



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Automatische Itemgenierung bei Aufgaben zum
visuellen Gedächtnis“

Verfasserin

Sandra Kogler

angestrebter akademischer Grad

Magistra der Philosophie (Mag.phil.)

Wien, 2010

Studienkennzahl lt. Studienblatt:

A 190 299 406

Studienrichtung lt. Studienblatt:

(UF Psychologie/Philosophie UF Mathematik)

Betreuer:

Univ.-Prof. Mag. Dr. Martin Arendasy

Inhaltsverzeichnis

1	Vorwort und Danksagung	1
2	Einleitung	3
3	Theorie des Gedächtnisses	7
3.1	Das Modell des Arbeitsgedächtnisses von Baddeley	8
3.1.1	Die zentrale Exekutive.....	9
3.1.2	Die phonologische Schleife.....	9
3.1.3	Der bildhaft-räumliche Notizblock	16
3.1.4	Der episodische Puffer.....	17
3.2	Unterschied zwischen Kurzzeitgedächtnis und Arbeitsgedächtnis	18
3.3	Arbeitsgedächtnis und Intelligenz	20
3.4	Arbeitsgedächtnis und Legasthenie	24
3.5	Weiter Zusammenhänge mit Arbeitsgedächtnis	29
4	Gedächtnisdiagnostik	31
4.1	Material zur Messung des Gedächtnisses	31
4.2	Darbietung des Materials	32
4.3	Paradigma des verbalen Wiedererkennens	33
4.3.1	Verbales Wiedererkennen.....	33
4.3.1.1	Methoden zum verbalen Wiedererkennen	33
4.3.1.2	Verbales Wiedererkennen und Geschlecht	34
4.3.1.3	Verbales Wiedererkennen und Alter	35
4.4	Weitere Einflussfaktoren auf die Gedächtnisleistung	36
4.4.1	Der serielle Positionseffekt.....	36
4.4.2	Hypothese außer-experimenteller Interferenz	37
4.5	Psychometrische Gedächtnisdiagnostik	38
4.5.1	Intelligenzmodelle.....	38
4.5.1.1	Ein-Faktor-Konzeption: Globale Intelligenz von Binet	38
4.5.1.2	Die Zwei-Faktoren-Theorie von Spearman	39
4.5.1.3	Das Modell mehrerer gemeinsamer Faktoren von Thurstone.....	40
4.5.1.4	Das „ <i>Structure of Intellect</i> “-Modell von Guilford.....	42
4.5.1.5	Das Modell der »fluid« und »crystallized general intelligence« von Catell.....	42
4.5.1.6	Erweiterung von Catells Modell durch Horn.....	44
4.5.1.7	Das Berliner Intelligenzstrukturmodell von Jäger.....	44
4.5.1.8	Zusammenfassung und Überleitung	45

4.5.2	Darstellung ausgewählter psychometrischer Verfahren	46
4.5.2.1	Inventar zur Gedächtnisdiagnostik (IGD)	46
4.5.2.2	Wechsler Gedächtnistest – Revidierte Fassung (WMS-R)	49
4.5.2.3	Der Auditiv-Verbale Lerntest nach Rey und Lezak (AVLT)	51
4.5.2.4	Verbaler Lern- und Merkfähigkeitstest (VLMT)	53
4.5.2.5	Der Benton-Test (Wahlform).....	55
4.5.2.6	Wechsler-Intelligenztest für Erwachsene (WIE)	57
4.5.2.7	Hamburg-Wechsler-Intelligenztest für Kinder-IV (HAWIK-IV)	60
4.5.2.8	Ein Beispiel für ein computergestütztes Verfahren.....	61
4.5.3	Ausgewählte Links für eine Teilnahme an online-Untersuchungen.....	63
5	Automatische Itemgenerierung	67
5.1	Theoretischer Hintergrund der automatischen Itemgenerierung.....	67
5.2	Entwicklung von Itemgeneratoren	69
5.3	Vorteile der automatischen Itemgenerierung	70
5.4	Bereits durchgeführte Studien mittels AIG	72
5.5	Computerdiagnostik	74
5.5.1	Klassifikation von computergestützten Testverfahren	74
5.5.1.1	Vorgabe von Paper-Pencil-Verfahren	74
5.5.1.2	Simulation apparativer Verfahren.....	75
5.5.1.3	Genuine Computerverfahren.....	75
5.5.1.3.1	Computergestützte Instruktion.....	76
5.5.1.3.2	Erfassung von Latenzzeiten	76
5.5.1.3.3	Adaptive Testvorgabe	76
5.5.1.3.4	Multimediale Vorgabe.....	79
5.5.2	Vorteile vom computergestützten Testen	79
5.5.3	Nachteile vom computergestützten Testen.....	82
5.5.3.1	Computervorerfahrung und Einstellung zum Computer	83
5.5.3.1.1	Populationsspezifische Effekte bezüglich Einstellung zum Computer.....	83
5.6	Testmedium Internet	84
5.6.1	Vor- und Nachteile des Testens mittels Internet	86
5.6.1.1	Generalisierbarkeit der Ergebnisse	87
5.6.1.2	Ökologische Validität.....	88
5.6.1.3	Spezifische Zielpopulationen	88
5.6.1.4	Pragmatische Vorteile	89
5.6.1.5	Finanzielle Vorteile	89
5.6.1.6	Kontrolle.....	90
5.6.1.7	Technische Gesichtspunkte	92

5.6.1.8	Motivation und Dropout.....	92
5.6.2	Ethische Richtlinien für die Nutzung des Internets	93
5.6.3	Grenzen und Chancen der Experimentierens im Internet.....	94
5.6.4	Rekrutierungsmöglichkeiten.....	95
5.6.4.1	Aktive Rekrutierung	96
5.6.4.2	Passive Rekrutierung	96
6	Diskussion und Ausblick.....	97
7	Literaturverzeichnis.....	101
8	Anhang.....	117

1 Vorwort und Danksagung

Der Abschluss eines Studiums – und somit das Verfassen einer Diplomarbeit – ist ein mühevoller Prozess, der einem einiges an Durchhaltevermögen und Kraft abverlangt. Sehr viele Leute haben mich bei diesem Prozess, sowohl in emotionaler als auch in fachlicher Hinsicht, unterstützt.

Der größte Dank gilt zunächst meinen Eltern, die mir mein Studium ermöglicht haben und mir dabei – vor allem aber in der letzten Phase – immer eine große Stütze waren. Außerdem möchte ich meiner Tochter danken, die in dieser für mich sehr herausfordernden und anstrengenden Zeit zurückstecken musste und es trotzdem schaffte, Verständnis zu zeigen und mich zu motivieren.

Darüber hinaus möchte ich mich bei all meinen Freunden und Kollegen bedanken, die mich in einer Phase ständiger Hochs und Tiefs stets unterstützten und mir sehr viel Kraft gegeben haben.

Last but certainly not least gilt für sein Entgegenkommen sowohl in zeitlicher als auch in inhaltlicher Hinsicht mein besonderer Dank dem Betreuer und Begutachter der vorliegenden Arbeit Herrn Univ.-Prof. Mag. Dr. Martin Arendasy.

St. Peter/Au, im Mai 2010

Sandra Kogler

2 Einleitung

Ausgangspunkt dieser Arbeit ist das übergeordnete Ziel der Entwicklung eines neuen Verfahrens zur Messung des visuellen Arbeitsgedächtnisses mit Hilfe von automatischer Itemgenerierung.

Es wird in der heutigen Zeit immer schwieriger Testbatterien, Tests und sogar einzelne Items vor „Testknackern“ oder einer (weltweiten) Verbreitung zu schützen. Auch machen es neuere Entwicklungen – wie zum Beispiel ein adaptiver Testvorgang – immer wichtiger, einen großen Pool an Items zu Verfügung zu haben. Mit Hilfe eines automatisch generierten Tests ist es weitestgehend möglich – neben anderen Vorteilen – erstere Vorgänge zu unterbinden und einen solch großen Pool an Items zu schaffen.

Aus diesen Gründen soll in der vorliegenden Arbeit die Möglichkeit einer derartigen Messung des visuellen Arbeitsgedächtnisses einer theoretischen Betrachtung unterzogen werden.

Von einigen ersten Ideen ausgehend, nämlich das visuelle Arbeitsgedächtnis in Bezug auf verbales Material mittels Buchstabensequenzen, die automatisch generiert vorgegeben werden, zu testen, indem die Testperson im multiple choice-Format die richtige Sequenz wieder erkennen soll, wird die Möglichkeit einer online-Vorgabe ins Auge gefasst, eine Vorgehensweise, die klarerweise an eine computergestützte Diagnostik gebunden ist.

Somit werden in dieser Arbeit folgende Themen behandelt:

Im ersten Kapitel, welches der Theorie des Gedächtnisses gewidmet ist, werden das Modell des Arbeitsgedächtnisses von Baddeley (1986) und darauf aufbauende empirische Erkenntnisse und Zusammenhänge, vor allem zu Intelligenz und Legasthenie, vorgestellt.

Der nachfolgende Abschnitt beschäftigt sich mit dem großen Thema der Gedächtnisdiagnostik. Da es von großem Interesse ist, welche Verfahren zur Messung des Arbeitsgedächtnisses bereits existieren und wie bei diesen die gewünschte Messung erfolgt, werden einige als relevant für die vorliegende Arbeit angesehene Verfahren aus den unterschiedlichsten Gebieten der Psychologie beschrieben. Darüber hinaus werden am Ende des zweiten Kapitels URLs für online durchzuführende Testmöglichkeiten angeführt und kurz dargestellt.

Das dritte und letzte Kapitel beschäftigt sich mit dem Thema der Automatischen Itemgenerierung. Daneben handelt ein Teil dieses Kapitels von der Diagnostik mittels Computer, da eine Vorgabe von automatisch konstruierten Items an eine Computerdurchführung gebunden ist. Und da darüber hinaus eine Testvorgabe mittels dem Medium Internet vor Augen schwebt, setzt sich der Abschluss dieser Arbeit damit auseinander.

Zum Sprachgebrauch in dieser Arbeit

Aufgrund besserer Lesbarkeit wird in dieser Arbeit stets die männliche Form benutzt. Es sind selbstverständlich auch immer die weiblichen Personen mitgemeint, wenn beispielsweise von „Probanden“, „Teilnehmer“ oder „Legastheniker“ gesprochen wird.

3 Theorie des Gedächtnisses

Gegen Mitte der 60er-Jahre wurde die Sichtweise von Gedächtnisphänomenen einer grundlegenden Revision unterzogen und von einer neuen Perspektive, bei der Ablauf und Art der Informationsverarbeitung im Mittelpunkt stehen, abgelöst. Innerhalb dieses Rahmens lassen sich zwei Paradigmen unterscheiden: Die so genannten *Mehrspeichermodelle* bestehen relativ unabhängig, wenngleich auch teilweise konkurrierend neben dem *Mehrebenenansatz*.

Mehrspeichermodelle geben die Annahme eines einheitlichen Gedächtnissystems auf und postulieren mehrere verschiedene Gedächtnisarten, die sich hinsichtlich Kapazität, Dauer der Verfügbarkeit und Kodierung unterscheiden, die von den aufgenommenen Informationen durchlaufen werden können und für die Kodierung, die Speicherung und den Abruf der Information verantwortlich sind (Schermer, 2006).

Dieser strukturellen Betrachtung steht nach Schermer (2006) ein vertikal orientierter Ansatz, nämlich der des Mehrebenenansatzes, gegenüber. Im Zuge dessen wird das Gedächtnis nicht in verschiedene Komponenten unterteilt, sondern die unterschiedlichen Gedächtnisleistungen auf die unterschiedliche Tiefe der Informationsverarbeitung zurückgeführt. Craik und Lockhart (1972) haben mit dem „*levels of processing*“-Ansatz ein solches Mehrebenenmodell vorgestellt.

Die ersten Mehrspeichermodelle wurden in Form von Zweispeichermodellen entwickelt und unterscheiden ein Kurz- vom Langzeitgedächtnis, also zwei getrennte Gedächtnisarten für kurz- und langfristige Behaltensleistung. Es wird davon ausgegangen, dass beiden Gedächtnisarten unterschiedliche Systeme, nämlich der Kurzzeitspeicher und der Langzeitspeicher, zugeordnet sind, die für deren Leistung verantwortlich sind. Die Übertragung von Informationen aus dem Kurzzeitgedächtnis in das Langzeitgedächtnis erfolgt durch Memorieren des Lernstoffes, was einen Wiederholungsmechanismus (*rehearsal*) darstellt. Da jedoch die Kapazität des Kurzzeitgedächtnisses begrenzt ist, kann durch den eben beschriebenen Mechanismus nur eine kleine Informationsmenge verarbeitet werden. Diese duale Gedächtnistheorie wurde 1968 von Atkinson und Shiffrin um einen dritten Speicher, das dem

Kurzzeitgedächtnis vorgeschaltete sensorische Register, erweitert. Diese vertreten damit ein „Drei-Speicher-Modell“ (Schermer, 2006).

Zum Verhältnis von Mehrspeicher- und Mehrebenenansatz zueinander ist aus heutiger Sicht anzumerken, dass diese nach Schermer (2006) eher auf gegenseitige Ergänzung angelegt sind und dass durch die gleichzeitige Berücksichtigung sowohl von strukturellen als auch von funktionalen Merkmalen eine Möglichkeit geschaffen wird, gedächtnisbezogene Aspekte angemessener zu erforschen.

Wie bereits aus dem Titel der Arbeit zu erkennen ist, wird hier die Theorie des Mehrspeicheransatzes vertreten. Allerdings soll nur einem dieser Speicher, nämlich dem Kurzzeitgedächtnis vermehrt Aufmerksamkeit geschenkt werden. Dieses schließt aber wiederum mehrere Komponenten ein. Denn es existieren inzwischen gute Belege dafür, dass das bewusster mentaler Aktivität unterliegende Gedächtnissystem, das Kurzzeitgedächtnis, nicht eine einzelne Entität ist, sondern zumindest drei Komponenten umfasst (Parkin, 1993). Eben dieser Gedanke wurde im Arbeitsgedächtnismodell von Baddeley formalisiert.

3.1 Das Modell des Arbeitsgedächtnisses von Baddeley

Baddeley (2000) beschreibt sein Modell des Arbeitsgedächtnisses als eine Weiterentwicklung von bereits bestehenden Modellen zum Kurzzeitgedächtnis, solchen wie von Broadbent (vgl. Broadbent, 1958) oder Atkinson und Shiffrin (vgl. Atkinson & Shiffrin, 1968), das sich von diesen jedoch in zwei Punkten unterscheidet. Erstens verwirft es das Konzept eines einheitlichen Speichers zugunsten eines Mehrkomponenten-Systems und zweitens hebt es die Funktion eines solchen Systems bei komplexen Kognitionen hervor.

Die erste Fassung des Modells von Baddeley und Hitch von 1974 umfasste zunächst nur drei Komponenten: ein Kontrollsystem mit begrenzter Kapazität, bezeichnet als die *zentrale Exekutive*, welche durch zwei Speichersysteme (auch Sklavensysteme genannt) unterstützt wird, nämlich die *phonologische Schleife*, welche auf Klang und Sprache basiert und der *bildhaft-räumliche Notizblock* (Baddeley, 2003). Da es jedoch

schon immer Phänomene gab, die damit nicht erklärt werden konnten, fand eine Neuformulierung des theoretischen Rahmens statt. Diese führte Baddeley im Jahre 2000 zu der Annahme einer neuen Komponente des Arbeitsgedächtnisses, nämlich der des *episodischen Puffers*. Diese einzelnen Komponenten sollen im Folgenden dargestellt werden, wobei aufgrund der Bedeutung für die vorliegende Arbeit der Schwerpunkt auf die phonologische Schleife gelegt wird.

3.1.1 Die zentrale Exekutive

Nach Baddeley (2003) ist die zentrale Exekutive die wichtigste, jedoch auch die am wenigsten verstandene Komponente des Arbeitsgedächtnisses. Ihre Hauptaufgabe soll nach Schermer (2006) in der Lenkung von Aufmerksamkeit sowie der Kontrolle, Koordination und Integration von insbesondere aus den im Anschluss dargestellten Hilfssystemen stammenden Informationen bestehen. Ferner obliegt diesem Aufmerksamkeitssystem mit begrenzten Ressourcen die Kontrolle nicht-routinierter Aktivitäten und der Abruf von Informationen aus dem Langzeitgedächtnis (Hamm, Junglas & Bredenkamp, 2004).

3.1.2 Die phonologische Schleife

Die phonologische Schleife stellt die bis jetzt (Baddeley, 2009) am weitesten entwickelte Komponente des Arbeitsgedächtnismodells dar. Baddeley (1986) versteht unter phonologischer Schleife ein Teilsystem des Arbeitsgedächtnisses, das die klangliche Verarbeitung sprachlicher Informationen leistet. Das Modell der phonologischen Schleife beinhaltet einen temporären phonetischen Speicher („*phonological store*“), in welchem auditive Gedächtnisspuren über einige wenige Sekunden bestehen bevor diese verblasen, und einen subvokalen artikulatorischen Kontrollprozess („*subvocal rehearsal*“), der dem inneren Wiederholen zur Aufrechterhaltung von Informationen analog ist. Anders ausgedrückt: Der phonetische Speicher kann auditiv-verbale Informationen für etwa 1,5 bis 2 Sekunden repräsentieren, während der Kontrollprozess durch eine Art „inneres Sprechen“ dafür sorgt, dass Informationen auch über diese zwei Sekunden hinaus im Zugriffsbereich der bewussten Verarbeitung bleiben (Hasselhorn, Grube & Mähler, 2000). Darüber

hinaus leistet dieser Kontrollprozess auch die phonetische Umkodierung visuell dargebotener Informationen, sodass er laut Hasselhorn et al. (2000) neben der Rehearsal-Funktion auch eine Übersetzungsfunktion vom Bild in die Sprache erfüllt.

Das unmittelbare Gedächtnis hat eine begrenzte Spanne (die so genannte Gedächtnisspanne), da bei Ansteigen der Anzahl der zu wiederholenden Items der Punkt erreicht wird, an dem das erste Item bereits verblasst ist, bevor es wiederholt werden kann (Baddeley, 2003).

Unter der Gedächtnisspanne versteht man die Anzahl von Items, die eine Person nach einmaliger (akustischer) Darbietung im Sekundenrhythmus in der vorgegebenen Reihenfolge reproduzieren kann (Hasselhorn et al., 2000). Diese gilt laut Hasselhorn et al. (2000) als Indikator für die funktionale Gesamtkapazität des phonologischen Arbeitsgedächtnisses.

Hasselhorn und Marx (2000) führen an, dass zwischen der Gedächtnisspanne und der Sprechrates für das gleiche Itemmaterial eine lineare Beziehung besteht, die im Übrigen weitgehend material- und altersinvariant ist (vgl. Hasselhorn, 1988). Weiters meinen diese, dass es sich eingebürgert hat, die Sprechrates als Schätzmaß für die Geschwindigkeit des artikulatorischen Kontrollprozesses im phonologischen Arbeitsgedächtnis zu interpretieren.

Als Messverfahren zur Bestimmung der Funktionstüchtigkeit des phonetischen Speichers schlagen Gathercole, Willis, Baddeley und Emslie (1994) – zumindest für Kinder – das Nachsprechen von Kunstwörtern („*nonword repetition*“) vor. Dabei handelt es sich nach Hasselhorn, Seidler-Brandler und Körner (2000, S. 119) um eine Aufgabe, bei der eine Lautfolge, die bedeutungsfrei ist, aber in der Regel Ähnlichkeit mit „richtigen“ Wörtern hat, akustisch mit der Aufforderung dargeboten wird, das gehörte Kunstwort nachzusprechen. Die Anzahl der richtig nachgesprochenen Kunstwörter halten Gathercole et al. (1994) für ein geeignetes Maß für die Funktionstüchtigkeit des phonetischen Speichers, da diese die Ansicht vertreten, dass mit dem Kunstwörter-Nachsprechen vor allem die Verarbeitungskapazität des Arbeitsgedächtnisses bzw. des phonetischen Speichers erfasst wird.

Gathercole et al. (1994) haben aus dieser Anforderung für den englischen Sprachraum einen Test entwickelt. Hasselhorn, Tiffin-Richards, Woerner, Banaschewski und Rothenberger (2000) betonen hingegen, dass dieser Test in der deutschsprachigen pädagogisch-psychologischen Diagnostik schon seit Jahrzehnten bekannt ist (vgl. Mottier, 1951) und seit langem (1974) einen Bestandteil des „Zürcher Lesetests“ (ZLT) darstellt. Auch die fünfte und überarbeitete Auflage des ZLT (Zürcher Lesetest - Förderdiagnostik bei gestörtem Schriftspracherwerb, Linder & Grisseemann, 1996) beinhaltet noch als Zusatzverfahren den Mottiertest.

Denn bereits in den 50er Jahren verwendete Mottier (1951, zitiert nach Hasselhorn, Seidler-Brandler & Körner, 2000, S. 119) die Leistung beim Nachsprechen von Kunstwörtern zur Erfassung des „Wort- und Sprachgedächtnisses“ als Teil der sprachlichen Fähigkeiten, dessen Störungen ihrer Auffassung nach fast immer mit Lese-Rechtschreibschwierigkeiten einhergehen.

Anfang der 80er Jahre häuften sich dann nach Angabe von Hasselhorn et al. (2000, S. 119) die empirischen Belege eines substantiellen Zusammenhangs zwischen der Leistung beim Nachsprechen von Kunstwörtern und Lese- sowie Rechtschreibleistungen. Seit Beginn der 90er Jahre, so Hasselhorn et al. (2000, S. 120) weiter, konnte auch die nicht unerhebliche Bedeutung des Kunstwörter-Nachsprechens für syntaktische Sprachleistungen sowie für Entwicklungsveränderungen im Wortschatz und für spezifische Sprachentwicklungsstörungen demonstriert werden. Und schließlich, so Hasselhorn et al. (2000) weiter, fanden Gathercole und Baddeley (1993) vielfältige empirische Belege für einen bedeutsamen Zusammenhang zwischen der Leistung beim Kunstwörter-Nachsprechen und herkömmlichen Maßen der Arbeitsgedächtniskapazität, wobei diese eine solche Leistung als einen Indikator für die Funktionstüchtigkeit des phonologischen Arbeitsgedächtnisses interpretieren (S.120).

Gathercole et al. (1994, zitiert nach Hasselhorn et al., S. 120) vertreten die Ansicht, dass mit dem Kunstwörter-Nachsprechen vor allem die Verarbeitungskapazität des Arbeitsgedächtnisses bzw. des phonetischen Speichers erfasst wird. Dieser

Auffassung entsprechend wird das unbekannte und bedeutungslose Kunstwort nur aufgrund seiner lautlichen Charakteristika verarbeitet und Fehler bei der Reproduktion des Kunstwortes spiegeln Fehler bei der Erstellung einer korrekten Repräsentation im phonologischen Arbeitsgedächtnis wider. Dazu passend ist der Befund, dass das Nachsprechen von Kunstwörtern umso schwerer fällt, je länger das Kunstwort ausfällt (Hasselhorn et al., S. 120).

Nach Baddeley (2009) wird das Nachsprechen von Kunstwörtern oft als Test zur Diagnose der Legasthenie eingesetzt, auch wenn eine reduzierte Kapazität der phonologischen Schleife nur eine von vielen Variablen darstellt, die das Lesenlernen beeinflussen.

Beispielsweise wird im Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC) von Jansen, Mannhaupt, Marx und Skowronek (1999), einem häufig verwendeten Verfahren um das Risiko für eine spätere Entwicklung einer Lese-Rechtschreibschwäche bei Vorschulkindern abzuschätzen, die Kapazität des Arbeitsgedächtnisses für sprachliches Material mit einer solchen Aufgabe zum Pseudowörter-Nachsprechen überprüft, indem dem Kind vom Kassettenrecorder je ein Wort (z.B. „Zippelzack“) vorgespielt wird, das von diesem nachgesprochen werden soll.

Auch das Münsteraner Screening (Müsc) zur Früherkennung von Lese-Rechtschreibschwierigkeiten von Mannhaupt (2006) beinhaltet eine Kurzzeitgedächtnisaufgabe (*Pseudowörter nachsprechen*).

Hasselhorn, Seidler-Brandler und Körner (2000) weisen darauf hin, dass beim Einsatz des Kunstwörter-Nachsprechens als diagnostisches Verfahren zur Erfassung der Funktionstüchtigkeit des phonologischen Arbeitsgedächtnisses das verwendete Kunstwortmaterial hinsichtlich seiner Wortähnlichkeit (die Ähnlichkeit zu aus der Muttersprache vertrauten Wörtern) zu kontrollieren bzw. konstant zu halten ist, wenn man die entsprechende Nachsprechleistung als solche des phonologischen Arbeitsgedächtnisses interpretieren will, da die Leistung beim Nachsprechen von

Kunstwörtern sowohl durch das phonologische Arbeitsgedächtnis als auch durch das lexikalische Vorwissen determiniert ist (vgl. dazu Gathercole, 1995).

Aufgrund des eben dargestellten Sachverhalts wird die Beziehung zwischen Arbeitsgedächtnis und Legasthenie im Kapitel 3.4 genauer behandelt.

Baddeley (2000) fasst einige Phänomene zusammen, die für eine Annahme eines Modells der phonologischen Schleife sprechen:

(1) Der phonologische Ähnlichkeitseffekt

Conrad und Hull konnten 1964 zeigen, dass im Fall größerer akustischer Ähnlichkeit zwischen den Items einer Wortfolge die Leistung der Versuchspersonen geringer ist (Baddeley, 1979, Salamé & Baddeley, 1986). Weiters konnten diese darlegen, dass der Faktor der akustischen Ähnlichkeit im Vergleich zum Faktor der Wortmenge der viel einflussreichere auf die Gedächtnisleistung ist.

Dies bedeutet, dass die Größe der Gedächtnisspanne nicht nur von der Anzahl der Brocken (chunks) abhängt – wie dies Miller (1956) behauptet, indem er den Umfang des Kurzzeitgedächtnisses mit plus/minus sieben chunks schätzt –, sondern durch zusätzliche Faktoren wie akustische Ähnlichkeit (vgl. Conrad und Hull 1964) und Wortlänge (vgl. Baddeley, 1979; Thompson und Buchanan 1975) beeinflusst wird.

Weiters untersuchten Conrad und Hull (1964) auch den Einfluss der akustischen Ähnlichkeit beim Merken von Buchstabenfolgen, indem sie Versuchspersonen visuell Buchstabenfolgen darboten, die sich im Grad ihrer akustischen Ähnlichkeit unterschieden. So konnten sie zeigen (vgl. auch Wickelgren, 1965b), dass Folgen hoher Ähnlichkeit (z.B.: DCBTPW) zu beträchtlich mehr Fehlern bei visueller Darbietung führten als unähnliche Folgen (z.B.: LVKFRT).

Wickelgren (1965a) beschreibt, dass auch Conrad 1962 und 1964 diesen Effekt zeigen konnte, indem er Listen von jeweils sechs von zehn Buchstaben (B, C, P, T, V, F, M, N, S, X) visuell darbot, die die Versuchspersonen sofort nach Ende der Liste

wiederholen mussten. Conrad fand heraus, dass Buchstaben, deren (englische) Aussprache mit einem [i] endet (B, C, P, T, V) untereinander zu Verwechslungen führen und dasselbe für Buchstaben gilt, deren (englische) Aussprache mit einem [e] beginnt (F, M, N, S, X).

Wickelgren (1965a) machte Beobachtungen, die mit denen von Conrad übereinstimmen und wies darüber hinaus darauf hin, dass die akustische Ähnlichkeit in erster Linie das Behalten der Reihenfolge beeinträchtigt, mit der die Buchstaben dargeboten wurden und nicht das Behalten der Buchstaben als solches.

Zusammenfassend kann man daher sagen, dass die Arbeiten von Conrad (1962, 1964) und Wickelgren (1965a) darlegen, dass das Kurzzeitgedächtnis weitestgehend auf einer akustischen Kodierung aufbaut. Der akustische Ähnlichkeitseffekt bedeutet also grob gesprochen, dass es schwieriger ist, sich an Items wie Buchstaben oder Wörter zu erinnern, die sich akustisch ähnlich sind, als an Items, die sich akustisch nicht ähneln. So fällt (bei englischsprachigen Personen) die unmittelbare serielle Reproduktion einer Sequenz aus den Buchstaben G, C, B, T, V, P im Vergleich zu einer Sequenz aus den Buchstaben F, W, K, S, Y, Q schlechter aus (Baddeley, 2000). Gleiches gilt bei unmittelbarer serieller Reproduktion klangähnlicher Items (z.B. Schwan, Kran, Bahn, Zahn) im Verhältnis zu klangunähnlichen (z.B. Top, Schuh, Baum, Zahn) (Hasselhorn et al. 2000).

(2) Der Wortlängeneffekt

Es wird von Personen als leichter empfunden, eine Sequenz von kurzen Wörtern (z.B.: wit, sum, harm, bag, top) als von langen Wörtern (z.B.: university, aluminium, opportunity, constitutional, auditorium) aus dem Gedächtnis abzurufen (Baddeley, 2000). Dies warf die Frage auf, ob dieser Wortlängeneffekt von der Anzahl der Silben bei kurzen und langen Wörtern abhängt oder von der unterschiedlichen Aussprechdauer. Deswegen verglichen Baddeley et al. die Gedächtnisspanne für Items mit gleicher Silbenzahl, jedoch mit unterschiedlicher Aussprechdauer (z.B.: „wicket“ versus „harpoon“). Dabei fanden sie heraus, dass die Gedächtnisspanne für Wörter mit längerer Aussprechdauer weniger Items umfasste als für Wörter mit

kürzerer Aussprechdauer. Daraus schlossen Baddeley et al., dass das für das Erinnern verantwortliche System artikulatorischer Natur ist (Parkin, 1993).

Weiters meint Baddeley (2000) aufgrund seiner experimentellen Ergebnisse von 1975, dass es mehr Zeit in Anspruch nimmt, ein vielsilbiges Wort einzuprägen und es während des Abrufs auszusprechen, wodurch mehr Zeit für den Zerfall der Gedächtnisspur zur Verfügung steht.

(3) Der Effekt der artikulatorischen Suppression

Werden Personen daran gehindert, sich die zu erinnernden Items einzuprägen, indem sie aufgefordert werden, kontinuierlich eine bedeutungsfreie Sequenz (wie z.B. „the“, „the“, „the“,...) auszusprechen, fällt die Erinnerungsleistung merklich ab. Außerdem verschwindet unter artikulatorischer Suppression der Wortlängeneffekt.

Hervorzuheben ist, dass die Leistung bei kurzen Wörtern auf das gleiche Niveau sinkt wie bei langen Wörtern, wenn verbales Einprägen verhindert wird, so dass es keinen Unterschied macht, wie lange man braucht, diese Wörter zu artikulieren. Nach Parkin (1993) ist dies ein geeigneter Beleg dafür, dass der Vorteil beim Wiedergeben kurzer Wörter bei Gedächtnisspannenaufgaben entscheidend vom Rückgriff auf ein artikulatorisches Kodierungssystem abhängt. Ist dies nicht möglich – wie bei artikulatorischer Suppression –, hängt das Gedächtnis für kurze und lange Wörter von denselben Prozessen ab, sodass diese auch in vergleichbarem Ausmaß erinnert werden.

(4) Informationstransfer zwischen Codes

Aufgrund der Effizienz des phonologischen Speichers beim seriellen Abruf tendieren erwachsene Personen normalerweise dazu, visuell dargebotene Items zu benennen und sich durch subvokales Sprechen einzuprägen; die Information wird also von einem visuellen in einen auditiven Code transferiert. Die oben beschriebene artikulatorische Suppression verhindert diesen Transfer mit dem Effekt, dass der Effekt der phonologischen Ähnlichkeit für visuell dargebotenes Material

verschwindet, jedoch nicht für auditiv präsentiertes, da dieses automatisch im phonologischen Speicher registriert wird (Baddeley, 2000).

(5) Neuropsychologische Belege

Neben den gerade dargestellten Phänomenen sprechen auch neuropsychologische Befunde – vor allem von Patienten mit spezifischen Defiziten des phonologischen Kurzzeitgedächtnisses – für die Annahme eines Zwei-Komponenten Modells der phonologischen Schleife.

Patienten mit Defiziten des phonologischen Kurzzeitgedächtnisses, die jedoch keine generelle Sprachbeeinträchtigung aufweisen, besitzen normalerweise Läsionen der linken tempoparietalen Region (*Brodman area*). Werden ihnen Wörter visuell präsentiert, zeigen diese weder einen phonologischen Ähnlichkeitseffekt noch einen Wortlängeneffekt, ein Bild also, das sich mit der Vermeidung des Gebrauchs des defizitären phonologischen Speichers deckt (Baddeley, 2003). Aufgrund der Komplexität und des nötigen Vorwissens wird auf eine detaillierte Ausführung dieser neuropsychologischen Belege jedoch verzichtet.

Zusammenfassend lässt sich sagen, dass die phonologische Schleife vor allem für die Wiederholung von sequentieller Information geeignet ist und ihre Funktion sehr klar in den so genannten Gedächtnisspannenaufgaben, bei denen eine Itemsequenz sofort nach Darbietung in derselben Reihenfolge wie bei der Präsentation wiederholt werden muss, zum Vorschein kommt.

3.1.3 Der bildhaft-räumliche Notizblock

Der bildhaft-räumliche Notizblock stellt neben der phonologischen Schleife das zweite so genannte Sklavensystem der zentralen Exekutive dar. Baddeley (2000, 2002) geht davon aus, dass dieses System im Stande ist, räumliche und visuelle Informationen temporär aufrechtzuerhalten und zu verarbeiten, was eine sehr wichtige Rolle bei der räumlichen Orientierung und bei der Lösung von bildhaft-räumlichen Problemen spielt. Im Gegensatz zur phonologischen Schleife bedarf es bei der

Bereithaltung von Informationen im bildhaft-räumlichen Notizblock keiner kontinuierlichen Wiederholung (Schermer, 2006). Zudem stellt der bildhaft-räumliche Notizblock ein Verbindungsstück zwischen visueller (bildhafter) und räumlicher Information dar, welche über Wahrnehmungsprozesse oder das Langzeitgedächtnis zugeführt werden.

Sowohl neuropsychologische Befunde als auch Ergebnisse von funktionellen bildgebenden Verfahren unterstützen die Annahme des bildhaft-räumlichen Notizblockes als ein Mehrkomponenten-System, wobei die Aktivierung der Okzipitallappen die visuelle Komponente widerspiegelt, die parietalen Regionen das Räumliche repräsentieren und frontale Aktivierung für deren Koordination und Kontrolle zuständig sind.

3.1.4 Der episodische Puffer

Obwohl mithilfe des visuellen (bildhaft-räumlicher Notizblock) und des verbalen (phonologische Schleife) Sklavensystems des Arbeitsgedächtnisses für eine große Anzahl der bisherigen Daten eine plausible Erklärung gefunden werden kann, sagt Baddeley (2000), dass Belege von Patienten mit Kurzzeitgedächtnisstörungen für die Annahme eines weiteren back-up Speichers sprechen.

Belege für eine eingebundene Speicherung von Information aus verschiedenen Modalitäten und Systemen stammen von dem kleinen aber signifikanten Einfluss von visueller Ähnlichkeit beim verbalen Abruf und von dem sehr beachtlichen Einfluss der Wortbedeutung beim sofortigen Abruf von Sätzen und Prosapassagen. Deswegen besteht für Baddeley (2000) ein klares Bedürfnis, einen Prozess oder Mechanismus anzunehmen, der die Informationen aus den verschiedenen Subsystemen in eine Form der temporären Repräsentation bringt. Eben diese vierte Komponente des Arbeitsgedächtnisses nennt Baddeley „episodischer Puffer“.

Der episodische Puffer stellt ein temporäres Speichersystem dar, das in seiner Kapazität begrenzt ist, und das fähig ist, Informationen von den verschiedensten Quellen zu integrieren und diese mithilfe eines multidimensionalen Codes zu

speichern. Es stellt so ein Verbindungsstück zwischen den Sklavensystemen – also phonologische Schleife und bildhaft-räumlicher Notizblock – und dem Langzeitgedächtnis dar.

Es wird angenommen, dass der episodische Puffer von der zentralen Exekutive kontrolliert wird, welche dafür verantwortlich ist, dass die Informationen von einer Reihe von Quellen in eine zusammenhängende Episode verbunden werden. Solche Episoden gelten als bewusst abrufbar.

Der Puffer fungiert als ein vom Langzeitgedächtnis getrennter „Raum“, der aber eine wichtige Stufe beim episodischen Langzeitlernen darstellt. Durch bewusste Aufmerksamkeit kann die zentrale Exekutive auf den episodischen Puffer zurückgreifen, welche darüber hinaus den Speicherinhalt durch Hinwendung zu einer bestimmten Informationsquelle – entweder zu anderen Komponenten des Arbeitsgedächtnisses oder zum Langzeitgedächtnis – beeinflussen kann. Somit stellt der episodische Puffer nicht nur einen Mechanismus zur Verarbeitung, sondern auch für das Erschaffen neuer kognitiver Repräsentationen dar, welche wiederum das Problemlösen erleichtern könnten (Baddeley, 2000).

Baddeley (2009) betont, dass sich die Entwicklung des Konzepts des episodischen Puffers noch immer in einer frühen Phase befindet, auch wenn sich die Annahme des episodischen Puffers schon einige Male als brauchbar herausgestellt hat.

3.2 Unterschied zwischen Kurzzeitgedächtnis und Arbeitsgedächtnis

Da aus dem oben dargestellten Modell von Baddeley der Unterschied zwischen Arbeitsgedächtnis und Kurzzeitgedächtnis nicht hervorgegangen ist, soll dies an dieser Stelle nachgeholt werden.

Baddeley (2009) meint, dass der Begriff des Kurzzeitgedächtnisses ein eher „schlüpfriger“ ist. Die Allgemeinheit versteht darunter etwas anderes als Psychologen. Umgangssprachlich bezieht sich der Begriff des Kurzzeitgedächtnisses auf Dinge, an die man sich nach ein paar Stunden oder Tagen erinnert. Für einen

Psychologen stellen dies Inhalte des Langzeitgedächtnisses dar. Die Erinnerung über ein paar Minuten, Stunden oder Jahre scheint von ein und dem selbem Langzeitgedächtnissystem abzuhängen (Baddeley, 2009).

Baddeley (2009) benützt den Begriff Kurzzeitgedächtnis um den Speicherprozess einer kleinen Anzahl von Informationen über ein kurzes Intervall zu beschreiben. Er meint weiters, dass das Gedächtnissystem oder -systeme, die für das Kurzzeitgedächtnis verantwortlich sind, einen Teil des Arbeitsgedächtnisses darstellen. Der Begriff des Arbeitsgedächtnisses stellt nach Baddeley (2009) jenen Begriff dar, der für ein System benutzt wird, das nicht nur temporär Informationen speichert, sondern diese auch verarbeitet, damit so komplexe Aktivitäten wie Lernen und Schlussfolgern möglich werden.

Von Aster, Neubauer und Horn (2006, S. 16) sind der Meinung, dass das Konzept des Arbeitsgedächtnisses das Konzept des Kurzzeitgedächtnisses ersetzt bzw. modifiziert hat. Ihrer Ansicht nach beschreibt das Arbeitsgedächtnis eher den aktiven Anteil des informationsverarbeitenden Systems, während das traditionelle Konzept des Kurzzeitgedächtnisses eher den passiven Pufferspeicher beschreibt. Als Gemeinsamkeit des Arbeitsgedächtnisses und des Kurzzeitgedächtnisses führen von Aster et al. (2006) an, dass beide eine zeitlich begrenzte Zwischenspeicherung von eingehender Information ermöglichen und in ihrer Kapazität begrenzt sind. Allerdings geben sie einen zentralen Punkt an, in dem sich beide Konzepte unterscheiden: Das Kurzzeitgedächtnis beschreibt eine passive Gedächtnisform, das Arbeitsgedächtnis hingegen eine aktive. Während sich das Kurzzeitgedächtnis auf die passive Zwischenspeicherung von Information bezieht, die dann entweder ins Langzeitgedächtnis übertragen oder anschließend vergessen wird, beschreibt das Arbeitsgedächtnis eine zeitlich begrenzte Verarbeitung eingehender Informationen (von Aster et al. 2006, S. 16).

Weiters sind von Aster et al. (2006, S. 17) der Meinung, dass – auch wenn spezifische Aspekte der Theorie des Arbeitsgedächtnisses noch kontrovers diskutiert werden – die meisten kognitiven Psychologen dahingehend übereinstimmen, dass jede Definition

des Arbeitsgedächtnisses die zeitlich begrenzte Speicherung von Inhalten umfasst, die aktiv bearbeitet werden.

Kail und Hall (2001) sind der Meinung, dass zwar die Begriffe Kurzzeitgedächtnis und Arbeitsgedächtnis zentrale Konstrukte moderner Theorien über Gedächtnis und Kognition darstellen, aber nur wenige Forscher begonnen hätten, sich mit deren Relation zueinander zu beschäftigen. Um herauszufinden, ob Kurzzeitgedächtnis und Arbeitsgedächtnis voneinander abgegrenzt werden können, wenngleich das Kurzzeitgedächtnis einen Teil des Arbeitsgedächtnisses darstellt, führten diese zwei Studien durch. Kail und Hall (2001) konnten zeigen, dass sowohl mittels explorativer als auch mittels konfirmatorischer Faktorenanalyse Kurzzeitgedächtnisaufgaben und Aufgaben zum Arbeitsgedächtnis auseinander gehalten werden können.

Auch die Ergebnisse einer Studie von Engle, Laughlin, Tuholski und Conway (1999) zeigten, dass das Kurzzeit- und das Arbeitsgedächtnis unterscheidbare Konstrukte sind, auch wenn diese hoch korrelieren. Aus diesem Grund meinen diese, dass eine dichotome Sicht von Kurzzeitgedächtnisaufgaben und Arbeitsgedächtnisaufgaben zu simpel wäre, da – auch wenn diese beiden Konstrukte unterscheidbar sind – Aufgaben, die vermeintlich eines dieser Konstrukte messen, auch immer das andere Konstrukt reflektieren.

Die von Engle et al. (1999, S. 313) (auf rein konzeptueller Ebene) angeführte Gleichung fasst etwas anschaulicher zusammen:

Arbeitsgedächtniskapazität = Kurzzeitgedächtniskapazität + zentrale Exekutive oder kontrollierte Aufmerksamkeit + Messfehler.

3.3 Arbeitsgedächtnis und Intelligenz

Es existiert keine einheitliche und allgemeingültige Definition der Intelligenz. Sehr viele Wissenschaftler haben versucht, dieses Konstrukt zu definieren und diese Definitionen bestehen beinahe nebeneinander. Gittler und Arendasy (2003, S. 55) beschreiben den Umstand dermaßen, dass nach alltagspsychologischer Auffassung

Intelligenz eine relativ einheitliche Fähigkeit ist, es jedoch schwer fällt, zu beschreiben, was intellektuelle Leistungen eigentlich sind. Auch Psychologen können, so Gittler und Arendasy (2003, S. 55) weiter, Intelligenz nicht durch eine klare Umreißung von Aufgabentypen bzw. speziellen Fähigkeiten definieren.

Stern (1912, zitiert nach Gittler & Arendasy, 2003, S. 56): *„Intelligenz ist die allgemeine Fähigkeit des Individuums, sein Denken bewusst auf neue Forderungen einzustellen; sie ist die allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens.“*

Boring (1923, zitiert nach Gittler & Arendasy, 2003, S. 55): *„Intelligenz ist das, was Intelligenztests messen.“*

Wechsler (1956, S. 13) definierte Intelligenz wie folgt: *„die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umwelt wirkungsvoll auseinander zu setzen.“*

Anastasi (1958, zitiert nach Gittler & Arendasy, 2003, S. 55) behauptet, dass *„unsere Intelligenztests nur die Fähigkeit messen, in unserer speziellen Kultur erfolgreich zu sein.“*

Hofstätter (1977, zitiert nach Gittler & Arendasy, 2003, S. 56): *„Intelligenz ist die Fähigkeit zur Auffindung von Ordnungen (Redundanz) in der Welt.“*

Gardner (1991, zitiert nach Vidonyi, 2005, S. 9) betont die Wichtigkeit des kulturellen Aspekts. Was in der einen Kultur wichtig ist und als intelligent gilt, muss nicht unbedingt als intelligent in einer anderen Kultur gelten.

Guthke (1999, zitiert nach Vidonyi, 2005, S.9): *„Intelligenz ist der Oberbegriff für die hierarchisch strukturierte Gesamtheit jener allgemeinen geistigen Fähigkeiten (Faktoren, Dimensionen), die das Niveau und die Qualität der Denkprozesse einer Persönlichkeit bestimmen und mit deren Hilfe die für das Handeln wesentlichen Eigenschaften einer Problemsituation in ihren Zusammenhängen erkannt und die*

Situation gemäß dieser Einsicht entsprechend bestimmten Zielstellungen verändert werden kann.“

Die so genannten höheren mentalen Prozesse werden nach Schweizer und Moosbrugger (2004) als besonders wichtig für das Bearbeiten von Aufgaben eines Intelligenztests erachtet, da diese mentale Aktivitäten – wie Problem lösen und Schlussfolgern, die eng mit Intelligenz verbunden sind – ermöglichen. Forschungen, die sich auf diese Prozesse ausrichten, enthüllten Aufmerksamkeit und Arbeitsgedächtnis als wichtige Komponenten der kognitiven Grundlage.

Schweizer und Moosbrugger (2004) zu Folge wurde das Verhältnis von Aufmerksamkeit und Arbeitsgedächtnis bereits des Öfteren untersucht (neuere Untersuchungen stammen beispielsweise von: Schweizer & Moosbrugger, 1999; Schweizer, Zimmermann & Koch, 2000; Rockstroh & Schweizer, 2001). Jedoch sind die Resultate solcher Studien nach Meinung von Schweizer und Moosbrugger (2004) uneinheitlich, wobei diese finden, dass diese Uneinheitlichkeit vermutlich auf unterschiedlichen Verfahren, die zur Aufmerksamkeitsmessung herangezogen worden sind, und auf unterschiedlichen Arten von Aufmerksamkeit, mit denen diese Tests, assoziiert sind, basiert. Da diese Aufmerksamkeitsarten als eher unabhängig voneinander angesehen werden, gibt es Grund dafür, verschiedene Arten von Zusammenhängen zu Intelligenz zu erwarten. Für eine genauere Darstellung siehe Schweizer und Moosbrugger (2004).

Neben dem eben dargestellten Zusammenhang zwischen Arbeitsgedächtnis und Aufmerksamkeit gibt es zahlreiche Untersuchungen zur Beziehung zwischen Arbeitsgedächtnismaßen und Intelligenz (z.B.: Schweizer, 1996a, 1996b; Engle et al., 1999; Schweizer & Koch, 2001; Conway, Cowan, Bunting, Theriault & Minkoff, 2002; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002), von denen nun lediglich einige Resultate dargestellt werden sollen, da diese mit den restlichen weitgehend übereinstimmen, indem diese einen substantiellen Zusammenhang zwischen Arbeitsgedächtnis und Intelligenz abbilden (vgl. Heitz, Unsworth & Engle, 2005).

Die bereits erwähnte Studie von Engle et al. (1999) liefert überzeugende Hinweise für einen Zusammenhang der Kapazität des Arbeitsgedächtnisses mit fluider Intelligenz (siehe Kapitel 4.5.1.5), wobei diese jedoch die Ansicht vertreten, dass für die starke Korrelation die zentrale Exekutive oder Aufmerksamkeits-Kontroll-Komponenten verantwortlich sind, da kein Zusammenhang zwischen Kurzzeitgedächtnis und fluider Intelligenz gefunden werden konnte. Diese Ansicht teilen De Jong und Das-Smaal (1995) mit Engle et al. (1999).

Auch Conway et al. (2002) erhielten vergleichbare Resultate, da ihre Ergebnisse das Arbeitsgedächtnis, und nicht Kurzzeitgedächtniskapazität oder Verarbeitungsgeschwindigkeit, als guten Prädiktor für fluide Intelligenz bei jungen Erwachsenen und somit als einen essentiellen Aspekt von fluiden Fähigkeiten, hervorbringen.

Auch Geuß (1985) fand in seiner Untersuchung eine relativ enge Beziehung zwischen Intelligenz (gemessen mit den Progressive Matrices nach Raven) und dem „Zahlennachsprechen“ (vor- und rückwärts nach Wechsler), auch wenn anzuführen ist, dass es sich bei der verwendeten Stichprobe um eine anfallende Stichprobe handelte und somit die Ergebnisse somit keinesfalls sorglos generalisiert werden dürfen.

In einer von Schweizer und Moosbrugger (2004) durchgeführten Studie konnten Aufmerksamkeit und Arbeitsgedächtnis mehr als die Hälfte (61%) der Varianz der Intelligenz auf latenter Ebene vorhersagen, was als beachtliche Aufklärung zu interpretieren ist.

Im Gegensatz zu den bisher vorgestellten Studien, welche das Arbeitsgedächtnis als einheitliches System definiert haben und dessen Beziehung zu fluider Intelligenz untersuchten, gingen Haavisto und Lehto (2004) einen etwas anderen Weg. Diese trennten für ihre Studie das Arbeitsgedächtnis bereichsspezifisch in verbales und visuell-räumliches Arbeitsgedächtnis (vgl. Kapitel 3.1.1. und 3.1.2).

Grob zusammengefasst zeigt das Hauptergebnis von Haavisto und Lehto (2004), dass das visuell-räumliche Arbeitsgedächtnis mit fluider Intelligenz und das verbale Arbeitsgedächtnis mit kristalliner Intelligenz (siehe Kapitel 4.5.1.5) in Beziehung stehen, wobei diejenige zwischen visuell-räumlichem Arbeitsgedächtnis den stärkeren Zusammenhang aufzuweisen scheint (vgl. Shah & Miyake, 1996).

3.4 Arbeitsgedächtnis und Legasthenie

Nach der „Internationalen Klassifikation psychischer Störungen“ (ICD-10) der Weltgesundheitsorganisation (Dilling, Mombour & Schmidt, 2008) wird die Legasthenie oder Lese-Rechtschreibstörung den Entwicklungsstörungen zugeordnet und unter „umschriebene Entwicklungsstörungen schulischer Fähigkeiten“ (F81) eingegliedert. Störungen diese Kategorie werden wie folgt umschrieben:

„Es handelt sich um Störungen, bei denen der normale Erwerb von Fertigkeiten von frühen Entwicklungsstadien an beeinträchtigt ist. Dies ist nicht einfach Folge eines Mangels an Gelegenheit zu lernen, und nicht durch eine erworbene Hirnschädigung oder Krankheit verursacht.“ (S.294)

Die Lese-Rechtschreibstörung (F.81.0) wird demnach folgendermaßen definiert:

„Das Hauptmerkmal dieser Störung ist eine umschriebene und bedeutsame Beeinträchtigung in der Entwicklung der Lesefertigkeiten, die nicht allein durch das Entwicklungsalter, durch Visusprobleme oder unangemessene Beschulung erklärbar ist. Das Leseverständnis, die Fähigkeit, gelesene Worte wiederzugeben, vorzulesen und die Leistungen bei Aufgaben, für welche Lesefähigkeit benötigt werden, können sämtlich betroffen sein. Mit Lesestörungen gehen häufig Rechtschreibstörungen einher. Diese persistieren oft bis in die Adoleszenz, auch wenn im Lesen einige Fortschritte gemacht wurden.“ (S. 298)

Auch wird im ICD-10 (Dilling et al., 2008) darauf hingewiesen, dass es schon zuvor in der Entwicklung zu Auffälligkeiten kommt, vor allem zu Störungen beim Sprechen und der Sprache und dass die Störung in allen bekannten Sprachen auftreten kann. Folglich heißt es:

„Die Leseleistungen des Kindes müssen unter dem Niveau liegen, das auf Grund des Alters, der altersgemeinen Intelligenz und der Beschulung zu erwarten ist. (...) In der späteren Kindheit und im Erwachsenenalter sind die Rechtschreibprobleme meist größer als Defizite in der Lesefähigkeit“ (S. 299)

Klicpera, Schabmann und Gasteiger-Klicpera (2007, S. 119) geben zu bedenken, dass dies nicht die einzig gültige Definition der Legasthenie ist. Vielmehr verhält es sich so, dass verschiedene Versionen nebeneinander bestehen und unterschiedliche Definitionskriterien angewendet werden, um Legasthenie zu umschreiben. Daher ist auch jene der WHO nach ICD-10 nur als eine Möglichkeit unter anderen anzusehen.

Warnke et al. (2004, S. 2) weisen darauf hin, dass das ICD-10 von einem „doppelten Diskrepanzkriterium“ ausgeht, nämlich

1. die Diskrepanz zwischen der tatsächlichen Lese- bzw. Rechtschreibleistung und jener, die aufgrund der Intelligenz zu erwarten wäre sowie
2. jener zwischen tatsächlich erbrachter Leistung und der zu erwartenden aufgrund des Alters.

Laut WHO soll diese Leistungsdifferenz mehr als zwei Standardabweichungen betragen, um eine Legasthenie zu diagnostizieren (Dilling et al., 1994, zitiert nach Klicpera et al., 2007, S. 120). Dieses Kriterium zur Klassifikation ist allgemein nicht anerkannt und wird oft kritisiert, da es als zu einschränkend empfunden wird.

Nach Metz, Marx, Weber und Schneider (2003, S. 128) werden Kinder, die sowohl hinsichtlich ihrer intellektuellen Fähigkeiten als auch im Lesen und Rechtschreiben unterdurchschnittliche Leistungen zeigen, von den Legasthenikern abgegrenzt; sie werden als allgemein lese-rechtschreibschwach bezeichnet.

In einer Studie beschäftigten sich Metz et al. (2003) mit der Legitimation der Diskrepanzdefinition der Legasthenie. Bevor dessen Ergebnisse dargestellt werden können, muss der Begriff der *phonologischen Informationsverarbeitung* definiert

werden. Wagner und Torgesen (1987, zitiert nach Metz et al., 2003, S. 128) verstehen darunter die Nutzung von Informationen über die Lautstruktur von gesprochener und geschriebener Sprache. Darunter werden von diesen drei Bereiche subsumiert: die phonologische Bewusstheit, das phonologische Arbeitsgedächtnis und der Abruf aus dem semantischen Lexikon. Bei der phonologischen Bewusstheit handelt es sich um die Fähigkeit, die Lautstruktur der gesprochenen Sprache zu durchschauen. Das phonetische Rekodieren im Arbeitsgedächtnis bezeichnet die Fähigkeit, sprachliche Informationen im Arbeitsgedächtnis speichern und bearbeiten zu können. Beim Abruf aus dem semantischen Lexikon geht es in erster Linie um die Geschwindigkeit, mit der sich Kinder durch die Rekodierung schriftlicher Symbole in die lautlichen Entsprechungen Zugang zum Langzeitgedächtnis verschaffen können.

In der oben genannten Studie konnten Metz et al. (2003) zeigen, dass sowohl Kontrollkinder als auch Kinder, die erwartungswidrig positive Schülerleistungen bei geringem IQ erbrachten, so genannte *Overachiever*, gegenüber Legasthenikern bzw. allgemein-rechtschreibschwachen Kindern über eine eindeutig bessere phonologische Informationsverarbeitung verfügten und einen geringeren Anteil an Wahrnehmungsfehlern im Rechtschreibtest aufwiesen. Nach Metz et al. (2003) gibt es zwar einen Einfluss der intellektuellen Kapazität. Dieser ist aber weitaus weniger bedeutungsvoll als die Diskrepanzdefinition impliziert, was sich beispielsweise daran zeigt, dass der phonologischen Informationsverarbeitung eine weitaus größere Bedeutung als Prädiktor für schriftsprachliche Leistungen zukommt (S. 133).

Zusammenfassend lieferten die Ergebnisse Hinweise gegen die Annahme, dass bei intelligenten lese-rechtschreibschwachen Kindern ein qualitativ anderes Ursachenmuster gegeben ist als bei unterdurchschnittlich intelligenten lese-rechtschreibschwachen Kindern. Daher meinen Metz et al. (2003, S. 133), dass Gründe für eine Unterscheidung lese-rechtschreibschwacher Kinder nach ihrem Intelligenzniveau nicht wissenschaftlicher sondern pragmatischer Natur sei.

Esser, Wyschkon und Ballaschk (2008, S. 47) weisen in der Testbesprechung des BUEGA (Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter) darauf hin, dass der Diagnostiker vor einer Anwendung dieses

Verfahrens vor der grundsätzlichen Entscheidung stehe, ob er die im ICD-10 festgeschriebene Diskrepanzdefinition akzeptieren will. Denn im Manual dieses Verfahrens wird die vielfältige und gut begründete Kritik an dieser Konzeption nicht einmal erwähnt. Darüber hinaus wird dem Testanwender eine Auseinandersetzung mit dieser Diskussion empfohlen.

Hasselhorn, Tiffin-Richards, Woerner, Banaschewski und Rothenberger (2000) führen an, dass sich aus den Analysen von Mayringer und Wimmer (1999) Hinweise für einen Unterschied zwischen englischem und deutschem Sprachraum entnehmen lassen. In beiden Sprachräumen gehört der Abruf phonologischer Codes aus dem Langzeitgedächtnis zu den Defizitbereichen von lese-rechtschreibwachen Kindern. Allerdings scheint im englischen Sprachraum der phonologischen Bewusstheit eine größere Rolle zuzukommen, im deutschen Sprachraum aber dem phonologischen Arbeitsgedächtnis (S. 152).

Auch Wimmer (1996) beschäftigte sich mit der Untersuchung dieser beiden Sprachräume in Bezug auf Legasthenie. Er ging der Frage nach, ob legasthene Kinder beim Lesenlernen des Deutschen dieselben „*nonword*“-Leseschwierigkeiten zeigen, die für legasthene Kinder beim Lernen des Lesens der englischen Sprache charakteristisch sind.

Er ist der Meinung, dass bis zur Durchführung seiner Studie die meisten Untersuchungen die Schwierigkeiten der lese-rechtschreibschwachen Kinder beim „*nonword*“-Lesen in der englischen Sprache behandelten.

In einem Review von Rack, Olson und Snowling (1992) finden sich 16 solcher englischsprachiger Studien. In 10 (von sieben verschiedenen Forschergruppen durchgeführten) wird über eine beeinträchtigte Leistung im Lesen von „*nonwords*“ bei legasthenen Kindern im Gegensatz zu jüngeren Kontrollkindern mit demselben Leseniveau (*studies of nonword reading in dyslexics and reading-level-matched normal readers*) berichtet. Rack et al. (1992) weisen jedoch darauf hin, dass einige Studien möglicherweise u. a. deswegen kein Defizit beim Lesen von „*nonwords*“ finden konnten, da die Teilnehmerauswahl auf einer einzigen Messung der

Leseleistung basierte, was die Gefahr einer Regression zur Mitte mit sich bringt. Eine sehr anschauliche Erklärung dieses Regressionseffekts findet sich bei Gittler und Arendasy (2003, S. 35ff); für eine genauere Beschreibung dieser Untersuchungen siehe Rack et al. (1992).

Kurz gesagt konnten die Resultate der Untersuchung von Wimmer (1996) den Nachweis von spezifischen Schwierigkeiten beim Pseudowörterlesen bei legasthenen Kindern erbringen, die das Lesen der deutschen Sprache lernen.

Auch Hasselhorn, Tiffin-Richards, Woerner, Banaschewski und Rothenberger (2000) führten diesbezüglich eine Untersuchung durch, mit dem Ergebnis, das die Vermutung der Relevanz der Funktionstüchtigkeit des Speichers im phonologischen Arbeitsgedächtnis im deutschen Sprachraum für die Charakterisierung bzw. die Diagnose von Lese-Rechtschreibstörungen durchaus bestätigte. Ihrer Meinung nach führen die Besonderheiten der deutschen Sprache gegenüber der Englischen dazu, dass dem Rekodieren der sprachlichen Klanggebilde im phonetischen Speicher des Arbeitsgedächtnisses eine vergleichsweise große Bedeutung zukommt. Sie schlussfolgern, dass Funktionsdefizite in dieser Speicherkomponente des phonologischen Arbeitsgedächtnisses im deutschen Sprachraum zu einem Verursachungsfaktor der Legasthenie werden (S. 163).

Nach Warnke et al. (2004) haben Personen mit Lese-Rechtschreibstörung Schwierigkeiten, phonologisch gleich lautende Worte und Pseudoworte durch orthographische Merkmale, also visuell zu unterscheiden.

Darüber hinaus spielt das Arbeitsgedächtnis beim Erlernen des Lesens und Rechtschreibens eine bedeutende Rolle. Denn neben einem automatisierten Abruf des semantischen Lexikons aus dem Langzeitgedächtnis ist auch das kurzfristige Behalten von Lauten, Buchstaben, Silben und Wörtern im Arbeitsgedächtnis notwendig. So muss die lautsprachliche Information so lange im Arbeitsgedächtnis bereitgehalten werden, bis der Lese- und Rechtschreibvorgang abgeschlossen ist. Weisen Kinder eine eingeschränkte sprachliche Merkfähigkeit auf, sind sie gefährdet, später eine Lese-Rechtschreibstörung zu entwickeln.

Auch Klicpera et al. (2007, S. 183f) weisen auf einen deutlichen Zusammenhang zwischen Leseschwierigkeiten und dem phonologischen Rekodieren im Arbeitsgedächtnis hin. Mittels phonologischen Rekodierens kann eine Lautfolge in die artikulatorische Schleife eingehen und dadurch aktiv gehalten, manipuliert oder gespeichert werden. Dies ist bei unbekanntem bzw. neuen Wörtern oder auch Pseudowörtern notwendig. In einer neueren fMRI-Studie (*functional magnetic resonance imaging*) konnten Vasic, Lohr, Steinbrink, Martin und Wolf (2008) funktionelle Unterschiede in Hirnregionen, die den Exekutivfunktionen und dem Arbeitsgedächtnis zugerechnet werden, zwischen legasthenen 16-21-jährigen Personen und einer entsprechenden Kontrollgruppe feststellen. Während in dieser Studie Arbeitsgedächtnis und Exekutivfunktionen konfundiert waren, stellten Schuchardt, Maehler und Hasselhorn (2008) eine Beeinträchtigung der artikulatorischen Schleife (Arbeitsgedächtnis) bei legasthenen Volksschülern fest; ein Defizit der Exekutivfunktionen zeigte sich hier nicht.

Klinische Studien zur Testanalyse des Hamburg-Wechsler-Intelligenztest für Kinder-IV (HAWIK-IV) (Deimann & Kastner-Koller, 2008; eine Darstellung dieses Verfahrens findet sich unter 4.5.2.7) zeigten erwartungsgemäße Ergebnisse, da die Gruppe der Kinder mit Lese-Rechtschreibstörung signifikant schlechtere Werte im Indexwert *Arbeitsgedächtnis* erzielten.

Auch die Ergebnisse zur Überprüfung der Kriteriumsvalidität des Wechsler-Intelligenztests für Erwachsene (WIE) (von Aster et al., 2006, S. 42; eine Darstellung dieses Verfahrens findet sich unter 4.5.2.6) sprechen für diesen Zusammenhang. Denn Testpersonen mit Leseproblemen während der Schulzeit weisen die geringste Gesamtintelligenz auf und die Beeinträchtigungen im Verbalteil sind größer als im Handlungsteil und insbesondere das Arbeitsgedächtnis scheint beeinträchtigt zu sein.

3.5 Weiter Zusammenhänge mit Arbeitsgedächtnis

Nach von Aster et al. (2006) wird in der Literatur darauf hingewiesen, dass das Arbeitsgedächtnis eine Schlüsselkomponente des Lernens darstellt (vgl. Woltz, 1988). Eine differentialpsychologische Betrachtungsweise des Arbeitsgedächtnisses sagt

voraus – so von Aster et al. (2006, S. 17) –, dass die Lernfähigkeit umso höher ist, je größer das Arbeitsgedächtnis ist. Unter diesem Gesichtspunkt ist das Arbeitsgedächtnis für Unterschiede in der Lernfähigkeit bei einer breiten Palette von Lernleistungen verantwortlich (von Aster et al., 2006, S. 17). Von Aster et al. (2006) meinen weiter, dass kognitive Psychologen postuliert haben, dass das Arbeitsgedächtnis eines der wesentlichen Prädiktoren der interindividuellen Unterschiede im Lernen, in den geistigen Fähigkeiten und in der Flüssigkeit des schlussfolgernden Denkens darstellt (S. 17). Laut Kyllonen (1987, zitiert nach von Aster et al., S. 17) bildet das Arbeitsgedächtnis – gemeinsam mit der Geschwindigkeit der Informationsverarbeitung, dem deklarativen und dem prozeduralen Wissen – die individuelle Fähigkeit, neue Informationen zu erlernen. Nach Aster et al. (2006) lassen die bisherigen Forschungsergebnisse zusammenfassend vermuten, dass das Arbeitsgedächtnis eine wesentliche Moderatorvariable des Lernens darstellt.

Darüber hinaus sind Engle et al. (1999) der Meinung, dass Messungen der Arbeitsgedächtniskapazität die Leistungen in breit gefächerten, praxisnahen kognitiven Aufgaben und Leistungsmessungen reliabel vorhersagen können. Da diese unter anderem von signifikanten Zusammenhängen zwischen Arbeitsgedächtniskapazität und Sprachverständnis (vgl. Daneman & Merikle, 1996), Lernen neuer Vokabel, Schreiben, komplexem Lernen und Schlussfolgern berichten.

Außerdem konnten diverse Studien zeigen, dass Messungen der Verarbeitungsgeschwindigkeit und des Arbeitsgedächtnisses signifikant mit schulischen Leistungen korrelieren (Colom, Excorial, Shih & Privado, 2007). Und Gathercole, Pickering, Knight und Stegman (2004, zitiert nach Colom et al., 2007) fanden signifikante Relationen zwischen Leistungen in den mathematischen und naturwissenschaftlichen Fächern einerseits und der Arbeitsgedächtniskapazität andererseits.

Die Anzahl bereits durchgeführter Studien zwischen Arbeitsgedächtnis und einer Vielzahl anderer Faktoren scheint so zahlreich zu sein, dass nur ein kleiner Einblick in das große Gebiet des Arbeitsgedächtnis und weiteren Zusammenhängen gegeben werden konnte.

4 Gedächtnisdiagnostik

Geuß (1985) ist der Meinung, dass nur wenige psychische Leistungsbereiche des Menschen bis heute so intensiv untersucht wurden, wie das Gedächtnis. Eine mögliche Erklärung dieses Umstandes liegt darin, dass die Fähigkeit, Informationen zu speichern und wieder abzurufen, wohl bei jeglichem Denken und Handeln eine bedeutsame Rolle spielt.

4.1 Material zur Messung des Gedächtnisses

Bereits Ebbinghaus, der laut Schermer (2006) das Gedächtnis noch als einheitliche, feste Größe interpretierte, arbeitete 1885 mit so genannten *sinnlosen Silben*, indem er aus den einfachen Konsonanten des Alphabets und den elf Vokalen und Diphthongen alle möglichen Konsonant-Vokallaut-Konsonant-Silben (KVK-Silben) bildete. Dadurch ergaben sich ca. 2300 Silben mit denen er seine Experimente durchführte (Ebbinghaus, 1992).

Aber so sinnlos wie der Name sagt, waren diese keineswegs. Denn nach Baddeley (1979) versuchte Glaze 1928 die Bedeutungshaltigkeit von sinnlosen Silben zu messen. 15 Versuchspersonen sollten bei jeder von 2019 KVK-Silben beurteilen, ob diese an ein englisches Wort erinnert. Da gezeigt werden konnte, dass eine Silbe umso leichter gelernt werden kann, je größer ihr Assoziationswert ist, sind Silbenreihen wie *cef*, *yim*, *zoy*, die bei keiner der Versuchspersonen von Glaze eine Assoziation hervorriefen, viel schwerer zu lernen als Silben wie *bal*, *vac*, *duc*, bei denen alle Versuchspersonen eine Assoziation zu einem englischen Wort angaben (Baddeley, 1979).

Nach Lehr (2007) bestehen beim Lernen von sinnlosem Material Altersunterschiede, da ältere Menschen bei sinnlosem Material schlechter lernen als jüngere.

Neben der gerade beschriebenen Bedeutungshaltigkeit beeinflussen folgende Faktoren die Lerngeschwindigkeit bei beziehungslosem sprachlichen Material (Baddeley, 1979): eingeschätzte Leichtigkeit der Aussprache der Silben, Häufigkeit des

Auftretens der drei Buchstaben der KVK-Silbe in der gegebenen Reihenfolge in der englische Sprache und die Vorhersagbarkeit. Unter der Vorhersagbarkeit einer Silbe versteht Baddeley (1979) den Grad der Annäherung ihrer Buchstabenstruktur an die Buchstabenstruktur der englischen Sprache.

Zur Messung der Schwierigkeit des Materials meint Baddeley (1979), dass die meisten Versuche, diese vorherzusagen, nur von einer Grundlage (also entweder Assoziation, Leichtigkeit der Aussprache, Vertrautheit oder Ähnlichkeit zum Englischen) ausgehen. Deswegen sei wenig überraschend, wenn bei einer einfallsreichen Versuchsperson kein Maß, das auf einer einzigen Dimension des Kodierens beruht, sich als reliabler Prädiktor des Lernens bewähren kann.

Seiner Meinung nach wäre die beste Methode zur Vorhersage der Schwierigkeit von sinnlosen Silben immer noch die, die Versuchspersonen selbst zu befragen, da diese ihre Begründung auf das gesamte Spektrum möglicher Strategien der Kodierung beziehen und nicht wie die meisten Maße auf eine einzige Kodierungsdimension beschränkt sind.

Schermer (2006) zu Folge nutzten Brown 1958 und Peterson und Peterson 1959 im Vergleich zu Ebbinghaus eine Drei-Konsonanten-Folge (ein so genanntes Trigramm) für ihre klassischen experimentellen Studien zum Nachweis des Kurzzeitgedächtnisses.

4.2 Darbietung des Materials

Laut Baddeley (1979) hat die Sinnesmodalität, in der das Material dargeboten wird, große Auswirkungen auf die Kurzzeitspeicher-Komponente. Diese Aussage stützt Baddeley (1979) auf zwei Experimente, in denen das Material jeweils akustisch und visuell dargeboten wurde und bei beiden die akustische Darbietung des Materials zu besseren Reproduktionsleistungen führte als die visuelle. Diese Ergebnisse lassen Baddeley (1979) getrennte visuelle und auditive Speichersysteme vermuten, die für verschiedene Arten der Interferenz durch den Reproduktionsprozess anfällig sind.

4.3 Paradigma des verbalen Wiedererkennens

Aufgrund der Verschiedenartigkeit der zu verarbeitenden Informationen und der Relationen zueinander wird in der Literatur oft ausgeführt, dass es das (alleinige) Forschungsparadigma der verbalen Lern- und Gedächtnispsychologie nicht geben kann. Aus diesem Grund sind verschiedene Paradigmen entwickelt worden, auf die zurückgegriffen werden kann. Nach Kintsch (1982) zählen zu den klassischen Paradigmen Paarassoziations-Lernen, Serielles Lernen, Freie Reproduktion, Wiedererkennen und Verbales Diskriminationslernen.

4.3.1 Verbales Wiedererkennen

Wiedererkennen wird von Mandler (1980) als ein Prozess beschrieben, der eine Entscheidung darüber verlangt, ob ein Ereignis oder Item schon früher aufgetreten ist, d.h. es müssen alte und neue Informationen voneinander unterschieden werden (vgl. Anderson, 2009). Zur Erklärung des Wiedererkennens werden oft die Signalentdeckungstheorie (*signal detection theory*) und die Zwei-Prozess-Theorien (*dual-process theories*) herangezogen (Anderson, 2009), auf die jedoch in dieser Arbeit nicht eingegangen werden soll.

4.3.1.1 Methoden zum verbalen Wiedererkennen

Wie bereits aus dem vorigen Kapitel hervorgeht, wird die Rekognitions- bzw. Wiedererkennensleistung meist mit Lernlistenverfahren geprüft. Darunter werden Verfahren verstanden, in denen eine Liste mit „Lernwörtern“ – visuell oder auditiv – präsentiert wird und im Anschluss eine Testphase erfolgt.

Ein mögliches Verfahren zur Prüfung der Wiedererkennensleistung ist die „*forced choice*“ („erzwungene Wahl“) oder „*multiple choice*“ („Mehrfachaufgabe“). Vom Paradigma der erzwungenen Wahl zur Erfassung der Wiedererkennensleistung wird dann gesprochen, wenn in der Testphase jedes der dargebotenen Items mit wenigstens einem Distraktor (Items, die in der Lernphase nicht dargeboten wurden) zusammen präsentiert wird und die Versuchsperson entscheiden soll, welches Item in der Lernphase vorgekommen ist (Fleischhacker, 2001). Bei der Bewertung der erbrachten

Rekognitionsleistung ist jedoch zu berücksichtigen, dass die Anzahl der korrekten Antworten von der Ähnlichkeit der Distraktoren mit den dargebotenen Items abhängt (vgl. Kapitel 3.1.2).

Hinzu kommt, dass bei „*forced choice*“- und „*multiple choice*“-Verfahren die Häufigkeit der richtigen Antworten nicht zuletzt von der „Qualität“ und der Anzahl der Distraktoren abhängt, was vor allem bei der Testkonstruktion zu bedenken ist.

Nun soll darauf eingegangen werden, ob beim verbalen Wiedererkennen geschlechtsspezifische Leistungsunterschiede vorliegen.

4.3.1.2 Verbales Wiedererkennen und Geschlecht

Im Mittelpunkt bei geschlechtsspezifischen Untersuchungen kognitiver Fähigkeiten standen stets die Bereiche der mathematischen, räumlichen und verbalen Fähigkeiten (Halpern, 2000). Einen Leistungsvorteil des weiblichen Geschlechts konnten Studien von Hofstätter (1977, zitiert nach Fleischhacker, 2001) bei semantisch orientierten Problemstellungen wie Analogien, verbales Gedächtnis und Wortflüssigkeit zeigen. Auch nach McGiven et al. (1997, zitiert nach Halpern, 2000) verfügen Frauen über ein besseres Wiedererkennungsgedächtnis als Männer und dies von Kindheit an bis ins Erwachsenenalter.

In einer 165 Studien umfassenden Metaanalyse untersuchten Hyde und Linn (1988) Geschlechtsunterschiede bei verbalen Fähigkeiten. Dabei konnten lediglich 56 signifikante Unterschiede zwischen Männern und Frauen festgestellt werden. Insgesamt belegten 44 dieser 165 Studien, dass Frauen bei verbalen Aufgaben bessere Leistungen erbringen als Männer. Es überwog jedoch die Anzahl der Studien, die keine signifikanten Unterschiede zwischen den Geschlechtern bei verbalen Fähigkeiten nachweisen konnten. Um genauere Ergebnisse zu erhalten, führten die Autoren auch für die verschiedenen Untergruppen von verbalen Fähigkeiten getrennte Metaanalysen durch, die eine leichte Überlegenheit der Frauen belegten. Nach Hyde und Linn (1988) sind diese Unterschiede jedoch zu gering ausgeprägt, um sie als Geschlechtsunterschiede bezüglich verbaler Fähigkeiten interpretieren zu können.

Auch in einer von Hedges und Nowell (1995) durchgeführten Metaanalyse konnten keine Geschlechtsunterschiede bei verbalen Fähigkeiten gefunden werden.

Es existieren bislang wenige empirische Befunde zu geschlechtsspezifischen Unterschieden beim verbalen Wiedererkennen. Einzig beim Lernen von Wortlisten oder Paarassoziationen konnten Leistungsvorteile des weiblichen Geschlechtes beobachtet werden, nicht hingegen bei Wiedererkennungsaufgaben (Trahan & Quintana, 1990).

In diesem Zusammenhang ist durchaus denkbar, dass (auch) das Alter eine wesentliche Rolle für die verbale Wiedererkennungslleistung spielt. Mit diesem Thema befasst sich daher das nächste Kapitel.

4.3.1.3 Verbales Wiedererkennen und Alter

Die Altersvariable hat nach Lehr (2007) durch Vermittlung altersbezogener Überzeugungen eine wichtige Funktion bei der Bestimmung der Gedächtniskapazität. Fleischmann (1982, zitiert nach Lehr, 2007) fand in einer Stichprobe von 15- bis 94-jährigen Personen eine Reduzierung des Merkmumfangs zwischen vier und 28%, wobei sich diese Reduktion der Kapazität des Kurzzeitspeichers am stärksten bei visuell aufgenommenen Informationen zeigte.

Lehr (2007) weist darauf hin, dass Gedächtnisuntersuchungen in vieler Hinsicht Anforderungen darstellen, welche jener in Schulen und Universitäten ähnlich sind. Da in sehr vielen Untersuchungen über Altersunterschiede beim Gedächtnis Studenten herangezogen werden, während die Älteren meist schon aus dem Berufsleben ausgeschieden sind, ist sie der Meinung, dass die erzielten Resultate auch unter dem Aspekt, ob die Prüfsituation für die verschiedenen Altersgruppen unterschiedliche Anforderungen darstelle, interpretiert werden müssen. Auch Zivian und Darjes (1983, zitiert nach Lehr, 2007) meinen aufgrund einer von ihnen durchgeführten Untersuchung, dass die Jahre, die in der Ausbildung verbracht werden, wie vor allem die aktuelle Ausbildungssituation, bessere Prädiktoren von Gedächtnisleistungen seien als die Anzahl der Lebensjahre.

Eine weitere Begründung für die schlechteren Gedächtnisleistungen bei älteren Personen besteht darin, dass Ältere oft ein negativeres Bild von der eigenen Gedächtnisleistung haben als Jüngere. Deswegen meint Lehr (2007), ihre in der Realität erbrachten schlechteren Leistungen seien häufig darauf zurück zu führen.

Eine Möglichkeit zur Erklärung besteht zudem darin, die Leistungsdefizite zwischen älteren und jüngeren Personen als motivationsbedingt anzusehen.

Nach Lehr (2007) lassen sich altersspezifische Unterschiede in der Gedächtnisleistung auf verschiedene Mängel der Prozesse des Lernens und Erinnerns, wie z.B. reduzierter Informationsverarbeitungsgeschwindigkeit oder Schwierigkeiten bei der Integration neuer Informationen, zurückführen. Sie betont vor allem die Rolle der Speed-Komponente von Gedächtnistests bei der Verarbeitungsgeschwindigkeit. Wird das Lernmaterial zu schnell angeboten, werden ältere Personen dadurch mehr behindert als jüngere. Bei Eliminierung dieser Speed-Komponente nivellieren sich die Altersunterschiede.

Zahlreiche Untersuchungen konnten keine altersspezifischen Unterschiede bei verbalen Wiedererkennungsaufgaben nachweisen. Im Gegensatz dazu zeigen ältere Personen bei Reproduktionsaufgaben stets schlechtere Leistungen als jüngere. Craik & McDowd (1987) geben als Begründung dieser Ergebnisse an, dass die Reproduktion vermehrt eigenständige Verarbeitung erfordert, was beim Wiedererkennen nicht der Fall ist.

4.4 Weitere Einflussfaktoren auf die Gedächtnisleistung

4.4.1 Der serielle Positionseffekt

Der Positionseffekt ist einer der klassischen Effekte der Gedächtnisforschung. Dieser zeigt sich beispielsweise, wenn eine Liste aus unzusammenhängenden Wörtern reproduziert werden soll. Für die ersten paar Wörter ist die Leistung sehr gut (*Primacy-Effekt*), auch für die letzten paar (*Recency-Effekt*). Im mittleren Bereich ist sie jedoch eher schlecht (Zimbardo, 1999).

Schermer (2006) erklärt diesen Effekt damit, dass zu Beginn des Lernens die ersten Wörter der Liste auf den noch leeren Kurzzeitspeicher treffen. Deshalb können diese bis zur Erreichung von dessen Kapazitätsgrenze noch häufig wiederholt werden. Nach der maximalen Auslastung des Speichers müssen sich die auf ein Wort entfallenden Wiederholungen reduzieren. Da durch Wiederholung die Wahrscheinlichkeit erhöht wird, dass ein Element (Wort) in das Langzeitgedächtnis übertragen wird, kann vor allem für die ersten Wörter der vorgegebenen Liste von einer Aufnahme im Langzeitgedächtnis ausgegangen werden. Diesen Überlegungen zu Folge geht die bessere Reproduktionsleistung für den Anfangsteil der Liste – also der *Primacy-Effekt* – auf dessen Übertragung in den Langzeitspeicher zurück.

Die nachfolgenden Wörter treffen hingegen auf den gefüllten Kurzzeitspeicher und müssen für ihre Aufnahme ein im Speicher enthaltenes Element verdrängen und so zu sagen mit weniger „Platz“ auskommen. Unter diesen Bedingungen können diese Wörter nicht mehr so oft wiederholt werden, was zur Folge hat, dass sie schlechter behalten werden. Dies trifft zwar für den Mittelteil der Liste zu, jedoch nicht für den Schlussteil, da diese zum Zeitpunkt der Reproduktion noch im Kurzzeitgedächtnis enthalten sind und somit direkt aus diesem abgerufen werden können. Der *Recency-Effekt* ist somit eine Funktion des Kurzzeitgedächtnisses. Stört man den Wiederholungsprozess, der für die Bereithaltung eines Inhaltes im Kurzzeitgedächtnis notwendig ist, kann der *Recency-Effekt* unterbunden werden, während der *Primacy-Effekt* erhalten bleibt (vgl. dazu auch Baddeley, 1979).

Kurz gesagt, bezieht sich der Positionseffekt auf die Tendenz, Items vom Beginn (*Primacy-Effekt*) oder vom Ende (*Recency-Effekt*) bevorzugt wiederzugeben (Heubrock, 1992). Nebenbei bemerkt stellt der Positionseffekt einen weiteren empirischen Beleg für die Annahme eines Mehrspeichermodells (der Annahme eines Kurz- und Langzeitgedächtnisses) des Gedächtnisses dar.

4.4.2 Hypothese außer-experimenteller Interferenz

Wegen der Interferenz aufgrund der Abfolge von Buchstaben macht diese Hypothese die Vorhersage, dass bei vergleichbarem Niveau anfänglichen Lernens wenig

wahrscheinliche Buchstabenfolgen schneller als wahrscheinliche vergessen werden (Baddeley, 1979).

4.5 Psychometrische Gedächtnisdiagnostik

Gedächtnisdiagnostik kommt heutzutage im Kontext unterschiedlichster Fragestellungen zur Anwendung. Baller, Brand, Kalbe und Kessler (2006) führen neben gutachterlichen Stellungnahmen, bei denen es um die Feststellung und Bewertung funktioneller Defizite und Ressourcen geht, auch die Statuserhebung für die Planung von Therapien und die Verlaufstestung im Therapieverlauf zur Evaluation des therapeutischen Vorgehens an. Weiters betonen sie, dass diagnostische Verfahren im Kontext wissenschaftlicher Fragestellungen relevant sind.

Darüber hinaus kommt Gedächtnisdiagnostik des Öfteren im Rahmen der Intelligenzdiagnostik zur Anwendung, weshalb nun einige ausgewählte, für die vorliegende Arbeit als relevant angesehene Intelligenzmodelle, kurz vorgestellt werden sollen. Eine ausführlichere Darstellung der zahlreichen Intelligenzmodelle findet sich etwa bei Amelang und Bartussek (2001).

Erwähnenswert an dieser Stelle ist, dass nach Vidonyi (2005) Intelligenzmodelle meist im Zusammenhang mit der Konstruktion entsprechender Messverfahren entwickelt wurden. Aus diesem Grund werden zu jedem Intelligenzmodell auch Verfahren genannt, die sich auf das jeweilige Modell beziehen.

4.5.1 Intelligenzmodelle

4.5.1.1 Ein-Faktor-Konzeption: Globale Intelligenz von Binet

Im Rahmen der einfachsten Modellvorstellung wird davon ausgegangen, dass es sich bei Intelligenz um eine ganzheitliche, homogene Fähigkeit handelt, weshalb ein globaler Intelligenzwert ohne weitere Differenzierung bestimmt wird (Vidonyi, 2005, S.11).

Diese Konzeption wurde dem von Binet und Simon 1905 publizierten Test, der weltweit als erster Intelligenztest gilt, zu Grunde gelegt. Auch wenn Binet nach Amelang und Bartussek (2001) innerhalb der Intelligenz unabhängige Komponenten wie Gedächtnis, Vorstellung, Schlussfolgerung u. Ä. unterschied, sah der entwickelte Test einen einzigen Kennwert für die Beschreibung der intellektuellen Leistung vor, nämlich das Intelligenzalter. Damit wird implizit Intelligenz als einheitliches Ganzes behandelt.

Von den insgesamt 30 Testaufgaben, die dieser Test enthielt, wurden vier dieser Aufgaben dem Gedächtnis gewidmet. Die Aufgaben, die ins Deutsche übersetzt *Wiederholung dreier Ziffern* und *unverzögliche Wiederholung von Ziffern* hießen, dienten u. a. der Überprüfung des unmittelbaren Gedächtnisses. Wie bereits aus dem Titel erkennbar ist, verwendeten diese Aufgaben als Material Ziffern. Im Gegensatz dazu kam bei der Aufgabe *Wiederholung von aus 15 Wörtern zusammengesetzten Sätzen* verbales Material zur Anwendung, da diese Aufgabe das unmittelbare Erinnern verbalen Materials prüfen sollte. Die vierte der Gedächtnisaufgaben, *Zeichnen aus dem Gedächtnis*, wurde u. a. dem visuellen Gedächtnis gewidmet (Funke, 2006).

4.5.1.2 Die Zwei-Faktoren-Theorie von Spearman

Präzise Modellvorstellungen im Hinblick auf die Struktur der Intelligenz entwickelte erstmals Spearman 1904. Der Name seiner Theorie kommt von der zentralen Behauptung, dass jedes Maß für Intelligenz auf zwei Faktoren beruhe, nämlich einem Anteil zu Lasten von „*general intelligence*“ (g) und einem anderen, der spezifisch (s) für den jeweiligen Test ist. Diese spezifischen Faktoren werden für die durch den Generalfaktor nicht aufgeklärte Restvarianz verantwortlich gemacht. Dementsprechend basiere die Korrelation von Leistungsmaßen nur auf derjenigen Komponente, die ihnen gemeinsam ist (g) (Amelang und Bartussek, 2001, S. 204). Dieser Generalfaktor stellt somit einen Bestandteil aller Intelligenzleistungen dar und stellt nach Meinung von Spearman die beste Schätzung für das intellektuelle Niveau eines Menschen dar. Der Theorie von Spearman entsprechend sollten für Messverfahren Skalen mit möglichst hoher Generalfaktorsättigung ausgewählt werden (Vidonyi, 2005, S. 12): z. B.: Ravens SPM (*standard progressive matrices*) in

deutscher Bearbeitung von Kratzmeier und Horn (1988). Nach Vidonyi (2005) gehen auf Spearman's Zwei-Faktoren-Theorie auch die späteren hierarchischen Intelligenzmodelle zurück und die Testverfahren von David Wechsler stellen Verfahren dar, die nach diesem Modell konstruiert worden sind. Dazu zählen u. a.:

- WAIS-III und dessen deutsche Adaption WIE (Wechsler-Intelligenztest für Erwachsene) von von Aster, Neubauer und Horn (2006; eine Darstellung des WIE findet sich unter 4.5.2.6)
- WISC-IV und dessen deutsche Version HAWIK-IV (Hamburg-Wechsler-Intelligenztest für Kinder-IV) von Petermann und Petermann (2007; eine Darstellung des HAWIK-IV findet sich unter 4.5.2.7)

4.5.1.3 Das Modell mehrerer gemeinsamer Faktoren von Thurstone

Diesem Modell zu Folge sind beim Lösen von Denkaufgaben immer mehrere Gruppenfaktoren in wechselnden Gewichtungsverhältnissen beteiligt. Thurstone wählte den Ausdruck „*Primary abilities*“ als äquivalent zu den Gruppenfaktoren. Um deren Zahl und Spezifität zu ermitteln, entwickelte Thurstone das Verfahren der Multiplen Faktorenanalyse. Dafür grundlegend war die Annahme, dass die Zahl von Gruppenfaktoren stets niedriger ist als die Zahl der in einer Untersuchung eingesetzten Testsreihe – die Verfahren also nicht nur spezifische Faktoren erfassen –, und dass Leistung in einer bestimmten Aufgabe nicht zugleich von allen vorhandenen Primärfaktoren bestimmt wird (Amelang & Bartussek, 2001, S. 207).

Das Ergebnis waren zunächst (1938) neun, später sieben Primärfaktoren, die wie gleichberechtigte Bausteine, jedoch mit unterschiedlicher Breite, nebeneinander stehen. Nach Thurstone sind diese Faktoren keine realen Gegebenheiten, sondern theoretische Konstrukte, die zum Teil erheblich interkorrelieren (Gittler & Arendasy, 2003, S. 64).

Am besten gesichert sind folgende sieben Primärfaktoren (Gittler & Arendasy, 2003, S. 64f):

- (1) *space*: Raumvorstellung
- (2) *number*: rechnerische Fähigkeiten (nicht mathematische Begabung)
- (3) *verbal comprehension*: Sprachliche Intelligenz (Kenntnis und Bedeutung von Wörtern und angemessene Verwendung im Gespräch)
- (4) *word fluency*: Wortflüssigkeit, rasche Verfügbarkeit von Sprache
- (5) *memory (associative)*: Behalten paarweise gelernter Assoziationen (mechanische Kurzzeitgedächtnisleistung)
- (6) *reasoning*: logisches Schließen (unabhängig von der Korrektheit des Inhalts)
- (7) *perceptual speed*: Wahrnehmungsgeschwindigkeit; rasches Erkennen von Details, die in irrelevantem Material eingebettet sind

Beispiele für neuere Verfahren im deutschsprachigen Raum, deren Konstruktionsgrundlage das Modell mehrerer gemeinsamer Faktoren von Thurstone darstellt, sind:

- Der WILDE-Intelligenz-Test (WIT) von Jäger und Althoff (1994)
- Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) von Amthauer, Brocke, Liepmann und Beauducel (2001)
- Intelligenz Struktur Analyse (ISA) von Blum, Didi, Fay, Maichle, Trost, Wahlen und Gittler (1998)

4.5.1.4 Das „*Structure of Intellect*“-Modell von Guilford

Dieses Modell geht explizit von einer großen Anzahl unabhängiger Einzelfaktoren aus, wobei die Annahme eines allgemeinen Intelligenzfaktors prinzipiell abgelehnt wird (Vidonyi, 2005, S. 14). Guilford nimmt drei Gruppen von Variablen an, die jeweils eine der drei Dimensionen eines Würfels aufspannen (Gittler & Arendasy, 2003, S. 66):

- (1) *Operationen* (Bewertung, konvergente Produktion, divergente Produktion, Gedächtnis, Erkenntnis)
- (2) *Produkte* (Einheiten, Klassen, Beziehungen, Systeme, Transformationen, Implikationen)
- (3) *Inhalte* (figurale, symbolische, semantische, verhaltensbezogene)

Aus der Kombination dieser Operations-, Inhalts- und Produktklassen resultiert ein System von 120 (5 x 6 x 4) theoretisch postulierten Intelligenzfaktoren, von denen ca. 80% empirisch bestätigt sind. Auch wenn diesem Modell ein großer heuristischer Wert zugeschrieben werden kann, gilt es als Kritikpunkt anzumerken, dass die postulierte Unabhängigkeit der 120 Faktoren nicht gegeben ist und sich somit die Faktorenzahl reduzieren lassen müsse (Gittler & Arendasy, 2003).

Weiters existiert keine standardisierte Testbatterie zur Erfassung der verschiedenen Komponenten (Vidonyi, 2005, S. 15).

4.5.1.5 Das Modell der „fluid“ und „crystallized general intelligence“ von Catell

Die von Catell seit 1941 konzipierten Modellvorstellungen können als eine Synthese der Zwei-Faktoren-Theorie von Spearman und dem Modell mehrerer gemeinsamer Faktoren von Thurstone aufgefasst werden, da wie bei Spearman ein Generalfaktor, dessen Existenz allerdings aus den in spezifischer Weise interkorrelierenden Primärfaktoren von Thurstone, erschlossen wird (Amelang & Bartussek, 2001, S. 215).

Kurz gesagt ergaben Sekundäranalysen mehrere Sekundärfaktoren, von denen die ersten beiden von Catell als „fluid“ bzw. „*crystallized general intelligence*“ interpretiert wurden (Gittler & Arendasy, 2003). Wobei erstere (g_f) mehr die Fähigkeit, sich neuen Problemen oder Situationen anzupassen, ohne dass es dazu im wesentlichen Ausmaß früherer Lernerfahrungen bedarf, widerspiegelt und letztere (g_c) kognitive Fertigkeiten, in denen sich die kumulierten Effekte vorangegangenen Lernens kristallisiert und verfestigt hätten, vereinigt (Amelang & Bartussek, 2001, S. 215).

Es laden mehrere Primärfaktoren sowohl auf g_f als auch auf g_c , weswegen nach Gittler und Arendasy (2003) ein Faktor dritter Ordnung mit noch größerem Allgemeingrad, im Wesentlichen der Generalfaktor von Spearman, extrahiert werden kann.

Für einen Zusammenhang dieser zwei Faktoren mit dem Alter spricht nach Catell (1963, zitiert nach Amelang & Bartussek, 2001, S. 216), dass die Beobachtung, dass die Entwicklung der fluiden Intelligenz um das 14. bis 15. Lebensjahr zum Stillstand kommt, während die kristalline Intelligenz ihren Höhepunkt durchschnittlich vier bis fünf Jahre später erreicht, wobei die besonders Leistungsfähigen ihren Endpunkt zeitlich noch später erreichen, mit den Vorhersagen übereinstimmt.

Ein Verfahren, das sich auf das eben vorgestellte Modell von Catell bezieht, ist beispielsweise die CFT 20-R (Grundintelligenztest Skala 2 Revision) von Weiß (2006).

Vidonyi (2005) weist außerdem darauf hin, dass das AID 2 (Adaptives Intelligenz Diagnostikum von Kubinger & Wurst, 2000) Untertests enthält, bei denen die jeweilige Leistung weitgehend von der bisher erhaltenen Förderung abhängig ist. Zu diesen zählen *Alltagswissen*, *Synonyme finden*, *Soziales Erfassen* und *sachliches Reflektieren*.

4.5.1.6 Erweiterung von Cattells Modell durch Horn

Horn konnte 1965 nachweisen, dass neben den beiden Sekundärfaktoren g_f und g_c noch weitere auf gleicher Abstraktionsebene angesiedelte Faktoren existieren (Gittler & Arendasy, 2003, S. 74):

- G_v : Visuelle Informationsverarbeitung
- G_a : Auditive Informationsverarbeitung
- G_s : Geschwindigkeit der Informationsverarbeitung als Bestandteil fast aller kognitiver Tätigkeiten
- G_q : Quantitative Fähigkeit
- G_{sm} : Kurzzeitgedächtnis
- G_{lm} : Langzeitgedächtnis
- CDS (*Correct decision speed*): Schnelligkeit, mit der Antworten auf nicht-triviale Fragen gegeben werden können

Diese Faktorenstruktur konnte von Gustaffson (1989, zitiert nach Gittler & Arendasy, 2003) bestätigt werden.

4.5.1.7 Das Berliner Intelligenzstrukturmodell von Jäger

Dieses Modell vereint nach Amelang und Bartussek (2001, S. 223) Elemente aus den Theorien von Spearman, Thurstone und Guilford unter Berücksichtigung allgemeiner Erkenntnisse der Intelligenzforschung und sehr gezielt durchgeführter gesonderter Erhebungen bzw. Analysen.

Von einem psycholexikalischen Ansatz ausgehend führten Jägers diverse Analysen zu **vier Operationsfaktoren** (*Bearbeitungsgeschwindigkeit, Gedächtnis,*

Einfallsreichtum, Verarbeitungskapazität) und **drei Inhaltsfaktoren** (*figural-bildhaft, verbal, numerisch*) durch, deren Allgemeinheitsgrad etwa dem Niveau von Sekundärfaktoren entspricht. Die zwölf möglichen Operations-Inhaltskombinationen repräsentieren als Ganzes einen Generalfaktor der Intelligenz, der als nicht weiter differenzierbare Einheit dem Modell vorangestellt ist (Gittler & Arendasy, 2003, S. 76).

Als Beispiel für ein Testverfahren, das auf dem Berliner Intelligenzstrukturmodell von Jäger basiert, ist der BIS-Test (Berliner Intelligenzstruktur-Test) von Jäger, Süß und Beauducel (1996) anzuführen.

4.5.1.8 Zusammenfassung und Überleitung

Aufgrund der großen Zahl an Intelligenzmodellen lässt sich leicht der Schluss ziehen, dass es kein allein gültiges Intelligenzmodell – und somit auch nicht *den* Intelligenztest – gibt.

Die Darstellung der Modelle erfolgte aus dem Grund, um erkennbar zu machen, dass unter sehr viele Intelligenzmodelle Gedächtnisaspekte subsumiert und daher bei vielen Intelligenzmessungen solche erfasst werden.

Geuß (1985) zu Folge sind zur Erfassung von Gedächtnisleistungen im Laufe der Zeit verschiedene Methoden entwickelt worden. Auf das unmittelbare Behalten sowie auf das Kurzzeitgedächtnis beschränkend meint dieser weiter, dass grundsätzlich zwischen der Bestimmung der Gedächtnisspanne (einmalige Vorgabe; Methode der erhaltenen Glieder) und der Untersuchung des Materialerwerbs (mehrmalige Vorgabe des Materials; Erlernens-, Transfer- und Ersparnisemethode) unterschieden werden kann und dafür meist Lernlisten aus Begriffen, Zahlen, sinnarmen Silben, Bildern oder verschiedenen Kombinationen solcher Materialien zusammen gestellt werden (Geuß, 1985, S. 131).

Im deutschen Sprachraum stehen zur psychometrischen Gedächtnisdiagnostik Verfahren zur Erfassung spezifischer Teilleistungen, aber auch Gedächtnisbatterien

zur Verfügung. Einige in diesem Zusammenhang wichtig erscheinende Testverfahren – teils spezielle Gedächtnisdiagnoseverfahren, teils Intelligenztests, aber auch Verfahren zum Thema Lese-Rechtschreibstörung – sollen nun kurz vorgestellt werden.

4.5.2 Darstellung ausgewählter psychometrischer Verfahren

4.5.2.1 Inventar zur Gedächtnisdiagnostik (IGD)

Beim Inventar zur Gedächtnisdiagnostik (IGD, Baller, Brand, Kalbe & Kessler, 2006) handelt es sich um eine Testbatterie, die für den klinischen und den leistungsdiagnostischen Anwendungsbereich konzipiert wurde. Es besteht aus drei Testmodulen zur Überprüfung der Lern- und Merkfähigkeit (Modul A), des semantischen und des autobiografischen Gedächtnisses (Modul B und C), wobei die einzelnen Module unabhängig voneinander durchgeführt und interpretiert werden können. Da sich die vorliegende Arbeit mit der Messung des Kurzzeit- bzw. Arbeitsgedächtnisses beschäftigt, wird auf die Darstellung der Module B und C verzichtet und nur die für diese Arbeit relevanten Untertests des Moduls A näher erläutert.

Die Autoren des IGD geben an, dass neben den auch in anderen Testbatterien erfassten Gedächtnisleistungen auch ausführlich Arbeitsgedächtnisleistungen, prospektives Gedächtnis, Priming und inzidentielles Lernen (unbeabsichtigtes Lernen bzw. Lernen ohne Lernabsicht, Kubinger, 2006) überprüft werden. Im Manual der IGD ist explizit angeführt, dass das Testmodul A aufgrund der guten Differenzierungsmöglichkeiten auch im höheren Leistungsbereich zur Überprüfung der Lern- und Merkfähigkeit bei Gesunden für leistungsdiagnostische Fragestellung eingesetzt werden kann. Des Weiteren stellt der IGD ein geeignetes Instrument zur Untersuchung von Fragestellungen der neurowissenschaftlichen und psychologischen Forschung dar (Baller et al., 2006).

Das Testmodul A zur Überprüfung der Lern- und Merkfähigkeit umfasst 12 Untertest und prüft prospektive Gedächtnisleistungen, Priming, Arbeitsgedächtnis sowie Lern- und Merkfähigkeit für neue visuelle und verbale Stimuli. Ein Untertest erfasst

sprachliche Verarbeitungsprozesse des Einspeicherns (semantisch/phonematisch); ein weiterer Untertest überprüft das inzidentelle Lernen.

Nun werden die 12 Untertests des Testmoduls A angeführt und in Anlehnung an das Manual des IGD beschrieben:

A1 Prospektives Gedächtnis: Handlungsanweisung einspeichern, bereithalten und in einem bestimmten zeitlichen und situativen Kontext abrufen

A2 Zahlenspanne: Kurzfristiges Behalten von Zahlenreihen zunehmender Komplexität

A3 Verbales Arbeitsgedächtnis: Wörter mit einem bestimmten Merkmal aus einer Wortliste selektieren und speichern, inzidentelles Lernen wird durch die spätere Abfrage der nicht selektierten Wörter erfasst

A4 Visuelles Arbeitsgedächtnis: Lage und Ausrichtung von Linien in Objekten und gleichzeitig die Position der Objekte auf der Seite verarbeiten

A5 Exekutive Kontrolle: Aufmerksamkeitsshift zwischen zwei zu merkenden visuellen Reizmustern; für den Abruf Transkodierung in verbale Information erforderlich

A6 Verbales Lernen: Worterkennung aus einer Textvorlage mit semantisch ähnlichen Distraktoren

A7 Visuelle Lernen: Figurenerkennung aus ähnlichen Vorlagen und freier Abruf eines fehlenden Figurendetails

A8 Paarassoziationslernen: Form/Farb-Paare lernen, Abruf durch Vorgabe der Form – Zuordnen der Farbe

A9 *Verzögerte Rekognition*: Wortliste: Rekognition von bewusst und nicht bewusst gelernten Wörtern aus Untertest 3 aus einer Wortliste mit phonematischen und semantischen Distraktoren und neutralen Wörtern

A10 *Freier Abruf von Textinformation aus Untertest 6 mit Hinweisreizen durch Fragen*

A11 *Verzögerte Rekognition*: Figuren: Rekognition der Figuren aus Untertest 7 aus einer Anzahl von dem Zielitem ähnlichen und weniger ähnlichen Figuren

A12 *Priming*: Wortfragmente teilweise zuvor verarbeiteter Wörter des Untertests 3 ergänzen

Den eben dargestellten 12 Untertests des Testmoduls A werden fünf verschiedenen Skalen (Kurzzeitgedächtnis/Arbeitsgedächtnis, Lernen, Verzögerter Abruf, Verbales Gedächtnis, Visuelles Gedächtnis) zugewiesen. Bei der Skala, die für die vorliegende Arbeit von größter Bedeutung ist, handelt es sich um die Skala *Kurzzeitgedächtnis/Arbeitsgedächtnis*, welche die Untertests *Zahlenspanne*, *Visuelles/Verbales Arbeitsgedächtnis* und *Exekutive Kontrolle* beinhaltet.

Da es von Interesse ist, wie mit Hilfe des IGD das Kurzzeit- bzw. Arbeitsgedächtnis gemessen wird, werden diese vier Untertest an dieser Stelle einer genaueren Betrachtung unterzogen.

A2 *Zahlenspanne*: Der Testperson werden vom Testleiter einzelne Zahlenreihen zunehmender Länge (insgesamt zwölf) vorgelesen, die diese im Anschluss aufschreiben soll. Die ersten zwei Zahlenreihen umfassen drei Ziffern, die nächsten zwei vier usw., so dass die letzten beiden sieben Ziffern umfassen. Dieser Untertest dient zur Überprüfung der Funktion des Einspeicherns und Haltens bzw. frei Reproduzierens.

A3 Verbales Arbeitsgedächtnis: Der Testleiter liest eine Wortliste mit 14 Wörtern vor, von denen sieben ein „r“ enthalten. Die Aufgabe der Testperson besteht darin, sich nur diese sieben Wörter, die ein „r“ enthalten, zu merken und anschließend zu notieren. Die Anforderung dieser Aufgabe liegt in der Unterscheidung zwischen relevanten und irrelevanten Wörtern und dem gleichzeitigen Einspeichern der selektierten Wörter. Dieser Untertest soll die Funktion des Selektierens und Haltens bzw. frei Reproduzierens erfassen.

A4 Visuelles Arbeitsgedächtnis: Es werden sieben Kästchen, die Linien verschiedener Lage und Ausrichtung enthalten, über eine Seite verteilt dargeboten. Die Testteilnehmer sollen sich die Lage und Ausrichtung der Linien und die Position auf der Seite merken, die durch die Kästchen vorgegeben ist. In der Abfrage werden die Kästchen in gleicher Positionierung auf der Seite dargeboten, und die Testpersonen sollen die dazugehörigen Linien in die Kästchen einzeichnen. Diese Aufgabe stellt eine Mehrfachanforderung hinsichtlich der räumlich-visuellen Verarbeitung dar. Die einzuprägenden Linien müssen bezüglich ihrer Merkmale Ausrichtung und Lage frei reproduziert und zudem richtig im Raum positioniert werden. Dieser Untertest dient zu Erfassung der Funktion des Haltens und gleichzeitig Verarbeitens bzw. frei Reproduzierens.

A5 Exekutive Kontrolle: Dargeboten werden Kästchen von 1 bis 9, die eine unterschiedliche Anzahl von Punkten und Dreiecken enthalten. Aufgabe des Testteilnehmers ist es, die Punkte und Dreiecke in abwechselnder Reihenfolge zu zählen und die so entstehende Zahlenfolge im Anschluss aufzuschreiben. Dieser Untertest soll der Erfassung der Funktion des Haltens und Manipulierens bzw. frei Reproduzierens dienen.

4.5.2.2 Wechsler Gedächtnistest – Revidierte Fassung (WMS-R)

Der Wechsler Gedächtnistest in seiner revidierten Form (WMS-R, Calabrese, Deisinger, Härting, Kessler, Markowitsch & Neufeld, 2006) hat, da er neben einer breiten Palette verbaler und nichtverbaler Kurzzeit- und Langzeitgedächtnisaufgaben

auch noch Aufmerksamkeitsfunktionen miterfasst, breiten Eingang in die angloamerikanische Neuropsychologie gefunden (vgl. Parkin, 1999).

Die Autoren geben an, dass die Entwicklung der WMS-R zum Ziel hatte, eine Testbatterie für den klinisch-neuropsychologischen Gebrauch zu erstellen. Auch wenn der WMS-R einer Differenzierung im unteren mnestischen Leistungsbereich dienen soll und er nicht geeignet ist für Leistungsdifferenzierungen im oberen (gesunden) mnestischen Leistungsbereich, ist es von Interesse, wie hier vor allem die Messung des Kurzzeitgedächtnisses stattfindet.

Der revidierte Wechsler Gedächtnistest besteht insgesamt aus 13 Untertests, aus denen sich fünf Gesamt-Indizes errechnen lassen: Verbales Gedächtnis, Visuelles Gedächtnis, die zusammen den Allgemeinen Gedächtnis-Index ergeben, Aufmerksamkeit/Konzentration und Verzögerte Wiedergabe. Da für die vorliegende Arbeit nur zwei Untertests von Bedeutung sind, sollen nur diese kurz vorgestellt werden. Bei diesen zwei Untertests handelt es sich um die Untertests *Zahlenspanne* und *Blockspanne*.

Zahlenspanne: Wie auch schon im gleichnamigen Untertest des IGD werden beim Untertest *Zahlenspanne* des WSM-R dem Probanden Zahlenfolgen von wachsender Länge vorgelesen, die dieser im unmittelbaren Anschluss wiederholen muss. Ein Unterschied besteht jedoch darin, dass beim IGD die Testperson diese Zahlen aufschreiben muss und beim WSM-R diese die Zahlen verbal zu wiederholen hat und dass der WSM-R auch die *Zahlenspanne rückwärts* (die Ziffern müssen in umgekehrter Reihenfolge wiederholt werden) prüft. Gemeinsam ist beiden Untertests zwar, dass diese bei einer Länge von drei Ziffern beginnen, doch im Gegensatz zum IGD, bei dem die maximale Ziffernlänge bei sieben liegt, beträgt die größte Zahlenfolge des WSM-R beim Vorwärtswiederholen acht Ziffern.

Visuelle Merkspanne (Blockspanne): Auch dieser Untertest besteht aus zwei Aufgabentypen, der *Blockspanne vorwärts* und der *Blockspanne rückwärts*. Bei der *Blockspanne vorwärts* werden dem Probanden auf dem Blockspannenbrett (ein Brett auf dem sich mit Ziffern bedruckte Würfel befinden) Folgen wachsender Länge

gezeigt, die dieser im unmittelbaren Anschluss wiederholen, also auf dieselben Würfel in derselben Reihenfolge tippen muss. Bei der *Blockspanne rückwärts* werden dem Probanden ebenfalls Folgen wachsender Länge gezeigt, die dieser jedoch in genau umgekehrter Reihenfolge nachzutippen hat. Begonnen wird jeweils mit einer Tippfolge von zwei Ziffern, dann wird eine um eine Ziffer längere vorgegeben. Dies geht solange bis bei der *Blockspanne rückwärts*, genauso wie bei der *Zahlenspanne rückwärts*, die Anzahl der Ziffern bei sieben liegt und bei der *Blockspanne vorwärts*, abermals genauso wie bei der *Zahlenspanne vorwärts*, die Folge eine Länge von acht Ziffern hat.

4.5.2.3 Der Auditiv-Verbale Lerntest nach Rey und Lezak (AVLT)

Mit dem auditiv-verbale Lerntest nach Rey und Lezak (Heubrock, 1992) liegt ein diagnostisches Instrument vor, das möglichst verschiedene, abgrenzbare Aspekte der Merkfähigkeit erfassen soll und Lernbedingungen simuliert und Lernverläufe abbildet.

Auch wenn der AVLT eigentlich primär für den Einsatz in der Neuropsychologie entwickelt wurde, da nach Heubrock (1992) im deutschsprachigen Raum mit wenigen Ausnahmen keine neuropsychologischen Testverfahren existieren, mit denen verschiedene verbale Merkfähigkeitsparameter erfasst werden können, erscheint es aus mehreren Gründen wichtig, dieses Verfahren kurz vorzustellen.

Der erste dieser Gründe besteht darin, dass dieser Test – im Gegensatz zum IGD und WMS-R, die bereits oben kurz dargestellt worden sind – mit verbalem Material (Wörtern) arbeitet.

Ein weiterer Grund liegt darin, dass der AVLT – im Gegensatz zu den bereits vorgestellten Tests – auch das Wiedererkennen (*recognition*) von zuvor gelerntem Material prüft. Anzumerken ist an dieser Stelle, dass bei diesem Wiedererkennen mit einer Wortliste, aus der die zuvor gelernten Wörter wieder erkannt werden sollen, gearbeitet wird, die zusätzlich semantisch und phonematisch ähnliche Items enthält.

Als letzter Punkt spricht für die Darstellung des AVLTL, dass mit diesem neben dem Erfassen der unmittelbaren Merkspanne auch eine Aussage über Positionseffekte (*Primacy-* und *Recency-Effekt*) möglich ist. Hinzu kommt, dass der AVLTL laut Angaben von Elger, Hasse-Sander, Helmstaedter, Horn und Müller (1997) auch in der wissenschaftlichen Forschung rege Verwendung findet.

Der AVLTL besteht aus Wortlisten mit jeweils 15 Substantiven, wobei für eine Testdurchführung zwei solche Wortlisten benötigt werden. Zunächst werden vom Versuchsleiter die Wörter der ersten Liste (A) laut vorgelesen und anschließend die Versuchsperson aufgefordert, so viele Wörter wie möglich in beliebiger Reihenfolge wiederzugeben. Dieser Vorgang wiederholt sich mit derselben Liste und bei gleicher Versuchsanordnung vier weitere Male (A1-A5). Danach liest der Versuchsleiter eine zweite Wortliste (B) vor, die ebenfalls unmittelbar nach der Darbietung abgefragt wird. Im letzten Durchgang (A6) soll sich die Versuchsperson ohne erneute Vorgabe der Liste, an so viele Wörter wie möglich aus der ersten, mehrfach wiederholten Wortliste erinnern. Ergänzend dazu kann auch das Wiedererkennen zuvor gelernter Materials geprüft werden, indem eine Liste von 50 Substantiven (R) vorgelegt wird, von denen die Versuchsperson die Items der ersten Liste (A) identifizieren soll. Diese R-Liste enthält – wie bereits oben erwähnt – neben den Wörtern der Listen A und B auch semantisch und phonematisch ähnliche Wörter. Im Detail besteht diese R-Liste zusätzlich zu den jeweils 15 Substantiven der Listen A und B aus acht Wörtern, die eine semantische, aus sechs Wörtern, die eine phonematische, und einem Wort, das sowohl eine semantische als auch eine phonematische Beziehung zu einem Wort aus der Liste A aufweisen, weiters aus jeweils zwei Wörtern, die eine semantische bzw. phonematische Beziehung zu einem Wort aus der Liste B aufweisen und einem Wort, das sowohl eine semantische als auch eine phonematische Beziehung zu einem Wort aus der Liste B aufweist.

Heubrock (1992) führt folgende Parameter der verbalen Merk- und Lernfähigkeit an, die mittels AVLТ ermittelt werden können:

- die unmittelbare Merkspanne (A1),
- die Lernverlaufskurve über alle Durchgänge,
- die Tendenz zu proaktiver Hemmung (A5-B),
- die Tendenz zu retroaktiver Hemmung (B-A6),
- Positionseffekte (Primacy- und Recency-Effekt),
- Lernstrategien (z.B. semantische Cluster-Bildung) und
- die Fähigkeit, Gelerntes wieder zu erkennen (R).

Um eine Aussage über Positionseffekte machen zu können, wird vorgeschlagen, die 15 Items des AVLТ in Drittel zu gruppieren (1-5, 6-10, 11-15) und die Anzahl der aus jedem Drittel genannten Wörter jeweils über die Lerndurchgänge zu summieren und deren prozentuellen Anteil zu berechnen. In gleicher Weise lassen sich auch Positionseffekte für die Interferenzliste (B) und für die Wiedergabe nach Interferenz (A6) berechnen.

Entscheidend für das Vorliegen von Positionseffekten ist nach Heubrock (1992) das deutliche Überwiegen von Reproduktionen aus dem ersten und/oder letzten Drittel gegenüber dem mittleren Drittel der Items, wobei jedoch keine genaue Angabe besteht, ab welchem Prozentsatz von einem deutlichen Überwiegen gesprochen werden kann.

4.5.2.4 Verbaler Lern- und Merkfähigkeitstest (VLMT)

Der Ausgangspunkt des Verbalen Lern- und Merkfähigkeitstests (VLMT, Helmstaedter, Lendt & Lux, 2001) bildete, genauso wie beim soeben dargestellten

AVLT, der akustisch-verbale Lerntest von Rey (1964). Demzufolge sind sich diese zwei Verfahren sehr ähnlich, mit der Ausnahme, dass beim VLMT nach einer 30-minütigen Pause die erste Wortliste ein siebtes Mal zu produzieren ist. Auch das Überprüfen der Wiedererkennensleistung findet beim VLMT statt, und zwar genauso wie beim AVLT beschrieben.

Aber wozu wird dann der VLMT hier extra angeführt? Weil in einer Studie von Elger et al. (1997) der Frage nachgegangen wird, welche Gedächtnismessungen mit einem Vorgehen wie beim VLMT, daher höchstwahrscheinlich auch AVLT, wenn auch (noch) nicht empirisch belegt, möglich sind.

Elger et al. (1997) führen an, dass Heubrock (1992) und Vakil und Blachstein (1993) behaupten, dass die Leistung beim ersten Reproduzieren der ersten Wortliste (vgl. A1 des AVLT) die Gedächtnisspanne bzw. das Kurzzeitgedächtnis reflektiert; – wahrscheinlich weil normalerweise ungefähr sieben Items korrekt wiederholt werden können. Aber dieses Statement steht nach Elger et al. (1997) in Konflikt mit der oft geäußerten Annahme, dass sogar beim ersten Wiederholen einer Wortliste einige der Items aus dem Langzeitgedächtnis wiederholt werden.

Konkret wurden in der Studie von Egler et al. (1997) Strukturgleichungsmodelle verwendet, um die Variablen, die man mit Hilfe des VLMT erhält, mit der Unterscheidung zwischen Kurzzeitgedächtnis und Langzeitgedächtnis in Beziehung zu setzen. Die Autoren dieser Studie sind der Meinung, dass auf diese Weise eine empirische Antwort auf die Frage, ob das erste Reproduzieren der ersten Wortliste wirklich nur das Kurzzeitgedächtnis reflektiert oder nicht, gefunden werden kann. Außerdem sei es mit dieser Vorgehensweise auch möglich, zu sehen, ob die Unterscheidung zwischen Kurz- und Langzeitgedächtnis reicht, um die Kovarianzen zwischen allen erhaltenen Variablen zu erklären.

Die Ergebnisse der Studie lassen sich kurz zusammenfassen, indem man sagt, dass keine der durch VLMT erhaltenen Variablen ein reiner Indikator für die Kapazität des Kurzzeitgedächtnisses ist, d. h. dass bei jeder Variable, bei manchen mehr, bei manchen weniger, auch Langzeitgedächtniskomponenten miteinfließen. Darüber

hinaus zeigen die erhaltenen Resultate, dass die zwei latenten Variablen (Kurzzeitgedächtnis und Langzeitgedächtnis) ausreichen, um alle Kovarianzen zwischen den erhaltenen, empirischen Variablen, die mittels VLMT erhalten werden, statistisch zu erklären. Dieser Punkt validiert somit nach Angabe der Autoren nicht nur den VLMT, sondern auch die Annahme eines Kurz- bzw. Langzeitgedächtnisses.

Anzumerken ist an dieser Stelle, dass es sich bei den hier vorgestellten Verfahren um Paper-Pencil-Tests, also um keine Computervorgabe handelt und das zu merkende verbale Material immer nur auditiv präsentiert wurde.

Mit dem Verbalen Lerntest (VLT, Sturm & Willmes, 1999) liegt ein Testverfahren vor, das beide Bedingungen erfüllt, also als Computerversion verfügbar ist und in dem das verbale Material visuell präsentiert wird. Darüber hinaus wird mit der Methode des Wiedererkennes gearbeitet. Da die Autoren aber davon ausgehen, dass das Lernmaterial des VLT sukzessive im verbalen Langzeitgedächtnis abgespeichert wird und über die Methode des Wiedererkennens abgerufen wird, befasst sich dieses Verfahren nicht mit dem selben Teil des Gedächtnisses wie diese Arbeit (Kurzzeitgedächtnis) und es wird daher an dieser Stelle verzichtet, den VLT vorzustellen.

Mit dem Benton-Test (BT, Sivan & Spreen, 1996) liegt ebenfalls ein Testverfahren vor, das in einer Computerversion vorliegt und in dem das Testmaterial visuell dargeboten wird, auch wenn es sich bei diesem um geometrische Figuren handelt und nicht um verbales Material. Außerdem wird in einer der vorliegenden Testformen des BT, der Wahlform, mit der Methode des Wiedererkennens gearbeitet. Aus diesen Gründen soll auch die Wahlform des BT an dieser Stelle kurz vorgestellt werden.

4.5.2.5 Der Benton-Test (Wahlform)

Arthur Benton veröffentlichte 1946 die erste Form des Benton-Tests (Visual Retention Test), da er den Bedarf nach einem Gedächtnistest für visuelle Formen für den klinischen Gebrauch erkannte. Seither wurde dieser des Öfteren überarbeitet, erweitert und die Normwerte restandardisiert. Heute liegt daher die siebente,

vollständig überarbeitete Auflage des ursprünglichen Verfahrens vor (Sivan & Spreng, 1996). Der Benton-Test (BT) wird von den Autoren als Instrument der klinischen und der Forschungsarbeit beschrieben, mit dem die visuelle Wahrnehmung, das visuelle Gedächtnis und visuo-konstruktive Fähigkeiten gemessen werden können.

Der BT enthält drei alternative Formen (C, D, und E) und vier alternative Anwendungsformen (Instruktionen A, B, C, D). Jede Form besteht aus zehn Vorlagen mit einer oder mehreren Figuren; die meisten enthalten drei Figuren. Es existieren zwei Wahlformen (F und G), bei denen fünf verschiedene Instruktionen (Instruktionen M, N, O, P, CR) angewendet werden können. Für die zwei Wahlformen werden die gleichen Vorlagen verwendet, jedoch jeweils 15. Jeder dieser Vorlagen folgt eine Wahlseite, auf der vier Vorlagen abgebildet sind, aus denen die Versuchsperson die zuvor gesehene herausuchen muss. Im Gegensatz zu den anderen Formen, bei denen die Figuren nachgezeichnet werden sollen, sind bei den Wahlformen die richtigen Figuren wiederzuerkennen.

Kommt die Instruktion M zur Anwendung, wird jede Vorlage zehn Sekunden lang dargeboten, ehe die Versuchsperson die richtige aus den vier Vorlagen herausuchen soll. Bei der Instruktion N ist nur die Darbietungszeit auf fünf Sekunden verkürzt, ansonsten ist sie identisch mit der Instruktion M. Instruktion O benutzt eine Wartezeit von 15 Sekunden bevor die Wahl getroffen werden kann. Mittels der Instruktion P wird die Wahlform zu einer Formunterscheidungsaufgabe, denn die Stimulus- und Wahlkarte werden gleichzeitig dargeboten und die Versuchsperson muss die der Stimuluskarte entsprechende Vorlage lediglich auswählen. Instruktion CR ergänzt die Instruktion zum Abzeichnen (C) dadurch, dass die Versuchsperson nach Beendigung des Abzeichnens die von ihr gezeichnete Vorlage unter den auf der Wahlkarte vorgelegten Zeichnungen erkennen muss. An dieser Stelle ist zu erwähnen, dass Instruktion C die Gedächtniskomponente des Tests von der wahrnehmungsmotorischen Komponente zu trennen erlaubt. Es wird also bei Anwendung der Instruktion C bzw. CR keine Gedächtnisleistung überprüft, genauso wenig wie bei Instruktion P, bei der es sich um eine reine Formunterscheidungsaufgabe handelt.

4.5.2.6 Wechsler-Intelligenztest für Erwachsene (WIE)

Wie bereits aus dem Namen und der obigen Darstellung der Intelligenzmodelle zu entnehmen ist, handelt es sich beim WIE (Wechsler-Intelligenztest für Erwachsene, von Aster et al., 2006) um einen Intelligenztest. Er dient in erster Linie zu einem Einsatz in der klinischen Einzelfalldiagnostik. Darüber hinaus, auch wenn Wechsler seine Tests nicht explizit als neuropsychologische Untersuchungsinstrumente entwickelte, sind von Aster et al. (2006) der Meinung, dass ihm klar war, dass deren Ergebnisse einen Beitrag zum Verständnis der Beziehungen zwischen Hirnfunktionen und Verhalten liefern. Wechsler (1940, zitiert nach von Aster et al., 2006, S. 14) wies darauf hin, dass die Untersuchung geistiger Fähigkeiten und die Bestimmung des Intelligenzniveaus oft eine große Hilfe bei der Untersuchung von Gedächtnisstörungen, Sprachstörungen oder anderen Defiziten sein können. Weiters kann der WIE nach von Aster et al. (2006) auch durchaus für einen Einsatz zur Hochbegabungsdiagnostik geeignet sein, indem versucht wird, Probanden mit sehr stark ausgeprägten geistigen Fähigkeiten zu identifizieren.

Genauer gesagt handelt es sich beim WIE um die deutsche Version des WAIS-III. Sowohl der WIE als auch der WAIS-III weisen wesentliche strukturelle Veränderungen zu ihren Vorgängerversionen auf, da die Zahl der Interpretationsebenen von drei auf vier erhöht wurde. Bisher wurden die einzelnen Untertestergebnisse zu einem Verbal-IQ, einem Handlungs-IQ und dem Gesamt-IQ verrechnet. Mit den neueren Verfahren können nun zusätzlich vier verschiedene Index-Werte für das *Sprachliche Verständnis*, die *Wahrnehmungsorganisation*, das *Arbeitsgedächtnis* und die *Arbeitsgeschwindigkeit* bestimmt werden, was bedeutet, dass das Arbeitsgedächtnis u. a. als zentrale Leistungskomponente berücksichtigt wird. Werden der Verbal-IQ und der Handlungs-IQ berechnet, gehen alle Untertests zu *Arbeitsgedächtnis* in die Berechnung des Verbal-IQ ein, da diese ausschließlich aus verbalen Aufgaben bestehen.

Dieser Umstand sollte erkennbar machen, warum dieser Test hier vorgestellt werden soll.

Aus den insgesamt 14 Untertests des WIE sollen hier nur drei dargestellt werden, da nur diese zur Messung des Arbeitsgedächtnisses verwendet werden, nämlich (von Aster et al., 2006):

Rechnerisches Denken: Eine Serie von Rechenaufgaben (18 Aufgaben aus der WAIS-III und zwei neu entwickelte), die der Proband im Kopf lösen und mündlich beantworten muss (von Aster et al., 2006, S. 10).

Interessant ist, dass es für jede Aufgabe eine Zeitgrenze gibt, je nach Aufgabe zwischen 15 und 60 sec., wobei der Testperson bei Nennung der richtigen Lösung innerhalb von zehn Sekunden bei den letzten zwei Aufgaben Zeitpunkte gutgeschrieben werden.

Zahlennachsprechen: Serien von Ziffernfolgen unterschiedlicher Länge, die der Proband teils in derselben Reihenfolge und teils in der entgegengesetzten Reihenfolge, wie sie ihm vorgesprochen werden, nachsprechen muss (von Aster et al., 2006, S. 10).

Dieser Untertest setzt sich aus den gleichen Aufgaben wie sie im HAWIE-R zu finden sind zusammen. Es wurde jedoch für das *Zahlennachsprechen vorwärts* eine Zwei-Ziffern-Folge eingeführt, um im unteren Leistungsbereich besser differenzieren zu können (von Aster et al., 2006, S. 19).

Jede Aufgabe besteht aus zwei Durchgängen von gleicher Länge, wobei nach der Aufgabe die Serie um eine Ziffer verlängert wird. Begonnen wird sowohl beim *Zahlennachsprechen vorwärts* als auch beim *Zahlennachsprechen rückwärts* bei einer Länge von zwei Ziffern. Wird das bestehende Abbruchkriterium nicht erfüllt, reichen die Aufgaben für *vorwärts* bis zu einer Länge von neun Ziffern (es sind also insgesamt acht Aufgaben). Für *rückwärts* erreichen die Ziffernfolgen nur eine Länge von acht Ziffern (es gibt also sieben Aufgaben), da hier erfahrungsgemäß die Leistungen niedriger sind.

Buchstaben-Zahlen-Folgen: Folgen von Zahlen und Buchstaben unterschiedlicher Länge, die dem Probanden vorgelesen werden. Er muss sich diese merken und anschließend wiedergeben, indem er zunächst die Zahlen in aufsteigender Folge und anschließend die Buchstaben in alphabetischer Folge wiederholt (von Aster et al., 2006, S. 10).

Bei diesem Untertest handelt es sich um einen neuen Untertest, dessen Entwicklung teilweise auf den Vorarbeiten von Gold, Carpenter, Randolph und Weinberger (1997, zitiert nach von Aster et al., 2006, S. 19) basiert, die ähnliche Aufgaben zur Prüfung der Frage entwickelten, ob das Arbeitsgedächtnis bei Schizophrenen beeinträchtigt ist.

Jede der insgesamt sieben Aufgaben besteht aus drei Versuchen, wobei jeder Versuch aus einer anderen Kombination von Buchstaben und Zahlen besteht. Die erste Aufgabe besteht aus jeweils einer Ziffer und einem Buchstaben. Pro Aufgabe wird entweder eine Ziffer oder ein Buchstabe hinzugefügt, sodass die siebte und letzte Aufgabe aus acht zu merkenden Einheiten besteht.

All diese Aufgaben zur Untersuchung des Arbeitsgedächtnisses erfordern nach von Aster et al. (2006) die Beachtung (Aufmerksamkeit) für Informationen, das kurzzeitige Merken dieser Informationen und die Bearbeitung im Gedächtnis, um sie dann in eine Lösung umzuwandeln (S. 120).

Von Aster et al. (2006) erwähnen, dass die Drei-Komponenten-Theorie von Baddeley (Annahme der zentralen Exekutive, der phonologischen Schleife und des räumlich-visuellen Notizblocks) zwar populär geworden ist, aber andere Forscher die Unterscheidung zwischen einer phonologischen und einer visuellen Komponente nicht für bedeutsam halten. Aus diesem Grund wurde diese Unterscheidung im WIE ebenfalls nicht vorgenommen und somit wird das Arbeitsgedächtnis ausschließlich mit Aufgaben untersucht, die akustisch dargeboten werden.

Darüber hinaus weisen Carlson, Khoo, Yaure und Schneider (1990) darauf hin, dass es sich beim Arbeitsgedächtnis um einen isolierten Arbeitsbereich mit begrenzter

Kapazität handelt und dass dieses Modell gilt, unabhängig davon, ob man dieses in Teilsysteme gliedert oder nicht.

4.5.2.7 Hamburg-Wechsler-Intelligenztest für Kinder-IV (HAWIK-IV)

Die Definition von Wechsler (1956, siehe Kapitel 3.3) wurde Deimann & Kastner-Koller (2008) zu Folge lange Zeit als Begründung für die Annahme eines Verbal- und eines Handlungsteils in den Wechsler-Intelligenztest verstanden (vgl. von Aster et al., 2006). Im Hamburg-Wechsler-Intelligenztest für Kinder-IV (HAWIK-IV, Petermann & Petermann, 2007) wird diese Aufteilung des Verfahrens völlig aufgegeben und durch vier Indexwerte (vgl. 4.5.2.6) ersetzt, die sich gegenüber der Vorgängerversion, dem HAWIK-III in ihrer Bedeutung verändert haben. Neben dem sprachlichen Verständnis werden fluide kognitive Fähigkeiten, das Arbeitgedächtnis sowie die Verarbeitungsgeschwindigkeit erfasst (Deimann & Kastner-Koller, 2008, S. 161).

Dieses Testverfahren besteht insgesamt aus 15 Untertests, von denen 10 den so genannten Kernunters tests zuzuordnen sind und die restlichen fünf optional vorgegeben werden können und im Stande sind, falls nötig, Kernunters tests zu ersetzen. Die 10 Kernunters tests gehen in die Berechnung des Gesamt-IQ und der vier Indexwerte (*Sprachverständnis*, *Wahrnehmungsgebundenes Logisches Denken*, *Arbeitsgedächtnis* und *Verarbeitungsgeschwindigkeit*) ein.

Da wiederum nur die Messung des Indexwerts *Arbeitsgedächtnis* interessiert, wird auf eine Darstellung der anderen drei Indexwerte verzichtet.

Der Indexwert *Arbeitsgedächtnis* überprüft nach Deimann und Kastner-Koller (2006, S. 162) die Fähigkeit, Informationen im Gedächtnis zu behalten und zu manipulieren, wobei Aufmerksamkeit, Konzentration, mentale Kontrolle und Schlussfolgern die Leistung beeinflussen. Da die Vorgehensweise zur Messung des Arbeitsgedächtnisses derjenigen vom WIE sehr ähnlich ist, sollen an dieser Stelle nur noch wenige Unterschiede angeführt werden (vgl. 4.5.2.6).

Arbeitsgedächtnis setzt sich beim HAWIK-IV aus den zwei Kerntests *Zahlennachsprechen* und *Buchstaben-Zahlen-Folgen* zusammen, wobei vor einer Vorgabe des letztgenannten bei sehr jungen Kindern mit zwei Aufgaben überprüft wird, ob sie zählen und das Alphabet aufsagen können. Wenn dies nicht der Fall ist, wird auf dessen Vorgabe verzichtet und es kann der optionale Untertest *Rechnerisches Denken* vorgegeben werden. Bei diesem erfolgt, mit Ausnahme der ersten fünf Aufgaben, die Präsentation ausschließlich verbal, um die Anforderungen an das Arbeitsgedächtnis zu erhöhen (Deimann & Kastner-Koller, 2008). Es sei jedoch darauf hingewiesen, dass Deimann und Kastner-Koller (2008, S. 164) aufgrund eigener Testerfahrungen mit dem HAWIK-IV berichten, dass für den Indexwert *Arbeitsgedächtnis* höchst unterschiedliche Werte – von unterdurchschnittlich bis überdurchschnittlich – erzielt wurden, je nachdem, ob die beiden Kernunters tests *Zahlennachsprechen* und *Buchstaben-Zahlen-Folgen* miteinbezogen worden sind, oder ersatzweise das *Rechnerische Denken*.

4.5.2.8 Ein Beispiel für ein computergestütztes Verfahren

Im Folgenden wird ein von Singleton, Horne und Simmons (2009) entwickeltes adaptives Screeningverfahren zur Erfassung der Lese-Rechtschreibschwäche bei Erwachsenen vorgestellt. Es handelt sich um ein noch unveröffentlichtes, bisher namenloses Verfahren, das in Großbritannien entwickelt wurde und englischsprachig ist. Dennoch ist es aufgrund mehrerer Faktoren an dieser Stelle erwähnenswert.

Zum einen handelt es sich um ein Verfahren, das vollständig am PC vorgegeben und bearbeitet wird. Während dies nach Singleton et al. (2009) im englischsprachigen Raum keine Besonderheit darstellt, sind die meisten deutschsprachigen Testverfahren zur Diagnostik von Lese-Rechtschreibstörungen nach wie vor Papier-Bleistift Tests.

Neben anderen Vorteilen einer computergestützten Vorgabe (siehe Kapitel 5.2.2) ist bei Testverfahren zur Erfassung der Lese-Rechtschreibstörung vor allem die Testleiterunabhängigkeit hervorzuheben. So muss bei einigen Verfahren eine Textpassage vom Testleiter vorgelesen werden. Durch die computerisierte Vorgabe

kann ausgeschlossen werden, dass die Art der Vorgabe und die Geschwindigkeit die Testergebnisse beeinflussen (Singleton et al., 2009).

Ein weiterer wichtiger Vorteil ergibt sich nach Singleton et al. (2009) dadurch, dass durch eine computergestützte Vorgabe bei den Testpersonen allgemein und solchen mit Schwierigkeiten beim Lesen und Schreiben im Speziellen, Stress und Angst reduziert werden können.

Zum anderen erfolgt die Vorgabe bei diesem Verfahren adaptiv (siehe Kapitel 5.5.1.3.3).

Eine weitere Besonderheit besteht darin, dass es sich um ein Verfahren handelt, das speziell für Erwachsene entwickelt wurde, ganz im Gegensatz zu den bisher vorgestellten Verfahren zur Erfassung der Lese-Rechtschreibschwäche bei Kindern.

Das Screeningverfahren von Singleton et al. (2009) besteht aus drei Untertests:

Word Recognition: Von fünf am Bildschirm erscheinenden Wörtern muss die Testperson das reale Wort anklicken; die anderen stellen Pseudowörter dar.

Word Construction: In einer 3x3 Matrix werden der Testperson neun Silben angezeigt und über Lautsprecher ein Pseudowort dargeboten. Anschließend muss der Proband in der korrekten Reihenfolge jene Silben anklicken, die das vorgegebene Wort bilden.

Working Memory: Der Testperson werden über Lautsprecher Zahlenfolgen vorgegeben, die sie im Anschluss in umgekehrter Reihenfolge über die Tastatur wiedergegeben muss.

In dieser Arbeit wurde eine sehr beschränkte Auswahl an Verfahren getroffen. Eine sehr übersichtliche und ausführlichere Darstellung, zumindest der in der neuropsychologischen Diagnostik zur Anwendung kommenden Verfahren zur Gedächtnismessung, findet sich bei Lezak (1995).

4.5.3 Ausgewählte Links für eine Teilnahme an online-Untersuchungen

Abschließend soll noch dem Umstand Beachtung geschenkt werden, dass mittlerweile zahlreiche Experimente und Tests online angeboten werden.

Klarerweise ist es für die vorliegende Arbeit interessant, ob und wie bei online-Tests Messungen zum visuellen Arbeitsgedächtnis stattfinden. Daher werden hier einige, wenn auch nur wenige Beispiele vorgestellt. Das Wort „nur“ deutet bereits darauf hin, dass eine große Fülle an derartigen Tests und Experimenten im Internet angeboten werden (werden beispielsweise die Suchbegriffe „online psychological experiments memory“ in die Suchmaschine Google eingegeben, liefert diese ungefähr 370.000 Ergebnisse).

Darüber hinaus werden URLs (Uniform Resource Locator) angegeben, die einen guten Überblick und dazu auch gleich die Links zu sehr vielen aktuellen, psychologischen Experimenten liefern. Anzumerken ist jedoch, dass bei Angaben von URLs immer das Problem besteht, dass diese in Zukunft vielleicht nicht mehr aktuell sind und deswegen nicht mehr abgerufen werden können.

Zum Thema der Aktualität und Abrufbarkeit von URLs passend sei erwähnt, dass Francis, Neath und Surprenant (2000) „*The Cognitive Psychology Online Laboratory*“ vorstellen und zwei der dort verfügbaren Experimente näher beschreiben. Da eines dieser beiden sehr gut zum Thema des visuellen Kurzzeitgedächtnisses gepasst hätte, nämlich die online-Vorgabe des Brown-Peterson Memory Task (vgl. Schermer, 2006), ist äußerst schade, dass die von Francis et al. (2000) angeführte Adresse nicht mehr verfügbar ist.

<http://psych.hanover.edu/Research/exponnet.html>

Diese URL wurde bereits oben erwähnt. Hier sind sehr viele aktuelle, psychologische Experimente und deren Links nach Themengebieten geordnet (Cognition, Consumer Psychology, Cyber Psychology, Developmental, Emotions, Forensic Psychology, General, Health Psychology, Human Factors, Industrial/Organizational, Judgment and

Decision, Linguistics, Mental Health, Personality, Psychology and Religion, Relationships, Sensation and Perception, Sexuality, Social Psychology) zu finden.

<http://genpsylab-wexlist.unizh.ch>

Auch unter dieser Adresse findet sich eine sehr große Sammlung von zurzeit zur Verfügung stehenden Untersuchungen. Vorteilhaft ist, dass sowohl deutschsprachige als auch englischsprachige Untersuchungen angeführt sind. Zudem ist beim Großteil eine kurze Information angegeben ist (z.B. über Dauer, Teilnehmer, Javaunterstützung, Incentives).

<http://psychexps.olemiss.edu/Exps/labexperiments.htm>

Unter dieser Adresse der Universität von Mississippi können 26 verschiedene Experimente selbst durchgeführt werden. Zum Thema dieser Arbeit passt vor allem die Word Recognition Study.

http://memory.uva.nl/index_en

Diese Adresse der Universität von Amsterdam bietet die Möglichkeit, an sieben verschiedenen Gedächtnistests teilzunehmen.

www.essex.ac.uk/psychology/experiments/memtask.html

Hier ist u. a. ein leicht durchzuführendes Experiment zu den Positionseffekten (*Primacy-* und *Recency-Effekt*) beim Merken von Ziffernfolgen zu finden, wobei hier mit der Reproduktionsmethode gearbeitet wird.

<http://www.psych.uni.edu/psychexps>

Auch unter dieser Adresse kann an den verschiedensten Experimenten teilgenommen werden. Im Gegensatz zu den anderen Links können hier die Experimente auch nur angesehen werden, in dem auf *Demonstration* geklickt wird, wenn die Daten nicht für Forschungszwecke verwendet werden sollen.

<http://www.minddisorders.com/Pv-Z/Wechsler-adult-intelligence-scale.html>

Unter dieser Adresse ist ebenfalls eine ganze Reihe von Links für weitere Tests vorzufinden. Am interessantesten für diese Arbeit scheint der Word Memory Test zu sein, über den zwar unter www.wordmemorytest.com Informationen zur Verfügung gestellt werden, jedoch haben nur *qualified professionals* ein Recht, den Tests anzusehen. Interessant, wenn auch ein wenig am Thema vorbei, ist auch der angeführte Link (www.lumosity.com) zu wissenschaftlich entwickelten Spielen, mit denen das Gedächtnis verbessert werden kann. Dort findet man neben den Gedächtnistrainingsspielen auch viele weitere, beispielsweise zum Training der Geschwindigkeit, der Flexibilität und zur Steigerung der Problemlösefähigkeit.

Zum Abschluss dieses Kapitel und zur Überleitung zum nächsten Kapitel soll erwähnt werden, dass Baller et al. (2006) darauf hinweisen, dass speziell in der Gedächtnisdiagnostik zur Vermeidung von Lerneffekten entweder eine Parallelversion herangezogen oder ein größerer Abstand zwischen den Testungen eingehalten werden sollte. Ein Umstand, der abermals für eine (Gedächtnis-) Diagnostik spricht, die auf automatischer Itemgenerierung (siehe nächstes Kapitel) beruht.

5 Automatische Itemgenerierung

5.1 Theoretischer Hintergrund der automatischen Itemgenerierung

Nach Arendasy, Sommer und Hergovich (2007) sind im Rahmen der Angewandten Psychometrie im letzten Jahrzehnt bemerkenswerte Fortschritte in Bezug auf Testvorgabemodi, Testsicherheit und Itemkonstruktionstechnologien erzielt worden. Diese nennen u. a. die Einführung von fortgeschrittenen Methoden zur automatischen Itemgenerierung (AIG; für einen Überblick, Irvine & Kyllonen, 2002) als Beispiele solcher Entwicklungen.

Die AIG kann anhand ihres Generativitätsgrades, also der Kombination aus dem Ausmaß der theoretischen Fundierung ihrer Konstruktionsalgorithmen mit dem Automatisierungsgrad des Itemgenerierungsprozesses, in drei Ausprägungsgraden unterschieden werden (Arendasy et al., 2007, S. 119):

- (1) *Funktionale* Itemgenerierung
- (2) *Modell-basierte* Itemgenerierung
- (3) *Automatisierte* Itemgenerierung

Die *funktionale* Itemgenerierung ist dadurch gekennzeichnet, dass die Testaufgaben ohne genaue kognitionspsychologische Spezifikation des zu messenden Konstrukts automationsgestützt erstellt werden (Arendasy et al., 2007, S. 119). Die *modell-basierte* Itemgenerierung baut demgegenüber auf einer kognitionspsychologischen Analyse des zu messenden Bereichs auf, die es erlaubt, die Schwierigkeit von Items statistisch hinreichend genau zu erklären und vorherzusagen (Arendasy et al., 2007, S. 119). Auf der letzten Stufe, der *automatisierten* Itemgenerierung kann jegliche psychometrische Eigenschaft eines Items aufgrund eines syntaktischen, generativen Regelwerks mit höchster Präzision erschlossen werden (Arendasy et al., 2007, S. 119).

Im Rahmen der *modell-basierten* und *automatisierten* Itemgenerierung existieren zwei Beschreibungsparameter für automatisiert erstellte Testaufgaben, nämlich *radicals* (kontrollierende Elemente, Arendasy, Sommer, Gittler & Hergovich, 2006) und *incidentals* (nicht-kontrollierende Elemente, Arendasy, Sommer, Gittler & Hergovich, 2006). Unter *radicals* werden jene Aufgabenmerkmale verstanden, welche als „Haupteffekte“ einen statistisch signifikanten Beitrag zur Erklärung und Variation psychometrischer Itemeigenschaften leisten (Irvine, 2002 zitiert nach Arendasy, Sommer & Hergovich, 2007). *Incidentals* charakterisieren dagegen jene Aufgabeneigenschaften, die keinen diesbezüglichen Erklärungsbeitrag leisten können und daher als „Oberflächenmerkmale“ zur Itemgenerierung eingesetzt werden (Arendasy et al., 2007).

Bejar (2002, zitiert nach Arendasy et al., 2007) unterscheidet zwei prinzipielle Zugänge zur automatisierten Itemgenerierung: die Produktion von *Itemvarianten* und die Produktion von *Isomorphen* (vgl. Arendasy, 2003). *Itemvarianten* sind durch eine freie Kombination unterschiedlicher Itemmerkmale gekennzeichnet, weswegen es bei der Itemgenerierung zu einer Itemvermehrung (Erhöhung der Itemvielfalt) kommt (Arendasy, Sommer und Hergovich 2007, S.120). Die Produktion von *Isomorphen* ist durch die Fähigkeit eines Generatorsystems charakterisiert, auf Basis bestehender, bereits kalibrierter Itemschablonen neue Items zu generieren, welche hinsichtlich ihrer *radicals* jeweils identisch mit ihren „Mutteritems“ sind. In diesem Fall kommt es also psychometrisch gesehen nur zu einer Itemvervielfachung, auf Basis eines bereits bestehenden, kalibrierten Itempools (Arendasy et al., 2007, S.120, vgl. Arendasy, Sommer, Gittler & Hergovich, 2006).

Arendasy, Sommer und Hergovich (2007) weisen darauf hin, dass diese beiden Zugänge technologisch fundamentale Unterschiede aufweisen, indem sie schreiben:

Während Itemvarianten-Generatoren zumindest die modellbasierte Generierung von Testaufgaben voraussetzen und bei der Itemkonstruktion von der *Mikroebene*, d.h. von Itemmerkmalen, die in unterschiedlichen Ausprägungsgraden rekombiniert werden, ausgehen [...], können Isomorphengeneratoren bereits auf Stufe der funktionalen Itemgenerierung erstellt bzw. eingesetzt werden; der Generierungsprozess geht in diesem Fall von einer

Makroebene aus, d.h. er baut auf – bereits kalibrierten – Itemschablonen auf und verändert lediglich Oberflächenmerkmale in den Itemschablonen.

5.2 Entwicklung von Itemgeneratoren

Arendasy (2003) nennt zwei Gesichtspunkte, die bei einer Entwicklung eines Itemgenerators grundsätzlich zu beachten sind: die theoretische Fundierung des Generators und das psychometrische Modell, das mit dem Probandenverhalten modelliert werden soll.

Um auf den erstgenannten Punkt näher eingehen zu können, müssen die Betrachtungen über die Determinanten jeglicher Aktivität von Greeno, Moore & Smith (1993, zitiert nach Arendasy, 2003) kurz vorgestellt werden:

Aktivität ist ihrer Meinung nach allgemein anhand einer Person und einer Situation charakterisierbar, d.h. sie ist gekennzeichnet durch eine Interaktion von Eigenschaften der handelnden Personen und den Eigenschaften einer Situation. Den *Aufforderungscharakter*, den eine Situation in Bezug auf die in ihr vorkommenden Objekte und Materialien aufweist, bezeichnen Greeno et al. als „*affordances*“. Beispielsweise könnte eine in einem Raum stehende Person und ein Stuhl durch die affordance „Stuhl unterstützt Person“ verbunden sein. In diese affordance würden unterschiedliche, mögliche Handlungen wie z.B. auf dem Stuhl zu sitzen, stehen oder zu knien, fallen. Eine Situation weist aber in Abhängigkeit der in ihr vorkommenden Objekte und Personen auch eine Reihe von objekt- oder personenbezogenen Handlungseinschränkung auf, die Greeno et al. als „*constraints*“ bezeichnen und welche die Situation zusammen mit deren affordances genau definieren.“ (S.50)

Dies überträgt Arendasy (2003, vgl. Arendasy, Sommer & Hergovich, 2007) auf die Frage einer allgemeinen, theoretischen Fundierung von Itemgeneratoren und sagt, dass Itemgeneratoren von ihrer Implementierung her einer wohl definierten „*affordance-constraint*“-Struktur folgen sollten, indem die „*constraints*“ im Rahmen der AIG den zu vermeidenden Teil des Itemgenerierungsprozesses darstellen und die *affordance*-Struktur eines Items das definiert, was in der Situation „Item“ möglich sein soll (S.51). Es sollte bei genauer Definition der „*affordance-constraint*“-Struktur eines Aufgabentyps und deren Implementierung, so Arendasy (2003) weiter, möglich sein, interpretationseindeutige Testaufgaben vollautomatisch zu generieren.

Demzufolge nennt Arendasy (2003) vier Schritte für die Konzeption und Implementierung eines Itemgenerators, die nun verkürzt und an diesen stark angelehnt vorgestellt werden sollen.

Als erster Schritt sollte die vorhandene kognitionspsychologische Fachliteratur im Hinblick auf das zu erfassende Konstrukt genau untersucht und evaluiert werden, um relevante Eigenschaften eines Itemgenerators im Sinne von „*radicals*“ und „*incidentals*“ ableiten zu können.

Für den zweiten Schritt – die Untersuchung der „*constraints*“-Seite eines Itemgenerators im Sinne einer eingebetteten, psychometrischen Qualitätskontrolle – kann die meist umfangreiche, angewandt-psychometrische Literatur, in der das zu messende Konstrukt hinsichtlich des Zusammenhangs zwischen psychometrischer Qualität und bestimmter Itemeigenschaften untersucht wurde, herangezogen werden. Das Ergebnis dieses zweiten Schritts sind „*constraints*“ des Itemgenerators in Form eines automatisiert ablaufenden psychometrischen Expertensystems, das in den Generierungsprozess als prüfendes Regelwerk eingreift.

Der dritte Schritt stellt die Validierung des Itemgenerators im Hinblick auf die Erfüllung der „*affordance-constraint*“-Struktur dar.

Als vierten und letzten Schritt kann der Versuch einer nomothetischen Validierung (vgl. Embretson, 1983, 1998, zitiert nach Arendasy, 2003) der Produkte des Itemgenerators im Sinne des Zusammenhangs zu externen, konstruktnahen und konstruktfernen Kriterien unternommen werden.

5.3 Vorteile der automatischen Itemgenerierung

Holling, Blank, Kuchenbäcker & Kuhn (2008) zu Folge hat eine regelgeleitete Itemkonstruktion – wozu auch die AIG zu zählen ist – zwei Hauptvorteile. Der erste liegt in der Tatsache, dass explizit kognitive Prozesse für die Lösung der Items verantwortlich gemacht werden, sodass die Konstruktvalidität der Items geprüft und Items konstruiert werden können, die spezifische „Unterfähigkeiten“ (*sub-abilities*)

messen. Der zweite Vorteil besteht in der Möglichkeit, große Itempools ohne großen Aufwand konstruieren zu können.

Darüber hinaus nennen Arendasy, Sommer und Hergovich (2007, S. 129, vgl. Arendasy, Sommer, Gittler & Hergovich, 2006) vier Punkte, aufgrund welcher aus praktischer Sicht die AIG vorteilhaft sein könnte:

- (1) Präzise definierte generative und prüfende Regelwerke – vorausgesetzt der zugehörige Itemgenerator wurde fehlerfrei implementiert – können helfen, Fehlinterpretationen menschlicher Itemkonstruktoren zu vermeiden. Dies ist vor allem dann von Vorteil, wenn Testaufgaben konstruiert werden sollen, die die Fähigkeiten der Itemkonstruktoren übersteigen.
- (2) Da für jedes generierte Item genau operationalisiert und empirisch belegt werden kann, welche Intelligenzdimension bzw. welche ihrer Teilkomponenten gemessen wird, wäre eine inhaltlich validere Interpretierbarkeit von Testergebnissen gegeben.
- (3) Es könnte eine erhöhte Testsicherheit erzielt werden, da relativ rasch neue Itempools generiert werden könnten und so die Item-Exposure-Rate von Testaufgaben relativ niedrig ausfallen würde.
- (4) Da die Kosten des Itemkonstruktionsprozesses mittelfristig abgesenkt werden könnten, könnte der AIG eine hohe wirtschaftliche Bedeutung zukommen.

Nach Arendasy (2003) ist der Wert der AIG „insbesondere dann gegeben, wenn auch der höhere Anteil der Itementwicklungskosten, jener der Kalibrierung, nicht klassisch, also über zeit- und kostenintensive Untersuchungen anhand geeignet umfangreicher Stichproben gegeben wäre, sondern prädiktiv auf dem Boden modellbasierter, z.B. kognitionspsychologischer Zugänge erfolgen könnte“ (S.49). Arendasy (2003, S. 49) meint weiters, dass der Hauptschritt bei der Entwicklung eines psychologischen Tests dann nicht mehr in der Herstellung und aufwendigen Kalibrierung eines Itempools

liegen würde, sondern in der Grundlagenforschung über Itemmodelle zur Entwicklung von Problemschemata oder Isomorphen bekannter Schwierigkeit bzw. der theoriegeleiteten Untersuchung von Konstruktionsrationalen. Embretson & Reise (2000, zitiert nach Arendasy, 2003, S. 49) zu Folge liegt der Hauptvorteil eines solchen Vorgehens darin, „dass zusätzlich zum normorientiertem auch kriterienorientiertes Testen (vgl. Kubinger, 2006) angeboten werden könnte, da aufgrund der hinreichend spezifizierten, theoretischen Basis von Itemgeneratoren über jedes Item im wesentlichen ausgesagt werden könnte, *was* es tatsächlich misst“.

Zusammenfassend meint Arendasy (2003, S.49), dass der Wert moderner AIG, neben der Möglichkeit inhaltlich definiert zu messen, darin läge, die Effizienz von Itemkonstruktion sowohl hinsichtlich des Herstellungsprozesses, als auch hinsichtlich des psychometrisch bedingten Kalibrierungsaufwandes, zu erhöhen.

5.4 Bereits durchgeführte Studien mittels AIG

Da davon ausgegangen wird, dass der Bereich der automatischen Itemgenerierung ein Novum darstellen kann und daher nicht ausgeschlossen werden kann, dass die aktuelle Verbreitung des Einsatzes von automatisch generierten Items bekannt ist, sollen an dieser Stelle einige Arbeiten angeführt werden, die bereits in diesem Bereich durchgeführt wurden.

Fairon und Williamson (2002, zitiert nach Arendasy, 2003) stellen ein vollautomatisches System zur Herstellung von Textaufgaben (*Automatic Text Generation System*, ATG) vor, welches analytisches Schlussfolgern im sprachlichen Bereich messen soll.

Eines der nach Arendasy, Sommer, Gittler und Hergovich (2006) bekanntesten Beispiele für modell-basierte Itemgenerierung stellt die Arbeit von Hornke (2002) dar, welche sich mit der modell-basierten Konstruktion von Kurzzeitgedächtnis- und figuralen Matrizenitems beschäftigt.

Zudem werden von Arendasy (2003) Matrizen tests eingesetzt, die meist den Generalfaktor der Intelligenz nach Spearman (vgl. Kapitel 4.5.1.2) messen sollen. Die mittels AIG konstruierten Items werden erprobt und auch erste Studien zur Qualitätssicherung des Itemgenerators GeomGen werden durchgeführt.

Auch Arendasy und Sommer (2005) und Arendasy (2005) beschäftigen sich mit automatisch generierten Items bei solchen Aufgaben (figurale Matrizen).

Bei Arendasy, Sommer und Hergovich (2007) sind drei Studien angeführt, bei denen der von Arendasy entwickelte Itemgenerator NGen für den Fähigkeitsbereich „Numerische Flexibilität“ der Intelligenzdimension „Quantitatives Denken“ eingesetzt worden ist.

Sowohl Holling et al. (2008) als auch Arendasy, Sommer, Gittler und Hergovich (2006) widmen sich dem Thema quantitatives Schlussfolgern bei Textaufgaben (algebra word problems), wobei erstere „nur“ regelgeleitet vorgehen und letztere einen Schritt weiter in Richtung Vollautomatisierung gehen und den Itemgenerator AGen von Arendasy und ein Qualitätskontrollsystem einsetzen.

Auch zur Überprüfung der Intelligenz zugrunde liegender Theorien wurden automatisch generierte Items bereits eingesetzt (Arendasy, Hergovich & Sommer, 2008).

Zu diesem Zeitpunkt dürften keine Studien bekannt sein, die verbales Material zur Messung des Kurzzeitgedächtnisses mittels Verwendung automatisch generierter Items messen.

Da die Automatische Itemgenerierung (AIG) gezwungenermaßen durch eine Testung am Computer erfolgt, wird in Folge kurz eine Definition der Computerdiagnostik gegeben und auf die Vor- und Nachteile eingegangen, die eine Diagnostik mittels Computer nach sich zieht. Darüber hinaus meint Hambleton (2004), dass die größte Veränderung in der psychologischen Diagnostik in den nächsten 20 Jahren darin besteht, dass mehr Verfahren mittels Computer vorgegeben werden.

5.5 Computerdiagnostik

„Computerdiagnostik wird als eine strategische Variante innerhalb der Diagnostik verstanden, um psychologisch relevante Variablen zu erfassen, deren Auswahl zu steuern, die erhaltenen Informationen zu einem Urteil zu verdichten und gegebenenfalls schriftlich und/oder bildlich darzustellen.“ (Booth, 1992, S.186)

5.5.1 Klassifikation von computergestützten Testverfahren

Laut Schuhfried (2006) lassen sich Computertestverfahren grob in drei Gruppen unterteilen:

- Vorgabe von Paper-Pencil-Verfahren
- Simulation apparativer Verfahren
- Genuine Computerverfahren

5.5.1.1 Vorgabe von Paper-Pencil-Verfahren

Schuhfried (2006) ist der Meinung, dass die Umsetzung von Paper-Pencil-Verfahren in die computergestützte Form wissenschaftlich uninteressant ist und technisch keine Herausforderung darstellt. Ihm zu Folge liegt der Vorteil in der sicheren und raschen Auswertung der Tests. Es ist jedoch darauf zu achten, dass die *Normierung* („Ausmaß und Qualität der populationsspezifischen Bezugsgrößen zur Interpretation von Personenparametern, insbesondere zur Bestimmung der relativen Position einer Testperson in Bezug auf (u. a. verschiedene) Populationsverteilungen von Testwerten“, Amelang & Schmidt-Atzert, 2006, S. 137) und *Validierung* (s. u.) nicht ungeprüft auf die jeweilige Computerversion eines Tests übertragen werden dürfen. Weiters führt Schuhfried (2006) an, dass auch zu überprüfen ist, ob die *Testfairness* (s. u.) in der computerisierten Version erhalten bleibt.

Jedoch ist spätestens seit der Arbeit von Klinck (2002), in der diese zeigt, dass bei sachgemäßer Programmierung der Computerversion keine Unterschiede in der

Schwierigkeit zum jeweiligen Paper-Pencil-Verfahren feststellbar sind, dieses Äquivalenzproblem ausdiskutiert (Schuhfried, 2006, vgl. Kapitel 5.5.3.1).

5.5.1.2 Simulation apparativer Verfahren

Die Simulation von apparativen Verfahren stellt für Schuhfried (2006) eine viel interessantere Verwendung des Computers dar als die oben dargestellte Vorgabe von Paper-Pencil-Verfahren, da mit einem Computer durch entsprechende Programme und eventuell notwendige Eingabegeräte psychologische Testgeräte realisiert werden können. Als Vorteil führt Schuhfried (2006) die höhere Praktikabilität und die niedrigeren Kosten an, meint jedoch – was einen Nachteil darstellen kann –, dass Simulationen apparativer Verfahren technisch nicht trivial sind.

5.5.1.3 Genuine Computerverfahren

Unter genuinen Computerverfahren sind Verfahrensweisen gemeint, zu deren Durchführung der Computer zwingend notwendig ist (Schuhfried, 2006, S. 132).

Dazu zählen nach Schuhfried (2006) folgende Anwendungen:

- Computergestützte Instruktion
- Erfassung von Latenzzeiten
- Adaptive Testvorgabe
- Multimediale Vorgabe
- Zusammenfassung einzelner Testergebnisse einer Testbatterie zu einer Eignungsbeurteilung
- Internetttests

Einige für diese Arbeit relevante Punkte sollen nun einer näheren Betrachtung unterzogen werden.

5.5.1.3.1 Computergestützte Instruktion

Laut Schuhfried (2006) wird bei einem guten computergestützten Test auch die Instruktion über den Bildschirm vorgegeben, manchmal erfolgt dies auch akustisch. Diese Instruktionen folgen meist dem Prinzip des programmierten Lernens; einzelne Lernschritte werden vorgegeben und es wird abgeprüft, ob die Testpersonen diese verstanden haben. Ist dies nicht der Fall, muss eine neuerliche Lernschleife durchlaufen werden, und zwar so lange, bis sichergestellt ist, dass die Aufgabenstellung verstanden worden ist. Wird das nötige Instruktionsverständnis nicht erreicht, werden also einzelne Lernschleifen zu oft durchlaufen, wird die Testung automatisch abgebrochen (Schuhfried, 2006).

Eine solche Vorgehensweise impliziert nach Schuhfried (2006) eine höhere *Objektivität* (s. u.) als bei herkömmlicher Testvorgabe, da die Aufgabenstellung unabhängig vom Testleiter ist (Durchführungsobjektivität, vgl. Kapitel 5.5.2). Als weiteren Vorteil sieht Schuhfried (2006) an, dass eine höhere *Fairness* gewährleistet ist, da nur jene Personen in die Testphase gelangen, die die Aufgabenstellung (ausreichend) verstanden haben.

5.5.1.3.2 Erfassung von Latenzzeiten

Mit Hilfe eines Computers kann für jedes einzelne Item die Latenzzeit (Zeit vom Auftauchen eines Reizes bis zum Beginn der darauf folgenden Reaktion; Reaktionszeit) auf Millisekunden genau gemessen werden. Dies kann nach Schuhfried (2006) bei den verschiedensten Fragestellungen von Bedeutung sein (z.B.: Rückschluss auf hirnorganische Störungen, Psychosen, implizite Einstellungen).

5.5.1.3.3 Adaptive Testvorgabe

Bei einer linearen Testvorgabe (konventionelle Vorgabe) entspricht beim Großteil der Items die Schwierigkeit nicht der Fähigkeit des Probanden (Schuhfried, 2006). Nach

Amelang & Schmidt-Atzert (2006) tragen jedoch nur solche Items, deren Schwierigkeit mit der Fähigkeit des Probanden hinreichend übereinstimmen, wesentlich zur Testgesamtinformation bei, die anderen nicht, sodass diese zu leichten bzw. zu schwierigen Items bei der Testvorgabe einfach weggelassen werden könnten, ohne die Testgenauigkeit beträchtlich zu verringern. Und genau dieser Idee bedient sich das adaptive Testen. Ein Algorithmus schätzt den Fähigkeitsparameter der Testperson und wählt danach das geeignete Item (maximaler Informationsgewinn bei Items, bei dem die Lösungswahrscheinlichkeit für den jeweiligen Probanden 50% beträgt) aus. Mit jeder Antwort des Probanden steigt die Genauigkeit der Schätzung der Leistungsfähigkeit (Schuhfried, 2006).

Dem adaptiven Testen kann nach Schuhfried (2006) ein hohes Maß an *Ökonomie* (s. u.) zugeschrieben werden, da dieses mit Vorgabe von sehr wenigen Items – je nach Abbruchkriterium (gewünschte Genauigkeit der Fähigkeitsschätzung) – auskommt.

Im Gegensatz dazu haben nach Häusler (2006) frühere Studien gezeigt, dass sich die Probanden bei Vorgabe von Items mit einer konstanter Lösungswahrscheinlichkeit von 50% überfordert fühlen und deswegen zu keiner ausreichend hohen Antwortsicherheit in einem angemessenen Zeitrahmen kommen. Er meint weiters, dass bei einer adaptiven Vorgehensweise die Items in Form eines reinen Power-Tests (also ohne Zeitbeschränkung) vorgegeben werden müssen. Daraus folgend ist allein der Proband verantwortlich dafür, wie viel Zeit er in die Bearbeitung eines Items investiert. Häusler (2006) ist der Meinung, dass als Folge die Bearbeitungszeit pro Item steigen wird. Dies kann den Zeitvorteil, der erzielt wurde, ausgleichen, indem nur wenige informative Items vorgegeben werden. Das Resultat einer adaptiven Testvorgabe kann somit auch einen Verlust der Ökonomie mit sich bringen.

Wenngleich es den Rahmen dieser Arbeit sprengen würde, genauer auf die Item Response Theorie (IRT) bzw. das Rasch-Modell einzugehen, muss an dieser Stelle darauf hingewiesen werden, dass solche adaptiven Strategien einen IRT-homogenen Itempool erfordern; es sei u. a. auf Rost (2004) verwiesen.

Ist dieser Itempool gut und groß genug, so kann nach Gittler and Arendasy (2003) auch in Extrembereichen präziser gemessen werden. Als weitere positive Konsequenzen des computerisierten adaptiven Testens führen diese an, dass der Testabbruch auf die jeweilige diagnostische Anforderung eingestellt werden kann. So kann beispielsweise nur ein grobes Screening, aber auch eine präzise Messung durchgeführt werden. Weiters sei die Testsicherheit und die Testfairness erhöht, weil „Testknackermethoden“, wie z. B.: das Auswendiglernen eines Antwortvektors, weitgehend unwirksam sind, da jede Testperson andere Items vorgelegt bekommt, weswegen klarerweise auch das Abschauen bzw. Schummeln bei Gruppentestungen unterbunden wird. Als letzten Vorteil nennen Gittler und Arendasy (2003) die Möglichkeit einer Bekanntgabe des *Itemtypes* der im Pool enthaltenen Aufgaben zur Testvorbereitung, um die Testangst zu reduzieren und denjenigen, die den Test ernst nehmen, ein Training zu ermöglichen.

Durch eine adaptive Vorgabe ergeben sich jedoch nicht nur Vorteile. Häusler (2006) zu Folge beschreiben Testpersonen den adaptiven Testvorgang oft als überfordernd, zu anstrengend und übermäßig demotivierend, was klare Nachteile einer solchen Vorgabeweise kennzeichnet. Denn dies steht im klaren Kontrast zur eigentlichen Intention, nämlich eine herausfordernde und optimal motivierende Testsituation zu schaffen, in der sich die Probanden weder unter- noch überfordert fühlen (Häusler, 2006).

Doch auch wenn Schuhfried (2006) ebenfalls als Nachteile des adaptiven Testens anführt, dass dieser Vorgabemodus von den Testpersonen meist als sehr schwierig empfunden wird und die Komplexität der Erstellung eines adaptiven Tests eine Herausforderung für den Testkonstrukteur darstellt, meint er, dass grundsätzlich das adaptive Testen als effizienteste Art psychologischen Testens betrachtet werden kann (vgl. Hornke, 1993).

Darüber hinaus nennt Häusler (2006) neue Wege, in deren Richtung sich das adaptive Testen in Zukunft bewegen könnte und im Zuge derer die oben angeführten Nachteile überwunden – wenn nicht sogar genutzt – werden könnten. Als Beispiel sei hier angeführt, dass diese Methode verwendet werden kann, um standardisierte über- und

unterfordernde Situation zu schaffen, in denen das Verhalten der Probanden im Sinne eines objektiven Persönlichkeitstests betrachtet wird.

5.5.1.3.4 Multimediale Vorgabe

Dank der Leistungsfähigkeit heutiger Computer besteht auch die Möglichkeit, mehr oder weniger kurze Videosequenzen (wenn gewünscht mit akustischer Begleitung) bei einer Testung vorzugeben. Es wird die Möglichkeit geschaffen, Items zu konstruieren, die für die Probanden ein hohes Maß an Realitätsnähe besitzen und einen modernen und anschaulichen Eindruck erwecken (Schuhfried, 2006).

5.5.2 Vorteile vom computergestützten Testen

Computergestützte Testungen werden laut Hornke (1993) von den Testpersonen insgesamt gut aufgenommen, da sie eine besondere *Fairness* und *Objektivität* suggerieren. *Fairness* und *Objektivität* sind zwei der 1986 vom Testkuratorium der Föderation deutscher Psychologenverbände festgelegten Gütekriterien.

Kubinger (2006) definiert *Fairness* (sinngemäß nach dem Testkuratorium der Föderation deutscher Psychologenverbände 1986) wie folgt: „Ein Test erfüllt das Gütekriterium *Fairness*, wenn die resultierenden Testwerte zu keiner systematischen Diskriminierung bestimmter Testpersonen zum Beispiel aufgrund ihrer ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppenzugehörigkeit führen“ (S. 118).

Darüber hinaus zählt Lienert (vgl. Lienert & Raatz, 1998) die Objektivität eines Verfahrens neben der Reliabilität und der Validität zu den Hauptgütekriterien. Kubinger (2006) gibt folgende sehr frei nach Lienert formulierte Definitionen an:

Objektivität: „Unter *Objektivität* eines Tests ist zu verstehen, dass die mit ihm gewonnenen Ergebnisse unabhgänglich vom Untersucher sind.“ (S. 34)

Man unterscheidet:

- *Durchführungsobjektivität* („Das Ausmaß, in dem die Unabhängigkeit des Tests von der Person des Untersuchungsleiters durch die Vorschriften der Testinstruktion und aller übrigen Durchführungsbedingungen gesichert ist.“, Amelang & Schmidt-Atzert, 2006, S. 136)
- *Auswertungsobjektivität* („Das Ausmaß, in dem die Auswertung des Tests unabhängig von personenbedingten oder apparativen Störquellen ist.“, Amelang & Schmidt-Atzert, 2006, S. 137)
- *Interpretationsobjektivität* („Das Ausmaß, in dem aus gleichen Scores verschiedener Probanden identische Schlüsse gezogen werden (von einem oder von mehrere Auswerten).“, Amelang & Schmidt-Atzert, 2006, S. 141)

Reliabilität: „Die *Reliabilität* eines Tests beschreibt den Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeitsmerkmal misst, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht.“ (S. 45)

Validität: „Unter *Validität* eines Tests ist zu verstehen, dass er tatsächlich jenes Persönlichkeitsmerkmal misst, welches er zu messen behauptet.“ (S. 50)

Booth (1992) ist der Meinung, dass eine realistische Stellungnahme zur Computertechnologie drei positive Aspekte zu berücksichtigen hat:

- die sich immer wiederholenden, zeitaufwendigen Arbeiten (z.B. bei der Testauswertung) können überflüssig gemacht werden,
- eine systematische Analyse von vielschichtig miteinander verbundenen Variablen wird ermöglicht und
- die Reliabilität des diagnostischen Urteils wird erhöht (S. 187).

Ein weiterer Vorteil von computergestütztem Testen gegenüber herkömmlichen Verfahren liegt laut Vogt (1999) darin, dass aufgrund der technischen Präsentationsmöglichkeiten (Randomisierung der Antwortalternativen) etwaigen Antwortlistenreihenfolgeneffekten (vgl. Primacy- und Recency-Effekte) vorgebeugt werden kann. Vor allem bei Fragen mit visualisierter Antwortliste treten *Primacy-Effekte* auf, denen durch die Randomisierung entgegengewirkt wird. Durch die randomisierte Darbietung der Items ergibt sich ein weiterer Vorteil der Computertests. Denn es ergibt sich damit die Möglichkeit, mehrere gleichwertige Testformen eines Tests zu entwickeln, wodurch bei einer mehrmaligen Vorgabe ein Übungseffekt reduziert werden kann (Steger-Wuchse, 1999).

Positiv anzumerken ist außerdem, dass nach Kasten (1994) Prüfungsängste, wie sie oft in Testsituationen zu beobachten sind, bei Computertests in geringerem Maße auftreten.

Als weiteren Vorteil nennt Appl (1999), dass durch den Einsatz von Computern Verfahren möglich werden, die sonst nicht zu realisieren wären (z.B.: Computersimulationen, vgl. Funke, 1998). Hier ist auch zu betonen, dass dadurch auch die Möglichkeit des adaptiven Testens (vgl. Kapitel 5.5.1.3.3) geschaffen wird.

Hinzu kommt, dass nach Appl (1999) die durch den Computer aufgezeichneten Reaktionen unmittelbar und in maschinenlesbarer Form zur Weiterverarbeitung bereit liegen, die Auswertung also mittels Computer geschieht und im Falle von wissenschaftlichen Studien die Daten im Idealfall direkt in ein Statistikprogramm eingelesen werden können.

Darüber hinaus gelten Computerverfahren als sehr ökonomisch (Ökonomie: „Das Ausmaß, in dem ein Test bei der Durchführung, Auswertung und Anschaffung als kostengünstig zu bewerten ist.“, Amelang & Schmidt-Atzert, 2006, S. 137).

Trotz dieser Vorteile existieren jedoch auch Nachteile aufgrund eines Computereinsatzes in der Diagnostik.

5.5.3 Nachteile vom computergestützten Testen

Laut Kubinger (2006) findet man auch heutzutage noch Skepsis (also mangelnde *Akzeptanz*: „Ausmaß, in dem subjektive Meinungen, Bewertungen oder gesellschaftspolitische Überzeugungen gegen einen Test angeführt werden.“, Amelang & Schmidt-Atzert, 2006, S. 137) gegenüber der Computerdiagnostik, vor allem was deren Einsatz bei älteren Personen betrifft. Es ist davon auszugehen, dass es auch jetzt noch Personen gibt, die weniger Erfahrung mit dem Medium Computer aufweisen und eventuell eine negative Einstellung zu Computern haben. Aus diesem Grund besteht die Möglichkeit, dass das Gütekriterium der *Fairness* verletzt wird. Jedoch fanden Ivnik & Malek (1995) heraus, dass nur wenige ältere Personen gegenüber den computergestützten kognitiven Verfahren eine negative Einstellung besitzen.

Weiters berührt dieser Sachverhalt auch ein weiteres Gütekriterium, nämlich die *Zumutbarkeit* („Das Ausmaß, in dem ein Test (absolut oder relativ zu dem aus der Anwendung des Verfahrens resultierenden Nutzen) die getestete Person in zeitlicher, psychischer (insbesondere „energetisch“-motivational und emotional) sowie körperlicher Hinsicht beansprucht.“, Amelang & Schmidt-Atzert, 2006, S. 136). Allerdings sollte im Einzelfall entschieden werden, ob einer Testperson eine Computertestung zuzumuten ist. Aus diesem Grund wird im nächsten Kapitel kurz auf den Sachverhalt der Computererfahrung und der Einstellung zum Medium Computer eingegangen.

Weiters stellen nach Booth (1992) auch ethische Fragen ein Problem dar. Der Computer besitzt nämlich die außergewöhnliche Fähigkeit, während des Testens ergänzende Daten zu registrieren. Nach Stout (zitiert nach Booth, 1992) gehören diese Daten zwei Gruppen an: Zu den *passiven* Daten zählen beispielsweise die Reaktionszeit der Testperson pro Testfrage. *Aktive* Daten schließen zum Beispiel die vom Computer durchgeführten Überprüfungen von Widersprüchen bei Antworten ein.

Es kann darüber diskutiert werden, ob auf solche Weise erhobene Daten eine Verletzung der Persönlichkeitssphäre des Menschen darstellen. Nicht nur die

Persönlichkeitssphäre muss geschützt werden, sondern auch ein Schutz der Person muss sichergestellt sein. Der Diagnostiker ist aus diesem Grund dazu angehalten, alle notwendigen Schritte zu unternehmen, um Datenschutz zu gewährleisten.

5.5.3.1 Computervorerfahrung und Einstellung zum Computer

Die Befundlage bezüglich eines mediumsspezifischen Effekts von Computervorerfahrung und Einstellung zu Computern auf Testwerte in Computerversionen von Leistungstests ist laut Klinck (2002) uneinheitlich. Zudem meint Klinck (2002), dass vermutet werden kann, dass Personen mit höherer intellektueller Leistungsfähigkeit mehr Computererfahrung aufweisen und eine positivere Einstellung zu Computern zeigen als Personen mit geringerer intellektueller Leistungsfähigkeit. Dies erklärt das beobachtbare Phänomen, dass zwar Korrelationen der Testleistung mit der Computererfahrung und Einstellung zum Computer existieren, dies jedoch sowohl bei Computervorgabe, als auch bei Papier-Bleistift-Vorgabe. Deswegen weist Klinck (2002) darauf hin, dass dies ein Hinweis darauf ist, dass andere kausale Variablen diesem Zusammenhang zugrunde liegen.

5.5.3.1.1 Populationsspezifische Effekte bezüglich Einstellung zum Computer

Laut Klinck (2002) gibt es Anzeichen dafür, dass bezüglich der Einstellung zu Computern, Computererfahrung und des Erwerbs von Computerfertigkeiten Geschlechtsdifferenzen existieren, wobei die meisten Befunde zu Unterschieden zwischen Frauen und Männern bezüglich der Einstellung zu Computern vorliegen. Miles und King (1998, zitiert nach Klinck, 2002) haben den momentanen Forschungsstand folgendermaßen zusammengefasst:

The broad range of studies has shown that females report both more negative attitudes toward computers [...] and more computer-related anxiety [...]. However, some studies have failed to find a relationship between gender and computer-related attitudes [...] (Miles & King, 1998, S. 69, zitiert nach Klinck, 2002)

Die Beziehung zwischen Alter und der Einstellung zum Computer ist laut Klinck (2002) ebenfalls umstritten, da sich in einigen Untersuchungen ein Alterseffekt in der

Form findet, dass ältere Personen eine ungünstigere Einstellung zu Computer aufweisen. Jedoch ist dieser Effekt nicht durchgängig zu finden.

Da, wie bereits in der Einleitung erwähnt worden ist, eine online-Vorgabe angestrebt wird, soll sich das folgende Kapitel mit dem Internet als Testmedium beschäftigen und etwaige Vor- und Nachteile darstellen.

5.6 Testmedium Internet

Es wird in dieser Arbeit darauf verzichtet, einen historischen Überblick über das Testen mittels Internet zu geben; dieser ist bei Musch und Reips (2000) zu finden.

Bennett (1999, zitiert nach Bennett, 2001) ist der Meinung, dass drastische Neuerungen im Bereich der Diagnostik vom Fortschritt in folgenden drei Gebieten abhängt: der Technologie, der Messung und der Kognitionsforschung. Er meint, dass neue Technologien auf kurze Sicht den größten Einfluss haben werden, weil diese – im Gegensatz zu den anderen zwei Gebieten – unsere Gesellschaft durchfluten. Bei diesen technologischen Fortschritten handelt es sich hauptsächlich um das Internet. Laut Bennett (2001) ist bzw. wird das Internet in Zukunft *interactive*, *broadband*, *switched* und *standards-based* sein.

Bennett (2001) erklärt diese Begriffe folgendermaßen:

* Interactive means that we can present a task to a student and quickly respond to that student's actions.

* Switched means that we can engage in different interactions with different students simultaneously. In combination, these two characteristics (interactive and switched) make for individualized assessments.

* Broadband means that those interactions can contain lots of information. For assessment tasks, that information could include audio, video, and animation. Those features might make tasks more authentic and more engaging, as well as allow us to assess skills that cannot be measured in paper and pencil [...]. We might also use audio and video to

capture answers, for example, giving examinees choice in their response modalities (typing, speaking, or, for a deaf student, American Sign Language).

* Networked indicates that everything is linked. This linkage means that testing agencies, schools, parents, government officials, item writers, test reviewers, human scorers, and students are tied together electronically. That electronic connection can allow for enormous efficiencies.

* Finally, standards-based means that the network runs according to a set of conventional rules that all participants follow. That fact permits both the easy interchange of data and access from wide variety of computing platforms, as long as the software running on those platforms (e.g., Internet browsers), adheres to those rules too. (S.3)

Die Verwendung der Internettechnologie bei einem Experiment setzt naturgemäß voraus, dass diese mittels Computer durchgeführt wird. Der Einsatz des Computers rechtfertigt jedoch (noch) nicht eine Bezeichnung als Online-Experiment. Denn dazu ist die Verwendung eines Client-Server Netzwerkes (vgl. Janetzko, 2002) mit einem Webserver und einem Client notwendig, der über einen Browser (ein Programm, mit dem sich Dokumente des World Wide Web darstellen lassen) Dienste in Anspruch nimmt. Man kann Browser sowohl im Online- (Anbindung ans Internet) als auch im Offline-Betrieb (Zugriff auf die Festplatte) verwenden. Janetzko (2002) definiert weiters das browserbasierte Experiment, „wenn mit Hilfe von Internettechnologien realisierte Versuchssteuerungsprogramme unter Laborbedingungen verwendet werden“ (S. 103).

Für webbasierte Experimente konnte keine präzise Definition gefunden werden. Um jedoch einer begrifflichen Verwirrung vorzubeugen sei kurz Meyer (2002) zitiert:

Werden Experimente über eine Internet-Lösung durchgeführt, ist der Experimentalserver unbeschränkt mit dem Internet verbunden. Daraus folgt, dass Client-Rechner von beliebigen Standorten aus jederzeit auf die Experimentalprogramme zugreifen können. Dieser Fall ist normalerweise gemeint, wenn von Internet-Experimenten, Web-Experimenten, Online-Experimenten oder eben auch webbasierten Experimenten die Rede ist.“, S. 116)

Auch Reips (2007) vertritt diese Meinung, wenn er schreibt, dass die Bezeichnungen „Web Experiment“, „Internet Experiment“, „Online Experiment“, „webbasiertes Experiment“ und „WWW Experiment“ meist synonym gebraucht werden.

Alle Vorgänge, die für solche Experimente notwendig sind, müssen über ein Computerprogramm gesteuert werden. Ein dafür eingesetztes Programm muss unabhängig von bestimmten Hardwarekonfigurationen und unabhängig von Betriebssystemen funktionieren. Nach Meyer (2002) ist mit solch einem Programm exakt vorzuschreiben, wie sich die einzelnen Akte, aus denen sich ein Experiment konstituiert – Begrüßung, Zuweisung zu einer Versuchsbedingung, Instruktion, Darbietung des Versuchsmaterials, Erfassung und Speicherung der Reaktionen und schließlich die Aufklärung – zu einem gesamten Ergebnis verbinden (S. 118). Das Experiment muss sozusagen „vollautomatisiert“ laufen.

Oben wurde bereits versucht, zu verdeutlichen, dass ein Testen mittels Computer gewisse Vorteile, aber auch gewisse Nachteile mit sich bringt. Gleiches gilt für das Testen mittels der Technologie des Internets, weshalb nun dessen Vor- und Nachteile erörtert werden sollen.

5.6.1 Vor- und Nachteile des Testens mittels Internet

Reips (2000) führt als *Vorteile* von Web-Experimenten – teils in Abhebung von den Nachteilen der laborexperimentellen Methode – folgende Punkte an:

1. leichte Erreichbarkeit eines zahlenmäßig und geographisch fast nicht begrenzten Versuchspersonenpools, in dem auch sehr spezifische und bisher nicht zugängliche Zielgruppen vertreten sind;
2. das Experiment kommt räumlich zur Versuchsperson anstatt umgekehrt;
3. die hohe statistische Power durch eine große Stichprobengröße erlaubt das Beibehalten eines konventionellen α -Niveaus; und
4. reduzierte Kosten, da weder Versuchsräume noch anwesende Experimentatoren gebraucht werden. (S. 320)

Darüber hinaus erwähnt Gadeib (1999) den Vorteil der gesteigerten Datenqualität, denn sie ist der Meinung, dass durch die durchgängig digitale Übermittlung der Daten die Fehlerquote sinkt.

Batinic & Bosnjak (1997, zitiert nach Wagner, 2001) führen als Vorteil noch die Möglichkeiten eines sofortigen Feedbacks an, da für viele potentielle Teilnehmer in Aussicht gestellte Informationen über das eigene „Abschneiden“ in einer Untersuchung einen wesentlichen Teilnahmereiz darstellen und aus diesem Grund die Motivation einer Teilnahme steigen wird.

Ergänzend zu diesen existieren auch noch weitere Punkte, die nicht von vornherein einem der beiden zugeordnet werden können, weshalb nun deren Vor- bzw. Nachteile einer gesonderten, näheren Betrachtung unterzogen werden.

5.6.1.1 Generalisierbarkeit der Ergebnisse

Vor immerhin fast 10 Jahren schrieb Reips (2000), dass verschiedene Nutzeranalysen zeigen, dass alle Populationsparameter mit denen der allgemeinen Bevölkerung konvergieren und deswegen in naher Zukunft (also bereits jetzt?) die Gruppe der Internetbenutzer wohl so repräsentativ für die Allgemeinpopulation sein wird wie zu dieser Zeit die Fernsehkonsumenten.

Betrachtet man darüber hinaus noch den Umstand, dass Reips und Bächtiger (1999, zitiert nach Reips, 2000) ausführen, dass mehr als 80% aller Studien mit Studierenden durchgeführt werden, obwohl deren Anteil an der Allgemeinbevölkerung nur ungefähr 3% beträgt, kann davon ausgegangen werden, dass eine Internetdurchführung einer Studie der Generalisierbarkeit der Ergebnisse nicht schadet. Höchst wahrscheinlich wird dadurch nur deren Generalisierbarkeit erhöht.

Nichts anderes ist gemeint, wenn Krantz und Dalal (2000) von einer hohen externen Validität sprechen, denn eine „externe Validität liegt vor, wenn das in einer Stichprobenuntersuchung gefundene Ergebnis auf andere Personen, Situationen oder Zeitpunkten generalisiert werden kann“ (Bortz & Döring, 2006, S. 53).

5.6.1.2 Ökologische Validität

"Ökologische Validität oder Gültigkeit bezeichnet das Ausmaß, in dem die von den Versuchspersonen einer wissenschaftlichen Untersuchung erlebte Umwelt die Eigenschaften hat, die der Forscher voraussetzt." (Bronfenbrenner 1981, S. 46)

Reips (2000) betont zwar den *Vorteil*, dass beim Web-Experiment eine Situation geschaffen wird, in der das Experiment zur Versuchsperson kommt und nicht umgekehrt, schreibt aber weiters, dass sich dadurch ein ökologischer *Nachteil* des Experimentierens im Internet andeutet. Die Versuchsperson kann in einer gewohnten Umgebung verbleiben, am Schreibtisch zu Hause oder auch am Arbeitsplatz, die einzige Voraussetzung für eine Teilnahme ist ein Computer mit Internetanbindung. Sie kann frei entscheiden, zu welcher Tages- oder Nachtzeit, auch an welchem Wochentag sie am Experiment teilnehmen will, da solche Experimente rund um die Uhr laufen. Darüber hinaus können auch mehrere Personen gleichzeitig am Experiment teilnehmen. Somit ist nach Reips (2000) diese Form des Experimentierens noch stärker von seinem Werkzeug abhängig als andere Formen. Ein Teil dieses Werkzeugs, die aktuelle Verbindungsgüte im Internet, entzieht sich nämlich der Kontrolle der Experimentierenden, was im schlimmsten Fall, bei einer unterbrochenen Netzwerkverbindung, das Ende der Teilnahme einer Versuchsperson bedeuten kann.

5.6.1.3 Spezifische Zielpopulationen

Ein weiterer *Vorteil* der Durchführung von Studien mittels Internet ergibt sich aus dem Umstand, dass diese Durchführungsmethode Forschungen in Bereichen erlaubt, die anderen etablierten Methoden bisher nicht oder nur unter extrem großem Aufwand zugänglich waren (Reips, 2000); etwa wenn eine Studie mit einer ganz spezifischen Zielpopulation durchgeführt werden soll. Reips (2000) führt als Beispiel einer solchen Zielpopulation extrem extravertierte Personen mit Diabetes im Alter zwischen 20 und 30 an, die ein traumatisches Verbrechenserlebnis in ihrer Biografie aufweisen. Die weltweite Erreichbarkeit von Menschen mit Hilfe des Internets, so Reips (2000)

weiter, erlaubt, dass solche Studien (bald, oder auch schon jetzt – 2010) ebenso leicht durchgeführt werden können wie mit einer Durchschnittsstichprobe.

5.6.1.4 Pragmatische Vorteile

Erdfelder, Faul & Buchner (1996) zu Folge ist die Effektstärke der meisten psychologischen Experimente sehr niedrig, weil die Teilnehmerzahl entweder nicht hoch genug ist oder die Wahrscheinlichkeit des α -Fehlers auf einen zu kleinen Wert gesetzt wird. Deswegen meinen diese, dass es nur zwei Wege gibt, um die Effektstärke zu erhöhen, wenn sowohl die Nullhypothese als auch die Alternativhypothese und die Teststatistik bereits festgelegt sind. Der erste Weg besteht darin, die Stichprobengröße zu erhöhen, der zweite in der Vergrößerung der Wahrscheinlichkeit eines Fehlers erster Art (α -Fehler).

Dazu meint Reips (2000), dass es bei Experimenten, die mittels Internet durchgeführt werden, kein Problem sei, auf die ideale, vorher berechnete, Stichprobengröße zu kommen und so ein konventionelles α -Niveau beizubehalten, da der Pool potentieller Versuchspersonen bei dieser Methode der Durchführung von fast unbegrenzter Größe ist.

5.6.1.5 Finanzielle Vorteile

Wie bereits erwähnt laufen Web-Experimente rund um die Uhr und erlauben eine große Zahl gleichzeitiger Zugriffe. Nach Reips (2000) ist dies nur möglich, weil Versuchsleiter während des Experiments nicht gebraucht werden. Somit fallen – sobald ein Web-Experiment programmiert, vorgetestet und ins Internet gestellt ist – keine Löhne für diese an, was natürlich einen finanziellen *Vorteil* darstellt.

Daraus ergibt sich auch noch ein weiterer *Vorteil*, nämlich der, dass etwaige Versuchsleitereffekte („Die unbewusste Verzerrung der Datenlage aufgrund der Erwartungen eines Versuchsleiters“, Zimbardo & Gerrig, 1999, vgl. Rosenthal, 1976) ebenfalls drastisch reduziert werden (Reips, 2000). Es entsteht also die Möglichkeit zur Durchführung eines Dreifachblindversuchs.

Sabin (2004) nutzt den Ansatz von Reips für Spekulationen, in dem sie schreibt, der Versuchsleiter formuliert die Fragen und wählt das Layout der Untersuchung und könnte somit auf diesem Weg Einstellungen und Erwartungen transportieren. Sie meint jedoch auch, dass der Unterschied bzw. der *Vorteil* darin liegt, dass dies auf alle Teilnehmer wirkt und somit standardisiert ist, was bei einem Einsatz mehrerer Testleiter, um nur ein Beispiel zu nennen, nicht der Fall wäre.

Allerdings ergibt sich daraus für Reips (2000) auch ein *Nachteil*, denn durch die Abwesenheit des Versuchsleiters steht dem Versuchsteilnehmer keine Ansprechperson für auftretende Fragen oder Missverständnisse zur Verfügung.

Reips (2000) nennt noch weitere Umstände, die sich kostenreduzierend auswirken, so etwa der Wegfall von benötigten Versuchsräumen und die Verwaltungskosten für die Terminvereinbarungen, Versicherungen, Telefon usw.

5.6.1.6 Kontrolle

Bei der Durchführung eines Experimentes treten viele Fragen auf, die sich die Experimentatoren stellen. Reips (2000) führt folgende an: „Woher weiß ich, dass jemand nicht mehrfach teilnimmt?“, „Wenn jede Versuchsperson einen anderen Computer benutzt, ergibt das dann nicht eine hohe Fehlervarianz?“ und „Da messe ich doch Aufgabenschwierigkeit und nicht Lernleistung, oder?“ (S. 329).

Dazu meint Reips (2000), dass sich diese Fragen auf drei unterschiedliche Arten von Kontrolle beziehen: 1. Einschränkung von „Betrug“, 2. kontrollierende Variablen im Sinne traditionellen Experimentierens in einer beherrschbaren Laborumgebung und 3. Vermeidung von Konfundierungen.

Vor allem zur Betrachtung des ersten Punktes (Einschränkung von Betrug) nennt Reips (2000) Studien, die zeigen, dass sich Online-Versuchspersonen in vieler Hinsicht ganz ähnlich verhalten wie Labor-Versuchspersonen. Weiters sollte bedacht werden, dass ein „Betrug“ einer Versuchsperson immensen Aufwand kostet, wobei kein Nutzen für diese entsteht.

Außerdem besteht nach Reips (2000) eine Reihe von praktischen Maßnahmen, mit deren Hilfe sich unbrauchbare Daten vermeiden oder entdecken lassen. Man kann die Versuchspersonen bitten, eine Kontaktmöglichkeit anzugeben, was erstens vor Missbrauch abschreckt und zweitens die Möglichkeit schafft, verschiedenste Techniken anzuwenden, um die Qualität der Daten festzustellen. Eine dieser Techniken besteht darin, eine Stichprobe der Teilnehmer zu ziehen und einige der gemachten Angaben zu prüfen. Zudem besteht die Möglichkeit, Telefoninterviews durchzuführen oder auch nach einiger Zeit die Teilnehmer mit der Bitte zu kontaktieren, die gemachten biographischen Angaben zu wiederholen.

Schließlich kann der Experimentator auch die Entscheidung treffen, jeweils nur die erste Teilnahme von einer Computeradresse aus auszuwerten, um nur einige der Kontrollmöglichkeiten zu nennen.

Zur Kontrolle im zweiten Sinn (kontrollierende Variablen im Sinne traditionellen Experimentierens in einer beherrschbaren Laborumgebung) bemerkt Reips (2000), dass sich diese immer nur auf einen Teil möglicher Einflussfaktoren beziehen kann. Es ist bei Laborexperimenten Tradition, die Fehlervarianz zu minimieren, indem möglichst viele Einflussfaktoren kontrolliert werden, damit ein vermuteter Effekt überhaupt auftreten kann und nicht „im Rauschen verloren geht“.

Kommt jedoch konsequent die Randomisierungstechnik bei einer großen Stichprobe zur Anwendung, setzt sich laut Reips (2000) jeder systematische Effekt vom Fehlerrauschen ab. Darüber hinaus spricht für Reips (2000) sogar besonders für die Bestätigung einer Vermutung, wenn sich ein angenommener Effekt trotz hoher Fehlervarianz zeigt, da es auch die ökologische Validität (siehe oben) erhöht, wenn Versuchspersonen möglichst frei das tun, was sie immer tun. Und mit einem Web-Experiment ist es möglich, fast alle Einflussfaktoren außer der unabhängigen Variable unkontrolliert zu lassen.

Die dritte Form der Kontrolle ist die Vermeidung von Konfundierungen, also die Kontrolle von Variablen, die mit der unabhängigen Variable systematisch variieren. Dabei sollten Experimentatoren nach Reips (2000) die Kontrolltechniken, die in

Laborexperimenten zur Anwendung kommen (Balancierung, Randomisierung, Konstanthaltung und Beobachtung) auch bei der Durchführung mittels Internet anwenden.

5.6.1.7 Technische Gesichtspunkte

Naturgemäß sind Internet-Experimente abhängig von elektronischer Technik. Da die Information vom Server zur Versuchsperson und zurück durch eine ganze Reihe von Geräten und Kabeln fließt, ergeben sich drei potentielle Problembereiche. Der erste umfasst den Umstand, dass es irgendwo in dieser Kette zu Ausfällen kommen kann, der zweite entsteht durch eine mehr oder weniger große Varianz der Übertragungsgeschwindigkeiten und der dritte kommt dadurch zu Stande, dass die Information bei keinem der Versuchsteilnehmer genau den gleichen Weg geht (Reips, 2000).

Außerdem trägt nach Reips (2000) die technische Varianz zur Fehlervarianz bei, indem Versuchspersonen verschiedene Computer, Monitore, Web-Browser und Netzwerkverbindungen benutzen. Aber wie bereits ausgeführt kann diese Varianz als *Vorteil* angesehen werden, da sie die Generalisierbarkeit der Ergebnisse und somit die externe Validität erhöht.

5.6.1.8 Motivation und Dropout

Bei Internet-Experimenten können Versuchspersonen während des Experiments die Teilnahme jederzeit beenden. Zwar ist dies bei herkömmlichen Laborexperimenten auch der Fall, jedoch aufgrund des Situationsdrucks (z.B. Rechtfertigung gegenüber dem Versuchsleiter) seltener und damit nach Reips (2000) ein Dropout (Abbruch) bei Internet-Durchführung wahrscheinlicher als im Labor. Aus diesem Grund sind in diesem Fall motivationale Faktoren wichtiger als bei Laborexperimenten.

Reips (2000) meint, dass auf der einen Seite Versuchspersonen kaum aus anderen Gründen als tatsächlichem Teilnahmewillen in der Experimentalsituation verbleiben werden und deshalb weniger Fehlervarianz erzeugen. Denn bei den Personen, die dabei bleiben, kann davon ausgegangen werden, dass sie wirklich motiviert sind. Auf

der anderen Seite kann jedoch eine hohe Dropout-Quote ein Experiment entwerten, da es möglich ist, dass die Abbrecher das Experiment aus einem Grund verlassen, der mit dem Untersuchungsgegenstand zu tun hat, was das Ergebnis dieses Experiments weniger verallgemeinbar machen würde. Reips (2000) empfiehlt daher die Anwendung eines Kontrollgruppendesigns, das eine Bedingung enthält, die auf die Kontrolle dieses Dropouts hin angelegt ist.

Er nennt noch weitere Punkte, mit denen sich dieser allgemeine Dropout von Internet-Experimenten reduzieren lässt, nämlich eine Vorauswahl der Versuchspersonen, Belohnungen für die Teilnahme, hohe Datentransferraten, attraktive Web-Designs, interessantes Untersuchungsmaterial und eine bedingungsunabhängige Aufwärmphase. Bei Wagner (2001) und Reips (2007) findet sich eine genauere Ausführung.

An dieser Stelle sei zudem erwähnt, dass Reips (2007) noch (fast) dieselben Vor- und Nachteile anführt wie sieben Jahre davor.

5.6.2 Ethische Richtlinien für die Nutzung des Internets

Ethische Probleme in einer Forschung, die sich mit menschlichen Wesen beschäftigt, sind nichts Neues (Ess, 2007). Wie aus dem vorigen Kapitel erkennbar geworden ist, sollten ethischen Aspekten bei Untersuchungen, die die Technologie des Internets nutzen, eine besondere Beachtung geschenkt werden.

Beispielsweise veröffentlichte die AoIR (Association of Internet Researchers) 2002 die erste Version ihrer interdisziplinären und internationalen Richtlinien für Online-Forschung, welche nach Ess (2007) in der englisch-sprechenden Welt beträchtliche Anwendung findet.

Dzyek (2001, zitiert nach Sabin, 2004, S. 67) fasst wesentliche ethische Aspekte für die psychologische Forschung mittels Fragebogen oder in Form eines Experiments zusammen:

- Die Freiwilligkeit der Versuchsteilnahme
 - Die informierte Einwilligung
 - Die Information, die Aufklärung und das Debriefing
 - Die Anonymität von Daten der Teilnehmer / -innen
 - Die Zusicherung der vertraulichen Behandlung der erhobenen Daten
 - Deren ausschließliche Verwendung für Zwecke des Forschungsprojekts
- (wenn nichts anderes vereinbart wurde)

Darauf Bezug nehmend meint Sabin (2004), dass die Freiwilligkeit der Teilnahme gerade bei Online-Untersuchungen ohne Anwesenheit des Versuchsleiters sehr gut zu realisieren ist, da die Teilnehmer nur das Browserfenster schließen müssen um die Teilnahme zu beenden (vgl. Reips, 2000).

Klarerweise gilt, dass sowohl bei traditionellen Forschungssettings als auch bei Untersuchungen, die online statt finden, die Daten ausschließlich zu dem Zweck verwendet werden dürfen, der den Teilnehmern genannt wurde, worauf Sabin (2004) ausdrücklich hinweist. Ebenso sollte eine vertrauliche Behandlung der erhobenen Daten selbstverständlich sein. Nähere Ausführungen zu Anonymität, Datenschutz und ethischen Richtlinien bei Versuchsleiterabwesenheit sind bei Sabin (2004) zu finden.

5.6.3 Grenzen und Chancen der Experimentierens im Internet

Web-Experimente bieten der psychologischen Forschung viele Vorteile, wie aus dem vorangegangenen Kapitel hervorgeht. Reips (2000) meint, dass diese zwei

Geschmacksrichtungen aufweisen: manche dieser Experimente im Internet sind methodologisch viel versprechend, andere einfach praktisch. Für Reips (2000) stellen Web-Experimente jedoch in dem Falle eine Bereicherung der Methodenlandschaft dar, da sie im Stande sind, den Geltungsbereich psychologischen Wissens in neuer Weise zu überprüfen.

Auch wenn sich Internet-Experimente für viele Bereiche der Forschung eignen, gibt es derzeit noch Bereiche, in denen die Durchführung mittels Internet problematisch erscheint. Es ist jedoch wahrscheinlich, dass neuere Entwicklungen auch eine Anwendungen in diesen Bereichen ermöglichen. Darüber hinaus nennt Reips (2000) Neuentwicklungen im Internet, die vielfältige Möglichkeiten für kreative Experimentatoren versprechen.

Als Argument für eine Durchführung eines Experiments mittels Internet führt Reips (2000) an, dass sich einige der potentiellen Nachteile durch geeignete Maßnahmen vermeiden lassen.

Alles in allem scheinen die Vorteile die Nachteile mehr als auszugleichen (Reips, 2000).

Abschließend sollen nun noch zwei mögliche Wege zur Teilnehmerrekrutierung dargestellt werden.

5.6.4 Rekrutierungsmöglichkeiten

Prinzipiell können nach Sabin (2004) Teilnehmer für Online-Untersuchungen auf demselben Wege wie bei herkömmlichen Untersuchungen rekrutiert werden. Darüber hinaus bietet sich die Möglichkeit an, auch für die Rekrutierung das Internet zu nutzen. Sabin (2004) meint, dass diese Form viel versprechender sei, da kein Medienbruch stattfindet. Daher soll nun hier nur auf die Rekrutierungsmöglichkeiten via Internet eingegangen werden.

5.6.4.1 Aktive Rekrutierung

Sabin (2004) zu Folge erfordert die aktive Rekrutierung, dass der Versuchsleiter die Teilnehmer zufällig auswählt. Dafür kann dieser eine Liste der Teilnehmer erstellen. Es sollte jedoch zusätzlich ein persönlicher Aufruf zur Teilnahme erfolgen. Dies kann via E-Mail oder in so genannten „One-on-One-Chats“ (z.B.: AIM, ICQ) geschehen. Denkbar ist auch eine Rekrutierung in „Pannels“ (Sabin, 2004; vgl. Göritz, 2007).

5.6.4.2 Passive Rekrutierung

Nach Sabin (2004) werden die Teilnehmer jedoch häufiger passiv rekrutiert. Diese nennt sieben mögliche Wege, wie der Aufruf zu Teilnahme erfolgen kann:

- Werbebanner in Webseiten
- Verzeichnisse von Online-Untersuchungen
- Newsgroups
- Diskussionsforen
- „Many-on-Many-Chats“
- Eintrag in Suchmaschinen
- Via E-Mail im Schneeballsystem (die E-Mails werden an eine kleine bekannte Gruppe mit der Bitte verschickte, den Aufruf weiter zu senden)
(S.120)

6 Diskussion und Ausblick

Hambleton (2004) formuliert in seinem Artikel *Theory, methods, and practices in testing for the 21st century* sehr treffend die Richtung, in die sich die psychologische Diagnostik in näherer Zukunft entwickeln wird bzw. sollte wie folgt:

„Item response theory will replace classical test theory because it provides features such as model parameters invariance that can improve the construction and analysis of tests. Computers should replace paper and pencil testing because they provide more flexibility in test scheduling, scores can be provided immediately following test administration, and the capability to assess higher level thinking skills is available. New item types, stimulated by the availability of computers for testing, can be expected to increase the validity of testing.“ (S. 696)

Dies beinhaltet genau jene Punkte, die in dieser Arbeit einer näheren Betrachtung unterzogen worden sind und gibt Hoffnungsschimmer, dass man sich in die richtige Richtung bewegt.

Die Untersuchung und Messung des Gedächtnisses im Allgemeinen und des Arbeitsgedächtnisses im Speziellen ist ein bereits etabliertes Gebiet der Psychologie und besteht zumindest seit Ebbinghaus (1885). Die meisten Untersuchungen derselbigen erfolgten bzw. erfolgen mit sehr traditionellen Methoden.

Aufgrund dieser Arbeit sollte ein Schritt weiter gedacht, die herkömmliche Methode eines Papier-Bleistift-Verfahrens aufgegeben und eine online-Vorgabe angestrebt werden.

Die größte Neuerung besteht jedoch bei der Konstruktion des angestrebten Verfahrens, denn die Generierung der Items soll automatisch erfolgen. Zurzeit besteht (soweit ersichtlich) noch kein vergleichbares Verfahren zur Messung des visuellen Arbeitsgedächtnisses mittels verbalen Materials.

Deshalb wird in diesem Teil vor allem den Ideen und Vorschlägen in Bezug auf AIG Aufmerksamkeit geschenkt.

Wie bereits erwähnt, nennt Arendasy (2003) vier Schritte für die Konzeption und Implementierung eines Itemgenerators.

Die vorliegende Arbeit diene vor allem dem ersten Schritt, also der Untersuchung und Evaluierung der vorhandenen kognitionspsychologische Fachliteratur im Hinblick auf das zu erfassende Konstrukt um relevante Eigenschaften eines Itemgenerators im Sinne von „*radicals*“ und „*incidentals*“ ableiten zu können (Arendasy, 2003).

Es schweben nun folgende Möglichkeiten für „*radicals*“ (jene Aufgabenmerkmale, welche als „Haupteffekte“ einen statistisch signifikanten Beitrag zur Erklärung und Variation psychometrischer Itemeigenschaften leisten, Irvine, 2002 zitiert nach Arendasy, Sommer & Hergovich, 2007) vor:

Aufgrund der Ergebnisse von Conrad und Hull (1964) könnten zur Differenzierung der Schwierigkeiten der einzelnen Buchstabensequenzen, neben der Anzahl der einzelnen Buchstaben in einer Sequenz, die phonologische Ähnlichkeit der einzelnen Buchstaben herangezogen werden. Genauer könnten leichtere Sequenzen dadurch generiert werden, dass nur wenige der vorkommenden Buchstaben sich phonologisch ähneln. Je nach Anzahl akustisch ähnlicher Buchstaben müsste mit Anwachsen deren Anzahl auch die Schwierigkeit ansteigen; auch deswegen, weil auch die Reihenfolge gemerkt werden muss, um die Sequenz richtig wieder erkennen zu können. Denn auch Wickelgren (1965a) zeigte, dass akustische Ähnlichkeit in erster Linie das Behalten der Reihenfolge beeinträchtigt.

Dafür müssen zunächst die Buchstaben der deutschen Sprache nach ihrer phonologischen Ähnlichkeit „geordnet“ werden, was vier Gruppen ergibt:

1. jene Buchstaben, deren Aussprache mit einem [a] endet (a, h, k)
2. jene Buchstaben, deren Aussprache mit einem [e] endet (b, c, d, e, g, p, t, w)

3. jene Buchstaben, deren Aussprache mit einem [e] beginnt (f, l, m, n, r, s)
4. klangunähnliche Buchstaben (i, o, u, v, x, y, z)

Eine allgemein gültige Zuordnung der Buchstaben j und q kann nicht stattfinden, da keine einheitliche Aussprache besteht („je“/„jot“ und „ku“/„que“).

Vorstellbar ist weiters, mit Hilfe eines Itemgenerators die Anzahl der Buchstaben aus derselben Gruppe festzulegen und aus wie vielen verschiedenen Buchstabengruppen die Sequenz zusammengesetzt sein soll. Mit einer Abnahme der Anzahl der herangezogenen Gruppen wird vermutlich die Itemschwierigkeit erhöht.

Als weiterer Aspekt bietet sich die Position der phonologisch ähnlichen Items an. Nach den bereits vorgestellten seriellen Positionseffekten sind bekanntlich die ersten und letzten paar Elemente leichter zu erinnern als die mittleren. Es könnte beispielsweise festgelegt werden, an welcher Position sich die akustisch ähnlichen Buchstaben befinden. Der Theorie zu Folge könnten phonologisch ähnliche Buchstaben in der Mitte der wiederzuerkennenden Sequenz schwieriger zu merken sein, als am Anfang oder am Ende.

Wie bereits erwähnt worden ist, hängt bei „*multiple choice*“-Verfahren die Häufigkeit der richtigen Antworten nicht zuletzt von der „Qualität“ und der Anzahl der Distraktoren ab. Aus diesem Grund könnte die Schwierigkeit der Items auch aufgrund der Zusammensetzung der Distraktoren variiert werden. Als leichte Distraktoren bieten sich Buchstabensequenzen an, deren Buchstaben jenen der Zielsequenz akustisch nicht ähneln oder auch gar nicht enthalten. Schwierige Distraktoren würden somit aus Buchstaben bestehen, die jenen der vorgegebenen Sequenz phonologisch ähnlich sind.

Mit Hilfe eines Itemgenerators könnte man die Anzahl jener Buchstaben in den Distraktoren wählen, die diese mit der Zielsequenz teilen.

Auch wenn es bis dorthin noch ein weiter Weg zu sein scheint, sollte an dieser Stelle noch ein Vorschlag für den vierten und letzten Schritt, also zum Versuch einer Validierung gegeben werden:

Eine Möglichkeit für eine Validierung des konstruierten Verfahrens – genauer für die Überprüfung der Kriteriumsvalidität – besteht in der gezielten Vorgabe an Personen, die nachweislich an der Lese-Rechtschreibschwäche (Legasthenie) leiden, und deren Leistungen mit denen von nicht an Legasthenieleidenden zu vergleichen. Der Theorie zu Folge müssten signifikant schlechtere Ergebnisse der Legastheniker erkennbar sein.

Darüber hinaus scheint nach Generierung und erster explorativer Untersuchungen denkbar, den entstandenen Itempool auf Raschhomogenität zu prüfen, um danach auch eine adaptive Testvorgabe zu realisieren. Da dies die Heranziehung der Item Response Theorie (IRT) impliziert, ist nun allen drei von Hamleton (2004) genannten Punkten Beachtung geschenkt worden.

Zusammenfassend scheint also eine Messung des visuellen Arbeitsgedächtnisses mittels AIG realisierbar.

7 Literaturverzeichnis

- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R: I-S-T 2000 R*. Göttingen: Hogrefe.
- Appl, W. (1999). *Computer in der Neuropsychologischen Diagnostik am Beispiel der ersten Erprobung einer computerisierten neuropsychologischen Testbatterie (CNP von Gur & Gur) im deutschen Sprachraum*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Arendasy, M.E. (2003). *Automatisierte Itemgenerierung und psychometrische Qualitätssicherung am Beispiel des Matrizentests GEOM*. Habilitationsschrift, Universität Wien.
- Arendasy, M.E. (2005). Automatic generation of Rasch-calibrates items: Figural Matrices Test GEOM and Endless Loops Test Ec. *International Journal of Testing*, 5, 197-224.
- Arendasy, M.E., Hergovich, A. & Sommer, M. (2008). Investigating the 'g' – saturation of various stratum-two factors using automatic item generation. *Intelligence*, 36, 574-583.
- Arendasy, M.E. & Sommer, M. (2005). The effect of different types of perceptual manipulations on the dimensionality of automatically generated figural matrices. *Intelligence*, 33, 307-324.

- Arendasy, M.E., Sommer, M., Gittler, G. & Hergovich, A. (2006). Automatic Generation of Quantitative Reasoning Items – A Pilot Study. *Journal of Individual Differences*, 27, 2-14.
- Arendasy, M.E., Sommer, M. & Hergovich, A. (2007). Psychometrische Technologie – Automatische Zwei-Komponenten-Itemgenerierung am Beispiel eines neuen Aufgabentyps zur Messung der Numerischen Flexibilität. *Diagnostica*, 53, 119-130.
- Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advantages in research and theory* (pp. 89-105). NY: Academic Press.
- Baddeley, A.D. (1979). *Die Psychologie des Gedächtnisses*. Stuttgart: Klett-Cotta.
- Baddeley, A.D. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Science*, 4, 417-423.
- Baddeley, A.D. (2002). Is working memory still working?. *European Psychologist*, 7, 85-97.
- Baddeley, A.D. (2003). Working memory: Looking back and looking forward. *Neuroscience*, 4, 829-839.
- Baddeley, A.D., Eysenck, M. W. & Anderson, M. C. (2009). *Memory*. Hove: Psychology Press.
- Baddeley, A.D., Thomson, N. & Buchanen, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 475-589.
- Baller, G., Brand, M., Kalbe, E. & Kessler, J. (2006). *IGD – Inventar zur Gedächtnisdiagnostik*. Göttingen: Hogrefe.

- Bennett, R.E (2001). How the Internet will help Large-Scale Assessment reinvent itself. *Education Policy Analyses Archives*, 9 (5), 1-22.
- Bennett, R.E., Goodman, M., Hessinger, J. Ligget, J., Marshall, G. Kahn, H. & Zack, J. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*, 15, 283-294.
- Blum, F., Didi, H. J., Fay, E., Maichle, U., Trost, G., Wahlen, J. H. & Gittler, G. (1998). *Intelligenz Struktur Analyse (ISA). Ein Test zur Messung der Intelligenz*. Frankfurt: Swets & Zeitlinger B. V, Swets Test Services.
- Booth, J.F. (1992). Computerdiagnostik. In Jäger, R.S. & Petermann, F. (Hrsg.), *Psychologische Diagnostik – ein Lehrbuch* (S. 186-197). Weinheim: Psychologie Verlags Union.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Broadbent, D.E. (1958). *Perception and Communication*. London: Pergamon Press.
- Bronfenbrenner, U. (1981). *Die Ökologie der menschlichen Entwicklung*. Stuttgart: Klett-Cotta.
- Carlson, R.A., Khoo, B.H., Yaure, R.G & Schneider, W. (1990). Acquisition of a problem-solving skill: Levels of organisation and use of working memory. *Journal of Experimental Psychology: General* 119 (2), 193-214.
- Calabrese, P., Deisinger, K., Härting, C., Kessler, J., Markowitsch, H.J. & Neufeld, H. (Hrsg.). (2000). *WMS-R Wechsler Gedächtnistest – Revidierte Fassung*. Bern: Huber.
- Colom, R., Excorial, S., Shih, P.C. & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences*, 42, 1503-1514.

- Conrad, R. & Hull, A.J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, 55, 429-437.
- Conway, A.R.A, Cowan, N., Bunting, M.F., Therriault, D.J. & Minkoff, S.R.B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence, *Intelligence*, 30, 163-183.
- Craik, F.I.M. & McDowd, J.M. (1987). Age Differences in Recall and Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13 (3), 474-479.
- Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Daneman, M. & Merikl, P.M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3 (4), 422-433.
- Deimann, P. & Kastner-Koller, U. (2008). HAWIK-IV. Hamburg-Wechsler-Intelligenztest für Kinder-IV. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40 (3), 161-165.
- De Jong, P.F. & Das-Smaal, E.A. (1995). Attention and intelligence: The validity of the Star Counting Test. *Journal of Educational Psychology*, 87 (1), 80-92.
- Dilling, H., Mombour, W. & Schmidt, M.H. (2008). *Internationale Klassifikation psychischer Störungen – ICD-10. Kapitel V(F)*. Bern: Huber.
- Ebbinghaus, H. (1992). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

- Elger, C.E., Hasse-Sander, I. Helmstaedter, D., Horn, R. & Müller, H. (1997). Rey Auditory-Verbal Learning Test: Structure of a Modified German Version. *Journal of Clinical Psychology, 53*, 663-671.
- Engle, R.W., Laughlin, J.E., Tuholski, S.W. & Conway, A.R.A. (1999) Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *Journal of Experimental Psychology: General, 128* (3), 309-331.
- Erdfelder, E. Faul, F. & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments and Computers, 28*, 1-11.
- Ess, C. (2007). Internet research ethics. In A. N. Joinson, K.Y.A. McKenna, T. Postmes & U.D. Reips (Hrsg.), *The Oxford Handbook of Internet Psychology* (S. 487-502). Oxford: University Press.
- Esser, G. Wyszkon, A & Ballaschk, K. (2009). BUEGA. Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 41* (1), 45-48.
- Fleischhacker, V. (2001). *Validierung und Ersterprobung der Subtests CPW und CPWD (Computerized Penn Word Memory Test) aus der computerunterstützten neuropsychologischen Testbatterie (CNP) von Gur & Gur*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Francis, G., Neath, I. & Surprenant, A. (2000). The Cognitive Psychology Online Laboratory. In M. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 267-283). San Diego: Academic Press.
- Funke, J. (1998). Computer-based Testung ad Training with Scenarios from Complex Problemsolving Research: Advantages and Disadvantages. *International Journal of Selection and Assessment, 6*, 90-96.

- Funke, J. (2006). Alfred Binet (1857 – 1911) und der erste Intelligenztest der Welt. In G. Lamberti (Hg.), *Intelligenz auf dem Prüfstand - 100 Jahre Psychometrie* (S. 23-40). Göttingen: Vandenhoeck & Ruprecht.
- Gathercole, S.E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 23, 83-94.
- Gathercole, S.E., Willis, C., Baddeley, A.D. & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, 2, 103-127.
- Gadeib, A. (1999). Ansprüche und Entwicklung eines Systems zur Befragung über das World Wide Web. In Batinic, B. Werner A., Gräf, L. & Bandilla, W. (Hrsg.), *Online Research – Methoden, Anwendungen und Ergebnisse* (S. 103-112). Göttingen: Hogrefe.
- Geuß, H. (1985). Ein neues Gedächtnistestverfahren zur getrennten Erfassung von Speicher- und Prozessanteilen im unmittelbaren und kurzzeitigen Bereich: Testentwicklung und erste Validitätsanalysen. *Diagnostica*, 31, 130-139.
- Gittler, G. & Arendasy, M. (2003). *Differentielle Psychologie I. (3.Auflage)*. Skriptum zur Vorlesung: Differentielle Psychologie 1.
- Göriz, A.S. (2007). The methodology of Internet-based experiments. In A. N. Joinson, K.Y.A. McKenna, T. Postmes & U.D. Reips (Hrsg.), *The Oxford Handbook of Internet Psychology* (S. 473-485). Oxford: University Press.
- Haavisto, M.L. & Lehto, J.E. (2004). Fluid/spatial and crystallized intelligence in relation to domain-specific working memory: A latent-variable approach. *Learning and Individual Differences*, 15, 1-21.

- Halpern, D.F. (2000). *Sex Differences In Cognitive Abilities*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hambleton, R.K. (2004). Theory, methods, and practices in testing fort the 21st century. *Psicothema, 16*, 696-701.
- Hamm, S., Junglas, K. & Bredenkamp, J. (2004). Die Zentrale Exekutive als präartikulatorische Kontrollinstanz. *Zeitschrift für Psychologie, 212* (2), 66-75.
- Hasselhorn, M. (1988). Wie und warum verändert sich die Gedächtnisspanne über die Lebensspanne? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 20*, 322-337.
- Hasselhorn, M. Grube, D. & Mähler, C. (2000). Theoretisches Rahmenmodell für ein Diagnostikum zur differentiellen Funktionsanalyse des phonologischen Arbeitsgedächtnisses. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Lese-Rechtschreibschwierigkeiten* (S. 119-133). Göttingen: Hogrefe.
- Hasselhorn, M., Seidler-Brandler, U. & Körner, K. (2000). Ist das „Nachsprechen von Kunstwörtern“ für die Entwicklungsdiagnostik des phonologischen Arbeitsgedächtnisses geeignet?. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Lese-Rechtschreibschwierigkeiten* (S. 119-133). Göttingen: Hogrefe.
- Hasselhorn, M. & Marx, H. (2000). Phonologisches Arbeitsgedächtnis und Leseleistungen. Ein Vergleich zwischen dysphasisch-sprachentwicklungsgestörten und sprachunauffälligen Kindern. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Lese-Rechtschreibschwierigkeiten* (S. 119-133). Göttingen: Hogrefe.

- Hasselhorn, M., Tiffin-Richards, M., Woerner, W., Banaschewski, T. & Rothenberger, A. (2000). Spielt der phonetische Speicher des Arbeitsgedächtnisses eine bedeutsame Rolle für die Differentialdiagnose von Lese-Rechtschreib-Schwierigkeiten? Analysen zum „Kunstwörter-Nachsprechen“ bei Kindern mit LRS- und/oder HKS-Diagnose. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Lese-Rechtschreibschwierigkeiten* (S. 119-133). Göttingen: Hogrefe.
- Häusler, J. (2006). Adaptive success control in computerized adaptive testing. *Psychology Science*, 48, 436-450.
- Hedges, L.V. & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41-45.
- Heitz, R.P., Unsworth, N. & Engle, R.W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R.W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (S. 61-77). Thousand Oaks: Sage Publications.
- Helmstaedter, C., Lendt, M. & Lux, S. (2001). *Verbaler Lern- und Merkfähigkeitstest – VLMT*. Göttingen: Beltz.
- Heubrock, D. (1992). Der Auditiv-Verbale Lerntest (AVLT) in der klinischen und experimentellen Neuropsychologie. Durchführung, Auswertung und Forschungsergebnisse. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 161-174.
- Holling, H., Blank, H., Kuchenbäcker, K. & Kuhn, J.T. (2008). Rule-based item design of statistical word problems: A review and first implementation. *Psychology Science Quarterly*, 50, 363-378.
- Hornke, L.F. (1993). Mögliche Einspareffekte beim computergestützten Testen. *Diagnostica*, 2, 109-119.

- Hyde, J.S. & Linn, M.C. (1988). Gender differences in verbal Ability: A Meta-Analysis. *Psychological Bulletin*, 104, 53-69.
- Irvine, S.H. & Kyllonen, P.C. (2002). *Item Generation for Test Development*. Mahwah, New Jersey: Lawrence Erlbaum Associates
- Ivnik, R.J. & Malec, J.F. (1995). Older person's reactions to a computerized assessment battery in comparison to traditional testing with psychometrics. *Journal of International Neurosciences*, 1, 166-171.
- Jäger, A.O. & Althoff, K. (1994). *Der WILDE-Intelligenz-Test: (WIT); ein Strukturdiagnostikum*. Göttingen: Hogrefe.
- Jäger, A.O., Süß, H.-M. & Beauducel, A. (1996). *Berliner Intelligenzstruktur-Test: BIS-Test*. Göttingen: Hogrefe.
- Janetzko, D. (2002). Browserbasierte Experimente. In D. Janetzko, M. Hildebrandt & H. Meyer (Hrsg.), *Das experimentalpsychologische Praktikum im Labor und WWW (S. 101-112)*. Göttingen: Hogrefe.
- Jansen, H. Mannhaupt, G. Marx, H. & Skowronek, H. (1999). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)*. Göttingen: Hogrefe.
- Kail, R. & Hall, L.D. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition*, 29, 1-9.
- Kasten, E. (1994). Behandlung von zerebral-bedingten Gesichtsfeldausfällen: Entwicklung und Validierung von computergestützten Diagnose- und Trainingsmethoden. In E. Kasten, W. Janke & B. Sabel (Hrsg.), *Medizinische und biologische Psychologie (S.69-70)*. Königshausen: Neuman Verlag.
- Kintsch, W. (1982). *Gedächtnis und Kognition*. Berlin: Springer.

- Klicpera, C., Schabmann, A. & Gasteiger-Klicpera, B. (2007). *Legasthenie. Modelle, Diagnose, Therapie und Förderung*. München: Ernst Reinhardt Verlag.
- Klinck, D. (2002). *Computergestützte Diagnostik*. Göttingen: Hogrefe.
- Krantz, J.H. & Dalal, R. (2000). Validity of Web-Based Psychological Research. In M. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 35-60). San Diego: Academic Press.
- Kratzmeier, H. & Horn, R. (1988). *Raven-Matrizen-Test: Standard Progressive Matrices (SPM)*. Weinheim: Beltz.
- Kubinger, K.D. & Wurst, E. (2000). *AID 2 Adaptives Intelligenz Diagnostikum 2*, Göttingen: Beltz Test GmbH.
- Kubinger, K.D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Lehr, U. (2007). *Psychologie des Alterns*. Wiebelsheim: Quelle & Meyer Verlag.
- Lezak, M.D. (1995). *Neuropsychological Assessment*. Oxford: Oxford University Press.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: PVU.
- Linder, M. & Grisseemann, H. (1996). *Zürcher Lesetest (ZLT); Förderdiagnostik bei gestörtem Schriftspracherwerb*. Bern: Huber.
- Mandler, G. (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, 87 (3), 252-271.
- Mannhaupt, G. (2006). *Münsteraner Screening (Müsc) zur Früherkennung von Lese-Rechtschreibschwierigkeiten*. Berlin: Cornelsen.

- Metz, U., Marx, P., Weber, J. & Schneider, W. (2003). Overachievement im Lesen und Rechtschreiben: Folgerungen für die Diskrepanzdefinition der Legasthenie. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 35 (3), 127-134.
- Meyer, H.A. (2002). Webbasierte Experimente. In D. Janetzko, M. Hildebrandt & H. Meyer (Hrsg.), *Das experimentalpsychologische Praktikum im Labor und WWW (S. 113-126)*. Göttingen: Hogrefe.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Mottier, G. (1951). Der Mottier-Test. Über Untersuchungen zur Sprache lesegestörter Kinder. *Folia Phoniatica*, 3, 170-177.
- Musch, J. & Reips, U.D. (2000). A Brief History of Web Experimenting. In M. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 61-87). San Diego: Academic Press.
- Parkin, A. (1993). *Gedächtnis – Ein einführendes Lehrbuch*. Weinheim: Psychologie Verlags Union.
- Parkin, J. (1999). *Memory: A Guide for Professionals*. Chichester: Wiley.
- Petermann, F., & Petermann, U. (2007). *Hamburg-Wechsler Intelligenztest für Kinder – HAWIK-IV*. Bern: Huber.
- Rack, J.P., Snowling, M.J. & Olson, R.K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, 27, 29-53.
- Reips, U.D. (2000). Das psychologische Experimentieren im Internet. In Batinic, B. (Hrsg.), *Internet für Psychologen* (S. 319-343). Göttingen: Hogrefe.

- Reips, U. D. (2007). The methodology of Internet-based experiments. In A. N. Joinson, K.Y.A. McKenna, T. Postmes & U.D. Reips (Hrsg.), *The Oxford Handbook of Internet Psychology* (S. 373-390). Oxford: University Press.
- Rockstroh, S. & Schweizer, K. (2001). The contributions of memory and attention processes to cognitive abilities. *The Journal of General Psychology*, 128, 30-42.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Appleton Century Crofts.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Sabin, S. (2004). *Online Research in der psychologischen Forschung: Methoden und Techniken für valide Online-Erhebungen und –Experimente*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Salamé, P. & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, 24 (4), 263-265.
- Schermer, F.J. (2006). *Lernen und Gedächtnis*. Stuttgart: Verlag W. Kohlhammer.
- Schuchardt, K., Maehler, C. & Hasselhorn, M. (2008). Working Memory Deficits in Children with Specific Learning Disorders. *Journal of Learning Disabilities*, 41 (6), 514-523.
- Schuhfried, G. (2006). Das Zeitalter der computergestützten psychologischen Diagnostik. In G. Lamberti (Hg.), *Intelligenz auf dem Prüfstand - 100 Jahre Psychometrie* (S. 121-136). Göttingen: Vandenhoeck & Ruprecht.
- Schweizer, K. (1996a). The speed-accuracy transition due to task complexity. *Intelligence*, 22, 115-128.

- Schweizer, K. (1996b). Level of encoding, preattentive processing, and working-memory capacity as sources of cognitive ability. *Personality and Individual Differences, 21*, 759-766.
- Schweizer, K. & Koch, W. (2001). Perceptual processes and cognitive ability. *Intelligence, 31*, 211-235.
- Schweizer, K. & Moosbrugger (1999). Aufmerksamkeit, Intelligenz und Verarbeitungsgeschwindigkeit als Komponenten der mentalen Leistungsfähigkeit. *Zeitschrift für Differentielle und Diagnostische Psychologie, 20*, 126-132.
- Schweizer, K., Zimmermann, P. & Koch, W. (2000). Sustained attention, intelligence and the crucial role of perceptual processes. *Learning and Individual Differences, 12*, 271-286.
- Schweizer, K. & Moosbrugger (2004). Attention and working memory as predictors of intelligence. *Intelligence, 32*, 329-347.
- Shah, P. & Miyake, A. (1996). The Separability of Working Memory Resources for Spatial Thinking and Language Processing: An Individual Differences Approach. *Journal of Experimental Psychology: General, 125*, 4-27.
- Singleton, C., Horne, J. & Simmons, F. (2009). Computerised screening for dyslexia in adults. *Journal of Research in Reading, 32* (1), 137-152.
- Sivan, A.B. & Spreen, O. (1996). *Der Benton-Test Handbuch*. Bern: Huber.
- Steger-Wuchse, D. (1999). *Computerunterstützte neuropsychologische Testbatterie (CNP) von Gur & Gur – Eine Adaption für den deutschen Sprachraum*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Sturm, W. & Willmes, K. (1999). *Verbaler Lerntest (VLT)*. Göttingen: Hogrefe.

- Süß, H.M., Oberauer, K., Wittman, W.W., Wilhelm, O. & Schulze, R. (2002). Working-memory capacity explains reasoning ability – And a little more. *Intelligence*, 30, 261-288.
- Trahan, E. E. & Quintana, J. W. (1990). Analysis of gender effects upon verbal and visual memory performance in adults. *Archives of Clinical Neuropsychology*, 5, 325-334.
- Vasic, N., Lorh, C., Steinbrink, C., Martin, C. & Wolf, R.C. (2008). Neural correlates of working memory performance in adolescents and young adults with dyslexia. *Neuropsychologia*, 46 (2), 640-648.
- Vidonyi, E. (2005). *Der Zusammenhang zwischen der Selbsteinschätzung kognitiver Fähigkeiten und wahrer Fähigkeiten*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Von Aster, M., Neubauer, A. & Horn, R. (2006). *Wechsler-Intelligenztest für Erwachsene WIE - Übersetzung und Adaption der WAIS-III*. Frankfurt: Harcourt Test Services.
- Vogt, K. (1999). Verzerrungen in elektronischen Befragungen? In Batinic, B. Werner A., Gräf, L. & Bandilla, W. (Hrsg.), *Online Research – Methoden, Anwendungen und Ergebnisse* (S. 127-144). Göttingen: Hogrefe.
- Wagner, B. (2001). *Aspekte psychologischer Forschung im Internet*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Warnke, A., Hemminger, U. & Plume, E. (2004). *Lese-Rechtschreibstörungen. Leitfaden Kinder- und Jugendpsychotherapie*. Göttingen: Hogrefe.
- Wechsler, D. (1956). *Die Messung der Intelligenz Erwachsener*. Bern: Huber.

- Wechsler, D. (2003). *WISC-IV. Wechsler Intelligence Scale for Children. Administration and scoring manual*. San Antonio: The Psychological Corporation.
- Weiß, R. (2006). *CFT 20-R Grundintelligenztest Skala 2 Revision*, Göttingen: Hogrefe.
- Wickelgren, W.A. (1965a). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 70, 102-108.
- Wickelgren, W.A. (1965b). Short-Term memory for phonemically similar lists, *The American Journal of Psychology*, 78 (4), 567-574.
- Wimmer, H. (1996). The Nonword Reading Deficit in Developmental Dyslexia: Evidence from Children Learning to Read German. *Journal of Experimental Child Psychology*, 61, 80-90.
- Woltz, D.J (1988). An Investigation of the Role of Working Memory in Procedural Skill Acquisition. *Journal of Experimental Psychology: General*, 177 (3), 319-331.
- Zimbardo, P.G. & Gerrig, R.J. (1999). *Psychologie*. Berlin: Springer.

8 Anhang

Abstract

Ausgangspunkt dieser Arbeit stellt das übergeordnete Ziel einer Entwicklung eines neuen Verfahrens zur Messung des visuellen Arbeitsgedächtnisses mit Hilfe von automatischer Itemgenerierung dar. Die Möglichkeit einer derartigen Messung wird einer theoretischen Betrachtung unterzogen.

Von einigen bereits bestehenden Ideen ausgehend werden in dieser Arbeit folgende Themen behandelt: Theorie des Arbeitsgedächtnis, Gedächtnisdiagnostik und Automatische Itemgenerierung.

Im ersten Kapitel, welches der Theorie des Gedächtnisses gewidmet ist, werden das Modell des Arbeitsgedächtnisses von Baddeley (1986) und darauf aufbauende empirische Erkenntnisse und Zusammenhänge, vor allem zu Intelligenz und Legasthenie, vorgestellt.

Das nachfolgende Kapitel beschäftigt sich mit der Gedächtnisdiagnostik. Darin werden einige relevante Verfahren aus den unterschiedlichsten Gebieten der Psychologie beschrieben und Beispiele für online durchzuführende Testmöglichkeiten genannt.

Das dritte und letzte Kapitel widmet sich der Automatischen Itemgenerierung und der Computerdiagnostik; speziell einer Testvorgabe mittels dem Medium Internet.

Im abschließenden Diskussionsteil werden Ideen und Vorschläge für eine Konstruktion eines derartigen Verfahrens vorgestellt.

CURRICULUM VITAE

SANDRA KOGLER

AUSBILDUNG

1990 – 1994	Volksschule St. Peter/Au
1994 – 1999	Bundesgymnasium Amstetten
1999 – 2002	Bundesrealgymnasium Waidhofen/Ybbs
seit Oktober 2003	Lehramtsstudium UF Psychologie und Philosophie / UF Mathematik Universität Wien
seit März 2004	Diplomstudium der Psychologie Universität Wien

BERUFLICHE TÄTIGKEITEN

Sommerferien 2008	Leitung der Mathematikgruppe der Lernakademie Waidhofen/Ybbs 2 Wochen Intensivkurs zu jeweils 30h / Woche
Herbst 2008	Absolvierung des 6-Wochen-Pflichtpraktikums im Arbeitsbereich der Psychologischen Diagnostik unter der Betreuung von Frau Dr. Wagner-Menghin
seit September 2010	Lehrtätigkeit am BG/BRG Purkersdorf Standort Tullnerbach