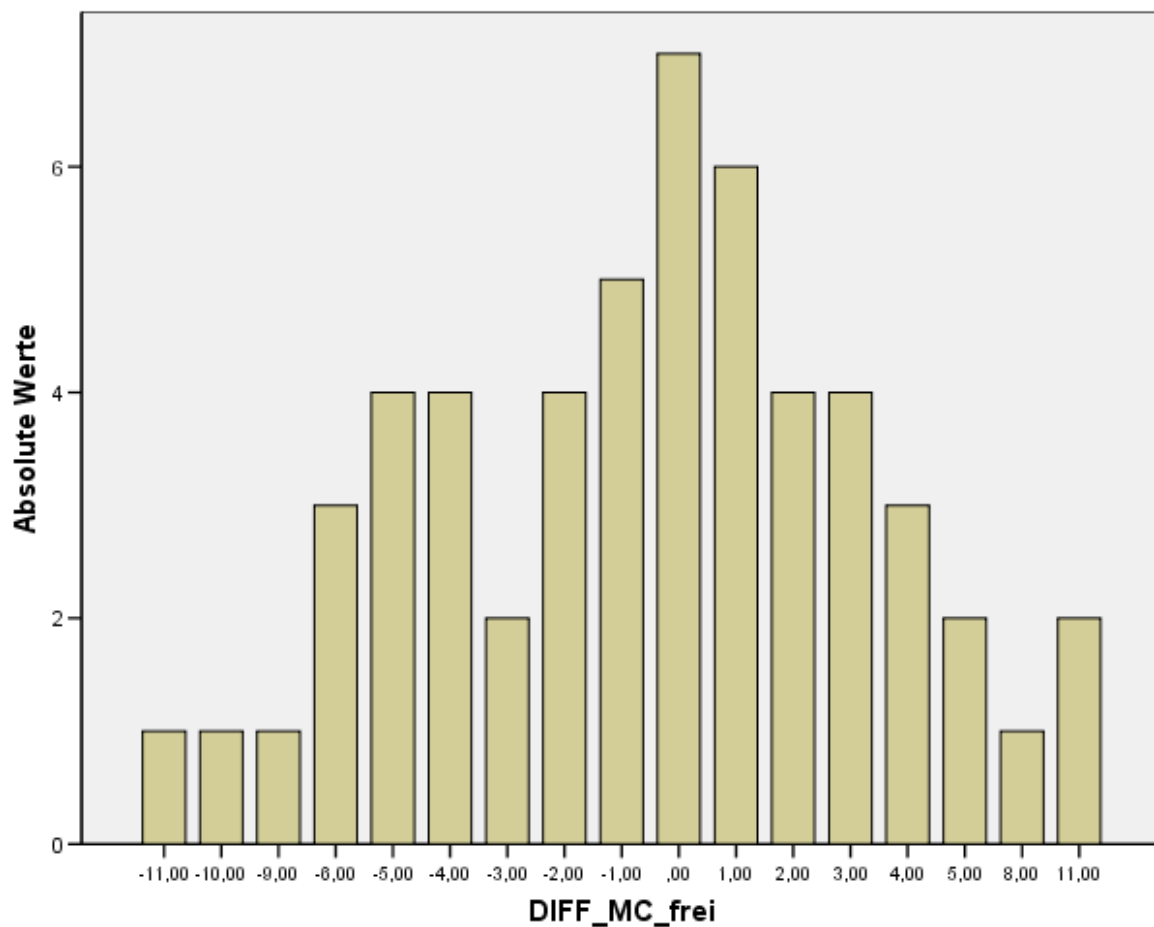


Abbildung: Grafische Modellkontrolle der Normalverteilung der Differenzen zwischen der MC-Version und freier Antwort des UT „Formale Folgerichtigkeit“; Version 6-7 J.;

Tabelle: SPSS Output des KS-Tests zur Überprüfung der Normalverteilung der Differenzen zwischen der MC-Version und des UT „Formale Folgerichtigkeit“; Version 6-7 J.;

Kolmogorov-Smirnov-Anpassungstest

		DIFF MC frei
N		54
Parameter der Normalverteilung(a,b)	Mittelwert	-,4630
	Standardabweichung	4,47936
Extremste Differenzen	Absolut	,082
	Positiv	,076
	Negativ	-,082
Kolmogorov-Smirnov-Z		,602
Asymptotische Signifikanz (2-seitig)		,862

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

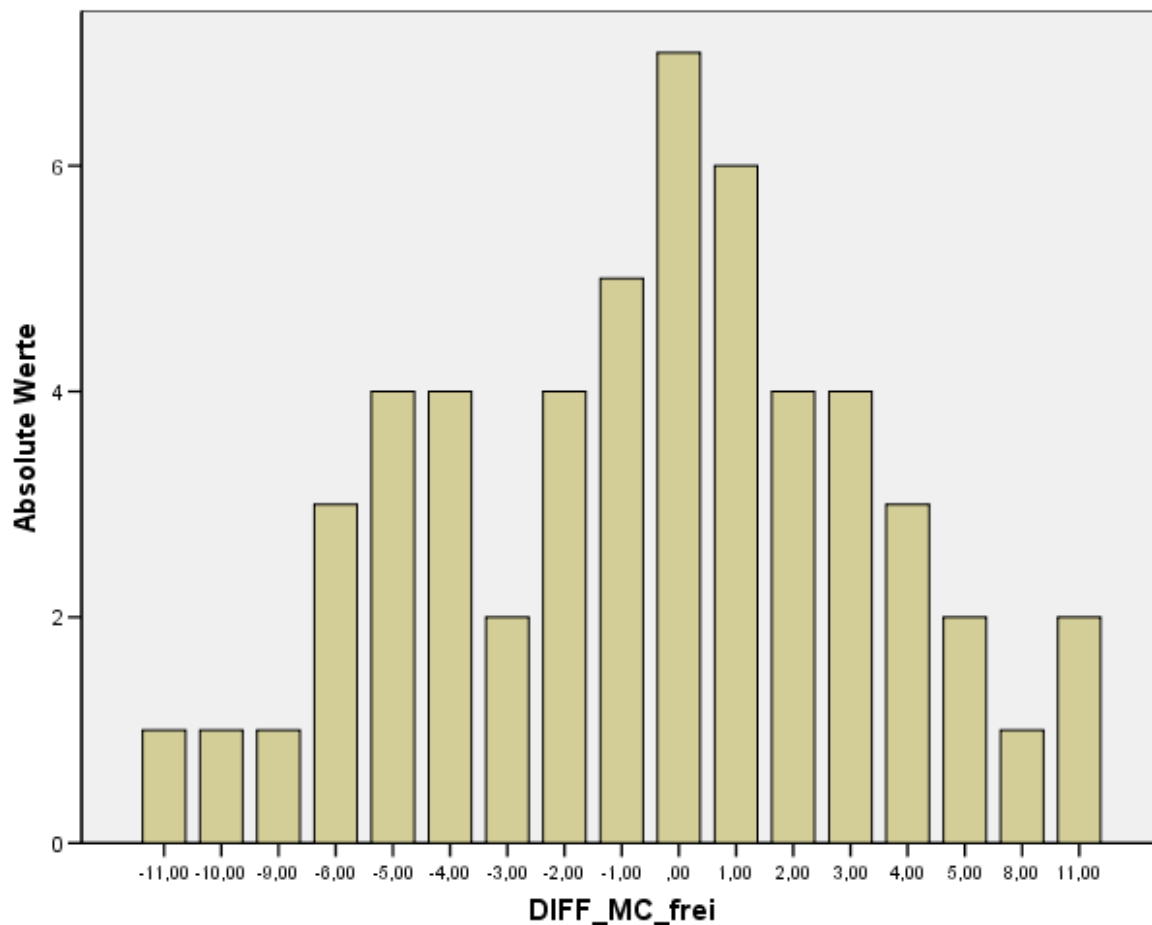


Abbildung: Grafische Modellkontrolle der Normalverteilung der Differenzen zwischen der MC-Version und freier Antwort des UT „Formale Folgerichtigkeit“; Version 8-11 J.;

Tabelle: SPSS Output des KS-Tests zur Überprüfung der Normalverteilung der Differenzen zwischen der MC-Version und des UT „Formale Folgerichtigkeit“; Version 8-11 J.;

Kolmogorov-Smirnov-Anpassungstest

		DIFF_MC_frei
N		54
Parameter der Normalverteilung(a,b)	Mittelwert	-,4630
	Standardabweichung	4,47936
Extremste Differenzen	Absolut	,082
	Positiv	,076
	Negativ	-,082
Kolmogorov-Smirnov-Z		,602
Asymptotische Signifikanz (2-seitig)		,862

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

Elternbrief



Johann Ecker
Diplomand an der Fakultät für Psychologie
Universität Wien
E-Mail: a0404089@unet.univie.ac.at
Tel.: +43 (0) 699 1922 5646

Carina Heuschober
Diplomandin an der Fakultät für Psychologie
Universität Wien
E-Mail: carina.heuschober@st.ursula-wien.at
Tel.: +43 (0) 650/4171917

Wien, Oktober 2010

Liebe Eltern!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Im Rahmen schulpsychologischer Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Schullaufbahnberatungen, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme, etc. kommt dabei meist u.a. ein Verfahren zur Erfassung des logischen Schlussfolgerns zum Einsatz. Diese Verfahren sind zumeist mit einem *multiple-choice* Antwortformat gestaltet, die sich dadurch auszeichnen, dass die Kinder aus einer Reihe von verfügbaren Antwort-Alternativen jene auswählen, die sie für richtig halten. Dem gegenüber stehen psychologisch-diagnostische Instrumente, bei denen ohne eine Auswahl von Antwortalternativen, die Lösung der Aufgabe von den Kindern gefunden werden muss – das sogenannte freie Antwortformat.

Im Rahmen unserer Diplomarbeit an der Fakultät für Psychologie, Universität Wien (betreut von Dr. Stefana Holocher-Ertl), führe ich nun in den nächsten Wochen mit Einverständnis der Direktion und der Lehrkräfte in der Klasse Ihres Kindes eine Testung mit einigen Aufgaben aus Verfahren zur Erfassung des logischen Schlussfolgerns (*PSB, Horn, W., 2003; AID-3 LB, Kubinger, 2009, Matritzentest in Erprobung; CFT 1, Cattell, Weiß und Osterland, 1997*) durch.

Ziel dieser Untersuchung ist es, herauszufinden, ob Leistungsunterschiede zwischen den Antwortformaten *multiple-choice* und *freie Antwort* bestehen.

Forschungen dieser Art sind wichtig, um den Standard psychologischer Diagnostik laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

Ich wende mich daher mit der Bitte an Sie, Ihr Kind an dieser Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, Ihr Kind ist einverstanden. Die Untersuchung findet zu zwei Zeitpunkten während der Schulzeit statt und dauert ca. jeweils eine Schulstunde. Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß.

Für eventuelle Rückfragen erreichen Sie uns gerne telefonisch oder per E-Mail:

Johann Ecker
E-Mail: a0404089@unet.univie.ac.at
Tel.: +43 (0) 699 1922 5646

Carina Heuschober
E-Mail: carina.heuschober@st.ursula-wien.at
Tel.: +43 (0) 650/4171917

Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke genutzt. Sämtliche Ergebnisse der Schüler(innen) werden (noch während der Testung) anonymisiert.

Ich bitten Sie, mit Ihrer Unterschrift auf dem beiliegenden Formular, Ihr Einverständnis zur Teilnahme Ihres Kindes an der oben beschriebenen Untersuchung zu erteilen.
Mit freundlichen Grüßen und herzlichen Dank im Voraus!

Carina Heuschober,
Johann Ecker

.....bitte hier abtrennen.....

Ich erkläre mich mit der Teilnahme meiner Tochter/meines
Sohnes _____, geboren am _____, an der *Testung*
für die Diplomarbeit von Carina Heuschober und Johann Ecker einverstanden.

.....
Datum, Unterschrift

LehrerInnenbrief



Johann Ecker
Diplomand an der Fakultät für Psychologie
Universität Wien
E-Mail: a0404089@unet.univie.ac.at
Tel.: +43 (0) 699 1922 5646

Carina Heuschober
Diplomandin an der Fakultät für Psychologie
Universität Wien
E-Mail: carina.heuschober@st.ursula-wien.at
Tel.: +43 (0) 650/4171917

Wien, Oktober 2010

Liebe Lehrende!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Im Rahmen schulpsychologischer Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Schullaufbahnberatungen, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme, etc. kommt dabei meist u.a. ein Verfahren zur Erfassung des logischen Schlussfolgerns zum Einsatz. Diese Verfahren sind zumeist mit einem multiple-choice Antwortformat gestaltet, die sich dadurch auszeichnen, dass die Kinder aus einer Reihe von verfügbaren Antwort-Alternativen jene auswählen, die sie für richtig halten. Dem gegenüber stehen psychologisch-diagnostische Instrumente, bei denen ohne eine Auswahl von Antwortalternativen, die Lösung der Aufgabe von den Kindern gefunden werden muss – das sogenannte freie Antwortformat.

Im Rahmen unserer Diplomarbeit an der Fakultät für Psychologie, Universität Wien (betreut von Dr. Stefana Holocher-Ertl), führe ich nun in den nächsten Wochen mit Einverständnis der Direktion und der Lehrkräfte in der Klasse Ihres Kindes eine Testung mit einigen Aufgaben aus Verfahren zur Erfassung des logischen Schlussfolgerns (PSB, Horn, W., 2003; AID-3 LB, Kubinger, 2009, Matritzentest in Erprobung; CFT 1, Cattell, Weiß und Osterland, 1997) durch.

Ziel dieser Untersuchung ist es, herauszufinden, ob Leistungsunterschiede zwischen den Antwortformaten multiple-choice und freie Antwort bestehen.

Forschungen dieser Art sind wichtig, um den Standard psychologischer Diagnostik laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

Ich wende mich daher mit der Bitte an die Eltern Ihrer Schüler(innen), diese an der Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, dass die Kinder damit einverstanden sind. Die Untersuchung findet zu zwei Zeitpunkten während der Schulzeit statt und dauert ca. jeweils eine Schulstunde. Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß.

Wir möchten Sie bitten, die Elternbriefe an Ihre Schüler(innen) zu übergeben und die Rückmeldungen der Eltern einzusammeln. Wenn sich Eltern Ihrer Schüler(innen) mit der Teilnahme an der Untersuchung einverstanden erklären, werden wir uns bei der Terminvereinbarung sehr bemühen, den Ablauf Ihres Unterrichts so wenig wie möglich zu stören.

Für eventuelle Rückfragen erreichen Sie uns gerne telefonisch oder per E-Mail:

Johann Ecker

Carina Heuschober

E-Mail: a0404089@unet.univie.ac.at
Tel.: +43 (0) 699 1922 5646

E-Mail: carina.heuschober@st.ursula-wien.at
Tel.: +43 (0) 650/4171917

Die gewonnen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke genutzt. Sämtliche Ergebnisse der Schüler(innen) werden (noch während der Testung) anonymisiert.

Mit freundlichen Grüßen und herzlichen Dank im Voraus!

Carina Heuschober,
Johann Ecker

Verhaltensbeobachtung

Testleiter:

Name/ Code des Probanden:

Geburtsdatum:

Testdatum:

Facette zum **reflexiven Arbeitsstil**:

Bearbeitet fordernde Aufgaben der Tests zum logischen Schlussfolgern erst nachdem es sich Zeit genommen hat für eine genaue Analyse des Problems. Achtet dann auf Details und macht wenige Fehler.
--

Tendiert dazu, Antworten zu überdenken und gegebenenfalls zu korrigieren.

Facette zum **impulsiven Arbeitsstil**:

Bearbeitet fordernde Aufgaben der Tests zum logischen Schlussfolgern direkt, ohne sich vorher Zeit für eine genaue Analyse des Problems zu nehmen. Achtet dann nicht auf Details und macht vermehrt Fehler.

Gegebene Antworten werden kaum überdacht. Verweilt nicht lange bei der Bearbeitung einzelner Items.



Beobachtungen während der Einzeltestung mit den **Logischen Blöcken** zum Arbeitsstil:

Wie ist das Vorbereitungs- und Antwortverhalten bei fordernden Aufgaben?

Reflexiv

Eine längere Betrachtungszeit und vereinzelte (richtige) Antworten deuten auf einen reflexiven Arbeitsstil hin. Wenn das Kind nach dem Nennen einer Lösung, diese noch einmal kontrolliert, so deutet dies ebenfalls auf einen reflexiven Arbeitsstil hin.

Impulsiv

Zeichnet sich durch sehr geringe Betrachtungszeiten mit sofortigem Antworten aus. Das Hantieren mit nur wenigen Bausteinen, sowie eine geringe Tendenz Antworten zu korrigieren, sind Hinweise auf einen impulsiven Arbeitsstil.

Beobachtungen während der Einzeltestung mit dem **PSB** zum Arbeitsstil:
Wie ist das Vorbereitungs- und Antwortverhalten bei fordernden Aufgaben?

Reflexiv

Eine längere Betrachtungszeit und vereinzelte (richtige) Antworten deuten auf einen reflexiven Arbeitsstil hin. Wenn das Kind nach dem zeichnen einer Antwort, dazu tendiert diese noch einmal zu kontrollieren und auszubessern, so deutet dies ebenfalls auf einen reflexiven Arbeitsstil hin.

Impulsiv

Zeichnet sich durch sehr geringe Betrachtungszeiten mit sofortigem Antworten aus. Achtet bei der Bearbeitung nicht auf Details. Eine geringe Tendenz Antworten zu korrigieren und auszubessern, deutet ebenfalls auf einen impulsiven Arbeitsstil hin.

Beobachtungen während der Einzeltestung mit dem **Matrizentest** zum Arbeitsstil:

Wie ist das Vorbereitungs- und Antwortverhalten bei fordernden Aufgaben?

Reflexiv

Eine längere Betrachtungszeit und vereinzelte (richtige) Antworten deuten auf einen reflexiven Arbeitsstil hin. Wenn das Kind nach dem zeichnen einer Antwort, dazu tendiert diese noch einmal zu kontrollieren und auszubessern, so deutet dies ebenfalls auf einen reflexiven Arbeitsstil hin.

Impulsiv

Zeichnet sich durch sehr geringe Betrachtungszeiten mit sofortigem Antworten aus. Achtet bei der Bearbeitung nicht auf Details. Eine geringe Tendenz Antworten zu korrigieren und auszubessern, deutet ebenfalls auf einen impulsiven Arbeitsstil hin.

Arbeitsstil

Unterdurchschnittlich reflexiv/ Überdurchschnittlich impulsiv	Durchschnittlich reflexiv/ Durchschnittlich impulsiv	Überdurchschnittlich reflexiv/ Unterdurchschnittlich impulsiv
--	---	--

Weil: _____

Notizen:

Testanordnung

RITA Randomisierungsergebnisse - Seite 1 von 1

So., 21.11.2010 - 21:13

Randomisierungsparameter:

Programmversion: RITA Version 1.23 [ACHTUNG! Testversion]

Randomisierungsverfahren: Vollstaendige Randomisierung

Zufallszahlen-Generator: MersenneTwister

Seed-Generator: echter Zufalls-Seed (www.random.org)

Seed: -487236357

Therapiegruppe 1: Einzel-Gruppe

Therapiegruppe 2: Gruppe Einzel

Anzahl Personen: 4

Absolute Differenz: 0

Klasse 1: Einzel-Gruppe (21.11.2010, 21:08)

Klasse 2: Gruppe Einzel (21.11.2010, 21:08)

Klasse 3: Gruppe Einzel (21.11.2010, 21:08)

Klasse 4: Einzel-Gruppe (21.11.2010, 21:08)

Generiert mit RITA - Copyright © A. Ziegler & I.R. König 2004-2009

Testanordnung der Einzeltestungen

RITA Randomisierungsergebnisse - Seite 1 von 4

So., 28.11.2010 - 19:41

Randomisierungsparameter:

Programmversion: RITA Version 1.23 [ACHTUNG! Testversion]

Randomisierungsverfahren: Urnenmodell UD(w,a,b) von WEI

Parameter 1: w: 1

Parameter 2: a: 2

Parameter 3: b: 4

Zufallszahlen-Generator: MersenneTwister

Seed-Generator: echter Zufalls-Seed (www.random.org)

Seed: 153153985

Therapiegruppe 1: PSB-LogischeBlöcke-Matrizen

Therapiegruppe 2: PSB-Matrizen-LogischeBlöcke

Therapiegruppe 3: LogischeBlöcke-Matrizen-PSB

Therapiegruppe 4: LogischeBlöcke-PSB-Matrizen

Therapiegruppe 5: Matrizen-LogischeBlöcke-PSB

Therapiegruppe 6: Matrizen-PSB-LogischeBlöcke

Anzahl Personen: 80

Absolute Differenz: 5

Person 1: Matrizen-PSB-LogischeBlöcke (28.11.2010, 19:27)

Person 2: PSB-Matrizen-LogischeBlöcke (28.11.2010, 19:27)

Person 3: PSB-Matrizen-LogischeBlöcke (28.11.2010, 19:27)

Person 4: PSB-LogischeBlöcke-Matrizen (28.11.2010, 19:27)

Person 5: Matrizen-LogischeBlöcke-PSB (28.11.2010, 19:27)

Person 6: Matrizen-LogischeBlöcke-PSB (28.11.2010, 19:27)

Person 7: Matrizen-LogischeBlöcke-PSB (28.11.2010, 19:27)

Person 8: Matrizen-PSB-LogischeBlöcke (28.11.2010, 19:27)

Generiert mit RITA - Copyright © A. Ziegler & I.R. König 2004-2009

So., 28.11.2010 - 19:41

Person 9:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 10:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 11:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 12:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 13:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 14:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 15:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 16:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 17:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 18:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 19:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 20:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 21:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 22:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 23:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 24:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 25:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 26:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 27:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 28:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 29:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 30:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 31:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 32:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 33:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 34:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 35:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 36:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)

Generiert mit RITA - Copyright © A. Ziegler & I.R. König 2004-2009

So., 28.11.2010 - 19:41

Person 37:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 38:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 39:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 40:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 41:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 42:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 43:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 44:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 45:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 46:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 47:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 48:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 49:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 50:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 51:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 52:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 53:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 54:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 55:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 56:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 57:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 58:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 59:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 60:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 61:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 62:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 63:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 64:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)

Generiert mit RITA - Copyright © A. Ziegler & I.R. König 2004-2009

So., 28.11.2010 - 19:41

Person 65:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 66:	Matrizen-LogischeBlöcke-PSB	(28.11.2010, 19:27)
Person 67:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 68:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 69:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 70:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 71:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 72:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 73:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 74:	PSB-Matrizen-LogischeBlöcke	(28.11.2010, 19:27)
Person 75:	LogischeBlöcke-PSB-Matrizen	(28.11.2010, 19:27)
Person 76:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 77:	Matrizen-PSB-LogischeBlöcke	(28.11.2010, 19:27)
Person 78:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)
Person 79:	LogischeBlöcke-Matrizen-PSB	(28.11.2010, 19:27)
Person 80:	PSB-LogischeBlöcke-Matrizen	(28.11.2010, 19:27)

Generiert mit RITA - Copyright © A. Ziegler & I.R. König 2004-2009

Testanordnung der Gruppentestung

Hier sind mit den Personen Schulklassen gemeint.

RITA Randomisierungsergebnisse - Seite 1 von 4

So., 21.11.2010 - 21:18

Randomisierungsparameter:

Programmversion: RITA Version 1.23 [ACHTUNG! Testversion]

Randomisierungsverfahren: Urnenmodell UD(w,a,b) von WEI

Parameter 1: w: 1

Parameter 2: a: 2

Parameter 3: b: 4

Zufallszahlen-Generator: MersenneTwister

Seed-Generator: echter Zufalls-Seed (www.random.org)

Seed: 1417763073

Therapiegruppe 1: PSB - Logische Blöcke - Matrizen

Therapiegruppe 2: PSB - Matrizen - Logische Blöcke

Therapiegruppe 3: Logische Blöcke - Matrizen - PSB

Therapiegruppe 4: Logische Blöcke - PSB - Matrizen

Therapiegruppe 5: Matrizen - Logische Blöcke - PSB

Therapiegruppe 6: Matrizen - PSB - Logische Blöcke

Anzahl Personen: 80

Absolute Differenz: 8

Person 1:	PSB - Matrizen - Logische Blöcke	(21.11.2010, 21:16)
Person 2:	PSB - Logische Blöcke - Matrizen	(21.11.2010, 21:16)
Person 3:	Logische Blöcke - PSB - Matrizen	(21.11.2010, 21:16)
Person 4:	Logische Blöcke - PSB - Matrizen	(21.11.2010, 21:16)
Person 5:	Matrizen - PSB - Logische Blöcke	(21.11.2010, 21:16)
Person 6:	Matrizen - Logische Blöcke - PSB	(21.11.2010, 21:16)
Person 7:	Logische Blöcke - Matrizen - PSB	(21.11.2010, 21:16)
Person 8:	Matrizen - Logische Blöcke - PSB	(21.11.2010, 21:16)

Adaptierte Testversion des UT 3 aus dem PSB mit freiem Antwortformat

PSB

Freie Antworten

Name:
Geburtsdatum:
Testdatum:

Auf den folgenden Seiten siehst du Reihen mit verschiedenen Symbolen.

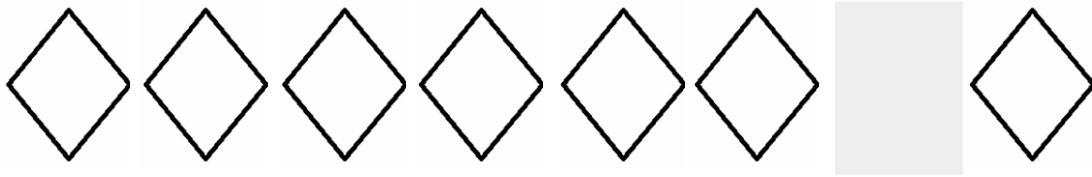
In jeder Reihe befindet sich ein graues Kästchen. In diesem fehlt ein Symbol, das die Reihe logisch ergänzt.

Versuche nun das Symbol zu finden, das am besten in die Reihe passt!

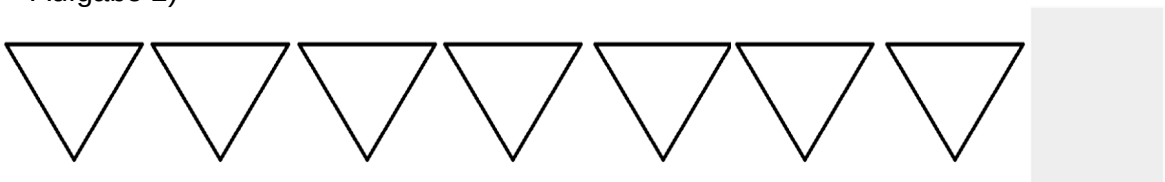
Wenn du die passende Antwort gefunden hast, dann zeichne diese bitte in das graue Kästchen.

Viel Erfolg und viel Spaß!

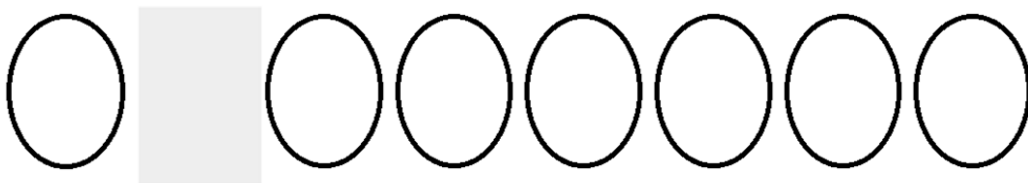
Beispielaufgabe:



Aufgabe 2)



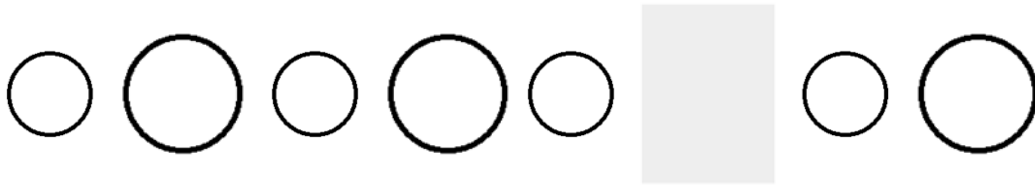
Aufgabe 3)



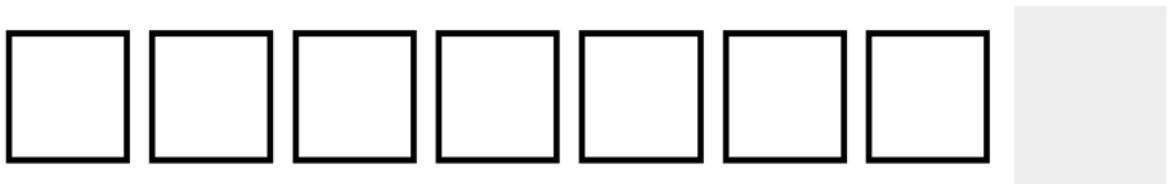
Aufgabe 4)



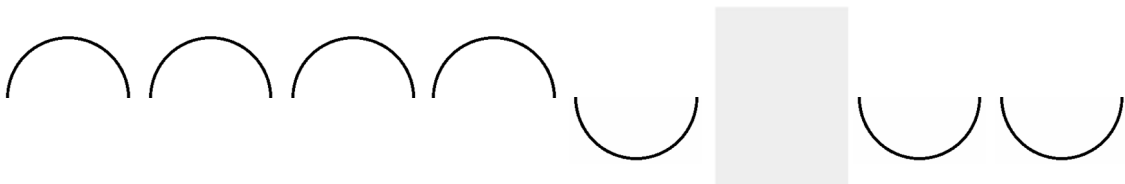
Aufgabe 5)



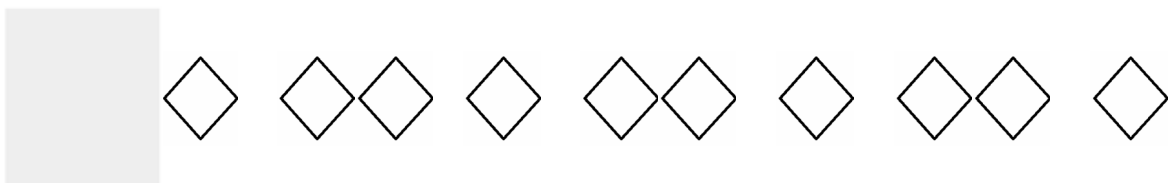
Aufgabe 6)



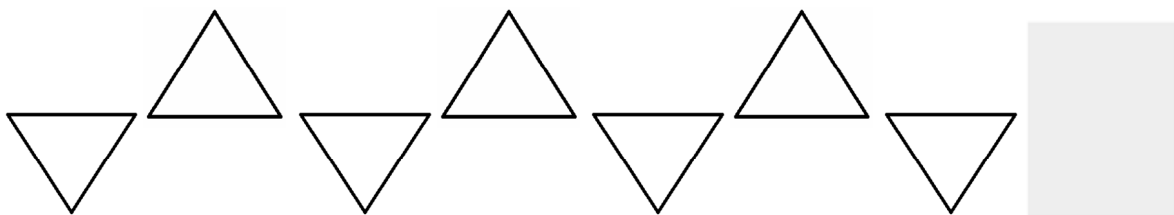
Aufgabe 7)



Aufgabe 8)



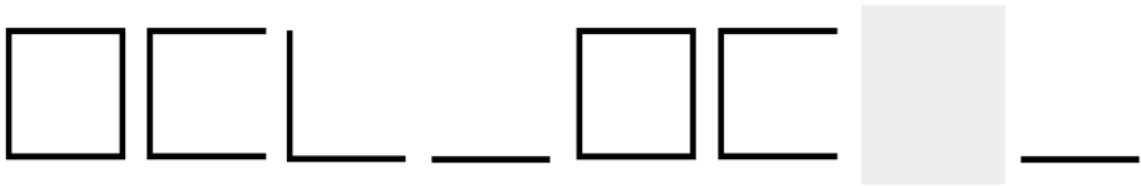
Aufgabe 9)



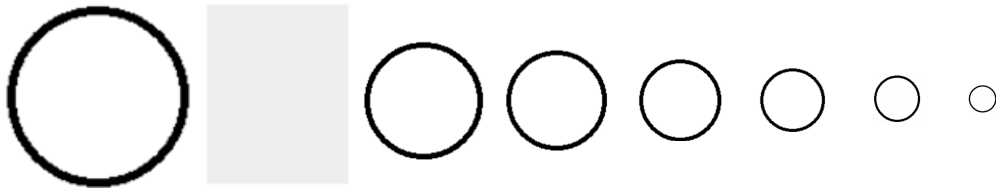
Aufgabe 10)



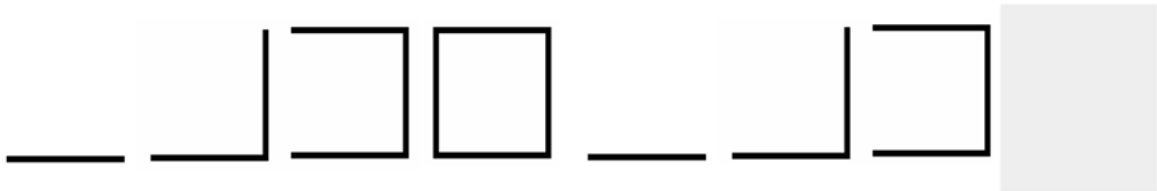
Aufgabe 11)



Aufgabe 12)



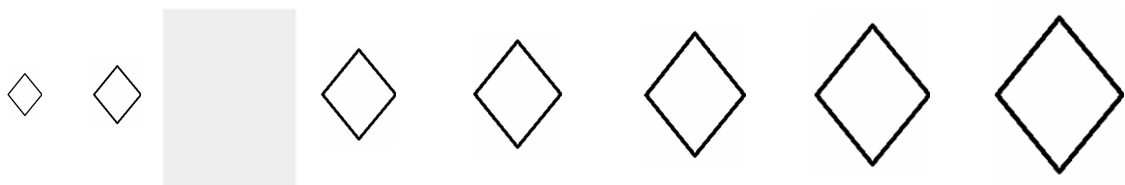
Aufgabe 13)



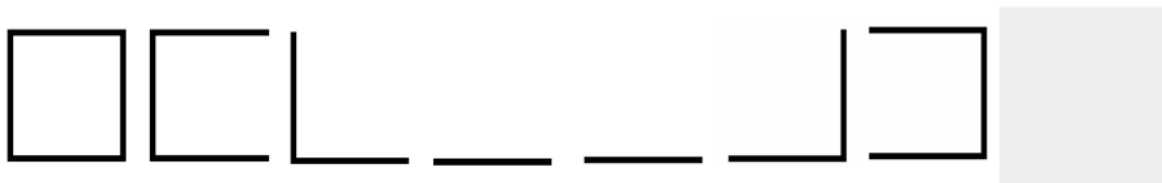
Aufgabe 14)



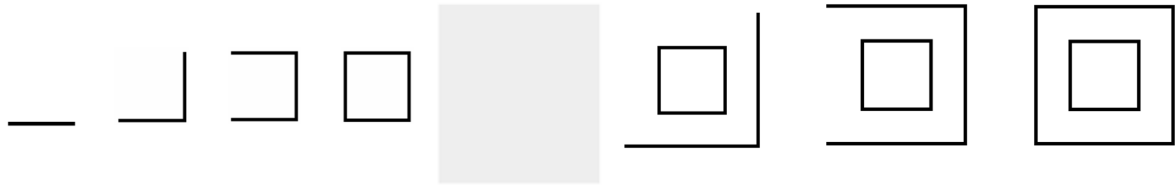
Aufgabe 15)



Aufgabe 16)



Aufgabe 17)



Aufgabe 18)



Aufgabe 19)



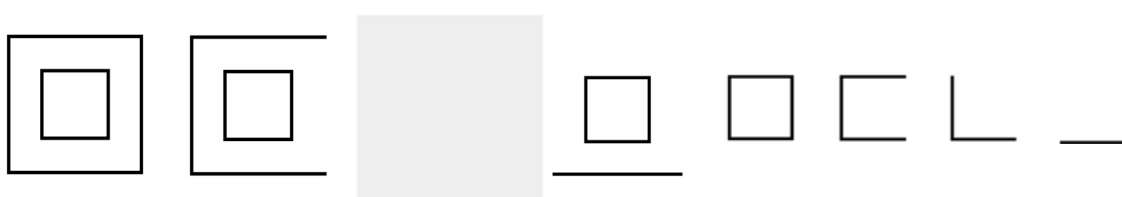
Aufgabe 20)



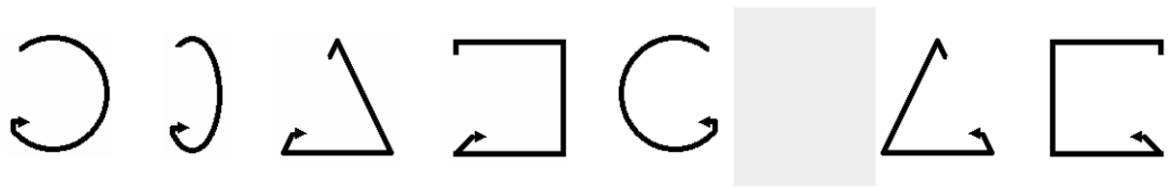
Aufgabe 21)



Aufgabe 22)



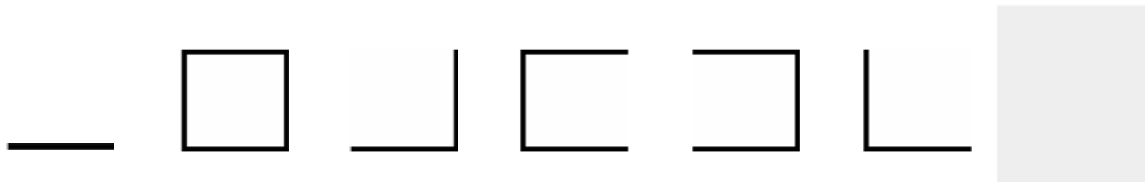
Aufgabe 23)



Aufgabe 24)



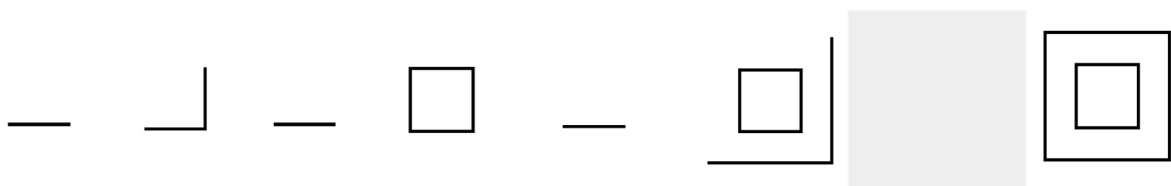
Aufgabe 25)



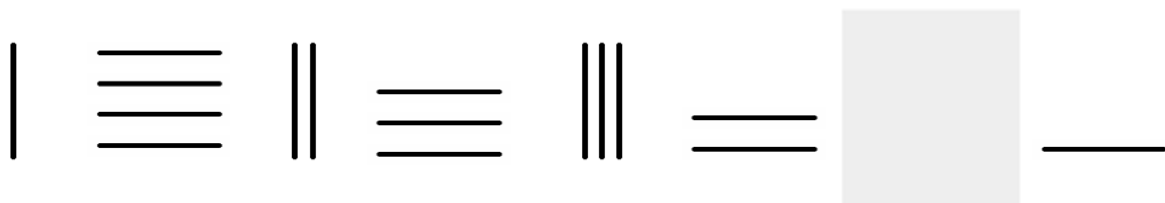
Aufgabe 26)



Aufgabe 27)



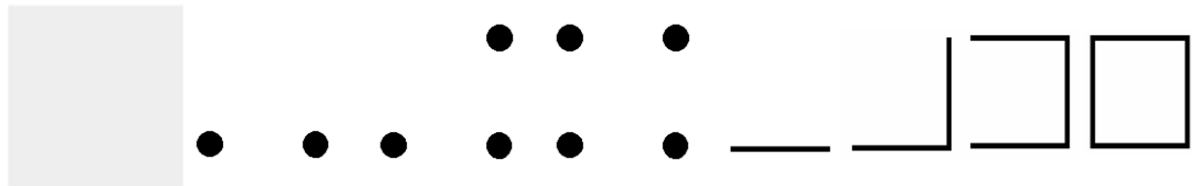
Aufgabe 28)



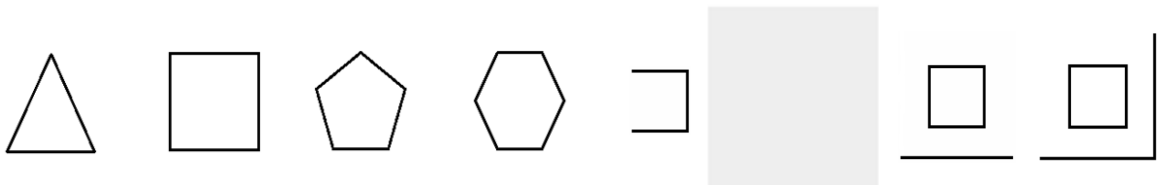
Aufgabe 29)



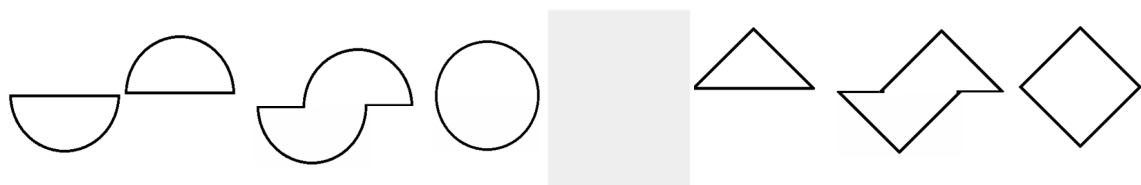
Aufgabe 30)



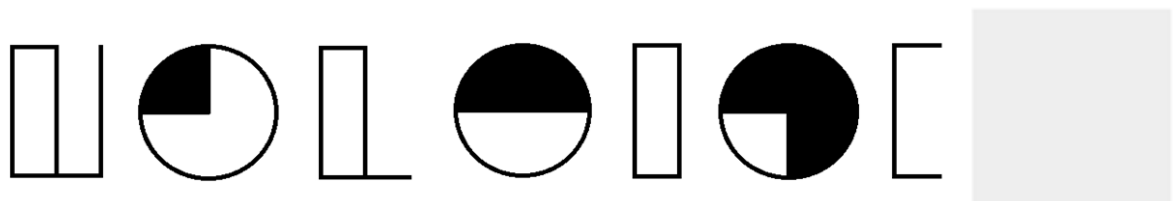
Aufgabe 31)



Aufgabe 32)



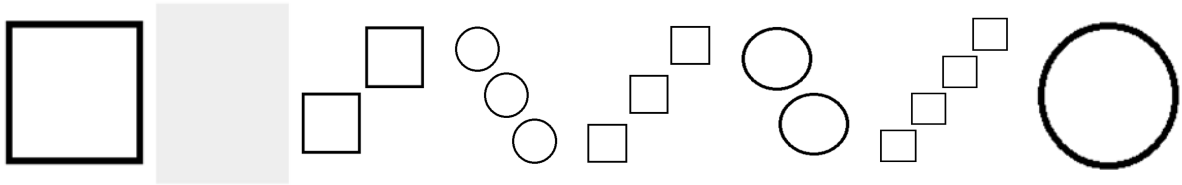
Aufgabe 33)



Aufgabe 34)



Aufgabe 35)



Adaptierte Testversion des UT 3 aus dem PSB im Multiple- Choice Antwortformat

PSB

MC Aufgaben

Name:
Geburtsdatum:
Testdatum:

Auf den folgenden Seiten siehst du Reihen mit verschiedenen Symbolen.

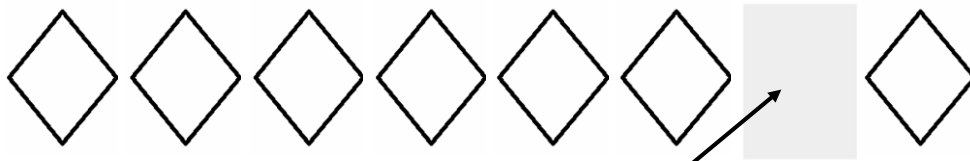
In jeder Reihe befindet sich ein graues Kästchen. In diesem fehlt ein Symbol, das die Reihe logisch ergänzt.

Versuche nun das Symbol zu finden, das am besten in die Reihe passt!

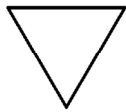
Wenn du die passende Antwort gefunden hast, dann kreise bitte die entsprechende Zahl auf dem Testheft ein.

Viel Erfolg und viel Spaß!

Beispielaufgabe



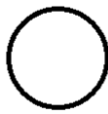
1.



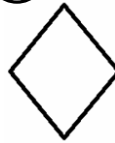
2.



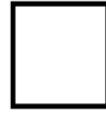
3.



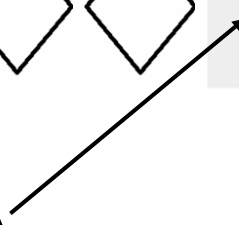
4.



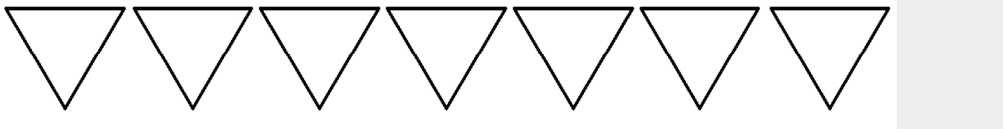
5.



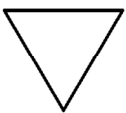
6.



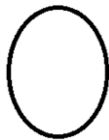
Aufgabe 2)



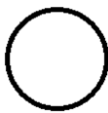
1.



2.



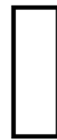
3.



4.



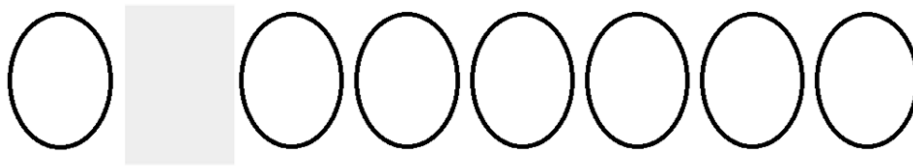
5.



6.



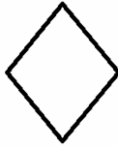
Aufgabe 3)



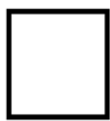
1.



2.



3.



4.



5.



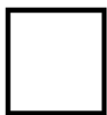
6.



Aufgabe 4)



1.



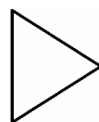
2.



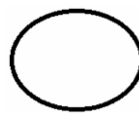
3.



4.



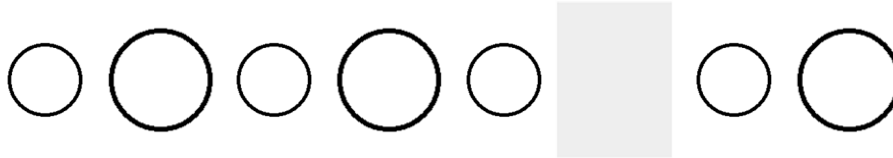
5.



6.



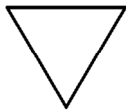
Aufgabe 5)



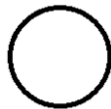
1.



2.



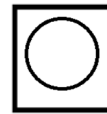
3.



4.



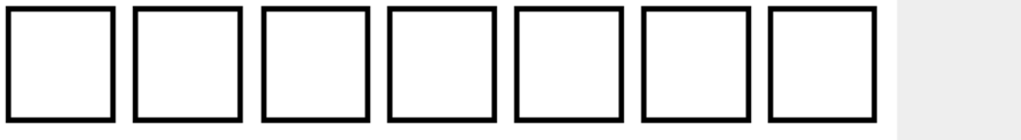
5.



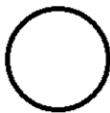
6.



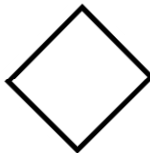
Aufgabe 6)



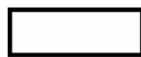
1.



2.



3.



4.



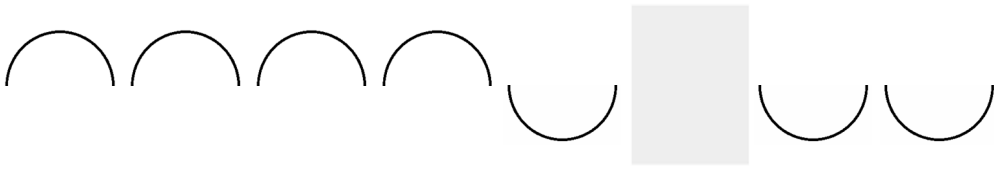
5.



6.



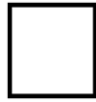
Aufgabe 7)



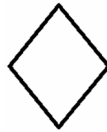
1.



2.



3.



4.



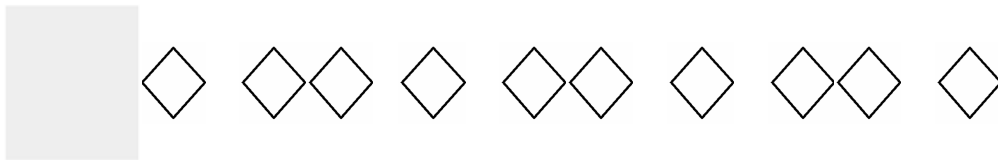
5.



6.



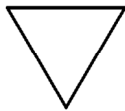
Aufgabe 8)



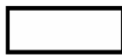
1.



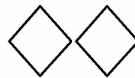
2.



3.



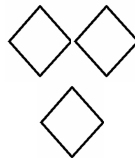
4.



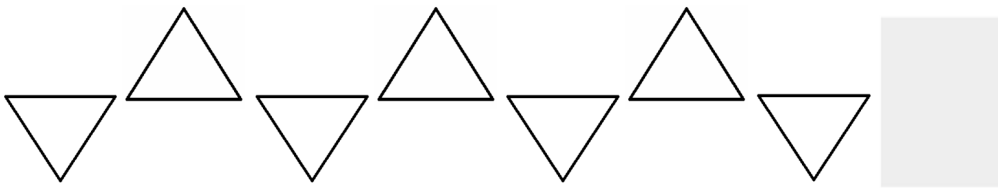
5.



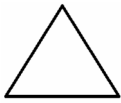
6.



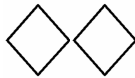
Aufgabe 9)



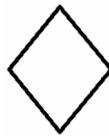
1.



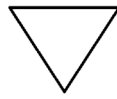
2.



3.



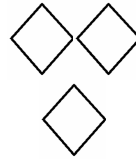
4.



5.



6.



Aufgabe 10)



1.



2.



3.



4.



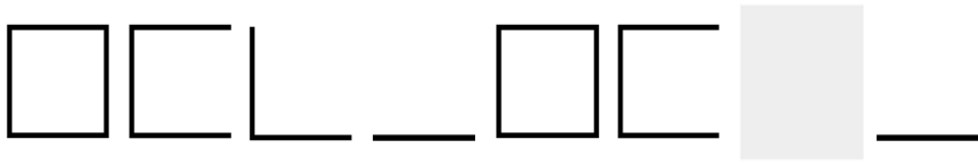
5.



6.



Aufgabe 11)



1.



2.



3.



4.



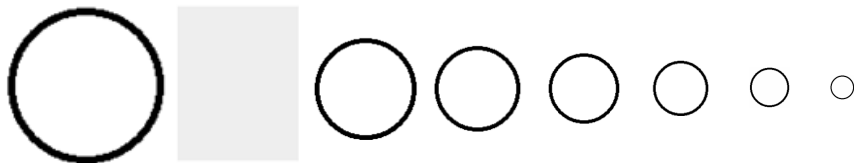
5.



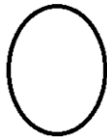
6.



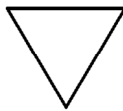
Aufgabe 12)



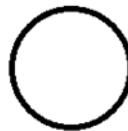
1.



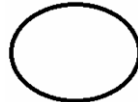
2.



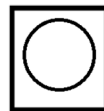
3.



4.



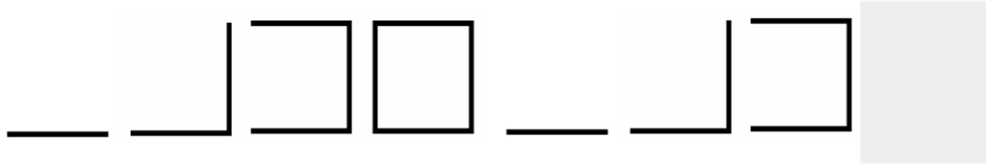
5.



6.



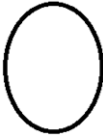
Aufgabe 13)



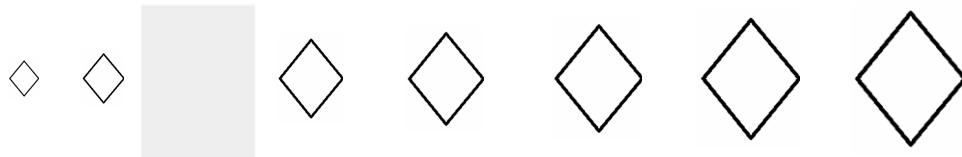
1.  2.  3.  4.  5.  6. 

Aufgabe 14)

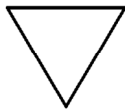


1.  2.  3.  4.  5.  6. 

Aufgabe 15)



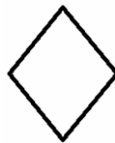
1.



2.



3.



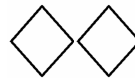
4.



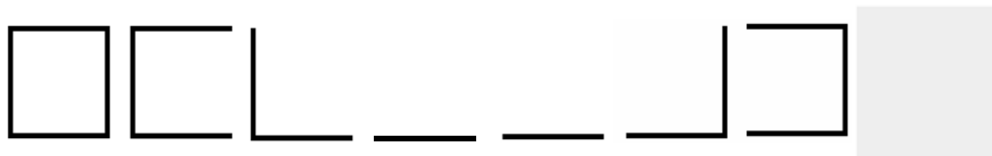
5.



6.



Aufgabe 16)



1.



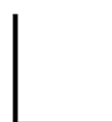
2.



3.



4.



5.



6.



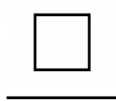
Aufgabe 17)



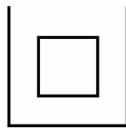
1.



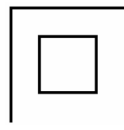
2.



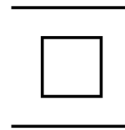
3.



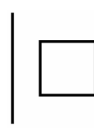
4.



5.



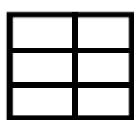
6.



Aufgabe 18)



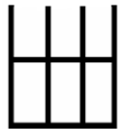
1.



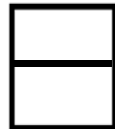
2.



3.



4.



5.



6.



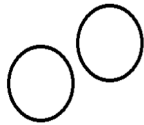
Aufgabe 19)



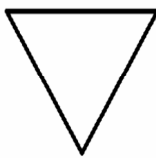
1.



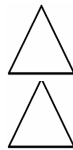
2.



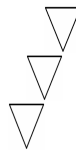
3.



4.



5.



6.



Aufgabe 20)



1.



2.



3.



4.



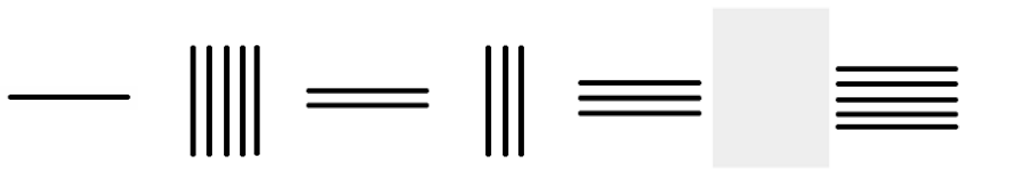
5.



6.



Aufgabe 21)



1.



2.



3.



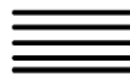
4.



5.



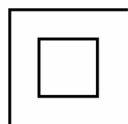
6.



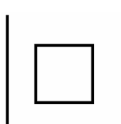
Aufgabe 22)



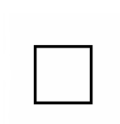
1.



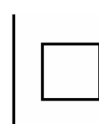
2.



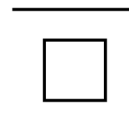
3.



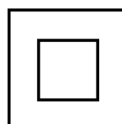
4.



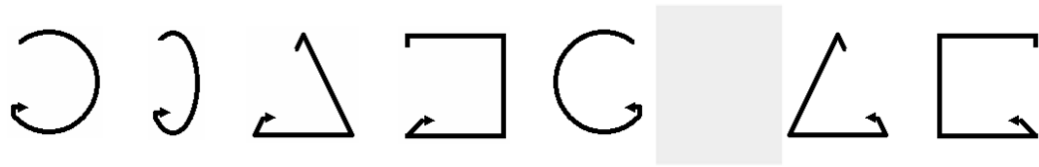
5.



6.



Aufgabe 23)



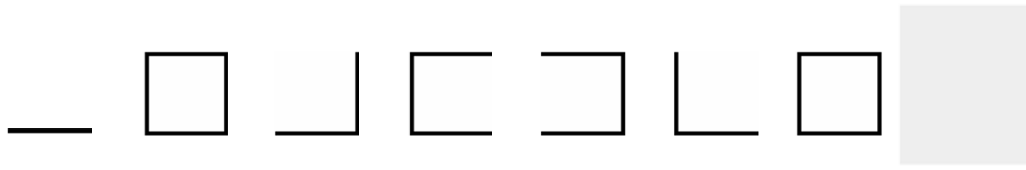
1. 2. 3. 4. 5. 6.
-

Aufgabe 24)



1. 2. 3. 4. 5. 6.
-

Aufgabe 25)



1.



2.



3.



4.



5.



6.



Aufgabe 26)



1.



2.



3.



4.



5.



6.



Aufgabe 27)



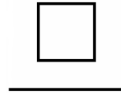
1.



2.



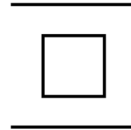
3.



4.



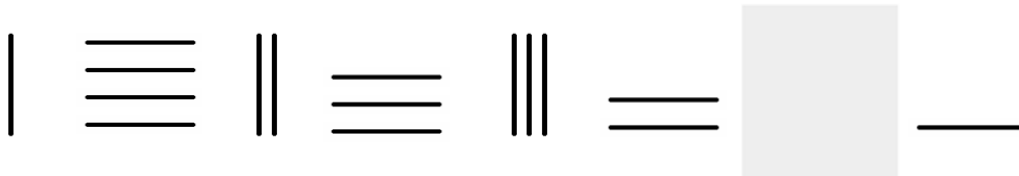
5.



6.



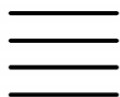
Aufgabe 28)



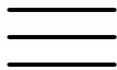
1.



2.



3.



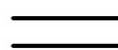
4.



5.



6.



Aufgabe 29)



1.

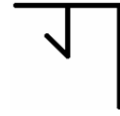
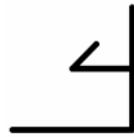
2.

3.

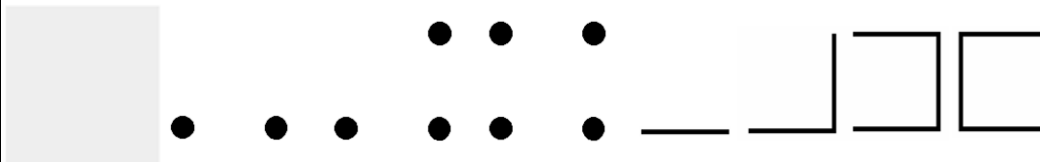
4.

5.

6.



Aufgabe 30)



1.

2.

3.

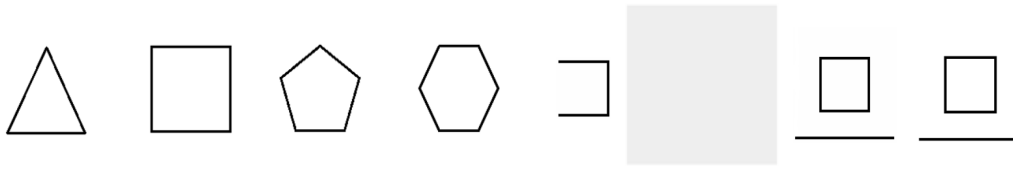
4.

5.

6.



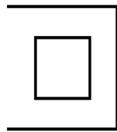
Aufgabe 31)



1.



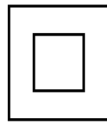
2.



3.



4.



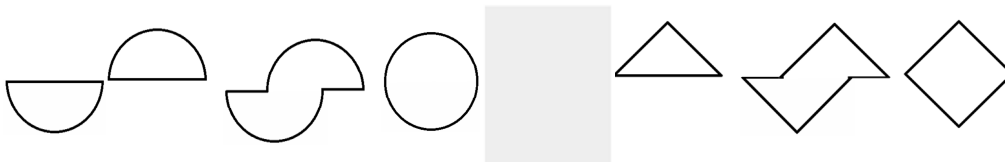
5.



6.



Aufgabe 32)



1.



2.



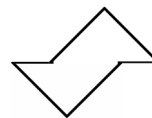
3.



4.



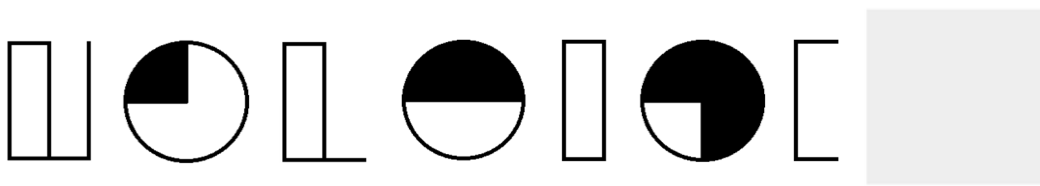
5.



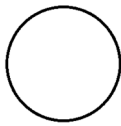
6.



Aufgabe 33)



1.



2.



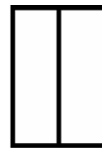
3.



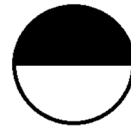
4.



5.



6.



Aufgabe 34)



1.



2.



3.



4.



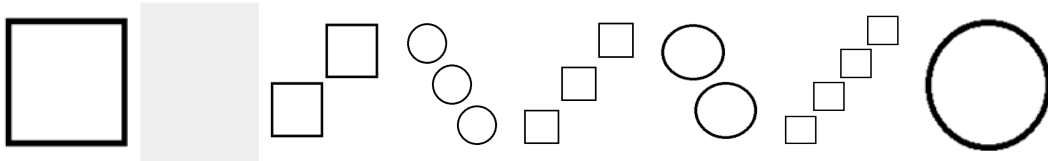
5.



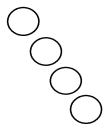
6.



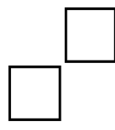
Aufgabe 35)



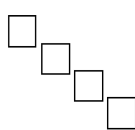
1.



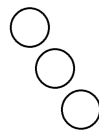
2.



3.



4.



5.



6.



Lebenslauf

Johann Ecker, geboren am 15.08.1985

Ausbildung

1992 - 1995	Volkschule; Christian Bucher Gasse
1996 - 1999	Unterstufe; AHS Bertha von Suttner
2000 - 2003	Oberstufe; AHS Bertha von Suttner Abschluss mit Reifeprüfung
Seit 2004	Diplomstudium Psychologie, Universität Wien

Berufliche Tätigkeiten

2005	Kriseninterventionszentrum Wien (Praktikum)
2009	6-Wochen Praktikum an der IV. Psychiatrischen Abteilung des Otto Wagner Spitals