



WESTFÄLISCHE  
WILHELMS-UNIVERSITÄT  
MÜNSTER

Fachgebiet: Psychologie

# Arbeitsgedächtniskapazität bei Kindern mit durchschnittlicher und hoher Intelligenz

Inaugural-Dissertation

zur Erlangung des Doktorgrades der

Philosophischen Fakultät der

Westfälischen Wilhelms-Universität zu Münster (Westf.)

vorgelegt von

Miriam Vock

aus Münster

August 2004

1. Gutachter: Prof. Dr. Holling
2. Gutachter: Prof. Dr. Schäfer

Tag der mündlichen Prüfung: 15.10.2004

Dekan: Prof. Dr. Tomasek

Referent: Prof. Dr. Holling

Korreferent: Prof. Dr. Schäfer

# Inhalt

|           |  |    |
|-----------|--|----|
| <b>1.</b> | <b>Einleitung</b> .....  | 1  |
| <b>2.</b> | <b>Diagnostik intellektueller Hochbegabung</b> .....   | 4  |
| 2.1       | Modellvorstellungen zu intellektueller Hochbegabung .....  | 4  |
| 2.2       | Intelligenzmodelle.....  | 6  |
| 2.3       | Diagnoseanlässe .....  | 10 |
| 2.4       | Probleme der Intelligenzmessung im Rahmen der Hochbegabungsdiagnostik .....  | 11 |
| 2.4.1     | Allgemeine Probleme bei der Messung von Intelligenz mit den üblichen Intelligenztestverfahren .....  | 13 |
| 2.4.2     | Spezifische Probleme bei der Messung hoher Intelligenz mit den üblichen Intelligenztestverfahren.....  | 15 |
| 2.5       | Spezielle Intelligenztests zur Messung hoher Intelligenz.....  | 18 |
| 2.6       | Überlegungen zur Entwicklung eines neuen Testverfahrens: Erfassung zentraler Komponenten der Intelligenz durch Arbeitsgedächtnisaufgaben ..... | 19 |
| <b>3.</b> | <b>Arbeitsgedächtnis</b> .....   | 22 |
| 3.1       | Das Konzept des Arbeitsgedächtnisses.....  | 22 |
| 3.2       | Das Arbeitsgedächtnis aus der Perspektive der differentiellen Psychologie .....  | 25 |
| 3.3       | Die Binnenstruktur des Arbeitsgedächtnisses.....   | 30 |
| 3.3.1     | Studien zu Zusammenhängen von Arbeitsgedächtnisaufgaben mit anderen kognitiven Leistungen.....   | 33 |
| 3.3.2     | Studien zu Zusammenhängen von verschiedenen Arbeitsgedächtnisaufgaben .....  | 35 |
| 3.3.3     | Gibt es einen separaten numerischen Faktor?.....   | 37 |
| 3.3.4     | Integration der Befunde zur Binnenstruktur .....   | 39 |
| 3.4       | Zusammenhänge des Arbeitsgedächtnisses mit allgemeiner Intelligenz und anderen kognitiven Variablen .....                                      | 41 |
| 3.4.1     | Intelligenz.....   | 42 |
| 3.4.2     | Geschwindigkeit der Informationsverarbeitung oder Speed.....   | 48 |
| 3.4.3     | Kurzzeitgedächtnis.....  | 51 |
| 3.4.4     | Schulleistungen .....  | 53 |
| 3.5       | Zusammenfassung.....   | 55 |
| <b>4.</b> | <b>Die Messung der Arbeitsgedächtniskapazität</b> .....  | 57 |
| 4.1       | Klassische Aufgabentypen zur Erfassung der Arbeitsgedächtniskapazität .....  | 57 |
| 4.1.1     | Aufgabentyp „Reading Span“ .....   | 58 |
| 4.1.2     | Aufgabentyp „Backward Digit Span“ .....  | 60 |

## Inhaltsverzeichnis

---

|           |  |           |
|-----------|--|-----------|
| 4.2       | Scoring von Arbeitsgedächtnisaufgaben.....   | 62        |
| 4.2.1     | Welche Leistungsdaten werden beim Scoring berücksichtigt?.....   | 62        |
| 4.2.2     | Zusammenfassung verschiedener Leistungen zu einem Score.....   | 63        |
| 4.2.3     | Vergleich verschiedener Scoringmethoden im Hinblick<br>auf die Aggregation von Aufgabenwerten.....     | 67        |
| 4.2.4     | Berücksichtigung der Reihenfolge der Antworten<br>( <i>item recall vs. serial recall</i> ).....        | 68        |
| 4.2.5     | Fazit zu den verschiedenen Scoringmethoden.....  | 69        |
| 4.3       | Individualdiagnostik der Arbeitsgedächtniskapazität<br>mit standardisierten Tests.....                 | 70        |
| 4.3.1     | Der „Star Counting Test“ (SCT).....  | 70        |
| 4.3.2     | Die „Working Memory Test Battery for Children“ (WMTB-C).....   | 71        |
| 4.3.3     | Der „Swanson Cognitive Processing Test“ (S-CPT).....   | 74        |
| 4.4       | Messgenauigkeit von Arbeitsgedächtnisaufgaben.....   | 76        |
| 4.5       | Zusammenfassung.....   | 78        |
| <b>5.</b> | <b>Aufgabenauswahl und Fragestellungen</b> .....   | <b>79</b> |
| 5.1       | Verwendung von Arbeitsgedächtnisaufgaben für die Diagnostik<br>hoher intellektueller Fähigkeiten.....  | 79        |
| 5.2       | Auswahl von Typen von Arbeitsgedächtnisaufgaben.....   | 82        |
| 5.3       | Fragestellungen.....   | 83        |
| 5.3.1     | Fragestellungen zum Konstrukt „Arbeitsgedächtnis“.....   | 84        |
| 5.3.2     | Fragestellungen zur Testentwicklung.....   | 95        |
| <b>6.</b> | <b>Untersuchung</b> .....  | <b>99</b> |
| 6.1       | Arbeitsgedächtnisaufgaben.....   | 99        |
| 6.1.1     | Aufgabe „Reading Span“.....  | 100       |
| 6.1.2     | Aufgabe „Verbal Span“.....   | 102       |
| 6.1.3     | Aufgabe „Computation Span“.....  | 104       |
| 6.1.4     | Aufgabe „Backward Digit Span“.....   | 105       |
| 6.1.5     | Aufgabe „Memory Updating Figural“.....   | 106       |
| 6.1.6     | Aufgabe „Spatial Working Memory“.....  | 109       |
| 6.2       | Weitere Materialien.....   | 111       |
| 6.2.1     | Fragebogen.....  | 111       |
| 6.2.2     | Kognitive Tests.....   | 112       |
| 6.3       | Stichprobe.....  | 115       |
| 6.3.1     | Stichprobe von Schülerinnen und Schülern aus regulären Schulen.....                                    | 116       |
| 6.3.2     | Stichprobe von Schülerinnen und Schülern aus Einrichtungen und<br>Organisationen für hoch Begabte..... | 116       |
| 6.3.3     | Überblick über die untersuchten Gruppen.....   | 118       |
| 6.4       | Durchführung der Untersuchung.....   | 122       |

|                        |  |            |
|------------------------|--|------------|
| <b>7.</b>              | <b>Ergebnisse</b> .....  | 125        |
| 7.1                    | Deskriptive Ergebnisse und Vergleich von Scorings.....   | 125        |
| 7.1.1                  | Fragebogen.....  | 126        |
| 7.1.2                  | Kognitive Tests .....  | 126        |
| 7.1.3                  | Arbeitsgedächtnisaufgaben .....  | 129        |
| 7.1.4                  | Empirischer Vergleich der Scorings der Arbeitsgedächtnis-<br>aufgaben auf Aufgaben- und auf Itemebene .....                          | 130        |
| 7.2                    | Untersuchung der Binnenstruktur des Arbeitsgedächtnisses .....   | 132        |
| 7.2.1                  | Bivariate und Partialkorrelationen der Arbeitsgedächtnisaufgaben.....  | 132        |
| 7.2.2                  | Exploratorische und konfirmatorische Faktorenanalysen.....   | 133        |
| 7.2.3                  | Prüfung der Invarianz der Faktorstruktur für verschiedene Alters-<br>und Begabungsgruppen sowie für Mädchen und Jungen.....          | 139        |
| 7.2.4                  | Untersuchung von mittleren Leistungsunterschieden zwischen<br>Gruppen.....   | 144        |
| 7.3                    | Prüfung der Zusammenhänge der Arbeitsgedächtnisfaktoren mit<br>kognitiven Fähigkeiten.....   | 147        |
| 7.3.1                  | Fluide und kristalline Intelligenz .....   | 148        |
| 7.3.2                  | Geschwindigkeit der Informationsverarbeitung.....  | 151        |
| 7.3.3                  | Gesamtmodell: Arbeitsgedächtnis und andere<br>kognitive Fähigkeiten.....   | 156        |
| 7.3.4                  | Zusammenfassung und Diskussion der Analysen der Zusammenhänge<br>zwischen Arbeitsgedächtnis und anderen kognitiven Fähigkeiten ..... | 158        |
| 7.4                    | Vorhersage von Schulleistungen anhand von Leistungen in<br>Arbeitsgedächtnisaufgaben? .....  | 161        |
| 7.5                    | Testtheoretische Analysen .....  | 165        |
| 7.5.1                  | Vorgehen bei der Analyse nach der Klassischen Testtheorie und<br>der Item Response Theorie .....                                     | 165        |
| 7.5.2                  | Dimensionalität der Arbeitsgedächtnisaufgaben .....  | 168        |
| 7.5.3                  | Bestimmung der Aufgabenkennwerte nach der Klassischen<br>Testtheorie und Schätzung der Aufgabenparameter nach<br>IRT-Modellen.....   | 171        |
| 7.5.4                  | Modellgeltungstests .....  | 182        |
| 7.5.5                  | Spezifische Modelltests für das 1PL-Modell .....   | 186        |
| 7.5.6                  | Reliabilität und externe Validität der Rasch-Modell<br>konformen Skalen .....  | 195        |
| 7.5.7                  | Zusammenfassung der testtheoretischen Analysen und Fazit.....  | 198        |
| <b>8.</b>              | <b>Zusammenfassung und Diskussion</b> .....  | <b>200</b> |
| <b>Literatur</b> ..... |  | <b>209</b> |
| <b>Anhang</b>          |  |            |
| A                      | Fragebogen   |            |
| B                      | Faktorstruktur des Arbeitsgedächtnisses im Gruppenvergleich (Abbildungen)  |            |

## Verzeichnis der Abbildungen

|   |     |
|---|-----|
| <i>Abbildung 2.1:</i><br>Berliner Intelligenzstrukturmodell (BIS) nach Jäger (1982) .....   | 9   |
| <i>Abbildung 3.1:</i><br>Hypothetisches Zwei-Facetten Modell der Arbeitsgedächtniskapazität<br>mit je einer der verwendeten Aufgabentypen pro Zelle<br>(Oberauer et al., 2000; Süß et al., 2002) .....            | 31  |
| <i>Abbildung 3.2:</i><br>Korrelationen der latenten Faktoren Arbeitsgedächtnis (AG),<br>allgemeines Wissen, Reasoning und Processing Speed in der<br>Untersuchung von Kyllonen und Christal (1990, Studie 1)..... | 44  |
| <i>Abbildung 3.3:</i><br>Korrelationen der latenten Faktoren Arbeitsgedächtnis (AG),<br>Schulleistung, fluide Intelligenz ( $g_f$ ) und Speed in der<br>Studie von de Jong und Das-Smaal (1995).....              | 45  |
| <i>Abbildung 3.4:</i><br>Korrelationen der latenten Faktoren Arbeitsgedächtnis (AG),<br>fluide Intelligenz ( $g_f$ ) und Processing Speed in der<br>Studie von Ackerman et al. (2002) .....                       | 47  |
| <i>Abbildung 6.1:</i><br>Übungsaufgabe 1 der Aufgabengruppe „Reading Span“ mit richtiger Lösung.....  | 101 |
| <i>Abbildung 6.2:</i><br>Übungsaufgabe 1 der Aufgabengruppe „Verbal Span“ mit richtiger Lösung .....  | 103 |
| <i>Abbildung 6.3:</i><br>Mögliche Positionen der Punkte und mögliche Pfeilrichtungen<br>in der Aufgabe „Memory Updating Figural“ .....  | 107 |
| <i>Abbildung 6.4:</i><br>Übungsaufgabe 3 der Aufgabengruppe „Memory Updating Figural“<br>mit richtiger Lösung .....   | 109 |
| <i>Abbildung 6.5:</i><br>Übungsaufgabe 1 der Aufgabengruppe „Spatial Working Memory“<br>mit richtiger Lösung.....   | 111 |

## Abbildungsverzeichnis

---

|  |     |
|--|-----|
| <i>Abbildung 6.6:</i><br>Hick-Aufgaben 0-3. Darstellung der Stimuli (Zahlen) und geforderten<br>Antworten (Pfeile), Beispielzeilen aus Hick-0 und Hick-3 ..... | 113 |
| <i>Abbildung 6.7:</i><br>Ablauf der Untersuchung.....  | 123 |
| <i>Abbildung 7.1:</i><br>Screeplot über die sechs Arbeitsgedächtnis-Scores.....  | 134 |
| <i>Abbildung 7.2:</i><br>Modell 1. Ein-Faktor-Modell des Arbeitsgedächtnisses (CFA).....   | 136 |
| <i>Abbildung 7.3:</i><br>Modell 2. Zwei-Faktoren-Modell des Arbeitsgedächtnisses (CFA).....  | 137 |
| <i>Abbildung 7.4:</i><br>Modell 3. Drei-Faktoren-Modell des Arbeitsgedächtnisses (CFA).....  | 138 |
| <i>Abbildung 7.5:</i><br>Zwei-Faktorenmodell der Intelligenz .....   | 149 |
| <i>Abbildung 7.6:</i><br>Zwei-Faktoren-Modell für Speed .....  | 152 |
| <i>Abbildung 7.7:</i><br>Modell latenter Faktoren.....   | 157 |
| <i>Abbildung 7.8:</i><br>Screeplot: Eigenwertverlauf der 20 größten Werte der<br>Faktorenanalyse über die 70 Aufgaben .....                                    | 169 |
| <i>Abbildung 7.9:</i><br>Testinformationsfunktion für die verbal-numerische Skala<br>nach dem 1PL-Modell .....   | 184 |
| <i>Abbildung 7.10:</i><br>Testinformationsfunktion für die verbal-numerische Skala<br>nach dem 2PL-Modell .....  | 184 |
| <i>Abbildung 7.11:</i><br>Testinformationsfunktion für die figurale Skala nach dem 1PL-Modell.....   | 185 |
| <i>Abbildung 7.12:</i><br>Testinformationsfunktion für die figurale Skala nach dem 2PL-Modell.....   | 185 |

## Abbildungsverzeichnis

---

|  |     |
|--|-----|
| <i>Abbildung 7.13:</i><br>Graphischer Modelltest für die verbal-numerische Skala,<br>Split-Kriterium: Score.....                     | 188 |
| <i>Abbildung 7.14:</i><br>Graphischer Modelltest für die verbal-numerische Skala,<br>Split-Kriterium: Geschlecht.....                | 189 |
| <i>Abbildung 7.15:</i><br>Graphischer Modelltest für die verbal-numerische Skala,<br>Split-Kriterium: Alter .....                    | 189 |
| <i>Abbildung 7.16:</i><br>Graphischer Modelltest für die verbal-numerische Skala,<br>Split-Kriterium: Intelligenz (HI vs. NHI) ..... | 189 |
| <i>Abbildung 7.17:</i><br>Graphischer Modelltest für die figurale Skala, Split-Kriterium: Score .....                                | 190 |
| <i>Abbildung 7.18:</i><br>Graphischer Modelltest für die figurale Skala, Split-Kriterium: Geschlecht .....                           | 190 |
| <i>Abbildung 7.19:</i><br>Graphischer Modelltest für die figurale Skala, Split-Kriterium: Alter .....                                | 191 |
| <i>Abbildung 7.20:</i><br>Graphischer Modelltest für die figurale Skala,<br>Split-Kriterium: Intelligenz (HI vs. NHI) .....          | 191 |



## Verzeichnis der Tabellen

|   |     |
|---|-----|
| <i>Tabelle 4.1:</i><br>Verschiedene Scoringverfahren im Vergleich anhand<br>eines Beispiel-Antwortmusters .....   | 65  |
| <i>Tabelle 4.2:</i><br>Bezeichnungen und gemessene Fähigkeiten der Untertests des S-CPT .....   | 75  |
| <i>Tabelle 6.1:</i><br>Struktur der Skala „Reading Span“ .....  | 102 |
| <i>Tabelle 6.2:</i><br>Struktur der Skala „Verbal Span“ .....   | 104 |
| <i>Tabelle 6.3:</i><br>Struktur der Skala „Computation Span“ .....  | 105 |
| <i>Tabelle 6.4:</i><br>Struktur der Skala „Backward Digit Span“ .....   | 106 |
| <i>Tabelle 6.5:</i><br>Struktur der Skala „Memory Updating Figural“ .....   | 108 |
| <i>Tabelle 6.6:</i><br>Struktur der Skala „Spatial Working Memory“ .....  | 110 |
| <i>Tabelle 6.7:</i><br>Anzahl der Pbn pro Altersgruppe (8;6 bis 13;11 Jahre) und der Anteil von<br>Mädchen und Jungen in den Substichproben sowie in der Gesamtstichprobe ..... | 119 |
| <i>Tabelle 6.8:</i><br>Anzahl der Pbn pro Klassenstufe nach Substichproben .....  | 119 |
| <i>Tabelle 6.9:</i><br>Mittelwerte und Standardabweichungen der IQ-Werte der Intelligenztests<br>für die Substichproben und die Gesamtstichprobe .....                          | 120 |
| <i>Tabelle 6.10:</i><br>Mittelwerte und Standardabweichungen der IQ-Werte des AG-Index<br>und des ZVT für die Substichproben und die Gesamtstichprobe .....                     | 120 |

## Tabellenverzeichnis

---

|   |     |
|---|-----|
| <i>Tabelle 7.1:</i><br>Mittelwerte und Standardabweichungen der Schulnoten<br>(Primar- und Sekundarstufe).....  | 127 |
| <i>Tabelle 7.2:</i><br>Mittelwerte, Standardabweichungen und Verteilungen der kognitiven<br>Tests für die Gesamtstichprobe und die Teilstichproben.....   | 128 |
| <i>Tabelle 7.3:</i><br>Mittelwerte, Standardabweichungen, Verteilungen und Reliabilität<br>der Skalenscores der Arbeitsgedächtnisaufgaben in der Gesamtstichprobe<br>und den Teilstichproben.....     | 129 |
| <i>Tabelle 7.4:</i><br>Kennwerte für die Aufgabenscores, die auf Itemebene mit<br>Einheitsgewichtung gebildet wurden (Gesamtstichprobe, $N = 374$ ) .....   | 131 |
| <i>Tabelle 7.5:</i><br>Korrelationen der Gesamtscores aus den Arbeitsgedächtnisaufgaben<br>(Korrelationen und Partialkorrelationen, Alter kontrolliert).....  | 133 |
| <i>Tabelle 7.6:</i><br>Modellpassung eines Ein-, Zwei- und Drei-Faktor-Modells des<br>Arbeitsgedächtnisses und Modellvergleich .....  | 138 |
| <i>Tabelle 7.7:</i><br>Passung verschiedener Faktormodelle des Arbeitsgedächtnisses<br>für jüngere und ältere Pbn.....  | 140 |
| <i>Tabelle 7.8:</i><br>Passung verschiedener Faktormodelle des Arbeitsgedächtnisses für<br>die Teilstichproben „Hoch Intelligente“ ( $N = 170$ ) und<br>„Nicht hoch Intelligente“ ( $N = 204$ ) ..... | 142 |
| <i>Tabelle 7.9:</i><br>Passung des Zwei-Faktoren-Modells des Arbeitsgedächtnisses<br>für Jungen ( $N = 205$ ) und Mädchen ( $N = 168$ ) .....   | 143 |
| <i>Tabelle 7.10:</i><br>Vergleich der Mittelwerte der Arbeitsgedächtnisaufgaben<br>von jüngeren ( $N = 196$ ) und älteren Pbn ( $N = 177$ ) .....   | 145 |

## Tabellenverzeichnis

---

|  |     |
|--|-----|
| <i>Tabelle 7.11:</i><br>Vergleich der Mittelwerte der Arbeitsgedächtnisaufgaben der Teilstichproben „hoch Intelligente“ (HI; $N = 170$ ) und „nicht hoch Intelligente“ (NHI; $N = 204$ ) .....   | 146 |
| <i>Tabelle 7.12:</i><br>Vergleich der Mittelwerte in den Arbeitsgedächtnis-Scores von Mädchen ( $N = 168$ ) und Jungen ( $N = 205$ ) in der Gesamtstichprobe .....   | 147 |
| <i>Tabelle 7.13:</i><br>Vergleich der Mittelwerte in den Arbeitsgedächtnis-Scores von Mädchen ( $N = 84$ ) und Jungen ( $N = 88$ ) in der Teilstichprobe „Unausgelesene Pbn“ .....   | 147 |
| <i>Tabelle 7.14:</i><br>Partialkorrelationen der Arbeitsgedächtnisaufgaben mit den Intelligenztests (Rohwerte, Alter kontrolliert) für die Gesamtstichprobe ( $N = 374$ ) und die Teilstichprobe „hoch Intelligente“ ( $N = 170$ ) .....                               | 150 |
| <i>Tabelle 7.15:</i><br>Partialkorrelationen der Arbeitsgedächtnisaufgaben mit Speed-Aufgaben (Alter kontrolliert).....  | 153 |
| <i>Tabelle 7.16:</i><br>Mittelwerte, Standardabweichungen, Range der Rohwerte, Verteilungen und Retest-Reliabilitäten der Aufgaben Hick-0 bis Hick-3 .....   | 154 |
| <i>Tabelle 7.17:</i><br>(Partial-)Korrelationen der Hick-Aufgaben (0 bis 3) untereinander und Partialkorrelationen der Hick-Aufgaben mit den standardisierten Speed-Tests (Rohwerte, Alter kontrolliert) .....   | 155 |
| <i>Tabelle 7.18:</i><br>Vergleich der Mittelwerte der Hick-Aufgaben der Teilstichproben „hoch Intelligente“ (HI; $N = 170$ ) und „nicht hoch Intelligente“ (NHI; $N = 203$ ) .....   | 155 |
| <i>Tabelle 7.19:</i><br>Korrelationen der Faktoren des Gesamtmodells .....   | 157 |
| <i>Tabelle 7.20:</i><br>Korrelationen der Arbeitsgedächtnisaufgaben mit den Schulnoten in den Hauptfächern und dem Notenschnitt sowie die standardisierten Beta-Gewichte der Arbeitsgedächtnis-Scores für die einzelnen Noten und den Notenschnitt (Primarstufe) ..... | 162 |

## Tabellenverzeichnis

---

|  |     |
|--|-----|
| <i>Tabelle 7.21:</i><br>Korrelationen der Arbeitsgedächtnisaufgaben mit den Schulnoten in den<br>Hauptfächern und dem Notenschnitt, die standardisierten Beta-Gewichte<br>der Arbeitsgedächtnis-Scores für die einzelnen Noten und den Notenschnitt<br>sowie die semipartiellen Korrelationen (Sekundarstufe)..... | 164 |
| <i>Tabelle 7.22:</i><br>Faktorladungen der 70 Arbeitsgedächtnisaufgaben auf zwei Faktoren<br>(Hauptachsenanalyse, Rotation: oblimin; Mustermatrix).....  | 170 |
| <i>Tabelle 7.23:</i><br>Klassische Schwierigkeiten und Trennschärfen der Aufgaben der<br>verbal-numerischen Skala.....   | 172 |
| <i>Tabelle 7.24:</i><br>Aufgabenparameter und Fit der verbalen und numerischen<br>Aufgaben für das 1PL- und 2PL-Modell .....   | 174 |
| <i>Tabelle 7.25:</i><br>Klassische Schwierigkeiten und Trennschärfen der figuralen Skala.....  | 178 |
| <i>Tabelle 7.26:</i><br>Aufgabenparameter und Fit der figuralen Aufgaben für das<br>1PL- und 2PL-Modell.....   | 179 |
| <i>Tabelle 7.27:</i><br>Modellvergleiche für die zwei Arbeitsgedächtnis-Skalen anhand<br>der Log-Likelihood und informationstheoretischer Maße.....  | 182 |
| <i>Tabelle 7.28:</i><br>Likelihoodquotiententests nach Andersen mit verschiedenen<br>Teilungskriterien für die beiden Skalen.....  | 192 |
| <i>Tabelle 7.29:</i><br>Modelltests nach Martin-Löf zur Prüfung der Aufgabenhomogenität .....  | 194 |
| <i>Tabelle 7.30:</i><br>Partialkorrelationen der Scores der verbal-numerischen und der figuralen<br>Skala mit den Intelligenztests und Merkfähigkeitsaufgaben (Rohwerte,<br>Alter kontrolliert) und bivariate Korrelationen der Skalen mit den<br>Schulnoten in den Kernfächern.....                               | 196 |

# 1. Einleitung

Diese Arbeit beschäftigt sich mit der Erfassung der Arbeitsgedächtniskapazität von durchschnittlich und hoch intelligenten Kindern. Dabei werden im Wesentlichen zwei Ziele verfolgt: Das erste Ziel bezieht sich auf die Diagnostik hohen intellektuellen Potentials im Kindesalter. Bei der Diagnostik sehr hoch ausgeprägter Fähigkeiten ergeben sich bestimmte besondere Schwierigkeiten, die bei der Messung von Begabungen im durchschnittlichen Bereich nicht bestehen. Diese sind vor allem darauf zurückzuführen, dass Intelligenztests in aller Regel für Messungen im mittleren Begabungsbereich konstruiert wurden und daher bei Probanden mit sehr hohen Fähigkeiten nur eine eingeschränkte Messgüte aufweisen. Bisher verfügbare Intelligenztests sind somit nur bedingt für die Messung hoher und sehr hoher Intelligenz geeignet.

Ein Anliegen dieser Arbeit besteht daher darin, die Grundlagen für die Entwicklung eines neuen Testverfahrens bereitzustellen, das auch für die Begabungsdiagnostik im höheren Fähigkeitsbereich verwendet werden kann. Bestimmte Schwächen, die bisher verfügbare Testverfahren aufweisen, sollen dabei vermieden werden. So soll beispielsweise ein angemessenes Schwierigkeitsniveau Deckeneffekte verhindern und aussagekräftige Messungen auch im hohen Begabungsbereich erlauben. Angestrebt wird weiterhin, dass eine Potentialdiagnose im Vordergrund steht, und auf eher fluide Intelligenzkomponenten fokussiert wird. Die Begabung sollte möglichst unabhängig von (schulischem) Wissen erfasst werden.

Für die Entwicklung der Testaufgaben wurde mit dem Arbeitsgedächtnis auf ein Konzept zurückgegriffen, das aus der experimentellen Denk- und Gedächtnispsychologie stammt und das in der Differentiellen Psychologie bereits ausführlich untersucht wurde, in der Individualdiagnostik jedoch bisher kaum Verwendung findet. Typische Arbeitsgedächtnisaufgaben stellen stets eine doppelte Anforderung an den Probanden: Es müssen bestimmte Informationen gespeichert und gleichzeitig in bestimmter Weise aktiv kognitiv verarbeitet werden. Entsprechende mentale Prozesse werden heute als grundlegend für die Informationsverarbeitung und somit auch für viele komplexe kognitive Leistungen angesehen. Alltagsbeispiele für diese Anforderung sind etwa das Behalten einer Adresse, während man den Weg dorthin erklärt bekommt, oder das Behalten von Sätzen einer Geschichte, die erzählt wird, während man versucht, den Sinn der ganzen Geschichte zu verstehen (Swanson, 1996). Solche Leistungen des Arbeitsgedächtnisses unterscheiden sich daher deutlich von typischen Leistungen des Kurzzeitgedächtnisses, die lediglich aus einer Speicherung von Informationen bestehen, ohne dass diese oder andere Informationen gleichzeitig verarbeitet werden müssen.

Bereits seit den ersten Untersuchungen individueller Unterschiede in diesen Aufgaben ist ihre Eigenschaft bekannt, die Ausprägung höherer kognitiver Leistungen wie

Lesen oder Rechnen (z. B. Daneman & Carpenter, 1980; Turner & Engle, 1989) oder schulische Leistungen (z. B. Gathercole & Pickering, 2000a, 2000b) gut vorhersagen zu können. Auch konnten enge Zusammenhänge zu den kognitiven Leistungen in Intelligenztests festgestellt werden, so dass heute viele Forscher davon ausgehen, dass die Kapazität des Arbeitsgedächtnisses einer Person Grundlage und begrenzender Faktor ihrer intellektuellen Fähigkeiten ist. Weiter gehende Ansätze nehmen an, dass die Kapazität des Arbeitsgedächtnisses in hohem Maße dem entspricht, was in der Intelligenzforschung als allgemeine Intelligenz oder  $g$  (Spearman, 1927) bezeichnet wird (z. B. Kyllonen, 1996; Swanson, 1996). Der Vorteil in der Verwendung des Konstrukts des Arbeitsgedächtnisses wird unter anderem darin gesehen, dass es nicht vorwiegend auf den Ergebnissen von Faktorenanalysen beruht, wie es bei Intelligenzfaktoren der Fall ist, sondern auf theoretischer Grundlage und auf der Basis experimenteller Befunde entwickelt wurde (Kyllonen, 1996). Neben ihrer kognitionspsychologischen Fundierung weisen Arbeitsgedächtnisaufgaben verschiedene weitere Vorteile auf, die sie für das hier verfolgte Vorhaben besonders günstig erscheinen lassen. So erfordern sie kaum Vorwissen, so dass sie weitgehend fluide Fähigkeiten erfassen. Zudem haben sie eine einfache Struktur, die es erlaubt, die Schwierigkeiten von Aufgaben relativ leicht zu verändern und sie an das Fähigkeitsniveau verschiedener Gruppen anzupassen.

Trotz vielfältiger Untersuchungen zu individuellen Unterschieden des Arbeitsgedächtnisses sind verschiedene Aspekte des Konstrukts bisher nicht eindeutig geklärt. Das zweite Ziel dieser Studie ist daher, hierzu weitere Erkenntnisse beizusteuern. Zum einen besteht kein Konsens über die Binnenstruktur des Arbeitsgedächtnisses. Ebenso wie in der Intelligenzstrukturforschung ist umstritten, ob sich Arbeitsgedächtnisleistungen durch eine gemeinsame latente Dimension erklären lassen, oder ob hierfür mehrere, miteinander korrelierende Faktoren notwendig sind. Zum anderen sind die Befunde zum Zusammenhangsmuster von Arbeitsgedächtnis und anderen kognitiven Variablen wie fluide Intelligenz, Kurzzeitgedächtnis und Informationsverarbeitungsgeschwindigkeit bisher nicht eindeutig. Zwar ist bekannt, dass das Arbeitsgedächtnis mit diesen Variablen bedeutsam korreliert, die Angaben zu den Ausprägungen dieser Zusammenhänge fallen jedoch sehr unterschiedlich aus, so dass hierzu bisher keine übereinstimmenden Aussagen vorliegen. Die Faktorenstruktur des Arbeitsgedächtnisses wurde zwar bereits an verschiedenen Probandengruppen wie Kindern, jungen und älteren Erwachsenen oder Lernbehinderten untersucht (vgl. z. B. Miyake, 2001), bislang fehlt jedoch eine Überprüfung an intellektuell hoch Begabten.

Zusammenfassend bestehen die Ziele dieser Arbeit darin, die Grundlagen für einen neuartigen, computergestützten Test zur Erfassung durchschnittlicher und hoher fluider kognitiver Fähigkeiten zu entwickeln. Als neuer Aufgabentyp werden hierzu Arbeitsgedächtnisaufgaben verwendet und erprobt. Weiterhin werden die Faktorenstruktur des Arbeitsgedächtnisses sowie die Zusammenhänge des Arbeitsgedächtnisses mit anderen kognitiven Variablen bei Kindern mit durchschnittlicher und hoher Intelligenz untersucht.

Die vorliegende Arbeit gliedert sich wie folgt: Im zweiten Kapitel, das sich an diese Einleitung anschließt, werden zunächst die Besonderheiten und einige Schwierigkeiten bei der Intelligenzdiagnostik im hohen Fähigkeitsbereich dargestellt. Danach werden bisher existierende Testverfahren für die Hochbegabungsdiagnostik kurz vorgestellt, und es werden die theoretischen und methodischen Grundlagen der hier verfolgten Testentwicklung beschrieben.

Die Kapitel 3 und 4 beschäftigen sich mit dem Arbeitsgedächtnis. In Kapitel 3 werden die theoretischen Grundlagen des Arbeitsgedächtniskonstrukts dargestellt, wobei insbesondere auf Theorien und empirische Befunde aus der differentiellen Psychologie eingegangen wird. Kapitel 4 widmet sich der Messung der Arbeitsgedächtniskapazität. Hierzu werden prototypische Arbeitsgedächtnisaufgaben aus der Forschung und standardisierte Testverfahren zur Erfassung der Arbeitsgedächtniskapazität vorgestellt. Auch spezifische Fragestellungen, die bei der Messung eine Rolle spielen, wie die Berechnung von Skalenscores aus Arbeitsgedächtnisaufgaben und die Messgüte häufig eingesetzter Aufgabentypen werden thematisiert.

Im ersten Teil von Kapitel 5 wird die Verwendung von Arbeitsgedächtnisaufgaben für die Entwicklung eines Tests zur Erfassung hoher intellektueller Begabungen auf der Basis der vorangehenden Ausführungen zum Forschungsstand begründet. Im zweiten Teil des Kapitels werden verschiedene Fragestellungen abgeleitet, die sich aus dem in den Kapiteln 3 und 4 referierten Forschungsstand ergeben. Zunächst werden verschiedene Fragestellungen zum Konstrukt des Arbeitsgedächtnisses formuliert, die in der vorliegenden Studie untersucht werden. Anschließend wird dargestellt, welche Fragestellungen hinsichtlich der Eignung der ausgewählten computergestützten Arbeitsgedächtnisaufgaben für die Begabungsdiagnostik untersucht werden.

In den Kapiteln 6 und 7 wird die empirische Untersuchung mit 374 durchschnittlich und hoch begabten Kindern dargestellt. In Kapitel 6 werden die verwendeten Arbeitsgedächtnisaufgaben und alle zusätzlich eingesetzten Testverfahren und kognitiven Aufgaben, die untersuchten Stichproben und die Durchführung der Studie beschrieben. Kapitel 7 behandelt zunächst die Analysen zur Faktorstruktur des Arbeitsgedächtnisses in der Gesamtstichprobe sowie in den verschiedenen untersuchten Alters- und Begabungsgruppen. Weiterhin werden die geschätzten Zusammenhänge der Arbeitsgedächtniskapazität mit anderen kognitiven Fähigkeiten wie fluider und kristalliner Intelligenz, Geschwindigkeit der Informationsverarbeitung und Kurzzeitgedächtnis auf der Konstruktebene berichtet. Die Ergebnisse der testtheoretischen Analysen entsprechend der Klassischen Testtheorie und der Item Response Theorie folgen im zweiten Teil des siebten Kapitels. In Kapitel 8 schließlich wird die vorliegende Arbeit zusammengefasst und die empirischen Befunde werden zusammenfassend diskutiert.

## **2. Diagnostik intellektueller Hochbegabung**

Bei der psychologischen Diagnostik von hoher intellektueller Begabung ergeben sich bestimmte Schwierigkeiten, die in diesem Kapitel erläutert werden sollen. Bisher verfügbare Testverfahren sind aus verschiedenen Gründen, auf die im Folgenden näher eingegangen wird, nur bedingt für diese Aufgabe geeignet. Es stellt sich daher die Frage, wie Testverfahren so konstruiert werden können, dass sie dieser Aufgabe besser gerecht werden. In der vorliegenden Arbeit wird ein neuer Ansatz erprobt, intellektuelle Begabungen über eine Messung der Arbeitsgedächtniskapazität zu erfassen. Bevor jedoch auf das Konzept des Arbeitsgedächtnisses und Möglichkeiten der Messung der Arbeitsgedächtniskapazität näher eingegangen wird (siehe Kapitel 3 und 4), soll in diesem Kapitel zunächst die derzeitige Situation der (Hoch-) Begabungsdagnostik skizziert werden.

Hierzu wird in Abschnitt 2.1 ein kurzer Überblick über das Konzept „intellektuelle Hochbegabung“ gegeben. Anschließend werden zwei aktuelle theoretisch fundierte und empirisch entwickelte Intelligenzmodelle vorgestellt, die den gegenwärtigen Stand der Intelligenzstrukturforschung widerspiegeln und die als Grundlage für die Entwicklung neuer Testverfahren herangezogen werden können (Abschnitt 2.2). Es folgt in Abschnitt 2.3 eine kurze Beschreibung von möglichen Anlässen für eine genaue Intelligenzdiagnostik im hohen Fähigkeitsbereich. Anschließend werden verschiedene Probleme erläutert, die bei der Testung hoher und sehr hoher intellektueller Fähigkeiten bestehen (Abschnitt 2.4). Abschnitt 2.5 widmet sich der Darstellung bestehender spezieller Intelligenztests für die Messung hoher Intelligenz. Den Abschluss des Kapitels bilden in Abschnitt 2.6 Überlegungen zur Entwicklung eines neuen Testverfahrens zur Messung eines hohen intellektuellen Potentials, wobei insbesondere die Verwendung von Arbeitsgedächtnisaufgaben diskutiert wird.

### **2.1 Modellvorstellungen zu intellektueller Hochbegabung**

Bisher gibt es keine allgemein akzeptierte generelle Definition von Hochbegabung, stattdessen existieren verschiedene theoretische Hochbegabungsmodelle parallel (Feldhusen & Jarwan, 2000). Die aktuellen Modelle unterscheiden sich vor allem darin, ob ihnen ein fähigkeits- oder ein leistungsorientiertes Verständnis von Hochbegabung zu Grunde liegt (Hany, 1987; Holling & Kanning, 1999). Diese Unterscheidung soll im Folgenden erläutert werden.

Fähigkeitsbasierte Modellvorstellungen reichen bis in die Anfänge der empirischen Hochbegabungsforschung zurück und haben auch heute noch einen zentralen Stellenwert. Bereits in der groß angelegten und einflussreichen Langzeitstudie von Lewis Terman und Mitarbeitern, die 1921 in den USA begonnen wurde, wurden die als hoch begabt klassifizierten Probanden (Pbn) auf der Grundlage ihrer hohen Intelli-



genz ausgewählt. In der Folgezeit wurde intellektuelle Hochbegabung vorwiegend über eine hohe allgemeine Intelligenz im Sinne von Spearman's  $g$ -Faktor (Spearman, 1927) definiert und diagnostiziert. Seit den 60er Jahren wurden mehrdimensionale Modelle und Definitionen entwickelt, die neben der allgemeinen Intelligenz verschiedene andere Fähigkeiten, insbesondere Kreativität, aber auch künstlerische, motorische und andere Fähigkeiten als Grundlage von Hochbegabung annahmen (z. B. Marland, 1972; Taylor, 1978). Ausgangspunkt war die Hoffnung, bereichsspezifische Leistungen besser durch spezifische als durch generelle Fähigkeiten vorherzusagen zu können (Hany, 1987). Ein Beispiel für ein aktuelles mehrdimensionales und fähigkeitsbasiertes Modell stellt das Münchner Hochbegabungsmodell nach Heller, Perleth und Hany (1994) dar. In diesem Modell werden verschiedene Begabungsfaktoren wie intellektuelle Fähigkeiten, Kreativität und soziale Kompetenz angenommen, aus denen sich hohe Leistungen in verschiedenen Bereichen (z. B. Naturwissenschaften, Mathematik oder soziale Beziehungen) entwickeln können. Voraussetzungen für die Umwandlung der Begabungsfaktoren in hohe Leistungen sind jedoch eine positive Ausprägung bestimmter nicht-kognitiver Persönlichkeitsmerkmale (z. B. Leistungsmotivation) sowie das Vorliegen günstiger Umweltbedingungen (z. B. ein förderliches Familienklima).

Den fähigkeitsbasierten Modellen wurden von verschiedener Seite leistungsorientierte Modelle entgegengestellt. Ein erstes und einflussreiches Modell dieser Art ist das „Drei-Ringe-Modell“ von Renzulli (1979). In diesem Modell wird angenommen, dass Begabung als die Schnittmenge der drei Personenmerkmale „überdurchschnittliche Fähigkeiten“ (in verschiedenen kognitiven Bereichen), „Kreativität“ und „Aufgabenverpflichtung“ verstanden werden kann. Das Modell wurde später von verschiedenen Forschern aufgegriffen und weiterentwickelt (z. B. Mönks, 1990; Wiczerkowski & Wagner, 1985). Der Ansatzpunkt dieser Modelle war die Überlegung, dass hohe intellektuelle Fähigkeiten für die Entwicklung außergewöhnlicher und kreativer Leistungen nicht ausreichend sind, und daher weitere Personenmerkmale stärker in den Mittelpunkt gerückt werden sollten.

Zusammenfassend kann festgestellt werden, dass alle aktuellen Modelle davon ausgehen, dass für die Leistungsentwicklung das Vorhandensein eines hohen intellektuellen Potenzials zwar eine notwendige, aber keine hinreichende Bedingung ist. Unterschiede in den Modellen bestehen vor allem darin, ob sich die Hochbegabungsdiagnose allein auf die Disposition bezieht, oder ob zumindest zum Teil auch bereits gezeigte hohe Leistungen vorausgesetzt werden.

In allen Modellen spielt eine (mindestens) überdurchschnittlich hohe Intelligenz eine zentrale Rolle. In der Praxis der Hochbegabungsdiagnostik hat daher die Intelligenzdiagnostik in der Regel einen zentralen Stellenwert. Da es keine verbindliche Definition von intellektueller Hochbegabung gibt, wird bisher meist pragmatisch vorgegangen. Wie bei jeder psychologischen Diagnostik sollte es von der diagnostischen Fragestellung abhängen, ob neben der allgemeinen Intelligenz oder spezifischen intellektuellen Fähigkeiten weitere Personenmerkmale gemessen werden. Für die Entscheidung über die Aufnahme in bestimmte Förderprogramme etwa kann es sinnvoll sein, neben der Intelligenz die aktuelle Leistungsfähigkeit, die Motivation

etc. zu diagnostizieren, um eine erfolgreiche Teilnahme im Programm so gut wie möglich vorherzusagen. Für andere Zwecke, insbesondere die wissenschaftliche Forschung, kann häufig eine Fokussierung auf die allgemeine Intelligenz bzw. spezifische intellektuelle Fähigkeiten angemessen sein. Meist wird das Vorliegen einer deutlich überdurchschnittlichen Intelligenz als Bedingung für Hochbegabung gesetzt, häufig wird hier als Kriterium ein  $IQ \geq 130$  gewählt (siehe z. B. Rost, 2000). Ein IQ von 130 entspricht einer Begabung, die zwei Standardabweichungen über dem Mittel liegt, und wird definitionsgemäß von ca. 2% der Bevölkerung erreicht oder übertroffen. Es handelt sich bei diesem Wert um eine reine Konvention. Dieser Wert wurde festgelegt, damit das Phänomen Hochbegabung klarer fassbar und auch wissenschaftlich besser erforschbar ist. Intelligenz ist jedoch ein kontinuierlich verteiltes Merkmal. Bisher gibt es keine Belege dafür, dass sich oberhalb eines IQ von 130 ein qualitativer Sprung in der intellektuellen Fähigkeit ereignet, der sich z. B. in qualitativ anderen Denkprozessen oder Fähigkeitskonstrukten niederschlägt. Ebenso gut könnten daher auch die 5% (entspricht etwa einem  $IQ \geq 125$ ) oder 10% (entspricht etwa einem  $IQ \geq 120$ ) intelligentesten Menschen der Bevölkerung als „in hohem Maße“ intellektuell begabt angesehen werden.

## 2.2 Intelligenzmodelle

Die Grundlage von Intelligenztests bildet stets eine – mehr oder weniger explizite – Modellvorstellung von Intelligenz. Wesentlich ist dabei das jeweilige Konzept der Struktur der zu erfassenden intellektuellen Fähigkeiten. Einige Tests messen relativ spezifische kognitive Fähigkeiten, die zudem eine unterschiedliche Relevanz für die allgemeine Intelligenz aufweisen (z. B. Wortschatztests oder figurale Matrizentests), andere erfassen eine größere Bandbreite verschiedener Fähigkeiten (Intelligenzstrukturtests wie der BIS-HB; Jäger et al., 2005, s. u.). Für die Interpretation gemessener Werte spielt das dem Test zugrunde liegende Modell eine zentrale Rolle.

Über Strukturanalyseverfahren wie die Faktoren- oder Clusteranalyse oder die Multidimensionale Skalierung wurden in der Vergangenheit verschiedene Intelligenzmodelle entwickelt, die die Struktur kognitiver Fähigkeiten beschreiben. Klassische einflussreiche Modelle der Intelligenzstruktur sind etwa die Generalfaktorentheorie von Spearman (1904), das Modell mehrerer gemeinsamer Faktoren von Thurstone (1938) oder das Radex-Modell von Marshalek, Lohman und Snow (1983).

Verschiedene Modelle unterscheiden sich in der Anzahl und im Inhalt der angenommenen Komponenten der Intelligenzstruktur, es konnten sich jedoch über die Zeit zwei Eigenschaften der Intelligenzstruktur als relativ robust herauskristallisieren, die in den aktuellen Intelligenzmodellen zum Tragen kommen (Kyllonen, 1996). Die erste Eigenschaft ist, dass intellektuelle Fähigkeiten hierarchisch angeordnet sind. An der Spitze der Hierarchie intellektueller Fähigkeiten steht dabei die allgemeine Intelligenz (auch als „general factor“ oder kurz „g“ bezeichnet). Auf einer oder mehreren untergeordneten Hierarchieebene(n) werden spezifischere Fähigkeiten angeordnet, z. B. finden sich hier die verbalen Fähigkeiten oder die Geschwindigkeit der Informationsverarbeitung. Die Spezifizierung der untergeordneten Hierarchieebenen un-

terscheidet sich in verschiedenen Intelligenzmodellen, jedoch ähneln sich die angenommenen spezifischen Fähigkeiten in der Regel über verschiedene Modelle hinweg. Es wird angenommen, dass alle intellektuellen Leistungen jeweils durch die generelle Intelligenz und zusätzlich durch spezifische Fähigkeiten beeinflusst werden.

Ein zweiter robuster Befund aus der Intelligenzstrukturforschung besagt, dass sich im Bereich der spezifischen Faktoren drei Inhaltsklassen differenzieren lassen. Viele Studien zur Intelligenzstruktur finden je einen verbalen, einen numerischen und einen räumlichen Faktor auf einer untergeordneten Ebene.

Im Folgenden werden zwei aktuelle hierarchische Intelligenzmodelle kurz vorgestellt, die diese beiden Eigenschaften aufweisen: Die *Three Stratum Theory* von Carroll (1993) und das *Berliner Intelligenzstrukturmodell* (BIS) von Jäger (1982). Beide Modelle wurden sowohl theoriegeleitet als auch empirisch entwickelt und stellen relativ umfassende Beschreibungen der Struktur menschlicher kognitiver Fähigkeiten dar.

#### *Three Stratum Theory*

Carroll (1993) entwickelte die Three Stratum Theory aus einer umfangreichen Reanalyse von Datensätzen aus der Intelligenzforschung. Insgesamt reanalysierte er 461 Datensätze mit Daten aus sehr unterschiedlichen kognitiven Tests und Intelligenztestaufgabentypen. Carrolls Faktorenanalysen führten zu einem Strukturmodell der Intelligenz mit insgesamt drei Hierarchieebenen (oder „Strata“), die jeweils verschiedene Fähigkeiten beinhalten. Die höchste und generellste Ebene des Modells bildet die allgemeine Intelligenz. Die Ausprägung der allgemeinen Intelligenz wird bestimmt durch acht breite Faktoren mittlerer Generalität, die sich aus insgesamt 69 relativ spezifischen Fähigkeiten auf der untersten Modellebene ergeben.

Die acht relativ globalen Fähigkeiten auf der mittleren Ebene des Modells lassen sich zusammenfassend wie folgt beschreiben:

1. *Fluide Intelligenz*: Dieser Faktor umfasst grundlegende kognitive Prozesse wie quantitatives und sequentielles Reasoning und induktives Denken. Es handelt sich um Prozesse, von denen angenommen wird, dass sie von Lernprozessen und kulturellen Einflüssen weitgehend unabhängig sind.
2. *Kristalline Intelligenz*: Unter kristallinen Fähigkeiten werden geistige Prozesse wie Lesefähigkeit, Sprachverständnis etc. verstanden, die nicht nur von der fluiden Intelligenz, sondern auch von Erfahrungen, Lernprozessen und kulturellen Einflüssen abhängig sind.
3. *Allgemeine Gedächtnisfähigkeit*: Diese Fähigkeit beschreibt Fähigkeiten zum Lernen und Behalten neuer Informationen oder neuen Verhaltens.
4. *Visuelle Wahrnehmung*: Die Fähigkeit zur visuellen Wahrnehmung ist relevant für solche kognitiven Prozesse, in denen die Wahrnehmung visueller Formen, z. B. die Erkennung räumlicher Beziehungen, gefordert ist.

5. *Auditive Wahrnehmung*: Die Fähigkeit zur auditiven Wahrnehmung bezieht sich auf die Wahrnehmung oder Unterscheidung auditiver Klangmuster oder gesprochener Sprache.
6. *Abruffähigkeit*: Wenn Informationen oder Ideen aus dem Langzeitgedächtnis abgerufen werden müssen, ist die Retrieval- oder Abruffähigkeit betroffen. Relevant ist diese Fähigkeit z. B. für die Entwicklung vielfältiger oder origineller Ideen, wie es für kreative Prozesse erforderlich ist.
7. *Kognitive Geschwindigkeit*: Die kognitive Geschwindigkeit bezieht sich auf die Fähigkeit, komplexere Informationen schnell zu verarbeiten, z. B. Intelligenztestaufgaben schnell lösen zu können.
8. *Verarbeitungsgeschwindigkeit*: Die Verarbeitungsgeschwindigkeit bezieht sich auf die Fähigkeit, relativ simple Informationen, wie sie in einfachen oder Wahlreaktionszeitaufgaben enthalten sind, schnell zu verarbeiten.

Es wird dabei angenommen, dass die allgemeine Intelligenz den stärksten Einfluss auf die fluide Intelligenz ausübt, die kristalline Intelligenz folgt an zweiter Stelle. Weitere Studien konnten das von Carroll entwickelte Modell mittels konfirmatorischer Faktorenanalysen erneut empirisch stützen (z. B. Bickley, Keith & Wolfle, 1995, für eine Variante des Originalmodells). Es zeigte sich auch, dass insbesondere die Faktoren „fluide Intelligenz“ und „kristalline Intelligenz“ hoch auf der allgemeinen Intelligenz laden (Ladungen von .88 und .87), jedoch können diese Fähigkeiten nicht mit der allgemeinen Intelligenz gleichgesetzt werden. Eine Setzung dieser Faktorladungen auf einen Wert von 1 führte zu einer signifikant schlechteren Modellpassung (Bickley et al., 1995). Die hierarchische Gliederung in eine übergeordnete allgemeine Intelligenz und untergeordnete spezifischere Faktoren erscheint somit notwendig.

### *Berliner Intelligenzstrukturmodell (BIS)*

Ein weiteres, überwiegend empirisch entwickeltes und ebenfalls hierarchisch strukturiertes Modell ist das Berliner Intelligenzstrukturmodell (BIS) von Jäger (1982, siehe Abbildung 2.1). Das BIS basiert auf einer Reanalyse einer großen Zahl in der Literatur auffindbarer Aufgabentypen zur Intelligenzmessung. Das Modell geht davon aus, dass jede intelligente Leistung – in unterschiedlicher Gewichtung – durch sämtliche intellektuellen Fähigkeiten bedingt ist. Es werden im BIS (mindestens) zwei Modalitäten angenommen, um Intelligenzleistungen und intellektuelle Fähigkeiten zu klassifizieren. In der aktuellen Modellversion werden die beiden Modalitäten „Operationen“ und „Inhalte“ spezifiziert, insgesamt werden drei inhaltsgebundene und vier operative Fähigkeiten angenommen. Die vier operativen Fähigkeitskonstrukte des BIS werden als „Verarbeitungskapazität“, „Einfallsreichtum“, „Merkfähigkeit“ und „Bearbeitungsgeschwindigkeit“ bezeichnet. Unabhängig von diesen vier operativen Fähigkeiten werden drei verschiedene inhaltsgebundene Fähigkeiten spezifiziert. Diese beziehen sich auf das zahlengebundene Denken (Fähigkeit zum Umgang mit numerischem Material), das sprachgebundene Denken (Fähigkeit zum Umgang mit verbalem Material) sowie auf das anschauungsgebundene, figuralbildhafte Denken. Alle sieben BIS-Fähigkeiten sind miteinander korreliert; funktio-

nale oder kausale Beziehungen zwischen den einzelnen Fähigkeiten werden im BIS jedoch nicht spezifiziert.

Die Verarbeitungskapazität entspricht dem Konzept des „schlussfolgernden Denkens“, das im englischen Sprachraum auch als „Reasoning“ bezeichnet wird, und kann im Modell von Carroll am ehesten mit dem Faktor „fluide Intelligenz“ verglichen werden. Die Merkfähigkeit kann Carrolls Faktor „Allgemeine Gedächtnisfähigkeit“, die Bearbeitungsgeschwindigkeit den beiden Speed-Faktoren „Kognitive Geschwindigkeit“ und „Verarbeitungsgeschwindigkeit“ und der Einfallsreichtum der Retrievalfähigkeit zugeordnet werden. Der verbalen und der figuralen Fähigkeit im BIS entsprechen in der Three Stratum Theory am ehesten die Faktoren „Kristalline Intelligenz“ und „Visuelle Wahrnehmung“ (vgl. Süß, 2001).

Das BIS-Modell wurde bisher in verschiedenen Studien empirisch überprüft und konnte für verschiedene Probandengruppen (z. B. Schülerinnen und Schüler, Studierende, Berufstätige) als gültig bestätigt werden (Jäger, 1982). Für Jugendliche zwischen 12 und 16 Jahren konnte zudem gezeigt werden, dass das BIS-Modell für durchschnittlich und hoch begabte Pbn gleichermaßen gilt.

Sowohl das Modell von Carroll als auch das BIS-Modell nehmen somit verschiedene, relativ breite Fähigkeiten an, die durch die allgemeine Intelligenz beeinflusst werden. Ein wesentlicher Unterschied besteht darin, dass im BIS Operationen und Inhalte als zwei Facetten betrachtet werden, während im Modell von Carroll keine entsprechende Unterscheidung vorgenommen wird. Dennoch ähneln sich die in beiden Modellen angenommenen Fähigkeiten deutlich.

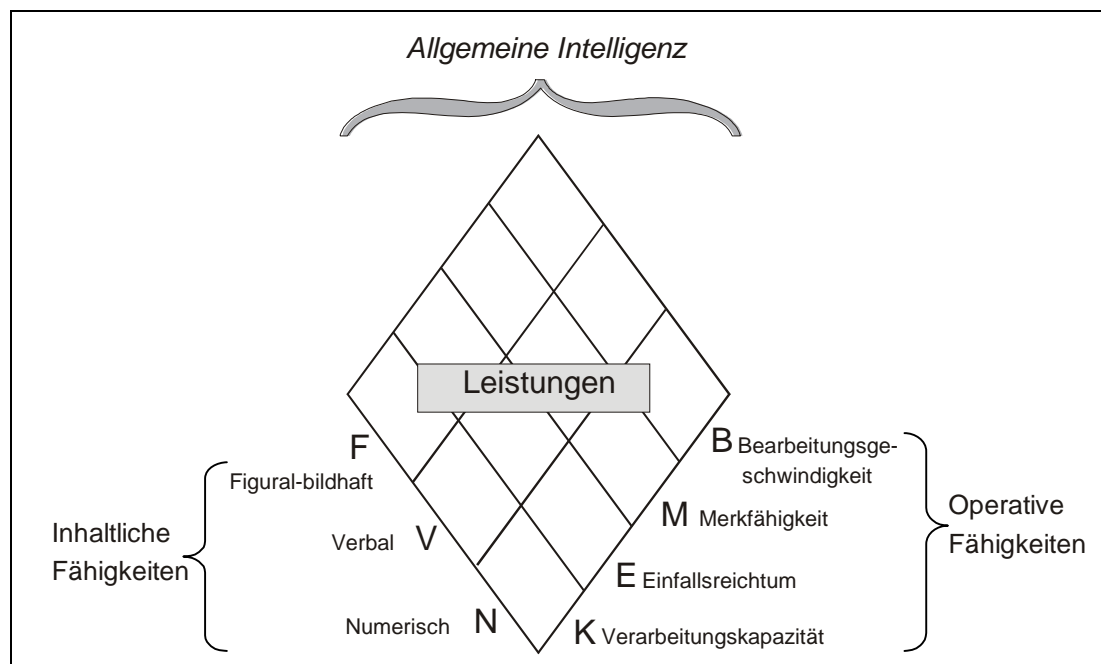


Abbildung 2.1: Berliner Intelligenzstrukturmodell (BIS) nach Jäger (1982)

Mit dem *Berliner Intelligenzstruktur-Test* (BIS-4; Jäger, Süß & Beauducel, 1997) und dem *Berliner Intelligenzstruktur-Test für Jugendliche: Begabungs- und Hochbegabungsdiagnostik* (BIS-HB; Jäger et al., 2005; siehe auch Abschnitt 2.5) liegen bereits zwei Testverfahren vor, die alle im BIS-Modell spezifizierten Fähigkeiten über insgesamt 42 Aufgabentypen erfassen (es gibt bisher keine Testbatterie, die alle von Carroll beschriebenen Fähigkeiten erfasst). Für Kinder unter zwölf Jahren existiert kein entsprechender Test, auch die Modellgeltung wurde für Kinder bisher nicht geprüft.

### 2.3 Diagnoseanlässe

Eine möglichst genaue und verlässliche Intelligenzdiagnose im Kindesalter bei einer vermuteten hohen Begabung kann in vielen Fällen notwendig oder wünschenswert sein. Ein zentraler Anwendungsbereich ist die Schaffung einer Entscheidungsgrundlage für die Aufnahme eines Kindes in ein bestimmtes Förderprogramm. Es kann davon ausgegangen werden, dass hohe intellektuelle Begabungen – ebenso wie z. B. musikalische oder sportliche hohe Begabungen – einer möglichst früh einsetzenden Förderung bedürfen, damit sich das Potenzial im weiteren Entwicklungsverlauf angemessen entfalten kann (Heller, 1992). Dies ist insbesondere relevant bei Kindern bestimmter Gruppen, die hinsichtlich ihrer Begabungsentfaltung erfahrungsgemäß bestimmten Risikofaktoren ausgesetzt sind. Risikofaktoren können z. B. ein benachteiligter sozioökonomischer familiärer Hintergrund, die Zugehörigkeit zu einer ethnischen Minderheit, eine Körperbehinderung oder das weibliche Geschlecht sein (Feger & Prado, 1998; Butler-Por, 1995).

Einen weiteren häufigen Anlass für Intelligenztestungen bei Kindern stellen Schullaufbahnentscheidungen dar. Die Erfassung der genauen Ausprägung der intellektuellen Begabung spielt hier eine wichtige Rolle, da an verschiedenen Punkten der Schullaufbahn weitreichende Entscheidungen getroffen werden. Zentral ist dabei im deutschen Schulsystem der Wechsel von der Grundschule auf die weiterführende Schule, der in den meisten Bundesländern nach der vierten Klasse erfolgt. Die fünfte und die sechste Klasse haben in der Regel den Status einer Erprobungsstufe, während der die Entscheidung für die Hauptschule, Realschule oder das Gymnasium überprüft wird. Insbesondere für hoch begabte Underachiever (s. u.) kann eine Begabungsdiagnose bei der Wahl einer angemessenen Schulform unterstützen. Eine Intelligenzdiagnostik ist daher in vielen Fällen insbesondere im Alter zwischen etwa 8 und 12-13 Jahren wichtig.

Auch eine individuelle Anpassung der schulischen Förderung an die Fähigkeiten von Schülerinnen oder Schülern erfordert häufig eine exakte Begabungsdiagnose (Holling, Vock & Preckel, 2001). Umfangreiche Evaluationsstudien konnten beispielsweise zeigen, dass die vorzeitige Einschulung und das Überspringen von Klassen Maßnahmen sind, die sich für intellektuell besonders begabte und entsprechend motivierte Kinder ausgesprochen positiv auswirken (z. B. Heinbokel, 1996). Werden sie jedoch bei einem weniger begabten Kind angewandt, können sich negative Effekte einstellen. Andererseits: Werden solche akzelerierenden Maßnahmen bei einem be-

gaben Kind unterlassen, so stellen sich häufig ebenfalls negative Auswirkungen ein (siehe z. B. Proctor, Black & Feldhusen, 1986; Heinbokel, 1996; Prado & Schiebel, 1996).

Verschärft stellt sich diese Situation bei den so genannten Underachievern oder Minderleistern dar. Es handelt sich hierbei nach der Definition von Hanses und Rost (1998) um Schülerinnen und Schüler, deren intellektuelle Fähigkeiten weit überdurchschnittlich sind, deren schulische Leistungen jedoch unterhalb des mittleren Leistungsniveaus ihrer Mitschüler liegen.

Rost (2000; Hanses & Rost, 1998) ermittelte in der großangelegten „Marburger Studie“ über hoch begabte Kinder einen Anteil von etwa 12% Underachievern unter hoch begabten Kindern. Die hohe intellektuelle Begabung dieser Schülergruppe ist für Lehrkräfte nur schwer zu erkennen. Auch die betroffenen Schülerinnen und Schüler selbst halten sich aufgrund des ständigen negativen Feedbacks aus der Umwelt oft für wenig befähigt. Diese negative Selbsteinschätzung kann die Leistungssituation zusätzlich verschlechtern sowie zu gravierenden psychischen Beeinträchtigungen führen. Um diesen Teufelskreis zu durchbrechen und dem Schüler oder der Schülerin eine angemessene Förderung zukommen lassen zu können, ist eine valide Intelligenzdiagnostik erforderlich.

#### **2.4 Probleme der Intelligenzmessung im Rahmen der Hochbegabungsdiagnostik**

Bei der Messung der Intelligenz im Rahmen einer Hochbegabungsdiagnostik bestehen jedoch verschiedene inhaltliche und methodische Probleme, die die Reliabilität und Validität der Diagnose gefährden. Eine grundlegende Schwierigkeit bei der Intelligenzmessung in diesem Zusammenhang ist das Fehlen spezifischer Aussagen zur Intelligenz in den Hochbegabungsmodellen. Die Modelle treffen in der Regel keine Aussagen darüber, welche Komponenten der Intelligenz relevant sind und über welche Verfahren diese gemessen werden sollten. Es besteht in der Hochbegabungsforschung keine Einigkeit darüber, welches Intelligenzmodell zugrunde gelegt werden soll.

In der Praxis der Hochbegabungsdiagnostik erfolgt die Messung der Intelligenz daher über alle gängigen Testverfahren, die auf jeweils unterschiedlichen Intelligenzmodellen basieren. Nach Stapf (1999) werden bei Schulkindern im Alter von sechs bis zwölf Jahren in Deutschland für eine Hochbegabungsdiagnostik am häufigsten Tests nach dem Wechsler-Konzept eingesetzt (HAWIK-R, Tewes, 1983; HAWIK III, Tewes et al., 2000; AID, Kubinger & Wurst, 2000), weniger häufig hingegen der Grundintelligenztest Skala 2 (CFT 20, Weiß, 1998) und Tests aus der Reihe der Kognitiven Fähigkeitstests (KFT 1-3, Heller & Geisler, 1983; KFT 4-13+, Heller, Gaedike & Weinläder, 1985). Es werden in der Praxis derzeit somit Profilverfahren (Wechsler-Tests) deutlich gegenüber solchen Tests bevorzugt, die nur eine Dimension erfassen (z. B. CFT 20).

Verschiedene Testverfahren erfassen unterschiedliche Dimensionen der Intelligenz oder dieselben Dimensionen in unterschiedlicher Qualität. Daraus resultieren deutliche Abweichungen in der Diagnose. Selbst Testverfahren, die ähnliche Konstrukte messen, korrelieren in der Regel nur moderat miteinander. Beispielsweise berichtet Weiß (1998) über eine Korrelation von  $r = .58$  bei Grundschulkindern zwischen dem CFT 20 und den Coloured Progressive Matrices (CPM, Raven, 1962b), die beide fluide Intelligenzkomponenten über figürliches Aufgabenmaterial erfassen. Die Ergebnisse aus Tests, die eher unterschiedliche Aspekte der Intelligenz messen und deutlich unterschiedliches Aufgabenmaterial verwenden, korrelieren tendenziell noch geringer. Kubinger und Wurst (2000) berichten für die Untertests des AID 2 Korrelationen im Bereich  $r = .03 - .49$  mit dem CFT 20 und im Bereich  $r = .00 - .64$  mit den Standard Progressive Matrices (SPM, Raven, 1962c).

Es kommt daher häufig vor, dass ein Kind nach dem Ergebnis in einem Intelligenztest als hoch begabt diagnostiziert wird, also einen bestimmten IQ-Grenzwert überschreitet, nach dem Ergebnis in einem zweiten Test jedoch nicht. Bei zwei Testergebnissen, die entsprechend der kritischen Differenzen signifikant unterschiedlich ausfallen, ist die Hochbegabungsdiagnose dann fraglich. Wenn die verwendeten Tests darüber hinaus nicht auf einem expliziten und empirisch überprüften Intelligenzmodell aufbauen, können solche abweichenden Befunde kaum inhaltlich interpretiert werden.

Bei der Messung hoher Intelligenz bestehen zudem verschiedene theoretische und messmethodische Probleme, die in den Abschnitten 2.4.1 und 2.4.2 näher erläutert werden sollen. Ein Grundproblem ist, dass Intelligenztests in aller Regel so konstruiert wurden, dass sie eine möglichst genaue Aussage über die Fähigkeit der meisten Menschen machen können. Sie erfassen daher vorwiegend die intellektuelle Leistungsfähigkeit im mittleren Bereich des Intelligenzspektrums. Bei Pbn, deren intellektuelle Fähigkeit um mehr als zwei Standardabweichungen vom Mittelwert abweicht, kann die Ausprägung der Intelligenz nicht mehr exakt festgestellt werden. Verschiedene in der psychologischen Praxis in Deutschland häufig eingesetzte Intelligenztestverfahren eignen sich somit auch nach den Angaben der Testautoren ausdrücklich nicht zur Messung der Intelligenz oberhalb eines IQ von 130 (z. B. HAWIK III, Tewes et al., 2000). Erst seit kurzer Zeit werden auch für den deutschsprachigen Raum spezielle Testverfahren für die Intelligenzdiagnostik im hohen Fähigkeitsbereich entwickelt (z. B. der Berliner Intelligenzstruktur-Tests für Jugendliche: Begabungs- und Hochbegabungsdiagnostik; BIS-HB; Jäger et al., 2005). Diese werden in Abschnitt 2.5 vorgestellt. Im Folgenden wird zunächst auf einige generelle messmethodische Probleme bei Intelligenztestungen eingegangen, im Anschluss daran auf einige spezifische Probleme bei der Messung hoher Intelligenz.



### 2.4.1 Allgemeine Probleme bei der Messung von Intelligenz mit den üblichen Intelligenztestverfahren

#### *Mangelnde theoretische Fundierung*

Einige der am meisten verwendeten Intelligenztests wurden entweder relativ intuitiv entwickelt oder stützen sich auf tradierte Aufgabentypen, die sich einmal als valide erwiesen haben. Häufig basieren die Tests jedoch nicht auf einem expliziten theoretischen Konzept (z. B. die Wechsler-Tests, vgl. Holling, Preckel & Vock, 2004). Ein theoretisch begründetes und empirisch überprüftes Intelligenzmodell ist jedoch eine Voraussetzung dafür, die gemessenen Werte inhaltlich sinnvoll interpretieren zu können. Insbesondere unterschiedliche Testwerte einer Person in verschiedenen Testverfahren oder Subtests eines Tests können nur unter Rückgriff auf die jeweiligen Intelligenzmodelle interpretiert werden.

#### *Messungenauigkeiten durch Testkonstruktion nach der Klassischen Testtheorie*

Fast alle derzeit in Deutschland für die Intelligenzdiagnostik bei Kindern eingesetzten Testverfahren wurden ausschließlich nach den Regeln der Klassischen Testtheorie oder Messfehlertheorie konstruiert (Ausnahme: AID 2, Kubinger & Wurst, 2000). Die Grundannahme der Klassischen Testtheorie besagt, dass sich jeder Messwert aus einem wahren Wert und einem Messfehler zusammensetzt. Weiterhin werden verschiedene Annahmen über die Art der Messfehler gemacht: Der Erwartungswert des Messfehlers über eine große Anzahl von Personen sei 0, das heißt der Fehler tritt unsystematisch auf, manchmal trägt er dazu bei, einen Wert zu über-, manchmal zu unterschätzen. Bei einer ausreichend großen Stichprobe mittelt sich der Wert heraus. Die zweite Annahme zum Messfehler besagt, dass er weder mit dem wahren Wert noch mit einer anderen Variable korreliert (Gulliksen, 1950).

Diese auf dem Messfehler basierende Theorie ist jedoch nach heutigem Verständnis nicht ausreichend für die Konstruktion reliabler und valider Instrumente (siehe z. B. Rost, 2004). Die alleinige Verwendung des klassischen Testmodells führt zu verschiedenen Problemen: Die Eigenschaften von Testaufgaben und die Fähigkeiten der Personen sind in der klassischen Testtheorie untrennbar miteinander verbunden. So wird die Schwierigkeit einer Aufgabe in der Klassischen Testtheorie über den Anteil der Probanden definiert, der die Aufgabe richtig lösen kann. Ob eine Aufgabe schwierig oder leicht ist, hängt somit von den Fähigkeiten der untersuchten Stichprobe ab. Jedoch hängt andererseits auch die gemessene Fähigkeit einer Person von der Schwierigkeit der Aufgaben ab. Klassisch konstruierte Tests erlauben keine spezifisch objektiven Vergleiche zwischen Personen oder Items, da die Messwerte von der jeweils vorgegebenen Aufgabenzusammenstellung abhängig sind. Würden der gleichen Person etwas andere Aufgaben desselben Aufgabentyps vorgelegt, so würde sie einen anderen Testwert erreichen.

Weiterhin problematisch ist beispielsweise, dass die Gütekriterien jeweils nur für das ganze Instrument, nicht aber für einzelne Aufgaben bestimmt werden können; häufig unterscheiden sich jedoch Aufgaben, die verschiedene Fähigkeitsbereiche erfassen,

in ihrer Validität und Reliabilität. Ebenso wie die Schwierigkeit sind in der klassischen Testtheorie auch die Trennschärfe, die Reliabilität und die Validität stichprobenabhängig. Der Standardmessfehler, der von der Reliabilität und der Varianz abhängt, wird stets für die gesamte Stichprobe berechnet. Tatsächlich fällt der Messfehler jedoch je nach Fähigkeitsniveau einer Person unterschiedlich aus. So liefert ein Test zum Beispiel weniger Informationen für einen Pb, der alle Aufgaben eines Tests richtig gelöst hat, als für einen Pb, der ungefähr die Hälfte der Aufgaben richtig bearbeitet hat. Während bei der zweiten Person im Test deutlich wird, wie viel sie leisten kann, ist dies bei der ersten Person nicht erkennbar. In klassischen Modellen werden direkte Beziehungen zwischen der Fähigkeit einer Person und ihrer Antwort zu einer bestimmten Aufgabe angenommen. Daher können keine Vorhersagen für die Lösungswahrscheinlichkeiten einzelner Aufgaben bezogen auf bestimmte Pbn getroffen werden. Dies ist beispielsweise relevant bei der Entwicklung von Aufgaben für bestimmte Populationen. Stattdessen werden Aussagen zur Schwierigkeit über alle Probanden hinweg getroffen.

Zusammenfassend kann festgehalten werden, dass klassische Messmodelle im Wesentlichen folgende Beschränkungen aufweisen (nach Hambleton, Swaminathan & Rogers, 1991):

- a. Itemabhängigkeit der Messwerte
- b. Stichprobenabhängigkeit der Itemstatistiken
- c. Standardmessfehler für die gesamte Stichprobe
- d. keine Wahrscheinlichkeitsaussagen auf der Aufgabenebene

Die Messgenauigkeit des Tests für bestimmte einzelne Pbn ist nicht bekannt. Insgesamt muss davon ausgegangen werden, dass klassisch konstruierte Testverfahren eine unterschiedliche Güte für unterschiedliche Subgruppen der untersuchten Stichproben aufweisen, an denen die Gütekriterien bestimmt wurden. Für die Testung sehr intelligenter Pbn folgt daraus, dass die Gütekriterien von klassisch konstruierten Testverfahren, die an überwiegend durchschnittlich intelligenten Personen entwickelt wurden, für hoch Begabte nicht unbedingt zutreffen müssen.

*Die Anpassung klassischer Testaufgaben an die steigende Leistungsfähigkeit einer Gruppe ist relativ aufwändig*

Der 1987 von Flynn entdeckte Effekt, dass die Intelligenztestleistung von Pbn in Industriegesellschaften jährlich um ca. 0.3 IQ-Punkte ansteigt, ist ein stabiles, wenn auch noch nicht völlig erklärtes Phänomen der Intelligenzforschung (Kane, 2000). Auch bewährte Intelligenztests müssen daher in regelmäßigen Abständen revidiert und neu normiert werden. Um Tests leichter veränderten Fähigkeitswerten innerhalb einer Population anpassen zu können, ist es hilfreich, solche Testaufgaben zu verwenden, von denen bekannt ist, welche Faktoren über ihre Schwierigkeit bestimmen. Auch für die Erstellung großer Item-Pools für adaptives Testen (s. u.) ist ein regelgeleitetes Vorgehen bei der Aufgabenkonstruktion von Vorteil. Einen entsprechenden Ansatz der regelgeleiteten Itemkonstruktion verfolgte zum Beispiel Preckel (2003)

bei der Entwicklung eines figuralen Matrizen-tests zur Messung hoher Intelligenz (siehe Abschnitt 2.5).

#### *Eingeschränkte Objektivität und Ökonomie*

Intelligenztests werden mit Kindern heute noch fast ausschließlich entweder als Individualtest in Form eines „standardisierten Dialogs“ (z. B. Wechsler-Tests; Tewes, 1991) oder als Individual- oder Gruppentest in Form eines Paper-Pencil-Tests durchgeführt. Wenn die Testdurchführung jedoch computergestützt stattfindet, kann die Durchführung- und Auswertungsobjektivität optimiert werden. Dies ist darauf zurückzuführen, dass z. B. der Einfluss des Testleiters geringer ausfällt und bei der Auswertung keine freien Antworten der Pbn bewertet werden müssen. Zudem ist die Auswertung von computergestützten Tests deutlich zeitsparender.

Insgesamt sollten Intelligenztests somit auf Modell- und auf Aufgabenebene theoretisch gut fundiert sein. Die Messqualität von Tests ist höher, wenn ihre Passung auch auf Modelle der Item Response Theorie gezeigt werden kann, als wenn ausschließlich die Kennwerte aus dem Rahmen der Klassischen Testtheorie ermittelt werden. Aufgaben, die regelgeleitet konstruiert wurden, sind flexibler auf veränderte Fähigkeiten der Population oder auf die Fähigkeiten bestimmter anderer Gruppen anpassbar. Im Hinblick auf die Objektivität und die Ökonomie sind computergestützte Tests vorteilhaft gegenüber Paper-Pencil-Tests.

### **2.4.2 Spezifische Probleme bei der Messung hoher Intelligenz mit den üblichen Intelligenztestverfahren**

#### *Deckeneffekte*

Sehr fähige Pbn können die Aufgaben häufig so gut lösen, dass der Test die Leistungen nicht mehr differenzieren und somit keine quantitativen Vergleiche zwischen sehr intelligenten Personen herstellen kann. Die Pbn stoßen mit ihrer Leistung an die „Testdecke“. Da insgesamt zu wenig schwierige Aufgaben im Test enthalten sind, ergibt der Test für Pbn mit sehr hohen Leistungen weniger Informationen. Die Messgüte wird dadurch geringer und der Messfehler aufgrund geringerer Ergebnisvarianz größer (Preckel, 2003). Über deutliche Deckeneffekte bei der Testung hoch Begabter bei den Advanced Progressive Matrices (APM, Raven, 1962a), die für die Erfassung überdurchschnittlicher Intelligenz konstruiert wurden, berichtet Preckel (2003). Weiterhin konnten Deckeneffekte im hohen Leistungsbereich für den CFT 20 und die K-ABC festgestellt werden (Baving, 2002; Van Melis-Wright & Stone, 1986). Ausgeprägte Deckeneffekte des Zahlenfolgentests des CFT 20 zeigten sich auch in der vorliegenden Studie (vgl. Abschnitt 7.1.2). Aufgrund von Deckeneffekten der aktuellen Testverfahren, die in den USA eingesetzt werden, empfehlen dort verschiedene Autoren, bei der Testung von Kindern bis zum Alter von zwölf Jahren, die sehr hohe Werte in einem Intelligenztest erreicht haben, zusätzlich auf die (veraltete) Version der *Stanford-Binet Intelligence Scale (Form L-M)* zurückzugreifen (z. B. Gross,

2000; Vernon, 1987). Es wird berichtet, dass sich diese Testversion aufgrund von deutlich schwierigeren Testaufgaben – eher als die Nachfolgeversionen – für die Messung sehr hoher Begabungen und die Bestimmung von Intelligenzquotienten von  $> 160$  eignet. Dieses Vorgehen ist jedoch nicht unproblematisch, und es kann nicht davon ausgegangen werden, dass auf diese Weise valide Ergebnisse erzielt werden. Der *Stanford-Binet Intelligence Scale (Form L-M)* wurde bereits 1972 normiert, die stark verbal-lastigen Aufgaben sind inhaltlich deutlich veraltet und die inzwischen über 30 Jahre alten Normen führen aufgrund des Flynn-Effekts (s. o.) zu einer Überschätzung der Intelligenz.

#### *Motivationsverluste durch zu einfache und wenig herausfordernde Aufgaben*

Besonders begabte Kinder müssen bei Intelligenztests eine Vielzahl von Aufgaben bearbeiten, die für sie zu leicht sind und die sie nicht herausfordern. Die Motivationsverluste, die sich auf diese Weise unter Umständen einstellen, können das Gesamtergebnis deutlich beeinträchtigen. Ein Beispiel hierfür ist der CFT 20. Der Gesamtest besteht aus zwei Teilen, die im Prinzip zwei Parallelformen desselben Tests darstellen und die nacheinander durchgeführt werden. Durch die wiederholte Vorgabe derselben Untertests (mit neuen Aufgaben) soll die Messgenauigkeit erhöht werden. Bei sehr begabten Pbn wurden jedoch regelmäßig Leistungseinbußen im zweiten Testteil festgestellt. Dieser Befund steht im Gegensatz zur Situation leistungsschwacher Pbn, die ihre Leistung im zweiten Testteil eher verbessern. Der Testautor vermutet, dass die häufig schlechtere Leistung von sehr intelligenten Pbn auf Motivationsverluste zurückzuführen ist. Es wird daher im Testhandbuch empfohlen, die Diagnose bei leistungsstarken Pbn nur auf den ersten Testteil zu stützen (Weiß, 1998). Dieses Vorgehen hat jedoch zur Folge, dass die Messgenauigkeit des Tests für hoch begabte Pbn geringer ist als für durchschnittlich begabte Pbn.

#### *Messungenauigkeiten im höheren Fähigkeitsbereich*

Bei der Messung von Extrembereichen der Begabung kann die Testkonstruktion nach der klassischen Testtheorie aufgrund der Stichprobenabhängigkeit der Kennwerte (s. o.) zu besonderen Problemen führen. Streng genommen darf ein Test nur bei solchen Personen eingesetzt werden, die das gleiche Fähigkeitsniveau aufweisen wie die Stichprobe, an der er entwickelt wurde. Aufgrund der geringen Anzahl hoch intelligenter Personen in der Population enthalten die Standardisierungsstichproben jedoch in der Regel nur einen sehr kleinen Anteil entsprechender Pbn. Einen IQ von 130 oder höher weisen definitionsgemäß nur etwa zwei von 100 Menschen auf, einen IQ von 145 oder höher erreicht nur ungefähr einer von 1000 Menschen und einen IQ von 160 erreichen nur noch etwa drei von 100.000 Menschen. Bei den in Deutschland üblichen Normierungsstichproben von ca. 1500 Pbn (z. B. im HAWIK III) kann somit nicht davon ausgegangen werden, dass ein substantieller Anteil sehr intelligenter Menschen darin enthalten ist. Für entsprechend intelligente Testpersonen, bei denen der Test angewendet wird, ist daher keine adäquate Vergleichsgruppe vorhanden und die Gütekriterien sind nicht gesichert.

*Strukturelle Äquivalenz der Intelligenzstruktur von durchschnittlich und hoch begabten Kindern?*

Eine valide Intelligenztestung beruht unter anderem darauf, dass das angenommene Intelligenzmodell für die Population gilt, aus der die jeweilige Testperson stammt. Es ist erforderlich, bei der Testkonstruktion im Sinne der Konstruktvalidierung die Gültigkeit des Modells für die untersuchte Stichprobe zu prüfen (Holling et al., 2004). Wenn das angenommene Modell für die Stichprobe oder für bestimmte Subgruppen, z. B. Alters- oder Begabungsgruppen, keine Gültigkeit hat, wäre eine Interpretation der Testergebnisse für die jeweiligen Personen stark beeinträchtigt. Ob generell dieselbe Intelligenzstruktur für durchschnittlich und hoch begabte Personen angenommen werden kann, oder ob für hoch Begabte prinzipiell eine andere Faktorenstruktur zutrifft, ist bislang nicht eindeutig geklärt (Preckel, 2003). Bisher wird dies bei der Konstruktion von Tests in der Regel nicht geprüft.

Verschiedene empirische Untersuchungen haben sich jedoch mit der Frage nach strukturellen Unterschieden zwischen Begabungsgruppen beschäftigt. Bereits seit den Untersuchungen von Spearman (1927) ist bekannt, dass die Interkorrelationen von verschiedenen Intelligenztestleistungen in Gruppen hoch intelligenter Personen geringer sind als in Gruppen mit durchschnittlicher Intelligenz. Dieser Befund konnte in neueren Studien bestätigt werden (z. B. Detterman & Daniel, 1989; Kane, 2000). Die Faktorenstruktur der Intelligenz bei durchschnittlich und hoch Begabten wurde verschiedentlich mit der amerikanischen Version des Wechsler-Tests für Kinder (WISC-R) untersucht. Die Befundlage hierzu ist jedoch widersprüchlich. Während einige Untersuchungen dieselbe Drei-Faktoren-Struktur für durchschnittlich und hoch Begabte fanden (Karnes & Brown, 1980; Sapp, Chissom & Graham, 1985), stellten andere Untersuchungen abweichende Faktorenstrukturen in der Hochbegabtengruppe fest (Willson, Reynolds & Gilbert, 1982; Speck, 1979). Für die Wechsler-Tests ist daher bis heute nicht zufriedenstellend geklärt, ob für durchschnittlich und hoch begabte Kinder dasselbe Modell angenommen werden kann.

Wenn ein Test bei hoch Begabten eingesetzt werden soll, sollte daher vorab geprüft worden sein, ob das zugrundeliegende Intelligenzmodell für hoch intelligente Pbn Gültigkeit aufweist. Eine entsprechende Überprüfung wurde bei der Entwicklung des BIS-HB (Jäger et al., 2005; s. u.) vorgenommen: In der Normierungsstichprobe des BIS-HB, die sich aus 921 unausgelesenen Pbn und 407 Pbn aus speziellen Schulen für hoch Begabte zusammensetzt, wurde die Gültigkeit des zugrundeliegenden Intelligenzmodells (BIS-Modell nach Jäger, 1982; s. u.) für beide Begabungsgruppen explizit geprüft. Über konfirmatorische Modelltests konnte gezeigt werden, dass eine strukturelle Invarianz für beide Gruppen gegeben ist, das BIS-Modell somit für durchschnittlich wie hoch Begabte gleichermaßen gilt.

Zusammenfassend kann festgehalten werden, dass für eine reliable und valide Testung hoher Ausprägungen von Intelligenz verschiedene Bedingungen in einem Test erfüllt sein müssen: Der Test muss eine genügende Anzahl ausreichend schwieriger Testaufgaben enthalten, um auch im überdurchschnittlichen Fähigkeitsbereich noch hinreichend zwischen Pbn differenzieren zu können und zuverlässige Messwerte zu

liefern. Andererseits dürfen nicht zu viele für die Pbn sehr einfache Aufgaben im Test enthalten sein, da sonst Leistungseinbußen durch Motivationsverluste entstehen können. Wünschenswert wäre hier der Einsatz adaptiver Testverfahren, die eine optimale Passung von Itemschwierigkeit und Personenfähigkeit realisieren. In adaptiven Tests wird die Auswahl der Testitems individuell auf die jeweils bereits gezeigte Leistung einer Person zugeschnitten („tailored testing“). Nach den ersten bearbeiteten Items wird eine Schätzung der Fähigkeit der Person vorgenommen. Im Folgenden werden der Testperson dann nur noch solche Aufgaben vorgelegt, bei der ihre individuelle Lösungswahrscheinlichkeit etwa 50% beträgt. Um die Möglichkeiten dieses Ansatzes optimal zu nutzen, ist jedoch ein großer Item-Pool erforderlich und die Testdarbietung sollte computergestützt erfolgen. Testverfahren, die für die Hochbegabungsdiagnostik eingesetzt werden, sollten auch an hoch begabten Pbn erprobt, optimiert und normiert worden sein. Die Passung des jeweils angenommenen Intelligenzmodells muss für die Gruppe der sehr intelligenten Pbn empirisch überprüft werden. Nur wenn das zugrunde gelegte Intelligenzmodell auch für die hoch Begabten gilt, können die Testwerte entsprechend des Intelligenzmodells interpretiert werden.

## 2.5 Spezielle Intelligenztests zur Messung hoher Intelligenz

Es gibt im deutschsprachigen Raum bisher ein normiertes Testverfahren und einen bereits erprobten, jedoch noch nicht normierten Test für eine exakte Messung sehr hoher Intelligenz. Die Tests wurden für Jugendliche und Erwachsene entwickelt. Für Kinder unter 12 Jahren existieren solche Testverfahren im deutschen Sprachraum bisher jedoch nicht.

Das erste dieser Verfahren ist der *Berliner Intelligenzstruktur-Test für Jugendliche: Begabungs- und Hochbegabungsdiagnostik* (BIS-HB; Jäger et al., 2005; eine detaillierte Testbeschreibung findet sich in Holling et al., 2004). Der BIS-HB kann bei durchschnittlich und hoch begabten Jugendlichen zwischen 12;6 und 16;5 Jahren eingesetzt werden und basiert auf dem Berliner-Intelligenz-Strukturmodell (BIS, Jäger, 1982). Im Aufgabenmaterial entspricht der BIS-HB weitgehend dem Berliner Intelligenzstruktur-Test 4 (BIS-4; Jäger et al., 1997), der für junge Erwachsene normiert ist. Über insgesamt 42 Aufgabengruppen werden im BIS-HB die vier operativen und drei inhaltlichen Fähigkeiten des BIS-Modells (s. o.) erfasst. Zusätzlich kann – als Integral der Fähigkeiten – ein Wert für die allgemeine Intelligenz der Testperson bestimmt werden. Der Test kann als Individual- oder Gruppentest durchgeführt werden. Normiert wurde der BIS-HB an insgesamt 1.328 Pbn aus Haupt- und Realschulen, Gymnasien und speziellen Schulen für intellektuell besonders begabte Schülerinnen und Schüler.

Ein zweites Verfahren für die Hochbegabungsdiagnostik ist ein figuraler Matrizen-test zur Erfassung hoher Ausprägungen der fluiden Intelligenz von Preckel (2003). Dieser Test wurde regelgeleitet konstruiert und an 429 hoch begabten Pbn im Alter von 13 bis 58 Jahren ( $M = 16.9$  Jahre) erprobt. Anhand der Kennwerte der Klassi-

schen Testtheorie sowie der Item Response Theorie weist der Test insgesamt eine gute Messqualität auf. Der Test ist noch nicht normiert.

In den USA gibt es erste Ansätze, spezielle Intelligenztests für Kinder zu entwickeln, die auch im oberen Intelligenzbereich exakt messen. Eine Recherche in der Testdatenbank des amerikanischen *BUROS Institute for Mental Measurement*, das auch das amerikanische *Mental Measurements Yearbook* herausgibt, führte zu zwei aktuell verfügbaren speziellen Testverfahren für die Messung hoher intellektueller Begabungen bei Kindern. Eines dieser beiden Testverfahren ist das *Screening Assessment for Gifted Elementary and Middle School Students - Second Edition* (SAGES-2, Johnsen & Corn, 2001). Der SAGES-2 misst neben dem schulischen Wissen im mathematisch-naturwissenschaftlichen sowie im sprachlichen und sozialwissenschaftlichen Bereich Reasoningfähigkeiten über bildliche Analogieaufgaben. Der Test ist einsetzbar für Kinder zwischen 5;0 und 14;11 Jahren, wobei für die jüngeren Kinder (Kindergarten bis dritte Klasse) und die älteren Kinder (vierte bis achte Klasse) separate Aufgabenzusammenstellungen vorliegen. SAGES-2 wurde an 5.313 Kindern aus 28 US-Staaten normiert, davon waren 2.290 Kinder bereits als besonders begabt identifiziert worden, die restlichen 3.023 Kinder entstammten regulären Schulen. Für die Interpretation einzelner Testwerte bietet der Test jeweils eine „Normalnorm“ und eine „Begabtennorm“ an, deren Werte um eine Standardabweichung auseinander liegen. Der Test ist sowohl für das Screening eines gesamten Schüler-Pools für die Aufnahme in Begabtenförderprogramme als auch für die Identifikation einzelner hoch begabter Schülerinnen und Schüler konzipiert. Die testtheoretischen Analysen erfolgten anhand der Klassischen Testtheorie und der Item Response Theorie. Der zweite Test für die Messung hoher Begabung bezieht sich auf die Erfassung hoher mathematischer Fähigkeiten bei Kindern zwischen sechs und zwölf Jahren (*Test of Mathematical Abilities for Gifted Students*, TOMAGS; Ryser & Johnsen, 1998). Da es sich bei TOMAGS um einen speziellen Begabungstest handelt, der keine Aussagen über die allgemeine intellektuelle Fähigkeit von Kindern macht, wird hier auf eine ausführlichere Darstellung verzichtet.

Vergleichbare Verfahren für Kinder bis etwa zwölf Jahre fehlen für den deutschsprachigen Raum bisher. Testverfahren aus anderen Ländern können jedoch nicht ohne weitere Prüfung der Inhalte und ohne eine Neunormierung „importiert“ werden, da Intelligenztests – trotz vielfältiger Bemühungen in diese Richtung – nie unabhängig von Sprache, Kultur und Bildungssystem eines Landes sind (siehe z. B. Greenfield, 1997). Daher ist es erforderlich, für den deutschen Sprach- und Kulturraum eigene adäquate Verfahren zu entwickeln und zu normieren.

## **2.6 Überlegungen zur Entwicklung eines neuen Testverfahrens: Erfassung zentraler Komponenten der Intelligenz durch Arbeitsgedächtnisaufgaben**

Ein Ziel der vorliegenden Arbeit besteht darin, die Grundlagen für die Konstruktion eines neuen Testverfahrens zu erproben, das sich zur Messung hoher Fähigkeitsaus-

prägungen bei Kindern eignet. Es sollen dabei möglichst viele Schwächen vermieden werden, die bisherige Testverfahren aufweisen.

Optimal wäre die Entwicklung einer umfassenden Testbatterie zur Messung verschiedener Aspekte intellektueller Fähigkeiten im Kindesalter. Da dies den Rahmen dieser Arbeit jedoch sprengen würde, muss sich der Versuch einer Testentwicklung auf einen Ausschnitt des Fähigkeitsspektrums beschränken. In der gegenwärtigen Intelligenzforschung werden insbesondere die Verarbeitungskapazität und die fluide Intelligenz als zentrale Konstrukte intellektueller Fähigkeiten betrachtet. Die Operation „Verarbeitungskapazität“ bzw. Reasoning entspricht auch am ehesten der Fähigkeit, die traditionell als Intelligenz verstanden wird (Carroll, 1993). Fluide Intelligenzfähigkeiten, die nach Carroll insbesondere Reasoningfähigkeiten beinhalten, laden am höchsten auf einem generellen Intelligenzfaktor. Zudem konnten substantielle Zusammenhänge zwischen der Verarbeitungskapazität und dem komplexen Problemlösen gefunden werden (z. B. Süß, 2001). Es kann daher davon ausgegangen werden, dass die Verarbeitungskapazität eine wesentliche Grundlage menschlicher intellektueller Fähigkeiten darstellt. Auch zeigte sich in verschiedenen Untersuchungen, dass Aufgaben zur Verarbeitungskapazität am besten dazu in der Lage sind, zwischen durchschnittlich und hoch Begabten zu differenzieren (z. B. Jäger et al., 2005; Robinson & Janos, 1987; McClelland, 1982).

Für eine Fokussierung auf die Messung von Reasoningfähigkeiten spricht auch, dass diese fluiden Fähigkeiten relativ unabhängig von der (Schul-)Bildung und vom Wissensstand sind. So kann auch das intellektuelle Potenzial von minderleistenden Pbn oder von Pbn mit Teilleistungsschwächen besser abgebildet werden. Die Intelligenzstruktur von Kindern ist zudem möglicherweise noch weniger ausdifferenziert als die von Jugendlichen und Erwachsenen (Perleth, Schatz & Mönks, 2000; Thorndike & Hagen, 1987; siehe aber Bickley, Keith & Wolfle, 1995). Insgesamt erscheint somit eine vorläufige Beschränkung auf die Messung zentraler Intelligenzkomponenten bei Kindern gerechtfertigt. Das Aufgabenmaterial sollte Aufgaben mit verbalen, numerischen und figuralen Inhalten enthalten, da sich diese drei Inhaltsbereiche reliabel als robuste Faktoren in der Intelligenzforschung nachweisen ließen.

Die Verarbeitungskapazität kann über verschiedene Aufgabentypen gemessen werden. Klassischerweise wird sie über Aufgaben zum deduktiven und induktiven Denken, zur Konstruktion und Planung sowie zur Systemsteuerung erfasst. Neuere Studien zeigen jedoch, dass auch basalere Maße aus der Kognitionspsychologie wie die Kapazität des Arbeitsgedächtnisses für die Erfassung von Verarbeitungskapazität geeignet sein können (z. B. Kyllonen & Christal, 1990). Arbeitsgedächtnisaufgaben wurden sowohl in der experimentellen als auch in der psychometrischen Forschung bereits intensiv untersucht. Das Arbeitsgedächtnis nimmt inzwischen den Stellenwert eines zentralen Konstrukts in kognitionspsychologischen Theorien ein (z. B. Anderson, 1983; Baddeley, 1986). Die Kapazität des Arbeitsgedächtnisses wird von einigen Forschern als der „Flaschenhals“ der kognitiven Leistungsfähigkeit verstanden (siehe z. B. Süß, 2001). Damit ist gemeint, dass höhere kognitive Leistungen von der Arbeitsgedächtniskapazität abhängen und nur so gut sein können, wie es die Arbeits-



gedächtniskapazität erlaubt. Es wird heute davon ausgegangen, dass Arbeitsgedächtnisaufgaben grundlegende und zentrale Prozesse intelligenten Verhaltens erfassen.

Arbeitsgedächtnisaufgaben haben somit gegenüber vielen klassischen Intelligenztestaufgaben den Vorteil, dass sie kognitionspsychologisch fundiert sind und nicht lediglich aus älteren, häufig eher intuitiv entwickelten Testverfahren tradiert wurden. Es existieren bereits zahlreiche verbale, numerische und figurale Aufgabentypen, die sich ausgezeichnet für die Prognose und Erklärung höherer kognitiver Leistungen eignen. Für erwachsene Pbn konnten bereits substantielle Zusammenhänge zwischen der Verarbeitungskapazität im Sinne des BIS-Modells und der Kapazität des Arbeitsgedächtnisses nachgewiesen werden (Oberauer, Süß, Schulze, Wilhelm & Wittmann, 2000). Auch für Kinder konnten bedeutsame Zusammenhänge zwischen Arbeitsgedächtnis und Intelligenz gezeigt werden (z. B. de Jong & Das-Smaal, 1995).

Arbeitsgedächtnisaufgaben konnten sich zudem bereits im Rahmen der individuellen Begabungsdiagnostik bewähren. In den USA existiert ein normiertes Testverfahren, das über verschiedene Arbeitsgedächtnisaufgaben unter anderem die Messung der allgemeinen Intelligenz bei Kindern und Erwachsenen anstrebt (Swanson, 1996; vgl. Abschnitt 4.3.3).

Ein weiterer Pluspunkt von Arbeitsgedächtnisaufgaben ist ihre besondere strukturelle und inhaltliche Einfachheit. Arbeitsgedächtnisaufgaben erfordern nur in sehr geringem Ausmaß schulisches oder akademisches Wissen und sind daher insbesondere für die Erfassung fluider Fähigkeiten geeignet. Viele Arbeitsgedächtnisaufgaben sind zudem einfach zu erklären, und die Instruktionen sind bereits von Kindern gut zu verstehen. Trotz der einfachen Struktur lassen sich sehr schwierige Aufgaben konstruieren, die auch zwischen sehr fähigen Personen noch differenzieren. Deckeneffekte können daher vermieden werden. Die relativ feine Abstufbarkeit in der Schwierigkeit begünstigt zudem die Erstellung großer Aufgabenpools für mögliche adaptive Aufgabenvorgaben.

Schließlich können Arbeitsgedächtnisaufgaben computergestützt und in Gruppentestungen eingesetzt werden, so dass eine hohe Durchführungsobjektivität und Ökonomie gewährleistet werden können. Gleichzeitig erfordert eine computergestützte Durchführung dieser Aufgaben weder besondere Computerkenntnisse noch ausgeübte feinmotorische Fertigkeiten (z. B. Bedienen der Maus), über die Kinder noch nicht verfügen.

Im folgenden Kapitel wird das Konzept des Arbeitsgedächtnisses näher erläutert. Es wird die Faktorstruktur des Arbeitsgedächtnisses thematisiert, und es werden Befunde zum Zusammenhang des Arbeitsgedächtnisses mit anderen kognitiven Konstrukten referiert. Kapitel 4 beschäftigt sich im Anschluss mit Fragen der Messung der Arbeitsgedächtniskapazität.

### 3. Arbeitsgedächtnis

In diesem Kapitel erfolgt eine Darstellung der theoretischen Grundlagen von Arbeitsgedächtnisaufgaben und ihrer möglichen Eignung für die Begabungsdiagnostik. Dazu wird zunächst ein Überblick über das Konzept des Arbeitsgedächtnisses gegeben (Abschnitt 3.1). Anschließend wird auf Modellvorstellungen zu individuellen Unterschieden im Arbeitsgedächtnis und auf Befunde aus der differentiellen Psychologie eingegangen (Abschnitt 3.2). Abschnitt 3.3 widmet sich der Frage nach der Binnen- bzw. Faktorstruktur des Arbeitsgedächtnisses. Es ist bislang nicht eindeutig geklärt, ob es sich beim Arbeitsgedächtnis um ein eindimensionales Konstrukt handelt, oder ob sich verschiedene domänenspezifische oder funktionelle Komponenten des Arbeitsgedächtnisses separieren lassen. In Abschnitt 3.4 werden Befunde zum Zusammenhang des Arbeitsgedächtnisses mit fluider Intelligenz und anderen kognitiven Variablen berichtet. Abschließend wird der Forschungsstand in Abschnitt 3.5 noch einmal zusammengefasst.

#### 3.1 Das Konzept des Arbeitsgedächtnisses

Das Arbeitsgedächtnis (*working memory*) ist ein System für die kurzzeitige simultane Speicherung und Verarbeitung von Informationen. Es konnte in zahlreichen experimentellen Studien gezeigt werden, dass das Arbeitsgedächtnis eine zentrale Rolle in der Informationsverarbeitung spielt und an vielen kognitiven Prozessen beteiligt ist (z. B. Baddeley, 1986; Gilhooly, Logie, Wetherick & Wynn, 1993). Darüber hinaus gibt es bedeutsame Zusammenhänge zwischen individuellen Unterschieden in der Arbeitsgedächtniskapazität und vielen kognitiven Fähigkeiten, wie der Lesefähigkeit (z. B. Daneman & Merikle, 1996), der Rechenfähigkeit (z. B. Turner & Engle, 1989; Bull & Scerif, 2001), der Verarbeitungskapazität oder Reasoning (z. B. Kyllonen & Christal, 1990) und dem Problemlösen (z. B. Demetriou, Christou, Spanoudis & Platsidou, 2002). Mittlerweile existieren viele Variationen von psychologischen Modellen des Arbeitsgedächtnisses, einen Überblick geben Miyake und Shah (1999).

Das Konzept des Arbeitsgedächtnisses geht auf die Arbeiten von Miller, Galanter und Pribram (1960) zurück, die die Kontrolle und Ausführung von Handlungsplänen analysierten. Sie gelangten dabei zu dem Schluss, dass für die Umsetzung von Plänen eine Instanz nötig sei, die Kognitionen und Verhalten kontrolliert und die gleichzeitig als Kurzzeitspeicher fungiert. Ein Plan müsse während seiner Ausführung in einem Speicher abgelegt werden, auf den schnell zugegriffen werden kann. Für diese Instanz schlugen sie den Begriff „working memory“ vor. Atkinson und Shiffrin (1971) verwendeten in ihren Arbeiten zum Kurzzeit- und Langzeitgedächtnis ebenfalls den Begriff „Arbeitsgedächtnis“. Sie nahmen in ihrem Drei-Speicher-Modell des Gedächtnisses an, dass der Kurzzeitspeicher, in dem Informationen auf ihrem Weg zwischen dem sensorischen Speicher und dem Langzeitgedächtnis zwischenge-

lagert werden, wesentlich auch an der aktiven Verarbeitung von Informationen beteiligt ist. Daher bezeichneten sie den Kurzzeitspeicher auch als „Arbeitsgedächtnis“.

Die britischen Forscher Baddeley und Hitch entwickelten 1974 das Modell eines Arbeitsgedächtnissystems, das aus mehreren Komponenten besteht, und das von der Arbeitsgruppe um Baddeley bis heute kontinuierlich weiterentwickelt wird. Ausgangspunkt ihrer Untersuchungen war die Frage, welche Funktionen das Kurzzeit- oder Arbeitsgedächtnis für die Informationsverarbeitung hat. Es wurde beispielsweise vermutet, dass ein solches System wichtig ist für das Verstehen von Sprache, da die einzelnen Wörter eines Satzes gespeichert und gleichzeitig so verarbeitet werden müssen, dass ein Sinn entsteht. Sie gingen davon aus, dass die Kapazität des Kurzzeit- oder Arbeitsgedächtnisses begrenzt ist. Zusätzlich machten sie die folgende Annahme: Personen, die die gesamte Kapazität des Kurzzeitgedächtnisses für die reine Speicherung der Informationen benötigen, haben größere Schwierigkeiten, diese Informationen gleichzeitig zusätzlich zu verarbeiten. In Experimenten ließen sie Pbn Reasoning- und andere Aufgaben bearbeiten, während sie gleichzeitig bestimmte Informationen behalten und später wiedergeben sollten. Bei geringerer Gedächtnisbelastung gelang es den Pbn, sich stets einige der Informationen zu merken, ohne dass es zu Leistungseinbußen bei der Hauptaufgabe kam. Wurde die Gedächtnisbelastung jedoch zu groß, verschlechterte sich die Leistung in der Hauptaufgabe.

Baddeley und Hitch schlussfolgerten, dass das Kurzzeitgedächtnis an wichtigen Prozessen wie Schlussfolgern und Verstehen beteiligt ist, der Zusammenhang jedoch nicht absolut ist. Auf der Basis dieser Befunde entwickelten sie ihr modulares Modell eines Arbeitsgedächtnisses, von dem sie annahmen, dass es aus mindestens zwei Komponenten besteht: einer generellen zentral-exekutiven Komponente (die die Bearbeitung der Hauptaufgabe in den Experimenten von Baddeley und Hitch ausführte), sowie einer „Hilfskomponente“, die eine Speicherung von verbalem Material vornimmt. Erst wenn die Speicherkapazität der Hilfskomponente überschritten wird, greift das System auf die zentral-exekutive Komponente zurück. Baddeley und Hitch (1974) vermuteten zudem, dass weitere Hilfskomponenten, insbesondere eine Komponente zur Speicherung visuell-räumlicher Informationen, existieren könnten. Daher schlugen sie ein modular aufgebautes Modell vor, das sich aus drei Teilen zusammensetzt: einer zentralen Exekutive (*central executive*), einer phonologischen Schleife (*phonological loop*) und einem visuell-räumlichen Skizzenblock (*visual-spatial sketchpad*). Die zentrale Exekutive wird dabei als eine zentrale Steuer- und Kontrollinstanz verstanden, die phonologische Schleife und der visuell-räumliche Skizzenblock als nachgeordnete Hilfssysteme (*slave systems*).

Die phonologische Schleife speichert und verarbeitet Informationen in Form von gesprochener Sprache und besteht aus zwei Komponenten: einem passiven phonologischen Speicher und einem aktiven artikulatorischen Kontrollprozess. Der passive Speicher kann phonologische Informationen nur etwa 1,5 bis 2 Sekunden aufrechterhalten, bevor sie zerfallen. Durch den artikulatorischen Kontrollprozess kann die Information jedoch erhalten werden. Dies geschieht über subvokale Wiederholungen (*Rehearsal*). Werden die Informationen rechtzeitig vor dem Zerfall aktiv wiederholt,

gehen sie erneut in den phonologischen Speicher ein. Über den artikulatorischen Kontrollprozess können auch nicht-phonologische Informationen (z. B. geschriebene Wörter, Abbildungen) in den phonologischen Speicher eingelesen werden.

Der visuell-räumliche Skizzenblock speichert und verarbeitet visuell-räumliche Informationen. Auch hier kann der Speicher die Informationen nur für kurze Zeit behalten, daher müssen sie für eine längere Speicherung durch aktives Rehearsal aufrechterhalten werden. Es gibt Hinweise darauf, dass sich im visuell-räumlichen Skizzenblock zwei Komponenten voneinander unterscheiden lassen, die jedoch eng interagieren (Logie & Pearson, 1997): a) eine Komponente der visuellen Erkennung statischer Muster (*inner eye*) und b) eine Komponente der räumlichen Speicherung von Bewegungsfolgen (*inner scribe*).

Die zentrale Exekutive steuert und koordiniert das Arbeitsgedächtnis als Gesamtsystem. In neueren Modellversionen verfügt sie – im Gegensatz zur ursprünglichen Modellversion – nicht über einen eigenen Speicher (Baddeley & Logie, 1999). Stattdessen nutzt sie die Speicher der beiden slave systems sowie möglicherweise weitere Systeme. Die zentrale Exekutive unterscheidet sich somit fundamental von einem Kurzzeitspeicher. Es wird angenommen, dass die Kapazität der zentralen Exekutive begrenzt ist. Sie dient der Ausführung einfacher kognitiver Operationen, die bei der kognitiven Verarbeitung von Informationen anfallen. Darüber hinaus ist sie beteiligt an der Planung von Handlungen und dem Treffen von Entscheidungen. Möglicherweise setzt sich die zentrale Exekutive aus mehreren Komponenten zusammen (Baddeley & Logie, 1999). Baddeley (2000) schlug als Erweiterung des klassischen Drei-Komponenten-Modells einen „episodischen Speicher“ (*episodic buffer*) als zusätzliche Komponente vor, die für die Verknüpfung des Arbeitsgedächtnisses mit dem Langzeitgedächtnis zuständig ist.

Die Messung der Kapazität des Arbeitsgedächtnisses kann nach Baddeley (1986) über Aufgaben erfolgen, die eine simultane Speicherung und Verarbeitung von Informationen erfordern und verschiedene Inhalte erfassen. Dieses Aufgabenkonzept wurde in der Folgezeit von vielen aufgegriffen – und auch die vorliegende Arbeit orientiert sich daran. Im nächsten Abschnitt wird dieses Konzept noch näher beleuchtet. Das Baddeley-Modell hat die experimentelle Forschung zum Arbeitsgedächtnis im europäischen Raum stark beeinflusst. Inzwischen liegen zahlreiche Befunde zum Baddeley-Modell aus verschiedenen Forschungsansätzen vor, insbesondere aus experimentellen Studien (z. B. Baddeley, Emslie, Kolodny & Duncan, 1998), aber auch aus Studien an Pbn mit spezifischen neuropsychologischen oder neurologischen Beeinträchtigungen (z. B. Della Sala, Gray, Baddeley, Allamano & Wilson, 1999; Hanley, Young & Pearson, 1991) sowie aus entwicklungspsychologischen Studien (z. B. Towse, Hitch & Hutton, 1998). Die Idee eines kapazitätsbegrenzten, aktiv an der Informationsverarbeitung beteiligten Arbeitsspeichers wurde bald auch in der Forschung zu interindividuellen Unterschieden aufgegriffen. Wenn es zutrifft, dass sich Individuen in der Kapazität ihres Arbeitsgedächtnisses unterscheiden, liegt der Gedanke nahe, zu versuchen, über diese basalen individuellen Unterschiede auch Unterschiede in höheren kognitiven Prozessen zu erklären.

### 3.2 Das Arbeitsgedächtnis aus der Perspektive der differentiellen Psychologie

Ausgangspunkt zur Untersuchung individueller Unterschiede im Arbeitsgedächtnis waren die Arbeiten von Daneman und Carpenter (1980) zum Leseverstehen bei Erwachsenen. Damals gingen viele Forscher davon aus, dass die Kapazität des Arbeitsgedächtnisses für die Lesefähigkeit relevant sei, da während des Lesens vielfältige Informationen zwischengespeichert werden müssen, um den Sinn eines Textes erschließen zu können (z. B. Just & Carpenter, 1980; Kintsch & van Dijk, 1978). Menschen mit einem größeren Arbeitsgedächtnis sollten somit bessere Leseleistungen aufweisen. Jedoch korrelierten die zu dieser Zeit verwendeten Maße – in der Regel „einfache Spannenmaße“, die eine reine Reproduktion von Informationen verlangten – nicht oder nur sehr schwach mit dem Leseverstehen (z. B. Perfetti & Goldman, 1976).

Daneman und Carpenter (1980) gelangten daraufhin zu der Ansicht, dass die Verwendung von Kurzzeitgedächtnisaufgaben, die nur sehr geringe Verarbeitungsanforderungen stellen, für die Vorhersage von Leseverstehensleistungen unzureichend ist. Sie entwickelten mit der „Reading Span“ einen Aufgabentyp, der zusätzlich zur Speicherung von Informationen deren Verarbeitung explizit erforderte. Damit nahmen sie Bezug auf die zentrale Exekutive im Modell von Baddeley und Hitch (1974). Die „Reading Span“ ist eine Aufgabe mit doppelter Anforderung: Zum einen werden die Pbn aufgefordert, sich voneinander unabhängige Wörter zu merken und diese korrekt wiederzugeben, zum anderen erfordert sie eine aktive Verarbeitung von verbalen Informationen. Eine ausführliche Beschreibung des Aufgabentyps findet sich in Abschnitt 4.1.1. Mit der „Reading Span“ gelang es Daneman und Carpenter, erstmals bedeutsame Zusammenhänge zwischen dem Arbeitsgedächtnis und komplexen Kognitionen, hier der Lesefähigkeit (erfasst über den *Verbal Scholastic Aptitude Test*), nachzuweisen (im Mittel:  $r = .66$ ). In einer Meta-Analyse konnten Daneman und Merikle (1996) den Zusammenhang von Reading Span-Aufgaben und Leseverstehen mit einer größeren Datenbasis erneut belegen ( $N = 38$  Studien;  $n = 2961$  Pbn). Aus den nach den Stichprobengrößen der einzelnen Studien gewichteten Effektstärken ergab sich ein mittlerer Zusammenhang zwischen der „Reading Span“ und Leseverstehentests von  $r = .41$ .

Andere Untersuchungen konnten in der Folge zeigen, dass der Zusammenhang nicht darauf beruht, dass es sich bei der „Reading Span“ und beim Leseverstehen um einander ähnliche, verbale Aufgaben handelt. Turner und Engle (1989) beispielsweise verwendeten strukturell äquivalente Aufgaben mit numerischem Material, die sich ebenfalls als gute Prädiktoren für das Leseverstehen erwiesen. Weitere einflussreiche Studien konnten später zeigen, dass die Leistungen in Arbeitsgedächtnisaufgaben in hohem Maße mit verschiedenen höheren kognitiven Leistungen korrespondieren (z. B. Kyllonen & Christal, 1990).

Bei der Untersuchung individueller Unterschiede wird das Arbeitsgedächtnis seitdem meist als eine globale kognitive Ressource konzipiert, die in die Aufrechterhaltung von Informationen bei simultaner Verarbeitung derselben oder anderer Information

involviert ist. Die Kapazität des Arbeitsgedächtnisses stellt zudem einen begrenzenden Faktor für viele verschiedene kognitive Leistungen dar (z. B. Engle, Kane & Tuholski, 1999; Just & Carpenter, 1992). Es wird dabei angenommen, dass individuelle Unterschiede in der Kapazität des Arbeitsgedächtnisses relativ stabil über die Zeit sind (Kyllonen & Christal, 1990). Auf das Modell von Baddeley und Hitch (1974) wird in unterschiedlicher Weise Bezug genommen. Häufig beziehen sich Studien zu individuellen Unterschieden ausschließlich oder vornehmlich auf die zentrale Exekutive des Arbeitsgedächtnisses (siehe z. B. Just & Carpenter, 1992).

Bisher gibt es keine allgemein akzeptierte Theorie darüber, warum komplexe Spannaufgaben wie die „Reading Span“-Aufgabe so gute Prädiktoren für verschiedene kognitive Leistungen sind (Miyake, 2001; Conway & Engle, 1996). Jedoch wurden in verschiedenen Forschergruppen Modellvorstellungen entwickelt, die diese Zusammenhänge beschreiben. Das klassische und bislang dominierende Modell stammt von Daneman und Carpenter (1980) und kann als *Resource-Sharing-Modell* bezeichnet werden (Hitch, Towse & Hutton, 2001). Es nimmt an, dass die Arbeitsgedächtniskapazität eine begrenzte Ressource ist, die für die Verarbeitungsfunktion und die Speicherfunktion – oder beide Funktionen gemeinsam – verwendet werden kann. Bei der Bearbeitung einer Arbeitsgedächtnisaufgabe erfolgt ein trade-off zwischen Speicher- und Verarbeitungsanforderungen. Gemessen wird somit die Kapazität kognitiver Ressourcen, die flexibel auf Verarbeitungs- und Speicheraktivitäten im Arbeitsgedächtnis verteilt werden kann. Das bedeutet, wenn eine Person beispielsweise über gute sprachliche Fähigkeiten verfügt, so benötigt sie für die Verarbeitungskomponente einer Reading-Span-Aufgabe relativ wenig kognitive Ressourcen. Die noch freien kognitiven Ressourcen kann sie daher für die Speicherung der Information aufwenden. Wenn sie jedoch über schwächere sprachliche Fähigkeiten verfügt, verbraucht sie bereits einen großen Teil ihrer Ressourcen bei der Verarbeitung der Information, so dass nur noch wenige Ressourcen für die Speicherung übrig bleiben. Die Behaltensleistung in einer Spannaufgabe ist daher ein Indikator dafür, über wie viel Ressourcen eine Person noch verfügen kann, wenn die Verarbeitung erfolgt ist. Diese Ressourcenteilung ist jedoch nur erforderlich, wenn in beiden Aufgaben Informationen aus derselben Domäne (verbal, numerisch oder räumlich) verarbeitet werden. Die Verarbeitung von Informationen aus unterschiedlichen Domänen greift hingegen nicht auf dieselben Ressourcen zurück. Entsprechend wird dieser Ansatz auch als „aufgabenspezifische Hypothese“ bezeichnet (Conway & Engle, 1996).

Seit Mitte der 90er Jahre wurden jedoch mehrere alternative Modellvorstellungen entwickelt, die die Gültigkeit dieser klassischen Sichtweise in Frage stellen (vgl. Miyake, 2001). Drei besonders einflussreiche aktuelle Ansätze werden im Folgenden kurz vorgestellt:

In der *Task-Switching-Hypothese* (Towse et al., 1998; Towse, Hitch & Hutton, 2000; Hitch et al., 2001) wird angenommen, dass komplexe Spannaufgaben ein Hin- und Herwechseln zwischen Verarbeitungs- und Speicherungsanforderungen erfordern. Die Effizienz, mit der die Verarbeitungsanforderung bewältigt wird, beeinflusst die erforderliche Dauer, während derer die Informationen aktiv gehalten werden müssen.

Je effizienter die Verarbeitung, desto kürzer müssen die Informationen behalten werden. Bei länger dauernden Prozessen steigt hingegen das Risiko, relevante Informationen zu vergessen. Es wird hier also nur von einer indirekten Beziehung zwischen Verarbeitung und Speicherung ausgegangen, es erfolgt – im Gegensatz zum Resource-Sharing-Modell – kein direkter trade-off.

Im *Ansatz der kontrollierten Aufmerksamkeit* (Engle, Kane et al., 1999; Kane, Bleckley, Conway & Engle, 2001) wird angenommen, dass das Arbeitsgedächtnis gemeinsam durch das Kurzzeitgedächtnis und durch „kontrollierte Aufmerksamkeit“ bestimmt wird. Letztere wird verstanden als domänen-unspezifische, begrenzte Aufmerksamkeitskapazität für die Ausführung kontrollierter Verarbeitung bzw. als die Fokussierung auf aufgabenrelevante Informationen in Gegenwart von interferierenden Stimuli. In den Begriffen des Baddeley-Modells wird die Arbeitsgedächtniskapazität hier definiert als Kapazität der *slave systems* plus der Effizienz der zentralen Exekutive. Dieser Ansatz fokussiert auf die Rolle exekutiver Kontrolle bzw. kontrollierter Aufmerksamkeit, also nicht wie das Resource-Sharing-Modell auf die Verarbeitung von Informationen. Leistungsunterschiede in komplexen Spannaufgaben werden in diesem Ansatz somit durch das Kurzzeitgedächtnis und die Fähigkeit, sich auf relevante Informationen zu konzentrieren, erklärt.

Nach dem *inhibitionsbasierten Ansatz* (Hasher & Zacks, 1988; siehe auch May, Hasher & Kane, 1999; Lustig, May & Hasher, 2001) ist die Annahme einer begrenzten Verfügbarkeit von Ressourcen, wie sie das Resource-Sharing-Modell macht, überflüssig. Die Leistungsfähigkeit des Arbeitsgedächtnisses wird hier verstanden als Ausdruck der Fähigkeit, irrelevante Informationen zu hemmen bzw. zu unterdrücken. Der Inhalt des Arbeitsgedächtnisses kann so auf relevante Informationen beschränkt werden. Nicht das Ausmaß verfügbarer Ressourcen bestimmt somit die Leistungsfähigkeit in kognitiven Aufgaben, sondern die Fähigkeit, nur die richtigen Inhalte in das Arbeitsgedächtnis einzulesen. Als Belege für diesen Ansatz werden Befunde angeführt, dass die Leistungen in Spannaufgaben zumindest teilweise davon abhängen, wie gut es einer Person gelingt, die Effekte proaktiver Interferenz zu bewältigen (May, Hasher & Kane, 1999).

Das Resource-Sharing-Modell und die verschiedenen neueren Ansätze schließen sich nicht unbedingt gegenseitig aus, da sich jeder der Ansätze jeweils nur mit einem spezifischen Aspekt individueller Unterschiede im Arbeitsgedächtnis beschäftigt. Die Leistung in komplexen Spannaufgaben wird vermutlich jedoch durch multiple Faktoren bestimmt. Miyake (2001) gibt zu bedenken, dass die Faktoren, die die Spannenwerte beeinflussen, nicht unbedingt dieselben sein müssen, die auch das Ausmaß des Zusammenhangs zwischen Arbeitsgedächtnisleistung und kognitiven Aufgaben beeinflussen. Lustig et al. (2001) konnten beispielsweise zwar zeigen, dass eine experimentelle Manipulation von Interferenz die Leistung in der Spannaufgabe beeinflusst. Aber es ist nicht klar, ob dieser Effekt auch für interindividuelle Unterschiede verantwortlich ist. Auch in der Studie von Hitch et al. (2001) konnte zwar experimentell gezeigt werden, dass die Länge der zu merkenden Sätze einen Einfluss auf die Leistung in der Spannaufgabe hat. Das bedeutet, dass eine längere Verarbeitung von Informationen auch ein Behalten der Information über längere Zeit

erfordert. Jedoch ist auch hier nicht klar, ob dieser Faktor für interindividuelle Unterschiede verantwortlich ist. Es ist daher wichtig, diejenigen Faktoren, die die Spannenwerte beeinflussen, konzeptuell von denjenigen Faktoren zu trennen, die als Quelle individueller (oder Gruppen-) Unterschiede fungieren.

Trotz unterschiedlicher Modellvorstellungen zu individuellen Unterschieden besteht weitgehender Konsens hinsichtlich einer operationalen Definition des Arbeitsgedächtnisses:

Das Arbeitsgedächtnis wird genutzt, wenn eine Aufgabe eine gleichzeitige Verarbeitung und Speicherung von Informationen erfordert. Weiter wird angenommen, dass die Kapazität, diese Operationen auszuführen, begrenzt ist (z. B. Tuholski, Engle & Baylis, 2001; siehe auch Baddeley, 1986; Kyllonen & Christal, 1990).

Dabei wird unter „Speicherung“ das Behalten neuer Informationen verstanden, die kurz zuvor für einige Sekunden präsentiert wurden. „Verarbeitung“ bezieht sich auf die Manipulation oder Transformation von Informationen (Oberauer, 2005; Oberauer et al., 2000). Es kann dabei entweder die Verarbeitung der zuvor gespeicherten oder anderer Informationen erforderlich sein.

Entsprechende Aufgaben werden als „komplexe Spannenaufgaben“ bezeichnet (in Abgrenzung zu „einfachen Spannenaufgaben“, die keine simultane Verarbeitung der Informationen verlangen). Komplexe Spannenaufgaben werden meist dazu verwendet, andere kognitive Leistungen vorherzusagen (z. B. Leseverstehen, Reasoning). Jedoch weisen verschiedene Studien darauf hin, dass Aufgaben, die dieser operationalen Definition entsprechen, zwar gute Arbeitsgedächtnisaufgaben sind, dass es jedoch auch Arbeitsgedächtnisaufgaben mit guter prädiktiver Validität geben kann, die von dieser Definition abweichen (z. B. Oberauer, 2005; Oberauer et al., 2000). Baddeley (1995) beispielsweise schließt aus seinen Untersuchungen, dass valide Arbeitsgedächtnisaufgaben nicht unbedingt eine Speicherung von Informationen erfordern müssen. Er konnte zeigen, dass Aufgaben mit Gedächtnisanforderung und solche ohne Gedächtnisanforderung ein relativ kohärentes Ergebnismuster zeigen. Andere Autoren hingegen ziehen die Notwendigkeit einer Verarbeitungskomponente in Arbeitsgedächtnisaufgaben in Zweifel. Dieser Position zufolge sind einfache Spannenaufgaben, die keine Verarbeitung der Informationen erfordern, ebenfalls gut dazu in der Lage, individuelle Unterschiede im Arbeitsgedächtnis abzubilden (z. B. de Jonge & de Jong, 1996; Oberauer, 2005). Weiter gibt es Hinweise darauf, dass insbesondere solche Speicheraufgaben ohne Verarbeitungsanforderung gute Indikatoren für das Arbeitsgedächtnis sind, in denen eine räumliche Komponente eine Rolle spielt (z. B. Oberauer, 1993). Eine ausführlichere Beschreibung typischer Arbeitsgedächtnisaufgaben, die bei der Untersuchung individueller Unterschiede verwendet werden, folgt in Kapitel 4.

In der differentiellen Psychologie werden Theorien (z. B. Intelligenztheorien, vgl. z. B. Carroll, 1993; Jäger, 1982) zunächst häufig über bestimmte Zuordnungsregeln beschrieben, die ein theoretisches Konstrukt mit spezifischen Verhaltensweisen in bestimmten Situationen verknüpfen (vgl. Brocke & Beauducel, 2001). Für das



Arbeitsgedächtnis lauten entsprechende Zuordnungsregeln etwa, dass eine Person bei einer Arbeitsgedächtnisaufgabe, an der sie teilnimmt, eine bestimmte Spanne erreicht, wenn sie über eine durchschnittlich ausgeprägte Arbeitsgedächtniskapazität verfügt. Diese Zuordnungsregeln werden in der Regel probabilistisch verstanden, das heißt, es werden die Korrelationen zwischen Variablen oder zwischen Variable(n) und Trait(s) betrachtet und theoretisch eingeordnet. Zwar existieren für das Arbeitsgedächtnis verschiedene theoretische Modellvorstellungen (s. o.), die experimentell überprüft wurden und werden, so dass die differentielle Forschung nicht nur auf die Faktorenanalyse angewiesen ist, um das Konstrukt „Arbeitsgedächtniskapazität“ identifizieren zu können (Kyllonen, 1996). Jedoch konzentriert sich ein großer Teil der differentiellen Forschung zum Arbeitsgedächtnis – ebenso wie in der Intelligenzforschung – auf die Exploration und Überprüfung der Faktorenstruktur. Hierbei wird zunächst explorativ von den Zusammenhängen verschiedener Arbeitsgedächtnisaufgaben ausgegangen, jedoch wird vielfach auch auf bereits gesicherte Befunde aus der Intelligenzforschung in Form von Hypothesen für mögliche ähnliche Zusammenhänge beim Arbeitsgedächtnis zurückgegriffen. So wurde beispielsweise verschiedentlich angenommen, dass sich die Inhaltsfaktoren verbal, numerisch und räumlich bzw. figural (vgl. Kapitel 2) auch für das Arbeitsgedächtnis nachweisen lassen sollten (z. B. Oberauer et al., 2000). Im Rahmen der Testkonstruktion stellt die Untersuchung der Faktorenstruktur des zu messenden Konstrukts zudem einen wichtigen Beitrag zur Konstruktvalidierung von Testskalen dar. Befunde zur Faktoren- oder Binnenstruktur des Arbeitsgedächtnisses werden daher im Folgenden in Abschnitt 3.3 beschrieben und diskutiert.

Für bereits weiterentwickelte Theorien in der differentiellen Psychologie gilt, dass sie mit umfassenden nomologischen Netzwerken verknüpft sein sollten (Brocke & Beauducel, 2001). Beiträge zu einem solchen nomologischen Netzwerk des Arbeitsgedächtnisses sind beispielsweise Aussagen über den Zusammenhang von Arbeitsgedächtnis und Intelligenz sowie anderen kognitiven Konstrukten oder Befunde zur Entwicklung des Arbeitsgedächtnisses in der Kindheit und über die Lebensspanne. Analysen des Zusammenhangs von Arbeitsgedächtnis (sowie ggf. weiteren kognitiven Variablen) und Intelligenz erlauben auch Aussagen zu kognitiven Prozessen im Sinne des Kognitive-Korrelate-Ansatzes (vgl. Pellegrino & Glaser, 1979). In diesem Ansatz der differentiellen Kognitionspsychologie wird die Frage gestellt, inwiefern individuelle Unterschiede in bestimmten Aspekten der Informationsverarbeitung dazu geeignet sind, individuelle Unterschiede in Intelligenztestleistungen zu erklären. Häufig gezeigt werden konnte bereits, dass Arbeitsgedächtnis- und Intelligenztestleistungen gemeinsame Varianz aufweisen. Umstritten ist jedoch die Größe des Varianzanteils von Intelligenz, der sich durch die Leistungsfähigkeit des Arbeitsgedächtnisses erklären lässt (z. B. Ackerman, Beier & Boyle, 2002). Auch im Hinblick auf eine Testkonstruktion sind Zusammenhangsanalysen von Arbeitsgedächtnisaufgaben und anderen kognitiven Tests von Interesse. Die kognitiven Tests werden hier als Außenkriterien verstanden, und die Korrelationen mit den Arbeitsgedächtnisaufgaben werden als Belege für die Kriteriumsvalidität der entsprechenden Aufgaben interpretiert. Entsprechende Befunde zu den Zusammenhängen von Arbeitsgedächtnis, Intelligenz und anderen kognitiven Variablen werden in Kapitel 3.4 dargestellt.

### 3.3 Die Binnenstruktur des Arbeitsgedächtnisses

Der differentielle Ansatz in der Arbeitsgedächtnis-Forschung beschäftigt sich seit den Anfängen mit der Frage nach der Binnenstruktur des Arbeitsgedächtnisses. Unumstritten ist, dass verschiedene Arbeitsgedächtnisaufgaben, die eine simultane Verarbeitung und Speicherung von Informationen erfordern, in der Regel bedeutsam miteinander korrelieren. Faktorenanalysen über die Daten aus verschiedenen Arbeitsgedächtnisaufgaben zeigen bisher jedoch ein inkonsistentes Ergebnismuster. Auch die Zusammenhänge von Arbeitsgedächtnisaufgaben und anderen kognitiven Leistungen fallen in verschiedenen Studien unterschiedlich aus. Es wird daher kontrovers diskutiert, ob es sich beim Arbeitsgedächtnis um einen einzigen, globalen Faktor handelt, oder ob verschiedene separierbare Faktoren unterschieden werden können. Bereits Baddeley und Hitch (1974) hatten aufgrund der Befunde aus experimentellen Studien ein einheitliches Modell abgelehnt und stattdessen ein modulares Modell vorgeschlagen (vgl. Abschnitt 3.1). In diesem Modell verarbeitet die phonologische Schleife sprachliche Informationen und der visuell-räumliche Skizzenblock visuelle Informationen.

Im differentiellen Ansatz wird der Frage nach der Binnenstruktur meist mithilfe der traditionellen Methoden der Differentiellen Psychologie – korrelative Daten und Analyse latenter Faktoren – nachgegangen. Vergleichbar mit dem Vorgehen in der psychometrischen Intelligenzforschung werden die Zusammenhänge zwischen den Leistungen von Probanden (Pbn) in verschiedenen Arbeitsgedächtnisaufgaben faktorenanalytisch untersucht. In den meisten Studien werden erwachsene Pbn untersucht, in einigen Studien auch Kinder. Eine explizite Untersuchung der Binnenstruktur für verschiedene Altersgruppen im Kindesalter wurde bislang jedoch noch nicht vorgenommen. Ebenfalls liegen bisher keine Befunde zur Binnenstruktur des Arbeitsgedächtnisses von hoch begabten Kindern vor.

Es existieren viele verschiedene Arbeitsgedächtnisaufgaben, die vor allem in der experimentellen Forschung entwickelt wurden und die unterschiedliche kognitive Funktionen und Domänen ansprechen (siehe Kapitel 4). Die Ergebnisse von Faktorenanalysen hängen jedoch stets davon ab, welche Aufgaben und welche Pbn untersucht werden. Oberauer et al. (2000) sowie Süß, Oberauer, Wittmann, Wilhelm und Schulze (2002) kritisieren an bisherigen Studien eine relativ unsystematische Operationalisierung des Arbeitsgedächtnisses, die der Breite und Differenziertheit des Konzepts nicht gerecht wird. Auf der Grundlage von Befunden aus der Intelligenzstrukturforschung formulierten sie zunächst – in Analogie zum Berliner Intelligenzstrukturmodell (BIS; Jäger, 1982, siehe Kapitel 2) – ein hypothetisches Zweifacetten-Modell mit je drei Kategorien. Die Funktionale Facette des Modells umfasst verschiedene Funktionen, die dem Arbeitsgedächtnis häufig zugeschrieben werden. Neben der klassischen und von vielen Forschern verwendeten Funktionszuschreibung des Arbeitsgedächtnisses der „simultanen Speicherung und Transformation“ enthält die Funktionsfacette mit „Koordination“ und „Supervision“ zwei weitere Kategorien (siehe Abbildung 3.1). Unter „Koordination“ verstehen Oberauer et al. (2000) die Integration verschiedener Informationen, die den simultanen Zugriff auf verschiedene Elemente erfordert. Dabei können Speicherprozesse relevant sein, dies

ist aber nicht zwingend. Ein Koordinationsprozess erfolgt z. B. bei der mentalen Konstruktion eines Gesamtbildes aus mehreren Teilbildern. Die Funktion „Supervision“ umfasst die Überwachung und Kontrolle von mentalen Operationen sowie die selektive Aktivierung und Hemmung von Schemata. Oberauer et al. (2000) legten ihr Modell somit deutlich breiter an als es in Studien zu individuellen Unterschieden im Arbeitsgedächtnis sonst üblich ist. In der Regel erfolgt eine Beschränkung auf die Kategorie „Simultane Speicherung und Transformation“. Die Inhaltsfacette beinhaltet die Domänen des Aufgabenmaterials (verbal, numerisch und räumlich). Aufgrund der engen Verwandtschaft von Intelligenz und Arbeitsgedächtnis (z. B. Kyllonen & Christal, 1990; vgl. Kapitel 3.4) stellten Oberauer et al. (2000) die Hypothese auf, dieselbe inhaltliche Gliederung könnte auch für das Arbeitsgedächtnis zutreffen.

| Inhalte \ Funktionen                            | <i>verbal</i>       | <i>numerisch</i>                         | <i>räumlich</i>                        |
|---|---------------------|--|--|
| <b>Simultane Speicherung und Transformation</b> | Reading Span        | Backward Digit Span                      | Spatial Working Memory                 |
| <b>Koordination</b>                             | Verbal Coordination | Memory Updating Numerical (STM-Version)* | Memory Updating Spatial (STM-Version)* |
| <b>Supervision</b>                              | Switching Verbal    | Switching Numerical                      | Switching Figurative                   |

\* Reine Speicheranforderung, keine Verarbeitungsanforderung

Abbildung 3.1: Hypothetisches Zwei-Facetten Modell der Arbeitsgedächtniskapazität mit je einer der verwendeten Aufgabentypen pro Zelle (Oberauer et al., 2000; Süß et al., 2002)

Zur Prüfung des Zwei-Facetten-Modells analysierten Oberauer et al. (2000) und Süß et al. (2002) die Daten erwachsener, überwiegend studentischer Pbn ( $N = 128$ ), die eine Testbatterie mit 23 verschiedenen Arbeitsgedächtnisaufgaben bearbeitet hatten. Die Testbatterie war so zusammengestellt, dass alle Zellen des Modells durch jeweils mehrere Aufgaben repräsentiert wurden. Eine exploratorische Faktorenanalyse ergab drei Faktoren, die wie folgt interpretiert wurden:

- Faktor 1: „Speichern/Transformation und Koordination – verbal-numerisch“
- Faktor 2: „Speichern/Transformation und Koordination – räumlich“
- Faktor 3: „Supervision“ (möglicherweise auch „Bearbeitungsgeschwindigkeit“)

Es zeigte sich somit sowohl eine Differenzierung in der Inhaltsfacette (verbal-numerisch vs. räumlich) als auch in der Funktionsfacette. Die Aufgaben, die von den Autoren als Supervisions-Aufgaben klassifiziert worden waren, luden nicht gemeinsam mit den anderen Arbeitsgedächtnisaufgaben auf einem Faktor. In konfirmatorischen Faktorenanalysen wurden verschiedene Modelle getestet, die die hypothetische Facettenstruktur und Variationen davon abbildeten. Hier zeigte sich, dass eine Unterteilung in die funktionale und die Inhaltsfacette gut dazu geeignet scheint, die Daten zu erklären. Für beide Facetten ließen sich je zwei der drei angenommen Faktoren

belegen. Nach den Kriterien Datenfit und Sparsamkeit erwies sich ein Modell mit vier Faktoren insgesamt als das beste Modell (Funktionale Faktoren: „Speichern/Transformation und Koordination“, „Supervision bzw. Bearbeitungsgeschwindigkeit“; inhaltliche Faktoren: „verbal-numerisch“, „räumlich“). Eine weitere Reduktion der Faktoren (z. B. Fusion des verbal-numerischen und des räumlichen Faktors) ergab eine signifikant schlechtere Modellpassung. Die beiden inhaltlichen Faktoren wiesen jedoch eine beträchtliche Korrelation auf ( $r = .82$ ; funktionale Faktoren:  $r = .46$ ). Süß et al. (2002) analysierten dieselben Daten (jedoch ohne einige der Arbeitsgedächtnisaufgaben) für jede Facette separat. Sowohl ein funktionales Modell mit den beiden Faktoren „Speichern/Transformation und Koordination“ und „Supervision“ als auch ein inhaltliches Modell mit den beiden Faktoren „verbal-numerisch“ und „räumlich“ ergab eine gute Modellpassung. In beiden Modellen korrelierten die beiden latenten Faktoren bedeutsam (inhaltliches Modell:  $r = .80$ ; funktionales Modell:  $r = .53$ ).

Oberauer, Süß, Wilhelm und Wittmann (2003) testeten eine überarbeitete Version des Facetten-Modells an 133 Studierenden mit neu konstruierten Aufgaben. Aufgrund ihrer früheren Ergebnisse nahmen sie nun an, dass die Inhaltsfacette nur aus den Kategorien verbal-numerisch und räumlich besteht. Die Funktionsfacette wurde über einen neuen Ansatz erfasst: Die Kategorie „simultane Speicherung und Transformation“ wurde wie bisher über Dual-Task-Aufgaben gemessen. Die Idee einer reinen Erfassung der Funktionen „Koordination“ und „Supervision“ mittels jeweils spezifischer Aufgaben, die von Oberauer et al. (2000) angestrebt worden war, wurde nun jedoch verworfen. Es wird argumentiert, dass die reine Messung von Supervisions- oder Koordinationsfunktionen prinzipiell nicht möglich ist. Alle Aufgaben, die Supervision messen, erfordern zwangsläufig auch grundlegende Prozesse – solche, die supervidiert werden. Diese grundlegenden Prozesse fließen in die Gesamtleistung mit ein. Koordinationsaufgaben erfordern zusätzlich zu Koordinationsprozessen Speicherleistungen. Um dennoch beide Funktionen erfassen zu können, wurden nun die einzelnen Komponenten dieser Anforderungen über verschiedene Aufgaben gemessen. So wurde etwa die Supervisionsfunktion über eine Aufgabe gemessen, die sowohl die Verarbeitung von Informationen als auch Supervision erforderte, und zusätzlich über eine andere Aufgabe, die nur die Verarbeitung von Informationen erforderte. Das Ausmaß geleisteter Supervision wurde dann aus dem Verhältnis der Leistungen in beiden Aufgabentypen bestimmt. Die Koordinationsleistung wurde aus dem Verhältnis der Leistungen in einer Koordinations- und einer reinen Speicheraufgabe ermittelt.

Die über dieses Vorgehen ermittelten Daten wiesen einen guten Modell-Fit für ein Fünf-Faktoren-Modell auf (drei funktionale und zwei inhaltliche Faktoren;  $\chi^2 = 42.0$ ,  $df = 38$ ; CFI = .99, jedoch zum Teil nicht signifikante Ladungen auf den Inhaltsfaktoren). In einer vereinfachten Modellvariation wurden keine Inhaltsfaktoren spezifiziert ( $\chi^2 = 53.02$ ,  $df = 49$ ; CFI = .98), dieses Modell ergab keinen signifikant schlechteren Fit als das Fünf-Faktoren-Modell ( $\Delta\chi^2 = 13.02$ ,  $\Delta df = 13$ ;  $p = .29$ ). Der Faktor „simultane Speicherung und Transformation“ korrelierte hoch mit „Koordination“ ( $r = .78$ ) und schwächer mit „Supervision“ ( $r = .28$ ). „Supervision“ und „Koordination“ korrelierten zu  $r = .31$  miteinander. Es gelang somit eine Modellierung der

Funktionsfacette, auch eine Differenzierung der Inhaltsfacette war möglich. Die Ergebnisse weisen auf eine funktionale Differenzierung hin, die Belege für eine Unterteilung der Inhaltsfacette in zwei Kategorien sind jedoch eher schwach.

Wie zuvor erwähnt, werden in der überwiegenden Zahl der Studien zu individuellen Unterschieden ausschließlich solche Arbeitsgedächtnisaufgaben verwendet, die auf die Erfassung des Aspekts der „simultanen Speicherung und Transformation“ abzielen. Entsprechend konzentriert sich die Diskussion zur Binnenstruktur des Arbeitsgedächtnisses vor allem auf die Inhaltsfacette. Häufig wird die inhaltliche Struktur auch unter der Überschrift „Domänenspezifität vs. Domänengeneralität“ thematisiert. Da in vielen Studien nicht zwischen verbalen und numerischen Arbeitsgedächtnisaufgaben faktoriell differenziert werden konnte, ist die Diskussion inzwischen in der Regel zugespitzt auf die Frage: Gibt es unterschiedliche Arbeitsgedächtnissysteme für verbal-numerische und für visuell-räumliche Arbeitsgedächtnisaufgaben?

Die Untersuchung der Domänenspezifität bzw. -generalität erfolgt auf zwei Wegen. Der deutlich häufiger verfolgte Ansatz besteht in der Korrelation der Arbeitsgedächtnisaufgaben mit verschiedenen anderen kognitiven Leistungen. Es wird dabei der Frage nachgegangen, ob die Leistungen in Arbeitsgedächtnisaufgaben nur mit solchen kognitiven Leistungen korrespondieren, die derselben Domäne entstammen („Task-Specific-Hypothese“, z. B. Daneman & Carpenter, 1980; Just & Carpenter, 1992), oder ob bedeutsame Zusammenhänge auch über die Domänen hinweg bestehen („General Capacity-Hypothese“, z. B. Engle, Cantor & Carullo, 1992; Engle, Kane et al., 1999; Swanson, 1996). Dieser Ansatz entspricht dem Vorgehen bei einer Kriteriumsvalidierung. Ein weiterer Ansatz ist die Betrachtung der Zusammenhänge verschiedener Arbeitsgedächtnisaufgaben untereinander. Für die Analyse dieser Zusammenhangsdaten werden exploratorische und/oder konfirmatorische Faktorenanalysen (EFA bzw. CFA) verwendet. In der Regel werden hier ein genereller oder zwei korrelierende domänenspezifische Faktoren (verbal-numerisch vs. räumlich) erwartet. Dieser zweite Ansatz entspricht dem Vorgehen bei einer Konstruktvalidierung im Sinne der Prüfung von faktorieller Validität.

Zunächst werden im Folgenden die Befunde aus dem ersten Untersuchungsansatz berichtet, der sich mit den Zusammenhängen des Arbeitsgedächtnisses mit anderen kognitiven Variablen beschäftigt. Anschließend werden die Ergebnisse aus Studien des zweiten Untersuchungsansatzes referiert, in dem Strukturanalysen im Mittelpunkt stehen.

### **3.3.1 Studien zu Zusammenhängen von Arbeitsgedächtnisaufgaben mit anderen kognitiven Leistungen**

Bereits Turner und Engle (1989) stellten – in Erweiterung der Arbeiten von Daneman und Carpenter (1980) – fest, dass nicht nur die „Reading Span“, sondern auch die „Operation Span“ (ermittelt über Aufgaben mit einer numerischen Verarbeitungskomponente und einer verbalen Speicherkomponente) ein signifikanter Prädiktor für die Lesefähigkeit ist. Diese Befunde standen im Gegensatz zur ursprünglichen

Annahme von Daneman und Carpenter (1980), dass die Prozesse beim Bearbeiten der „Reading Span“ spezifisch für sprachliche Prozesse sind.

Eine Reihe weiterer Studien beschäftigte sich mit Unterschieden in der Erklärung verbal-numerischer versus räumlicher kognitiver Leistungen durch verschiedene Arbeitsgedächtnisaufgaben. So fanden etwa Daneman und Tardif (1987), dass verbale und numerische Arbeitsgedächtnisaufgaben signifikant mit verbalen Fähigkeitsmaßen korrelieren ( $r = .61$  bzw.  $r = .51$ ). Für eine räumliche Arbeitsgedächtnisaufgabe konnte hingegen kein signifikanter Zusammenhang zu verbalen Leistungen nachgewiesen werden ( $r = -.09$ ). Diese Befunde können als Belege für eine Domänenspezifität interpretiert werden.

Ebenfalls für eine Domänenspezifität sprechen die Ergebnisse einer Studie von Shah und Miyake (1996). Hier erwies sich die Aufgabe „Reading Span“ als ein guter Prädiktor für die Leistung im verbalen Scholastic Aptitude Test (SAT), zwei räumliche Aufgaben konnten die verbale SAT-Leistung hingegen nicht signifikant vorhersagen ( $r = .45$  bzw.  $r = .07$ , n.s. und  $r = .02$ , n.s.). Die Leistungen in verschiedenen visuell-räumlichen Tests konnten nur durch die beiden räumlichen Arbeitsgedächtnisaufgaben vorhergesagt werden, nicht jedoch durch die verbale Arbeitsgedächtnisaufgabe ( $r = .66$  und  $r = .62$  bzw.  $r = .12$ , n.s.). Schließlich ergab eine exploratorische Faktorenanalyse über die Arbeitsgedächtnisaufgaben, den SAT und die visuell-räumlichen Tests zwei Inhaltsfaktoren (verbal und räumlich). Zusätzlich konstruierten Shah und Miyake (1996) so genannte „domain-mismatching“-Tasks: Bei diesen Aufgaben bestanden die Speicher- und die Verarbeitungskomponenten aus verschiedenen Domänen; beispielsweise erforderte eine Aufgabe eine verbale Verarbeitung, aber eine Speicherung von räumlichen Inhalten. Diese Aufgaben wurden – ebenso wie „domain-matching“ Tasks – mit den anderen Tests korreliert. Es zeigte sich, dass für den Zusammenhang mit anderen Maßen vor allem die Domäne der Speicheranforderung relevant war, die Domäne der Verarbeitungsanforderung hatte hingegen nur einen geringen Einfluss auf die Zusammenhänge. Die Autoren schlussfolgerten, dass die Domänenspezifität insbesondere auf der Speicherkomponente beruht und Verarbeitungsprozesse eher domänen-unspezifisch sind.

Jedoch finden sich auch etliche Befunde, die eher eine Domänengeneralität nahe legen. In verschiedenen Studien (z. B. Süß et al., 2002; Bayliss, Jarrold, Gunn & Baddeley, 2003) erwiesen sich verbale Arbeitsgedächtnisaufgaben als genauso gute Prädiktoren für räumliche Fähigkeiten wie räumliche Arbeitsgedächtnisaufgaben – und umgekehrt. Es gibt bislang nur wenige Studien, die diesen Ansatz verfolgen, und die die Frage der Domänengeneralität oder -spezifität des Arbeitsgedächtnisses bei Kindern untersucht haben (Hitch et al., 2001). Eine der wenigen Untersuchungen dieser Art stammt von Swanson (1996) und deutet eher auf eine Domänengeneralität hin. Er konnte für Pbn im Alter von 5 bis 19 Jahren zeigen, dass die Beziehungen zwischen Arbeitsgedächtnis und Test- sowie Schulleistungen nicht abhängig von der Art der Arbeitsgedächtnisaufgabe waren. Sowohl verbale als auch visuell-räumliche Arbeitsgedächtnisaufgaben korrelierten signifikant mit verschiedenen Leistungs- und Intelligenzaufgaben.

### 3.3.2 Studien zu Zusammenhängen von verschiedenen Arbeitsgedächtnisaufgaben

In der Regel finden sich moderate bis hohe Korrelationen zwischen den verschiedenen Arbeitsgedächtnisaufgaben mit unterschiedlichen Inhalten, die in einer Untersuchung eingesetzt werden. Lässt sich dieser stabile Befund bereits als Beleg für eine Domänengeneralität interpretieren? Oberauer (2005) weist in diesem Zusammenhang darauf hin, dass die Interpretation entsprechender Befunde deutlich von den theoretischen Vorannahmen abhängt: Positive Korrelationen zwischen Arbeitsgedächtnisaufgaben aus verschiedenen Domänen können – je nach theoretischer Position – zur Stützung sowohl der Generalitäts- als auch der Spezifitätshypothese herangezogen werden. Die Generalitätshypothese wird gestützt, wenn die positiven Zusammenhänge als Beleg für die Konstruktvalidität verstanden werden. Wenn hingegen von vornherein von einer Domänenspezifität ausgegangen wird, können Korrelationen zwischen Aufgaben aus verschiedenen Domänen als Kriteriumsvalidität interpretiert werden. In letzterem Fall wird die Korrelation als Ausdruck des Zusammenhangs zweier ähnlicher Konstrukte verstanden. Letztlich hängt es somit von der Ausgangshypothese und der Höhe der gefundenen Zusammenhänge ab, wie Korrelationen interpretiert werden.

Eine Schwierigkeit im oben beschriebenen Ansatz (Abschnitt 3.3.1) besteht darin, dass in vielen Studien lediglich ein Aufgabentyp pro Inhaltsbereich verwendet wird. Die entsprechenden Konstrukte werden so jedoch nicht sehr reliabel und valide erfasst. Da einzelne Aufgaben zudem immer multiple Prozesse widerspiegeln, bleibt unklar, welche durch die Aufgabe hervorgerufenen Prozesse die Zusammenhänge mit anderen Fähigkeiten bestimmen. Es kann so nicht festgestellt werden, ob es sich um eine Domänenspezifität innerhalb des Arbeitsgedächtnisses handelt oder um eine Domänenspezifität in anderen kognitiven Prozessen (Kane et al., 2004). Wenn hingegen mehrere Aufgaben pro Konstrukt verwendet werden, können latente Variablen gebildet werden. So wird auf statistische Weise die Fehlervarianz, die jeder einzelnen Aufgabe anhaftet, entfernt. Behalten wird nur die Varianz, die alle Maße teilen. Diese geteilte Varianz repräsentiert das latente Konstrukt. Neuere Studien zur Binnenstruktur des Arbeitsgedächtnisses verwenden zunehmend mehrere Aufgabentypen pro erwartetem (oder möglichem) Inhaltsfaktor, so dass die Extraktion eines latenten Faktors bzw. mehrerer latenter Faktoren möglich ist.

Verschiedene Studien, die exploratorische Faktorenanalysen verwendeten, fanden über Arbeitsgedächtnisaufgaben aus verschiedenen Domänen einen Faktor (z. B. Colom, Flores-Mendoza & Rebollo, 2003). Aufschlussreicher sind jedoch konfirmatorische Faktorenanalysen, da hier die Passung bestimmter Faktorenstrukturen auf die Daten explizit geprüft werden kann.

Conway, Cowan, Bunting, Theriault und Minkoff (2002) verwendeten drei komplexe verbale und numerische sowie drei einfache verbale und numerische Spannaufgaben. Sie konnten mit konfirmatorischen Faktorenanalysen zeigen, dass sämtliche komplexen Spannaufgaben einen gemeinsamen Faktor bilden, ebenso sämtliche einfachen Spannaufgaben. Es findet sich somit eine innere Differenzierung nach

der Funktion (Speichern und Verarbeiten vs. nur Speichern), nicht aber nach dem Inhalt. Auch Kyllonen und Christal (1990) verwendeten verbale und numerische Arbeitsgedächtnisaufgaben und stellten fest, dass alle einen gemeinsamen Faktor bilden (siehe aber auch die Reanalyse von Jurden, 1995, vgl. Abschnitt 3.3.3). Jedoch fehlen in diesen Untersuchungen räumliche Aufgaben, so dass die Befunde keine Aussage über die oben genannte Aufteilung der Inhaltsfacette in verbal-numerische und räumliche Aufgaben erlauben.

Kane et al. (2004) analysierten die Arbeitsgedächtnisstruktur anhand von drei verbal-numerischen und drei räumlichen Arbeitsgedächtnisaufgaben. Sie testeten ein Ein-Faktor- und ein Zwei-Faktoren-Modell. Beide Modelle wiesen einen sehr guten Datenfit auf ( $N = 236$ ; Ein-Faktor-Modell:  $\chi^2 = 9.87$ ,  $df = 5$ ; CFI = .99; NFI = .98; NNFI = .98; Zwei-Faktoren-Modell:  $\chi^2 = 6.75$ ,  $df = 4$ ; CFI = 1.00; NFI = .99; NNFI = .98), der Fit des sparsameren Ein-Faktor-Modells war jedoch nicht signifikant schlechter als der des Zwei-Faktoren-Modells. Im Zwei-Faktoren-Modell korrelieren die beiden latenten Variablen sehr hoch miteinander ( $r = .93$ ; wenn keine Korrelationen der Residuen zugelassen wurden:  $r = .84$ ); beide latente Faktoren teilen somit etwa 70-85% ihrer Varianz.

Separate verbale, numerische und räumliche bzw. figurale Aufgaben verwendeten Ackerman et al. (2002) in ihrer Untersuchung mit 135 erwachsenen Pbn. Sie fanden in exploratorischen und konfirmatorischen Faktorenanalysen ebenfalls einen gemeinsamen Faktor (Modell-Fit:  $\chi^2 = 21.57$ ,  $df = 14$ ,  $p = .09$ ; RMSEA = .06; CFI = .97). Einen generellen Arbeitsgedächtnisfaktor fanden auch de Jong und Das-Smaal (1995) in ihrer Untersuchung mit 381 Kindern der vierten Klassenstufe (Modell-Fit:  $\chi^2 = 4.11$ ,  $df = 2$ ,  $p = .13$ ; NNFI = .98; CFI = .99). Sie verwendeten den „Star Counting Test“ (vgl. Kapitel 4) sowie drei auditiv vorgegebene numerische und räumliche Arbeitsgedächtnisaufgaben. In der Arbeit von de Jonge und de Jong (1996) mit Kindern der vierten bis sechsten Klasse fand sich ebenfalls ein gemeinsamer Arbeitsgedächtnisfaktor. Neben dem „Star Counting Test“ wurden hier zwei einfache und zwei komplexe Spannaufgaben verwendet, davon jeweils eine Aufgabe mit verbalem und eine Aufgabe mit numerischem Inhalt. In dem Modell, das am besten auf die Daten passte, wurden ein gemeinsamer Arbeitsgedächtnisfaktor und ein separater Faktor für den „Star Counting Test“ spezifiziert (Modell-Fit:  $\chi^2 = 69.83$ ,  $df = 47$ ,  $p = .04$ ; NNFI = .94; CFI = .93). Dieser Befund steht im Gegensatz zu den Ergebnissen von Conway et al. (2002), die für eine Erwachsenenstichprobe zwei separate Faktoren für einfache und komplexe Spannaufgaben fanden.

Gathercole und Pickering (2000a, 2000b) konstruierten eine Testbatterie für Kinder zur Erfassung des Arbeitsgedächtnisses nach dem klassischen Modell von Baddeley (vgl. Abschnitt 4.3.2). Dabei verwendeten sie verschiedene verbale und numerische Kurzzeitgedächtnisaufgaben zur Erfassung der phonologischen Schleife. Die visuell-räumlichen Aufgaben erforderten das Einprägen von statischen visuellen Mustern oder von Bewegungsabfolgen. Die zentrale Exekutive wurde über verbale und numerische komplexe Spannaufgaben erfasst. Bei sechs- und siebenjährigen Kindern fanden sie zunächst ein Drei-Faktorenmodell, in dem sämtliche Aufgaben, die theoretisch der zentralen Exekutive zugeordnet waren, auf dem ersten Faktor luden. Zu-



sätzlich hatten drei der vier visuell-räumlichen Aufgaben ihre höchsten Ladungen auf dem ersten Faktor. Somit vereinigt der erste Faktor verbale und numerische Arbeitsgedächtnisaufgaben sowie visuell-räumliche Kurzzeitgedächtnisaufgaben. Auf den zweiten und dritten Faktor luden die verbalen und numerischen Kurzzeitgedächtnisaufgaben, so dass diese als Ausdruck der phonologischen Schleife interpretiert wurden.

Jarvis und Gathercole (2003) setzten diese Arbeiten fort und analysierten die Testdaten von 55 11-jährigen und 73 14-jährigen Pbn mit konfirmatorischen Faktorenanalysen. Entsprechend dem klassischen Baddeley-Modell verwendeten sie erneut Aufgaben zur Erfassung der phonologischen Schleife, des visuell-räumlichen Skizzenblocks sowie der zentralen Exekutive und prüften verschiedene Faktorenstrukturen. Für die jüngeren Pbn wies ein Zwei-Faktoren-Modell mit einem verbalen und einem nonverbalen Faktor (Faktoreninterkorrelation:  $r = .53$ ) den relativ besten Fit auf ( $\chi^2 = 32.00$ ,  $df = 19$ ,  $p = .03$ ; CFI = .93; RMSEA = .11). Für die Älteren fand sich ein guter Modell-Fit ebenfalls für das Zwei-Faktoren-Modell (Faktoreninterkorrelation:  $r = .60$ ;  $\chi^2 = 11.40$ ,  $df = 14$ ,  $p = .58$ ; CFI = 1.00; RMSEA = .00), jedoch auch eine recht gute Modellpassung für ein Drei-Faktoren-Modell entsprechend dem Baddeley-Modell ( $\chi^2 = 14.70$   $df = 11$ ,  $p = .20$ ; CFI = .96; RMSEA = .07). Eine gemeinsame Schätzung für beide Altersgruppen wurde nicht vorgenommen. Die Befunde sprechen für eine Einteilung des Arbeitsgedächtnisses in verbale und räumliche Komponenten, jedoch wurden hier sowohl einfache als auch komplexe Spannaufgaben verwendet. Aufgrund der relativ kleinen Stichprobenumfänge sind die Parameterschätzungen der berichteten Modelle zudem möglicherweise nicht stabil.

### 3.3.3 Gibt es einen separaten numerischen Faktor?

Seit Turner und Engle (1989) zeigen konnten, dass ihre verbal-numerische Aufgabe „Operation Span“ ein guter Prädiktor für Lesefähigkeit ist, wird meist davon ausgegangen, dass Arbeitsgedächtnisaufgaben mit verbalen und numerischen Inhalten Maße für dieselbe Fähigkeit sind. Auch in verschiedenen Strukturanalysen, in die separate verbale und numerische Aufgaben eingingen (z. B. Oberauer et al., 2000), ergab sich ein gemeinsamer verbal-numerischer Faktor. Zudem passt die Unterteilung in eine verbal-numerische und eine visuell-räumliche Fähigkeit zu den Annahmen entsprechender separater Systeme im Baddeley-Modell. Diese Befundlage ist jedoch von daher erstaunlich, da in der Intelligenzstrukturforschung die drei Inhaltsbereiche verbal, numerisch und räumlich strukturanalytisch deutlich getrennt werden können (z. B. Carroll, 1993; Jäger et al., 2005; Ackerman et al., 2002). Möglicherweise ist die Trennung verbaler und numerischer Ressourcen einfach schwieriger als die Abgrenzung zu visuell-räumlichen Ressourcen, da numerische Informationen tendenziell auch verbal kodiert werden (Shah & Miyake, 1996).

Es gibt jedoch auch einige Befunde aus Studien, in denen Arbeitsgedächtnisaufgaben aus verschiedenen Domänen miteinander korreliert werden, die als Anhaltspunkt für drei inhaltlich separate Komponenten im Arbeitsgedächtnis (verbal, numerisch und räumlich) interpretiert werden können. Hacker et al. (1999) fanden für mehrere Stu-

dierendenstichproben ( $N = 27-45$ ) keine signifikanten Zusammenhänge zwischen den Arbeitsgedächtnisaufgaben „Satzspanne“ und „Rechenspanne“. Auch in der Studie von Jurden (1995) mit erwachsenen Pbn korrelierten die Aufgaben „Reading Span“ und „Computation Span“ nur schwach oder gar nicht miteinander (erste Untersuchung:  $r = .21, p < .05, N = 84$ ; zweite Untersuchung:  $r = .17, p > .05, N = 52$ ). Bei einem gemeinsamen verbal-numerischen Arbeitsgedächtnissystem wären hier deutliche positive Zusammenhänge zu erwarten. Die fehlenden Zusammenhänge in diesen Studien sprechen somit dagegen, dass es sich bei verbalen und numerischen Aufgaben um Indikatoren für dieselbe generelle Fähigkeit handelt. Von den Forschern, die latente Faktoren untersuchten, berichtet lediglich Kyllonen (1996) über separate Arbeitsgedächtnisfaktoren für alle drei Domänen, die jedoch hoch korreliert sind.

Jurden (1995) reanalyisierte die Daten von Kyllonen und Christal (1990, Studie 1) auf Basis der berichteten Korrelationen. Kyllonen und Christal hatten zwei verbale und drei numerische Arbeitsgedächtnisaufgaben verwendet. Das in der Originalstudie angenommene domänengenerelle Ein-Faktor-Modell ergab in der Reanalyse eine schlechte Datenpassung ( $N = 723$ ;  $\chi^2 = 119.21, df = 5, p < .001$ ; NFI = .88; RMSEA = .18). Jurden prüfte daher ein obliques Zwei-Faktoren-Modell mit einem verbalen und einem numerischen Faktor und fand eine gute Datenpassung ( $\chi^2 = 2.07, df = 4, p = .72$ ; NFI = 1.00; RMSEA = .00). Auch für das von Kyllonen und Christal berichtete Modell mit den latenten Faktoren Arbeitsgedächtnis, Reasoning, Wissen und Speed (siehe Abschnitt 3.4) fand Jurden eine signifikant bessere Passung, wenn zwei Arbeitsgedächtnisfaktoren spezifiziert wurden ( $\chi^2 = 108.91, df = 44, p < .0001$ ; NFI = .96; RMSEA = .05;  $\Delta\chi^2 = 129.29, \Delta df = 1, p < .0001$ ).

In zwei weiteren Studien mit Kindern wurde die Vorhersagegüte verbaler und numerischer Arbeitsgedächtnisaufgaben für die Lese- und Rechenleistung über hierarchische Regressionsanalysen geprüft. Es wurde der Frage nachgegangen, ob eine verbale Arbeitsgedächtnisaufgabe auch dann noch zusätzliche Varianz der Kriteriumsmaße aufklärt, wenn sie nach der numerischen Arbeitsgedächtnisaufgabe in die Regressionsgleichung eingeht (und umgekehrt). Eine vollständige Domänenspezifität wäre gegeben, wenn jede Spannenaufgabe nur ungeteilte Varianz schulischer Fähigkeiten aufklären würde. Vollständige Domänengeneralität könnte hingegen angenommen werden, wenn beide Aufgaben nur geteilte Varianz aufklären würden.

Leather und Henry (1994) testeten 71 Kinder im Alter von sieben Jahren mit verschiedenen Aufgaben, unter anderem mit den beiden komplexen Spannenaufgaben „Counting Span“ und „Listening Span“ (eine auditive Variante der „Reading Span“, siehe Erläuterung in Abschnitt 4.1.1). Beide Aufgaben korrelierten signifikant ( $r = .47$ ) und erwiesen sich als signifikante Prädiktoren sowohl für die Lesefähigkeit als auch für die Rechenfähigkeit. In hierarchischen Regressionen zeigten sich jedoch auch Unterschiede in der Vorhersageleistung beider Aufgaben: Die „Counting Span“ diente nur dann als signifikanter Prädiktor der Lesefähigkeit, wenn die „Reading Span“ noch nicht in die Regressionsgleichung eingeschlossen worden war, sie lieferte somit keinen Beitrag zur Varianzaufklärung der Lesefähigkeit über den Beitrag der „Listening Span“ hinaus. Für die Vorhersage von Rechenfähigkeit zeigte sich ein anderes Bild: Die „Listening Span“ konnte auch dann noch signifikant zur Varianz-

aufklärung beitragen, wenn sie erst nach der „Counting Span“ in die Regressionsgleichung aufgenommen wurde. Die Autorinnen schlussfolgern: „Diese Ergebnisse zeigen weder eindeutige Belege für ein Modell genereller Kapazität noch für ein domänenspezifisches Modell, stattdessen liegen sie irgendwo zwischen diesen beiden Extremen.“ (S. 106, Übersetzung durch die Verfasserin).

Auch Hitch et al. (2001) berichten über Ergebnisse, die für das Vorhandensein sowohl domänenspezifischer als auch domänengenereller Aspekte im Arbeitsgedächtnis sprechen. In ihrer Untersuchung mit neun- bis elfjährigen Kindern ( $N = 81$ ) fanden sie ebenfalls, dass die Arbeitsgedächtnisaufgaben „Reading Span“ und „Operation Span“ jeweils Varianzanteile der Lese- und Rechenleistung aufklären. Für die Vorhersage der Lesefähigkeit lieferten beide Spannaufgaben einen signifikanten Beitrag, und zwar unabhängig davon, in welcher Reihenfolge sie in die Gleichung eingingen. Beide Aufgaben erklären somit sowohl geteilte als auch ungeteilte Varianzanteile der Lesefähigkeit. Auch die Rechenfähigkeit konnte durch beide Spannaufgaben vorhergesagt werden, jedoch war jede der beiden Aufgaben nur dann ein signifikanter Prädiktor, wenn sie vor der jeweils anderen Aufgabe in die Gleichung einging. Beide Aufgaben klären somit geteilte Varianzanteile der Rechenfähigkeit auf. Sowohl „Reading Span“ als auch „Operation Span“ klären geteilte Varianz der Rechen- und der Lesefähigkeit und signifikante ungeteilte Varianz in der Lesefähigkeit auf. Diese Befunde weichen von den oben berichteten Ergebnissen von Leather und Henry ab – hier können beide Aufgaben ungeteilte Varianz nur der Leseleistung aufklären, bei Leather und Henry nur der Rechenleistung –, belegen jedoch ebenfalls das Vorhandensein domänenspezifischer Aspekte im Arbeitsgedächtnis. Da in beiden Studien ein bedeutsamer Teil geteilter Varianz aufgeklärt wird, können zusätzlich domänengenerelle Aspekte angenommen werden. Weitergehende Interpretationen sind aufgrund der inhaltlich deutlich voneinander abweichenden Ergebnisse jedoch nicht möglich.

### 3.3.4 Integration der Befunde zur Binnenstruktur

Die Befunde aus den oben dargestellten Ansätzen sind somit insgesamt widersprüchlich. Während einige der Studien, die Korrelationen des Arbeitsgedächtnisses zu externen Kriterien untersuchen, zu dem Schluss kommen, dass das Arbeitsgedächtnis domänenspezifisch ist, finden Studien, die latente Variablen verwenden, in der Regel entweder einen Faktor (z. B. Ackerman et al., 2002) oder zwei hoch korrelierende Inhaltsfaktoren (verbal-numerisch und räumlich, z. B. Kane et al., 2004; Süß et al., 2002; Kyllonen & Christal, 1990). Entsprechende Studien, die die Faktorenstruktur des Arbeitsgedächtnisses bei Kindern untersuchten, finden meist ebenfalls einen Faktor (z. B. de Jong & Das-Smaal, 1995; siehe aber auch Jarvis & Gathercole, 2003).

Verschiedentlich werden methodische Aspekte als Erklärung für die widersprüchlichen Befunde angeführt (z. B. Kane et al., 2004; Miyake, 2001). So verwendeten die Autoren der Studien, deren Befunde eine Domänenspezifität stützen, häufig eher kleine, relativ homogene Stichproben (Studierende selektiver Universitäten). Dies führt zu einer eingeschränkten allgemeinen Fähigkeitsvarianz (Shah & Miyake,

1996). Wenn in diesen Gruppen Fähigkeitsunterschiede auftreten, sind diese eher auf domänen-spezifische Fertigkeiten, Strategien und Fähigkeiten zurückzuführen, die Domänengeneralität wird in diesen Studien daher eher unterschätzt (Kane et al., 2004). Mackintosh und Bennett (2003) argumentieren, die Testung deutlich überdurchschnittlich intelligenter Pbn führe dazu, dass verschiedene Arbeitsgedächtnisaufgaben weniger stark miteinander korrelieren. Sie verweisen dazu auf Befunde, dass verschiedene Aspekte der Intelligenz bei überdurchschnittlich intelligenten Pbn weniger stark miteinander zusammenhängen als bei durchschnittlich oder unterdurchschnittlich intelligenten Pbn (z. B. Detterman & Daniel, 1989). Kintsch, Healy, Hegarty, Pennington und Salthouse (1999) merken zudem an, dass sich die Studien, die die jeweiligen Positionen stützten, unter anderem in der Auswahl der Aufgaben unterscheiden. Vertreter einer Generalitäts-Hypothese verwenden eher verbale und numerische Aufgaben und vernachlässigen dabei visuell-räumliche Aufgaben. Neuere Arbeiten zeigen jedoch, dass auch eine ausgewogenere Aufgabenauswahl zur Stützung einer Generalitäts-Hypothese führen kann (z. B. Ackerman et al., 2002).

Die Interpretation der Befunde der bisherigen Studien wird dadurch erschwert, dass die Auswahl der Aufgaben in allen Untersuchungen unterschiedlich gehandhabt wurde. Insbesondere bei der Untersuchung von Kindern werden zum Teil sehr heterogene Aufgaben verwendet. Teilweise wird eine Aufgabenbatterie eingesetzt, deren Aufgaben unterschiedliche Operationen erfordern (z. B. eher passive Kurzzeitgedächtnis- oder aktive Arbeitsgedächtnisleistungen). Einige Studien verwenden nur verbale und numerische Aufgaben, so dass keine Aussagen zu einer möglichen räumlichen Komponente des Arbeitsgedächtnisses getroffen werden können. In anderen Studien werden zwar räumliche Aufgaben eingesetzt, meist ist dann aber – aufgrund der Aufgabenauswahl – keine Unterscheidung verbaler und numerischer Komponenten mehr möglich. Bisherige Befunde legen zwar nahe, dass die Leistung in verbalen und numerischen Arbeitsgedächtnisaufgaben auf dieselbe Fähigkeit zurückgeführt werden kann, jedoch existieren auch entgegengesetzte Befunde. Tendenziell werden insgesamt zu wenig Aufgaben verwendet, um sowohl Differenzierungen in der Funktions- als auch in der Inhaltsfacette aufzufinden. Ein systematischer Ansatz der Aufgabenauswahl, wie er von Oberauer et al. (2000) und Süß et al. (2002) in ihrer Studie mit erwachsenen Pbn verfolgt wurde, fehlt bislang in Studien mit Kindern. Verschiedene Studien, die eine Aufgabenauswahl entsprechend dem klassischen Baddeley-Modell vornahmen, hatten bisher Schwierigkeiten, die angenommene Struktur faktoranalytisch bei Kindern aufzufinden (z. B. Jarvis & Gathercole, 2003). Für das Kindesalter können daher auf der Basis der vorliegenden Befunde noch keine gesicherten Aussagen zur Binnenstruktur des Arbeitsgedächtnisses getroffen werden.

Trotz der oben diskutierten möglichen methodischen Probleme gehen inzwischen viele Forscher davon aus, dass sich die Heterogenität der Befundlage nicht allein über Methodenartefakte erklären lässt. Da vielfältige Belege sowohl für eine Domänengeneralität als auch für eine Domänenspezifität vorliegen, nehmen z. B. Miyake (2001) und Kintsch et al. (1999) an, dass Arbeitsgedächtnisaufgaben sowohl domänen-spezifische als auch -unspezifische Komponenten enthalten. Auch die regressionsanalytischen Befunde von Leather und Henry (1994) sowie von Hitch et al. (2001) sprechen für die Annahme verschiedener Komponenten. Engle, Kane et al.

(1999) schlagen daher vor, das Arbeitsgedächtnis als ein hierarchisches Konstrukt zu konzipieren. Ebenso wie in der Intelligenzstrukturforschung gebe es deutliche Hinweise auf einen Generalfaktor, dieser ließe jedoch noch ein beträchtliches Ausmaß an Varianz unerklärt, so dass domänenspezifische sekundäre Faktoren angenommen werden können. Bardon (1999) fand eine gute Passung für ein entsprechendes hierarchisches Modell (CFA mit den beiden Primärfaktoren „verbal-numerisch“ vs. „räumlich“ und einem sekundären, generellen Arbeitsgedächtnisfaktor) für seine Daten, die er jedoch an einer sehr kleinen Stichprobe gewonnen hatte ( $N = 48$  Schüler/innen der vierten Klasse). Auch die Befunde von Kane et al. (2004) deuten in diese Richtung, da die Daten für zwei hoch korrelierende Inhaltsfaktoren sprechen.

Weiterhin gibt es Hinweise darauf, dass sich das Ausmaß, in dem domänenspezifische und -unspezifische Komponenten die Leistung beeinflussen, über die Lebensspanne verändert. Für ältere Menschen konnte gezeigt werden, dass sich die Arbeitsgedächtnis-Ressourcen weniger differenzieren lassen als für junge Erwachsene (Reuter-Lorenz et al., 2000). Vergleichbare Befunde liegen auch in der Intelligenzstrukturforschung vor. Diese werden dort unter anderem auf die größere Varianz der Testleistungen mit zunehmendem Alter zurückgeführt (z. B. Fleischmann, 1989; Oswald, 1998). Möglicherweise verändert sich daher die Arbeitsgedächtnisstruktur auch im Kindesalter, systematische Untersuchungen hierzu liegen jedoch noch nicht vor.

### **3.4 Zusammenhänge des Arbeitsgedächtnisses mit allgemeiner Intelligenz und anderen kognitiven Variablen**

Der Ansatz vieler bisheriger Studien zu individuellen Unterschieden in der Leistung des Arbeitsgedächtnisses bestand in dem Versuch, das Konstrukt „Arbeitsgedächtnis“ genauer einzugrenzen, die Bezüge zu anderen kognitiven Variablen zu explorieren und das Arbeitsgedächtnis in ein nomologisches Netzwerk kognitiver Fähigkeiten einzuordnen (siehe z. B. Ackerman et al., 2002; de Jong & Das-Smaal, 1995). Ein wesentlicher Zweck der Erfassung individueller Unterschiede im Arbeitsgedächtnis lag zudem bereits in den ersten Arbeitsgedächtnisstudien darin, mit diesen Maßen komplexe kognitive Leistungen erklären zu können (z. B. Daneman & Carpenter, 1980). Eine zentrale Motivation, die Zusammenhänge von Arbeitsgedächtnis und anderen kognitiven Leistungen zu erforschen, beruht auf dem Anliegen, den grundlegenden Prozessen, die hinter Intelligenztestleistungen und psychometrisch ermittelten latenten Faktoren stehen, auf die Spur zu kommen. Entsprechende Erkenntnisse sollen zusätzlich für die theoretische Fundierung und Optimierung des psychometrischen Ansatzes nutzbar gemacht werden (z. B. Kyllonen, 1996).

Inzwischen konnten in mehreren Studien signifikante Korrelationen zwischen den Konstrukten Arbeitsgedächtniskapazität, fluide Intelligenz, Geschwindigkeit der Informationsverarbeitung (oder Speed), Kurzzeitgedächtnis und Leistungen in verschiedenen intellektuellen Bereichen gezeigt werden. Die Höhe der Zusammenhänge variiert im Vergleich der einzelnen Studien jedoch zum Teil beträchtlich. Die neueren Untersuchungen zu diesen Fragestellungen folgen dabei dem Prinzip, die

Zusammenhänge auf der Konstruktebene über latente Faktoren zu untersuchen. Metaanalysen zu diesen Zusammenhängen liegen – bis auf eine Studie zum Zusammenhang von Arbeitsgedächtnis und sprachlichen Fähigkeiten von Daneman und Merikle (1996) – noch nicht vor.

### 3.4.1 Intelligenz

Wenn die Zusammenhänge des Arbeitsgedächtnisses mit verschiedenen anderen kognitiven Konstrukten betrachtet werden, erweist sich in der Regel die Korrelation zu Faktoren der allgemeinen oder fluiden Intelligenz bzw. Reasoning als der relativ stärkste Zusammenhang. Zu anderen Konstrukten wie z. B. Speed finden sich auch bedeutsame Zusammenhänge, diese fallen meist jedoch deutlich geringer aus. Der empirisch gefundene Zusammenhang von Intelligenz und Arbeitsgedächtnis wird üblicherweise so erklärt, dass bei jeder Reasoning-Aufgabe, z. B. einer figuralen Matrizen- oder Analogieaufgabe, Zwischenergebnisse gefunden und gespeichert werden müssen, während eine weitere Bearbeitung stattfindet (z. B. Carpenter, Just & Shell, 1990; Shah & Miyake, 1996). Eine geringe Kapazität des Arbeitsgedächtnisses führt nach dieser Sichtweise zu Fehlern bei der Lösung von Intelligenztestaufgaben. Carpenter et al. (1990) stellten fest, dass die Fähigkeit, Ziele im Arbeitsgedächtnis simultan zu generieren, zu bewerten und aufrechtzuerhalten wichtig ist für die richtige Lösung von Aufgaben in den Advanced Progressive Matrices (APM, Raven, 1962a). Verguts und de Boeck (2002) konnten darüber hinaus zeigen, dass die Arbeitsgedächtniskapazität nicht nur für das Ausmaß der Speicherung von Zwischenergebnissen innerhalb einzelner Items relevant ist, sondern auch für die Speicherung von Regeln oder Lösungsprinzipien über mehrere Items eines Tests hinweg. Wenn einmal eine richtige Regel entdeckt wurde, haben Menschen mit größerer Arbeitsgedächtniskapazität eine größere Wahrscheinlichkeit, diese Regel auch für die Lösung der folgenden Items im Kopf zu behalten – sie werden daher auch die weiteren Items des Tests eher lösen.

Weitergehende theoretische Ansätze nehmen an, dass das Arbeitsgedächtnis der zentrale Faktor sein könnte, der sämtlichen kognitiven Leistungen sowie der Lernfähigkeit von Menschen zugrunde liegt (z. B. Kyllonen & Christal, 1990; Kyllonen, 1996). In der psychometrischen Intelligenzforschung wird dieses zentrale Konstrukt in der Regel als *general intelligence* oder *g*-Faktor (nach den Arbeiten von Spearman, 1927) bezeichnet. Spearman beschäftigte sich mit dem Nachweis eines generellen Faktors, der so gut wie allen intellektuellen Leistungen zugrunde liegen sollte. Die Idee eines generellen Faktors erwuchs aus der Beobachtung, dass die Leistungen in verschiedenen kognitiven Tests bei einer Gruppe von Probanden in der Regel positiv miteinander korrelieren. Die Leistungen, die Menschen bei unterschiedlichen kognitiven Aufgaben zeigen, sollten daher auf eine gemeinsame Quelle zurückzuführen sein. Spearman schloss daraus, dass kognitiven Leistungen ein gemeinsamer Faktor zugrunde liegt. Diese Erkenntnis beruht auf den damals gerade entwickelten statistischen Verfahren der Korrelationsrechnung und der Faktorenanalyse. Die allgemeine Intelligenz im Sinne von *g* kann über verschiedene bewährte Aufgaben-

typen (z. B. figurale Matrizenaufgaben) relativ gut gemessen werden. Unklar ist bis heute jedoch, was die Grundlage von  $g$  – über die Korrelation verschiedener Aufgaben hinaus – ist und insbesondere ist die Frage ungeklärt, auf welche Verarbeitungsprozesse es zurückzuführen ist.

Kyllonen (1996) verfolgte daher den Ansatz, die Erkenntnisse der kognitiven Psychologie zu den Grundlagen menschlicher Informationsverarbeitung für die Weiterentwicklung des psychometrischen Ansatzes zu nutzen. Zwar existiert in der kognitiven Psychologie bislang kein von allen Forschern geteiltes Modell der Informationsverarbeitung, jedoch besteht Konsens hinsichtlich einiger zentraler Komponenten der Informationsverarbeitung. Diese zentralen Komponenten könnten, so nimmt Kyllonen an, wesentliche Quellen individueller Unterschiede in kognitiven Leistungen sein. Auf dieser Basis schlägt er das *Cognitive Abilities Measurement Model* (CAM-Modell oder auch „Vier-Quellen-Modell“, Kyllonen & Christal, 1990) vor. Es enthält als die vier „Quellen“ kognitiver Leistungen die Geschwindigkeit der Informationsverarbeitung (*processing speed*), das Arbeitsgedächtnis sowie die Breite des deklarativen und des prozeduralen Wissens (*declarative/procedural knowledge*), die alle bei der Verarbeitung von Informationen miteinander interagieren. Kombiniert werden diese vier Quellen im CAM-Modell mit den drei Inhaltsbereichen bzw. Domänen, die aus der Intelligenzforschung bekannt sind (verbale, numerische und räumliche Inhalte). Mit einer entsprechend den Vorgaben des CAM-Modells entwickelten Testbatterie konnte die Struktur des Modells empirisch bestätigt werden (Kyllonen, 1996).

Individuelle kognitive Leistungsunterschiede entstehen dieser Vorstellung zufolge durch Unterschiede in einer oder in mehreren der vier Komponenten des Modells. Ebenso wie die klassischen hierarchischen Intelligenzmodelle wird auch das Vier-Quellen-Modell als ein hierarchisches Modell konzipiert, an dessen Spitze ein zentraler Faktor steht. In den klassischen Intelligenzmodellen wird dieser Faktor häufig als Reasoning oder fluide Intelligenz interpretiert, im Vier-Quellen-Modell wird angenommen, dass die Arbeitsgedächtniskapazität dieser zentrale Faktor ist. Zur Prüfung dieser Annahme wurden unter anderem die Zusammenhänge des latenten Faktors „Arbeitsgedächtnis“ mit einem Reasoning-Faktor untersucht.

Die erste groß angelegte Untersuchung zum Zusammenhang von Arbeitsgedächtnis und Reasoning mittels latenter Faktoren stammt von Kyllonen und Christal (1990). Die Autoren verwendeten unterschiedliche Testaufgaben zur Messung der vier Quellen im CAM-Modell. In vier Studien mit großen Stichproben mit Rekruten der Air Force ( $N = 412 - 723$ ) prüften Kyllonen und Christal die Beziehungen zwischen Arbeitsgedächtnis, Reasoning, Geschwindigkeit der Informationsverarbeitung und allgemeinem Wissen. Das Arbeitsgedächtnis wurde über sechs verschiedene Aufgabentypen erfasst. Der Faktor „Reasoning“ wurde über 13 verschiedene Aufgabentypen gemessen, darunter klassische Aufgaben wie verbale Analogien und Rechenaufgaben, aber auch Aufgaben, die den Arbeitsgedächtnisaufgaben relativ ähnlich sind (z. B. „Arrow Grammatical Reasoning“). Auch die Faktoren „Allgemeines Wissen“ und „Speed“ wurden über jeweils mehrere Aufgabentypen erfasst.

Das Modell latenter Faktoren, das am besten auf die Daten passte, ergab in den verschiedenen Studien Korrelationen von  $r = .80$  bis  $r = .88$  zwischen Arbeitsgedächtnis und Reasoning. Die Ergebnisse hinsichtlich der Zusammenhänge latenter Faktoren der ersten Studie in der Arbeit von Kyllonen und Christal (1990) sind in Abbildung 3.2 skizziert. Kyllonen und Christal (1990) schlussfolgerten hieraus, dass es sich bei Reasoning und Arbeitsgedächtnis um sehr ähnliche Konstrukte handelt und dass individuelle Unterschiede in Reasoning letztlich auf individuelle Unterschiede im Arbeitsgedächtnis zurückgehen. Als Unterschiede zwischen beiden Konstrukten stellten sie fest, dass Reasoning stärker mit dem deklarativen Wissen, das Arbeitsgedächtnis hingegen stärker mit Speed verknüpft ist (dieser Befund ergab sich insbesondere in den Studien 2-4, die hier nicht näher dargelegt werden).

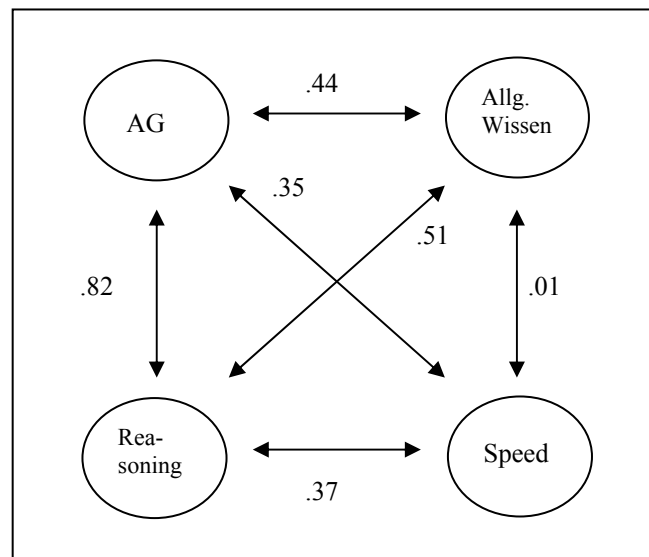


Abbildung 3.2: Korrelationen der latenten Faktoren Arbeitsgedächtnis (AG), allgemeines Wissen, Reasoning und Processing Speed in der Untersuchung von Kyllonen und Christal (1990, Studie 1)

Kyllonen (1996) berichtet über Ergebnisse einer weiteren vergleichbaren Studie (siehe auch Kyllonen, 1994), in der Arbeitsgedächtnis und Reasoning erneut am höchsten miteinander korrelierten. Die Zusammenhänge fielen hier jedoch mit Werten zwischen  $r = .94$  und  $r = .96$  noch einmal deutlich höher aus als in der oben berichteten Studie von Kyllonen und Christal (1990). Kyllonen erklärt diesen Befund so, dass die Fähigkeitsvarianz der Studie von 1990 eingeschränkt war, da ausschließlich Air-Force-Rekruten untersucht worden waren. In der neueren Studie hingegen war eine heterogenere Stichprobe getestet worden, so dass diese Ergebnisse die realistischere Schätzung des Zusammenhangs darstellen. Der Faktor „generelle Intelligenz/numerisch“ (ASVAB-G), gebildet durch zwei mathematisch-numerische Subtests der *Armed Services Vocational Aptitude Battery* (ASVAB), korrelierte gar



zu  $r = .99$  mit dem Arbeitsgedächtnisfaktor (Kyllonen, 1993). Eine Einschränkung der Gültigkeit dieser Befunde von Kyllonen und Christal (1990) sowie Kyllonen (1996) ist darin zu sehen, dass sich die verwendeten Arbeitsgedächtnis- und Reasoningaufgaben zum Teil sehr ähnelten, so dass die hohen Korrelationen vermutlich teilweise auch auf diesen Umstand zurückzuführen sind (siehe z. B. Süß, 2001).

Eine vergleichbare Studie führten de Jong und Das-Smaal (1995) mit 381 neunjährigen Kindern durch. Untersucht wurden hier die Konstrukte „Arbeitsgedächtnis“, „fluide Intelligenz“, „Speed“ und „Schulleistung“. Der Faktor „Arbeitsgedächtnis“ wurde über vier verschiedene Aufgabentypen erfasst, die jedoch nicht alle der strengen Definition im Sinne einer simultanen Speicherung und Transformation entsprachen. So wurden etwa sowohl die Vorwärts- als auch die Rückwärtsvariante der „Digit Span“ eingesetzt (vgl. Abschnitt 4.1.2). Die fluide Intelligenz wurde über klassische Reasoningaufgaben wie verbale Analogien und Zahlenreihen und die Geschwindigkeit der Informationsverarbeitung über drei verschiedene Speed-Aufgaben erfasst. Als Maße für die Schulleistung verwendeten de Jong und Das-Smaal standardisierte Lese- und Mathematiktests. Ebenso wie bei Kyllonen und Christal (1990) ergab sich auch hier ein bedeutsamer, wenn auch etwas geringerer Zusammenhang zwischen den Faktoren „Arbeitsgedächtnis“ und „Intelligenz“ ( $r = .66$ ; siehe Abbildung 3.3). Weiterhin zeigte sich ein starker Zusammenhang zwischen Arbeitsgedächtnis und Schulleistung ( $r = .72$ ). Zwischen fluider Intelligenz und Schulleistung fand sich eine Korrelation von  $r = .82$ ; dieser Zusammenhang war jedoch nicht signifikant größer als der Zusammenhang zwischen Arbeitsgedächtnis und Schulleistung.

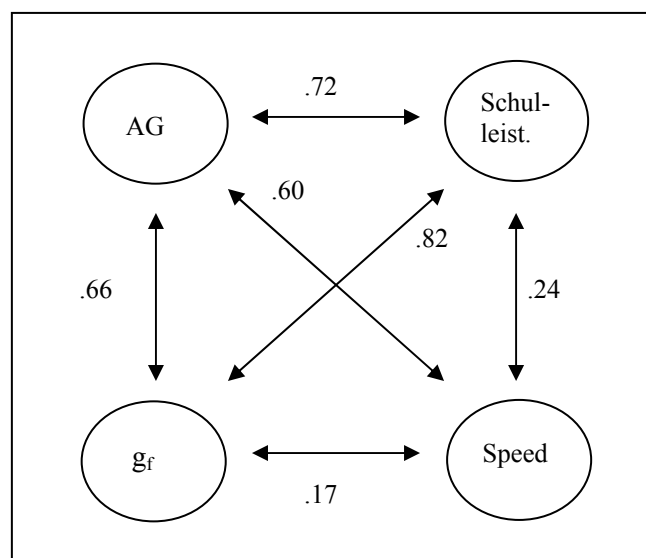


Abbildung 3.3: Korrelationen der latenten Faktoren Arbeitsgedächtnis (AG), Schulleistung, fluide Intelligenz ( $g_f$ ) und Speed in der Studie von de Jong und Das-Smaal (1995)

De Jong und Das-Smaal (1995) konnten somit zeigen, dass fluide Intelligenz und Arbeitsgedächtnis bereits im Kindesalter signifikant und substantiell zusammenhän-

gen. Relativ unklar ist bisher jedoch, ob und ggf. wie sich der Zusammenhang zwischen Arbeitsgedächtnis und Intelligenz im Laufe der Kindheit verändert. In bisherigen Studien wurde entweder nur eine Altersgruppe untersucht, oder es wurden relativ kleine Stichproben mehrerer Altersgruppen untersucht, so dass über die Veränderung oder Konstanz während der Entwicklung noch keine verlässlichen Aussagen getroffen werden können (Fry & Hale, 2000). Eine der wenigen Untersuchungen, die mehrere Altersgruppen mit einem Querschnittsdesign untersuchten, stammt von de Jonge und de Jong (1996). Sie stellten fest, dass die Zusammenhänge zwischen Arbeitsgedächtnis, Reasoning und Lesefähigkeit in den Klassenstufen vier, fünf und sechs sehr ähnlich ausfielen. Diese Befunde deuten darauf hin, dass die Beziehungen des Arbeitsgedächtnisses zu anderen Konstrukten zumindest ab etwa dem neunten Lebensjahr stabil bleiben.

In verschiedenen weiteren Studien konnten substantielle Korrelationen zwischen latenten Arbeitsgedächtnisfaktoren und Intelligenztestleistungen gefunden werden. Beispielsweise ermittelten Oberauer et al. (2000) Korrelationen im Bereich von  $r = .41$  bis  $r = .56$  zwischen ihren drei Arbeitsgedächtnisfaktoren (siehe Kapitel 3.3) und der Skala „Verarbeitungskapazität“ des BIS-4. Auch mit den figuralen Reasoningaufgaben anderer Tests, z. B. CFT oder Raven-Matrizen-Tests, korrelieren Arbeitsgedächtnisaufgaben in der Regel bedeutsam (z. B. Engle, Tuholski, Laughlin & Conway, 1999).

Der Befund, dass Reasoning bzw. die fluide Intelligenz und das Arbeitsgedächtnis signifikant miteinander zusammenhängen, scheint somit robust und nicht allein zurückzuführen auf die Verwendung bestimmter Intelligenz- und Arbeitsgedächtnisaufgaben oder die Untersuchung bestimmter Stichproben. Jedoch sind das Ausmaß der Zusammenhänge dieser Konstrukte und die Konsequenzen, die daraus folgen, weiterhin umstritten. So wurde die Schlussfolgerung, Reasoning sei „nur wenig mehr als die Arbeitsgedächtniskapazität“ (Kyllonen & Christal, 1990) bzw. die Vermutung, Arbeitsgedächtnis sei letztlich Spearman's  $g$  (Kyllonen, 1996) auch stark kritisiert. Die aktuellste Kritik in diese Richtung publizierten Ackerman et al. (2002). Auch Ackerman und Kollegen gehen zwar davon aus, dass Arbeitsgedächtnisleistungen positiv und signifikant mit Reasoning bzw. der fluiden Intelligenz verknüpft sind. Sie wenden jedoch ein, dass sich die Höhe des Zusammenhangs, die von Kyllonen und Christal (1990) postuliert worden war, weder auf der Aufgaben- noch auf der Konstruktebene replizieren lässt. Dazu verweisen sie auf Untersuchungen mit Studierendenstichproben, in denen Korrelationen zwischen *einzelnen* Arbeitsgedächtnisaufgaben und dem Raven-Test zwischen  $r = .15$  und  $r = .38$  (Conway et al., 2002) bzw. zwischen  $r = .28$  und  $r = .34$  (Engle, Tuholski et al., 1999) gefunden wurden. Bei Stichproben mit Erwachsenen mit einem breiteren Altersspektrum liegen die Korrelationen zwischen  $r = .20$  und  $r = .43$  (Jurden, 1995). Somit teilen der Raven-Test und Arbeitsgedächtnisaufgaben bei jungen Erwachsenen lediglich zwischen 2% und 14% gemeinsame Varianz, bei einer heterogeneren Altersgruppe bis zu 19%. Ackerman et al. (2002) schließen aus diesen Befunden: „Alles in allem ist das keine sehr beeindruckende Übereinstimmung angesichts von Behauptungen, es handele sich um Maße für dasselbe Konstrukt.“ (S. 568; Übersetzung durch die Verfasserin).

Jedoch fallen die Zusammenhänge deutlich höher aus, wenn das Arbeitsgedächtnis als latenter Faktor betrachtet wird und somit nur die allen Aufgaben gemeinsame Varianz berücksichtigt wird. So fanden beispielsweise Colom et al. (2003) eine Korrelation von  $r = .69$  zwischen Arbeitsgedächtnis (erfasst über acht verschiedene Aufgabentypen) und der Leistung im SPM. In zwei weiteren Studien wurde eine latente Variable der fluiden Intelligenz aus den Rohwerten im Raven-Test und im CFT gebildet. In einem Strukturgleichungsmodell, in dem die so gebildete fluide Intelligenz über verschiedene kognitive Faktoren vorhergesagt werden sollte, fand sich ein signifikanter Pfadkoeffizient des Arbeitsgedächtnisfaktors von  $.59$  bzw.  $.60$  (Engle, Tuholski et al., 1999; Conway et al., 2002).

Ackerman et al. (2002) prüften in einer eigenen Studie mit 135 Studierenden die Annahme, dass das Arbeitsgedächtnis einen großen Varianzanteil mit allgemeinen intellektuellen Fähigkeiten teilt, jedoch ein substantieller Teil der Varianz nicht erklärt werden kann. Sie nahmen somit an, dass Arbeitsgedächtnis und allgemeine Intelligenz nicht identisch sind. Weiterhin untersuchten sie den Zusammenhang zwischen Arbeitsgedächtnis und Speed, da sie davon ausgingen, dass auch hier substantielle Zusammenhänge bestehen (siehe Abschnitt 3.4.2).

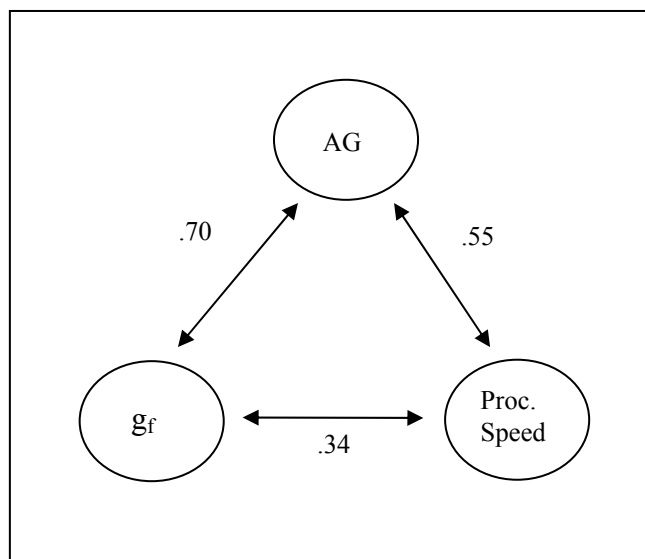


Abbildung 3.4: Korrelationen der latenten Faktoren Arbeitsgedächtnis (AG), fluide Intelligenz (gf) und Processing Speed in der Studie von Ackerman et al. (2002)

Sie fanden wie erwartet einen hohen, aber nicht perfekten Zusammenhang zwischen Intelligenz und Arbeitsgedächtnis ( $r = .70$ ; siehe Abbildung 3.4). Es zeigte sich ein ebenfalls bedeutsamer, aber geringerer Zusammenhang zwischen Arbeitsgedächtnis und einem Speed-Faktor ( $r = .55$ ) und ein eher schwacher Zusammenhang zwischen Speed und Intelligenz ( $r = .34$ ). Ein alternatives Modell, in dem kein Zusammenhang zwischen Arbeitsgedächtnis und Speed angenommen wurde, ergab eine signifikant schlechtere Datenpassung. Der Speed-Faktor korrelierte auch dann noch signifikant mit dem Arbeitsgedächtnis, wenn die Intelligenz auspartialisiert wurde.

Zusammenfassend lässt sich festhalten, dass auf der Konstruktebene ein substantieller Zusammenhang zwischen Arbeitsgedächtnis und fluider Intelligenz bzw. Reasoning besteht, der mehrfach mit verschiedenen Aufgaben und Probandenstichproben repliziert werden konnte. Werden sowohl das Arbeitsgedächtnis als auch die Intelligenz als latente Faktoren spezifiziert (hier sind höhere Zusammenhänge zu erwarten als bei der Betrachtung von Korrelationen auf Aufgabenebene), liegen die berichteten Korrelationen zwischen  $r = .66$  (de Jong & Das-Smaal, 1995) und  $r = .99$  (Kyllonen, 1993). Dieser beträchtliche Unterschied in den geschätzten Zusammenhängen lässt sich auf verschiedene Ursachen zurückführen, insbesondere die Zusammenstellung der Arbeitsgedächtnis- und Intelligenzaufgaben und der jeweiligen Probandenstichproben scheinen hier jedoch zentral. So finden sich höhere Korrelationen, wenn sich die Aufgaben des Arbeitsgedächtnis- und des Intelligenztests stärker ähneln oder wenn eine fähigkeitsheterogene Gruppe untersucht wird. Wie die Zusammenhänge bei verschiedenen Alters- oder Begabungsgruppen ausfallen, wurde bisher nicht systematisch untersucht. Zwar fanden de Jong und Das-Smaal (1995) bei ihren Pbn im Kindesalter einen geringeren Zusammenhang zwischen Intelligenz und Arbeitsgedächtnis als Kyllonen und Christal (1990) bei erwachsenen Pbn. Da sich beide Studien jedoch deutlich im Aufgabenmaterial unterscheiden, kann daraus kein Rückschluss auf altersbedingte Unterschiede gezogen werden.

### 3.4.2 Geschwindigkeit der Informationsverarbeitung oder Speed

Ebenso wie das Arbeitsgedächtnis gilt die Geschwindigkeit der Informationsverarbeitung oder (Processing) Speed als ein zentrales Konstrukt in Modellen der menschlichen Informationsverarbeitung und daher als bedeutsame Quelle individueller Unterschiede in kognitiven Leistungen (Kyllonen, 1996). Die Fähigkeit zur schnellen Verarbeitung von Informationen ist zudem Bestandteil der aktuellen Intelligenzstrukturmodelle (vgl. Abschnitt 2.2). Im BIS-Modell wird die Operation „Bearbeitungsgeschwindigkeit“ spezifiziert, die interpretiert wird als „Arbeitstempo, Auffassungsleichtigkeit und Konzentrationskraft beim Lösen einfach strukturierter Aufgaben von geringem Schwierigkeitsniveau“ (Jäger et al., 2005). Das Carroll-Modell unterscheidet innerhalb des zweiten Stratum zwei verschiedene Faktoren, die sich auf die Geschwindigkeit der Informationsverarbeitung beziehen: die kognitive Geschwindigkeit und die Entscheidungsgeschwindigkeit. Viele Intelligenztests enthalten entsprechend Aufgaben zur Erfassung der Verarbeitungsgeschwindigkeit (z. B. HAWIK III).

Die Geschwindigkeit der Informationsverarbeitung wurde häufig zur Erklärung von Prozessen sowohl im Arbeitsgedächtnis als auch beim Lösen von Intelligenztestaufgaben herangezogen. Baddeley (1986) etwa nimmt an, dass bei höherer Geschwindigkeit der Informationsverarbeitung die Rehearsal-Prozesse im Arbeitsgedächtnis schneller ablaufen, so dass mehr Informationen aktiv gehalten werden können und weniger Informationen durch Zerfall oder Interferenz verloren gehen. Jensen (1998) geht daher davon aus, dass eine schnellere Informationsverarbeitung die Wahrscheinlichkeit erhöht, eine Intelligenztestaufgabe zu beenden, bevor die notwendigen In-

formationen aus dem Arbeitsgedächtnis verlorengegangen sind. Nach dem Resource-Sharing-Modell sowie nach der Task-Switching-Hypothese (vgl. Abschnitt 3.2) führt eine schnellere bzw. effizientere Informationsverarbeitung zu besseren Leistungen in Arbeitsgedächtnisaufgaben. Demzufolge wird Speed häufig als ein vermittelnder Faktor zwischen dem Arbeitsgedächtnis und fluider Intelligenz verstanden (z. B. Kail & Salthouse, 1994; Jensen, 1998).

Der empirische Zusammenhang von Arbeitsgedächtnis und Speed ist bisher weniger gut belegt als der Zusammenhang von Speed und allgemeiner Intelligenz. Typischerweise werden auf der Konstruktebene bedeutsame Korrelationen gefunden, diese sind jedoch tendenziell geringer als die zwischen Arbeitsgedächtnis und allgemeiner oder fluider Intelligenz. Die beiden Arbeitsgedächtnisfaktoren aus der Studie von Oberauer et al. (2000, s. o.), auf denen die Aufgaben, die die Funktion „Simultane Speicherung und Transformation“ erfassen, hoch laden, korrelierten zu  $r = .19$  und  $r = .31$  mit der Skala „Bearbeitungsgeschwindigkeit“ des BIS-4 (Jäger et al., 1997). Kyllonen und Christal (1990) berichten Zusammenhänge zwischen  $r = .35$  und  $r = .48$  von Arbeitsgedächtnis und Speed auf der Konstruktebene. Für Kinder fanden de Jong und Das-Smaal (1995) einen Zusammenhang beider Faktoren von  $r = .60$ . Ackerman et al. (2002) verwendeten eine besonders umfangreiche Batterie von 16 Speed-Tests und fanden eine Korrelation von  $r = .55$  mit dem Arbeitsgedächtnis auf der Ebene latenter Faktoren. Ebenso wie für andere Intelligenzdimensionen ist auch für Speed nicht bekannt, ob und wie sich die Stärke der Beziehung zum Arbeitsgedächtnis im Verlauf der Entwicklung verändert (Fry & Hale, 2000).

Generell scheint das Konstrukt „Speed“ in bisherigen psychometrischen Studien häufig zu global konzipiert und zu undifferenziert operationalisiert worden zu sein (Ackerman et al., 2002). So gibt es Befunde, die zeigen, dass die Zusammenhänge von Speed zu anderen kognitiven Variablen in starkem Maße davon abhängen, welche Speed-Aufgaben verwendet werden. Speed-Aufgaben mit eher hohen Anforderungen an die kognitive Verarbeitung (z. B. der Zahlen-Symbol-Test aus dem Wechsler-Test) korrelierten in einer Studie von Babcock (1994) stärker mit dem Arbeitsgedächtnis als Speed-Aufgaben mit eher niedrigen Anforderungen (z. B. Line-Marking-Test;  $r = .29$  bzw.  $r = .59$ ).

Ackerman et al. (2002) untersuchten die Annahme, dass Aufgaben, die die Wahrnehmungsgeschwindigkeit (*perceptual speed*) erfassen, umso höher mit dem Arbeitsgedächtnis korrelieren, je komplexer die Anforderungen der jeweiligen Speed-Aufgaben sind. In einem ersten Ansatz analysierten sie hierzu verschiedene Speed-Aufgaben mit einer konfirmatorischen Faktorenanalyse und konnten vier Perceptual-Speed-Faktoren (PS) belegen: „PS-Pattern Recognition“ (Wiedererkennen einfacher Muster), „PS-Scanning“ (einfache Such- und Vergleichsprozesse), „PS-Memory“ (zusätzliche Anforderungen an das Kurzzeitgedächtnis) und „PS-Complex“ (zusätzliche kognitive und Gedächtnisanforderungen). Sie stellten die Hypothese auf, dass alle PS-Faktoren mit dem Arbeitsgedächtnis positiv korrelieren, dass jedoch die Faktoren „PS-Complex“ und „PS-Memory“ stärker mit dem Arbeitsgedächtnis assoziiert sind als die kognitiv weniger anspruchsvollen Faktoren

„PS-Pattern Recognition“ und „PS-Scanning“. Es zeigte sich, dass „PS-Complex“ und „PS-Memory“ tatsächlich stärker mit einem Arbeitsgedächtnisfaktor korrelierten als die beiden einfacheren PS-Faktoren ( $r = .48$  und  $r = .46$  vs.  $r = .23$  und  $r = .39$ ). Wenn der Einfluss der allgemeinen Intelligenz aus diesen Korrelationen auspartialisiert wurde, resultierte für den Faktor „PS-Pattern Recognition“ zudem kein signifikanter Zusammenhang mehr mit dem Arbeitsgedächtnis.

In einem zweiten Ansatz operationalisierten Ackerman et al. (2002) die Komplexität der Speed-Aufgaben über die mittlere Bearbeitungszeit, die für die Beantwortung eines einzelnen Items benötigt wurde (vgl. Kyllonen, 1985). Es fand sich für die mittlere Reaktionszeit über alle Aufgaben hinweg jedoch kein bedeutsamer Zusammenhang mit dem Arbeitsgedächtnis. Die Autoren schlussfolgern aus diesem Befund, dass die auf diese Weise definierte Komplexität der Speed-Aufgaben den Zusammenhang von Speed und Arbeitsgedächtnis nicht beeinflusst.

Als Speed-Aufgaben mit eher geringer Komplexität können Aufgaben, die nach dem Hick-Paradigma (Hick, 1952) konstruiert wurden, verstanden werden. Testaufgaben nach dem Hick-Paradigma werden zur Erfassung der Informationsverarbeitungsgeschwindigkeit im Rahmen der experimentellen Psychologie verwendet und gehören zur Gruppe der so genannten *Elementary Cognitive Tasks* (ECT), die basale kognitive Prozesse erfassen. Eine Besonderheit von Aufgaben nach dem Hick-Paradigma ist, dass die Menge zu verarbeitender Information in einer Aufgabe genau spezifiziert und kontrolliert ist.

Das Hick-Paradigma geht auf den britischen Experimentalpsychologen John Hick (1952) zurück, der die Beziehung zwischen der Anzahl von Reaktionsmöglichkeiten in einer einfachen Wahlaufgabe und der Reaktionszeit von Probanden untersuchte. Er stellte fest, dass die Reaktionszeit linear mit dem binären Logarithmus der Anzahl von Antwortalternativen steigt (*Hicksches Gesetz*).

$$\Delta RT = K \log_2 (n+1) \quad K = \text{Konstante}$$

Klassischerweise wird das Hick-Paradigma mit einem speziellen Versuchsaufbau untersucht, der einen besonderen Apparat erfordert (siehe z. B. Jensen, 1998). Dieser Apparat besteht aus einem Knopf (dem „Homebutton“), den der Pb mit dem Zeigefinger gedrückt halten soll. In einem Halbkreis um den Knopf herum sind acht weitere Knöpfe angeordnet, die aufleuchten können. Sobald ein Knopf aufleuchtet, lässt der Pb den Homebutton los und drückt den leuchtenden Knopf, der mit dem Drücken erlischt. Die Leuchtknöpfe können teilweise abgedeckt werden, so dass Aufgaben mit einer, zwei, vier oder acht Antwortalternativen entstehen (dies entspricht jeweils Null, ein, zwei oder drei Bits<sup>1</sup>). Mit diesem Versuchsaufbau kann die Reaktionszeit (Zeit zwischen Aufleuchten eines Knopfes und Verlassen des Homebuttons) separat von der Bewegungszeit (Zeit zwischen Verlassen des Homebuttons und Drücken des entsprechenden Knopfes) erfasst werden.

<sup>1</sup> Bits = *Binary Digits*; Menge an Information, die Unsicherheit um die Hälfte reduziert (Jensen, 1998)

In der *Processing Accuracy and Speed Test Battery* (PAST) von Roberts, Pallier und Stankov (1999) werden Hick-Aufgaben als Paper-Pencil-Version eingesetzt. Von den insgesamt vier Hick-Aufgaben erfordert je eine die Verarbeitung von Null (Hick-0), ein (Hick-1), zwei (Hick-2) oder drei (Hick-3) Bits. Anders als im klassischen Versuchsaufbau werden keine Reaktionszeiten erhoben, sondern es wird die Anzahl richtig ausgeführter (Wahl-)Reaktionen in einem bestimmten Zeitraum erfasst. Die Stimuli sind dabei Zahlen und die geforderten Antworten Pfeile, die in eine bestimmte Richtung zeigen (siehe auch Abbildung 6.6 in Abschnitt 6.2.2). Während die Aufgaben Hick-0 und Hick-1 nur sehr geringe Anforderungen an die kognitive Verarbeitung stellen, sind die Aufgaben Hick-2 und Hick-3 kognitiv anspruchsvoller und kommen in ihrer Anforderung den Speed-Tests, wie sie in Intelligenztests verwendet werden, nahe. Die Aufgabe Hick-0 erfordert lediglich schnelle Reaktionen (in der hier verwendeten Paper-Pencil-Version: schnelles Zeichnen von Pfeilen), die Aufgaben Hick-1 bis Hick-3 hingegen in zunehmendem Maße zusätzlich Entscheidungen zwischen verschiedenen Reaktionen (hier: die Auswahl des jeweils richtigen Pfeils).

Roberts, Stankov und Pallier (1999) untersuchten die Faktorstruktur von 25 Elementary Cognitive Tasks der PAST-Batterie im Paper-Pencil-Format und ihre Zusammenhänge mit Intelligenzfaktoren an 350 erwachsenen Pbn. In der exploratorischen Faktorenanalyse über alle 25 ECTs der PAST-Batterie zeigte sich, dass die Aufgaben Hick-0 und Hick-1 als einzige Aufgaben der Batterie hoch auf einem Processing-Speed-Faktor laden (.86 bzw. .88). Die komplexere Aufgabe Hick-2 lädt sowohl auf dem Processing-Speed-Faktor (.50) als auch auf dem Faktor „Complexity“ (.37), der durch ECTs bestimmt wird, die ein Vergleichen und Klassifizieren von Stimuli erfordern. Die komplexeste Aufgabe Hick-3 lädt substantiell nur auf dem Complexity-Faktor (.47). Innerhalb verschiedener ECTs erfassen somit die Aufgaben Hick-0 und Hick-1 die relativ simple Informationsverarbeitungsgeschwindigkeit, die Aufgaben Hick-2 und Hick-3 hingegen komplexere Verarbeitungsprozesse. Die Hick-Aufgaben korrelieren bedeutsam mit verschiedenen Intelligenzfaktoren. Es finden sich mittlere Korrelationen mit der Wahrnehmungsgeschwindigkeit ( $g_s$ -Faktor,  $r = .38$  bis  $r = .54$ ), aber nur moderate Zusammenhänge mit einem  $g_f$ -Faktor ( $r = .21$  bis  $r = .23$ ). Die Zusammenhänge der Hick-Aufgaben mit der kristallinen Intelligenz und einem Kurzzeitgedächtnisfaktor sind hingegen zu vernachlässigen. Zur Höhe des Zusammenhangs der Paper-Pencil-Version mit Maßen aus der klassischen Versuchsanordnung liegen keine Informationen vor.

### 3.4.3 Kurzzeitgedächtnis

Wie in Abschnitt 3.1 dargestellt wurde, stammt das Konzept des Arbeitsgedächtnisses unter anderem aus der Gedächtnisforschung. Bereits Baddeley und Hitch (1974) beschäftigten sich mit der Abgrenzung eines passiven Kurzzeitspeichers von einem komplexeren und aktiveren Arbeitsgedächtnissystem. Auch die Arbeit von Daneman und Carpenter (1980) wurde durch die Frage nach der Abgrenzung eines aktiveren Kurzzeitspeichers vom klassischen eher passiven Modell motiviert. Auf der anderen

Seite spielten Aufgaben zur Erfassung des Kurzzeitgedächtnisses in Intelligenzmessungen stets eine wichtige Rolle. Schon Binet verwendete einfache Spannenaufgaben in seinen Tests (z. B. Zahlenspanne; Binet & Simon, 1905), und früh in der Geschichte der Psychologie konnten positive Zusammenhänge des Kurzzeitgedächtnisses mit der allgemeinen Intelligenz und dem Schulerfolg gezeigt werden (Whipple, 1921). Auch in den aktuell verwendeten Intelligenztests sind häufig Kurzzeitgedächtnisaufgaben enthalten (z. B. HAWIK III, K-ABC).

Es scheint klar, dass Kurzzeit- und Arbeitsgedächtnis miteinander verwandte Konstrukte sind, denn beide erledigen kurzzeitige Behaltensprozesse. Andererseits unterscheiden sich Kurzzeit- und Arbeitsgedächtnisaufgaben deutlich in ihren Anforderungen und in ihren Bezügen zu anderen kognitiven Leistungen (vgl. z. B. Abschnitt 4.1.2). Brainerd und Kingma (1985) konnten beispielsweise zeigen, dass die Leistungen von Kindern in Reasoning-Aufgaben unabhängig von ihren Leistungen in Kurzzeitgedächtnisaufgaben sind. In verschiedenen Studien konnten jedoch Zusammenhänge zwischen Kurzzeitgedächtnisaufgaben und allgemeiner Intelligenz nachgewiesen werden (siehe Carroll, 1993). Engle, Tuholski et al. (1999) nehmen hierzu an, dass entsprechende Zusammenhänge letztlich auf Leistungen des Arbeitsgedächtnisses zurückzuführen sind. Werden sowohl das Arbeitsgedächtnis als auch das Kurzzeitgedächtnis einer Person gemessen, so sollte das Kurzzeitgedächtnis dieser Annahme zufolge keinen Beitrag zur Varianzaufklärung der allgemeinen Intelligenz über den Beitrag des Arbeitsgedächtnisses hinaus liefern (s. u.). Bisher besteht keine Einigkeit über die genaue Beziehung zwischen Kurzzeit- und Arbeitsgedächtnis (Engle, Tuholski et al., 1999).

Eine Modellvorstellung zur Beziehung von Kurzzeit- und Arbeitsgedächtnis stammt von Cowan (1995; siehe z. B. auch Engle, Tuholski et al., 1999; Engle, Kane et al., 1999): Es wird angenommen, dass das Kurzzeitgedächtnis eine Komponente des Arbeitsgedächtnisses ist. Zusätzlich enthält das Arbeitsgedächtnis eine Aufmerksamkeitskomponente. Das Arbeitsgedächtnis setzt sich somit aus einem Kurzzeitspeicher und einem Aufmerksamkeitssystem zusammen. Dies entspricht im Prinzip dem Originalmodell von Baddeley und Hitch (1974), wenn das Kurzzeitgedächtnis und die *slave systems* als äquivalent verstanden werden.

Engle, Tuholski et al. (1999) konnten in einer Studie mit latenten Variablen zwei separate Konstrukte belegen. Sie verwendeten je drei Kurzzeitgedächtnis- und drei Arbeitsgedächtnisaufgaben und fanden entsprechend einen Kurzzeitgedächtnis- und einen Arbeitsgedächtnisfaktor, die miteinander zu  $r = .68$  korrelierten. Ein Modell mit einem gemeinsamen Faktor aus Arbeitsgedächtnis- und Kurzzeitgedächtnisaufgaben ergab einen signifikant schlechteren Fit (Fit des Zwei-Faktoren-Modells:  $\chi^2 = 10.96$ ,  $df = 9$ ,  $p > .05$ ; CFI = .99; AGFI = .93; RMSEA = .05;  $\Delta\chi^2 = 22.64$ ,  $\Delta df = 1$ ).

In einem Strukturgleichungsmodell wurden daraufhin der Arbeitsgedächtnis- und der Kurzzeitgedächtnisfaktor dazu verwendet, die fluide Intelligenz (erfasst über den SPM und den CFT) vorherzusagen. Hier erwies sich nur der Pfad vom Arbeitsgedächtnis zur fluiden Intelligenz als signifikant, nicht jedoch der Pfad vom Kurzzeit-



gedächtnis zur fluiden Intelligenz. Ein sparsameres Modell ohne einen Pfad zwischen Kurzzeitgedächtnis und fluiden Intelligenz erbrachte keinen signifikant schlechteren Fit. Die Autoren schlussfolgerten daher, dass das Kurzzeitgedächtnis keine bedeutsamen Vorhersagen in Bezug auf die fluide Intelligenz über die Vorhersageleistung des Arbeitsgedächtnisses hinaus liefert.

Weiterhin analysierten Engle, Tuholski et al. (1999) die Vorhersageleistungen der beiden Faktoren im Hinblick auf die Leistungen im verbalen und quantitativen Scholastic Aptitude Test (VSAT und QSAT). Sie gingen davon aus, dass VSAT und QSAT sowohl fluide als auch kristalline Fähigkeiten erfordern und erwarteten, dass der Arbeitsgedächtnisfaktor die fluide Komponente und der Kurzzeitgedächtnisfaktor die kristalline Komponente vorhersagt. Es zeigte sich, dass beide Faktoren signifikante Pfade zum VSAT aufwiesen, der Pfadkoeffizient des Arbeitsgedächtnisses jedoch etwas größer ausfiel. Das heißt, dass das Kurzzeitgedächtnis – hier über verbale Aufgaben gemessen – zur Varianzaufklärung des VSAT über den Beitrag des Arbeitsgedächtnisses hinaus beiträgt. Der QSAT konnte hingegen signifikant nur durch den Arbeitsgedächtnisfaktor vorhergesagt werden, der Kurzzeitgedächtnisfaktor erklärte hier keine unique Varianz. Das Arbeitsgedächtnis und das Kurzzeitgedächtnis sind diesen Analysen zufolge somit zwei miteinander verwandte, aber deutlich voneinander trennbare Konstrukte, die unterschiedlich gut zur Vorhersage fluider und kristalliner Fähigkeiten geeignet sind.

### 3.4.4 Schulleistungen

Auf der Suche nach den individuellen Faktoren, die die Schulleistung beeinflussen, wurde verschiedentlich versucht, die schulischen Leistungen von Kindern und Jugendlichen anhand der Leistungsfähigkeit des Arbeitsgedächtnisses vorherzusagen. Es ist bekannt, dass neben Umweltbedingungen, motivationalen und Persönlichkeitsvariablen und den jeweiligen Entwicklungs- und Lernmöglichkeiten vor allem die kognitiven Fähigkeiten den Schulerfolg beeinflussen (z. B. Fraser, Walberg, Welch & Hattie, 1987; Wang, Haertel & Walberg, 1993). Der Zusammenhang zwischen Intelligenz und Schulnoten wurde bereits sehr häufig untersucht. Nach einer Zusammenstellung von Fraser et al. (1987) liegen die in verschiedenen Metaanalysen berechneten mittleren Korrelationen zwischen  $r = .34$  und  $r = .51$ .

Es kann daher angenommen werden, dass auch die Arbeitsgedächtniskapazität einen Einfluss auf die schulischen Leistungen hat. Häufig wurde der Zusammenhang zwischen Arbeitsgedächtnisaufgaben und solchen Aufgaben untersucht, die Voraussetzungen für bestimmte Schulleistungen erfassen, z. B. der Lesekompetenz oder der Rechenfähigkeit (z. B. Swanson, 1992; Turner & Engle, 1989). Einige Studien untersuchten jedoch auch den direkten Zusammenhang zwischen der Leistung in Arbeitsgedächtnisaufgaben und den schulischen Leistungen, die entweder über standardisierte Tests, Zeugnisnoten oder über eine Einschätzung durch die Lehrer erhoben wurden.

Wie oben bereits dargestellt, fanden de Jong und Das-Smaal (1995) einen beträchtlichen Zusammenhang zwischen den latenten Faktoren „Arbeitsgedächtnis“ und „Schulleistung“ ( $r = .72$ ), der nicht signifikant geringer ausfiel als der Zusammenhang zwischen Intelligenz und Schulleistung. Für die Vorhersage der Leistungen in standardisierten schulischen Leistungstests sind Arbeitsgedächtnismaße diesen Befunden zufolge ebenso gut geeignet wie Intelligenzmaße.

Eine weitere Studie zum Zusammenhang von Schulleistung und Arbeitsgedächtnis im Sinne des klassischen Modells von Baddeley stammt aus Großbritannien. Gathercole, Pickering, Knight und Stegmann (2004) testeten die Arbeitsgedächtniskapazität bei 40 7-jährigen Kindern der dritten Klasse und 43 14-jährigen Jugendlichen der zehnten Klasse. Als Maße für die schulischen Leistungen wurden die Ergebnisse aus den landesweiten standardisierten Leistungstests in den Fächern Englisch und Mathematik, in der zehnten Klasse zusätzlich Naturwissenschaften, herangezogen. Auf Basis dieser Ergebnisse wurden für jedes Fach drei Leistungsgruppen gebildet (niedrige, durchschnittliche und hohe Leistungen). Die Pbn bearbeiteten verbale und numerische Arbeitsgedächtnisaufgaben einer Vorläuferversion der *Working Memory Test Battery for Children* (Pickering & Gathercole, 2001, vgl. Abschnitt 4.3.2). Über einfache Spannaufgaben wurde die phonologische Schleife erfasst (PhS-Aufgaben) und über komplexe Spannaufgaben die zentrale Exekutive (ZE-Aufgaben). Visuell-räumliche Aufgaben wurden nicht verwendet, da diese in früheren Studien keine Vorhersagen von Schulleistungen erlaubten (Gathercole & Pickering, 2000a).

Die komplexen ZE-Aufgaben korrelierten signifikant mit der Schulleistung in beiden Altersgruppen (zwischen  $r = .32$  und  $r = .53$ ). Für die PhS-Aufgaben hingegen zeigte sich ein inkonsistentes Zusammenhangsmuster: Nur einige der Aufgaben korrelierten signifikant mit der Schulleistung (zwischen  $r = .32$  und  $r = .49$ ). Die Aufgabe „Digit Recall“ (entspricht der Aufgabe „Forward Digit Span“, vgl. Abschnitt 4.1.2) beispielsweise korrelierte signifikant mit der Mathematik- und der Englischleistung der siebenjährigen Kinder, nicht jedoch der Leistung der Jugendlichen.

Bei den siebenjährigen Kindern differenzieren die Aufgaben zur Erfassung der ZE gut zwischen den drei Leistungsgruppen sowohl im Fach Englisch als auch in Mathematik. Die jeweils leistungsstärksten Kinder erzielten die besten Leistungen in den ZE-Aufgaben, die leistungsschwächsten Kinder die schlechtesten Leistungen. Mit den PhS-Aufgaben gelang die Differenzierung weniger gut: Zwar erzielte die leistungsstärkste Gruppe die besten Werte, zwischen den durchschnittlichen und schwachen Schülern konnten die Aufgaben jedoch nicht differenzieren. Bei den Jugendlichen differenzierten die ZE-Aufgaben ebenfalls gut zwischen den drei Leistungsgruppen in den Fächern Englisch, Mathematik und Naturwissenschaften. Dieselbe Differenzierung gelang mit den PhS-Aufgaben für die Leistungen in Mathematik und in den Naturwissenschaften. Die PhS-Aufgaben differenzierten jedoch nicht zwischen den Leistungsgruppen im Fach Englisch – hier ergaben sich keine signifikanten Unterschiede zwischen den Mittelwerten der drei Leistungsgruppen. Lediglich für die einzelne ZE-Aufgabe „Backward Digit Span“ konnte ein signifikanter Gruppeneffekt im Leistungsbereich Englisch gefunden werden.

Die Befunde zeigen, dass komplexe Spannaufgaben in beiden Altersgruppen moderat mit der Schulleistung korrelieren, einfache Spannaufgaben hingegen tendenziell nur bei den jüngeren Schülern, nicht jedoch bei den Jugendlichen. Zu denselben Schlussfolgerungen hinsichtlich des Zusammenhangs einfacher und komplexer Spannaufgaben mit der Schulleistung (hier erfasst über Schulnoten) gelangte Lehto (1995) in seinen Untersuchungen mit 15- und 16-jährigen finnischen Schülern. Insgesamt konnten daher in den verschiedenen Schulsystemen unterschiedlicher Länder für Schüler unterschiedlicher Klassenstufen bedeutsame Zusammenhänge zwischen dem Arbeitsgedächtnis und der schulischen Leistung gezeigt werden. Wenn die Zusammenhänge auf der Konstruktebene betrachtet werden, dienen Arbeitsgedächtnismaße bisherigen Befunden zufolge als gleichwertige Prädiktoren für Schulleistungen wie Intelligenzmaße (siehe de Jong & Das-Smaal, 1995).

### 3.5 Zusammenfassung

Das Arbeitsgedächtnis wurde sowohl in der experimentellen Kognitions- und Gedächtnisforschung als auch in Studien zu individuellen Unterschieden vielfach untersucht. Es wurden viele verschiedene Aufgaben entwickelt und erprobt, die die Fähigkeit zur simultanen Speicherung und Transformation von Informationen erfassen und gut zwischen Personen mit unterschiedlichen Fähigkeiten differenzieren. Zwar gibt es einige theoretische Modellvorstellungen zu individuellen Unterschieden in der Kapazität des Arbeitsgedächtnisses, aus einer psychometrischen Perspektive existiert jedoch noch kein Modell des Arbeitsgedächtnisses, das allgemein geteilt wird. Die verschiedenen Arbeitsgedächtnisaufgaben, die in einer Studie eingesetzt werden, korrelieren zwar in aller Regel miteinander, die Faktorenstruktur ist jedoch noch umstritten. In der Kontroverse geht es überwiegend um die Inhaltsfacette des Arbeitsgedächtnisses; es liegen Belege sowohl für eine Domänengeneralität als auch für eine Domänenspezifität vor. Verschiedene methodische Besonderheiten der Studien schränken die Vergleichbarkeit der Befunde jedoch ein (z. B. die Verwendung sehr unterschiedlicher, oft relativ unsystematischer Aufgabenzusammenstellungen).

In den meisten Studien wurden junge erwachsene, meist studentische Pbn untersucht, nur einige Studien haben bisher die Arbeitsgedächtnisstruktur von Kindern untersucht. Für Kinder zeigt sich bislang kein systematisch anderes Bild als bei Erwachsenen. Mögliche Veränderungen der Struktur über verschiedene Altersgruppen im Kindesalter hinweg wurden bisher nicht untersucht. Es ist weiterhin keine Untersuchung bekannt, die Aussagen über die Binnenstruktur des Arbeitsgedächtnisses bei überdurchschnittlich oder hoch begabten Kindern macht. Auf der Grundlage von entsprechenden Befunden der Intelligenzforschung kann jedoch vermutet werden, dass sich die Faktorenstruktur des Arbeitsgedächtnisses von durchschnittlich und hoch begabten Kindern bedeutsam unterscheiden könnte (vgl. Mackintosh & Bennett, 2003). Es ist bekannt, dass in Stichproben mit hoher Intelligenz die Korrelationen zwischen Testaufgaben geringer ausfallen als in Stichproben durchschnittlicher oder niedriger Intelligenz (z. B. Detterman & Daniel, 1989).

Zum nomologischen Netzwerk des Arbeitsgedächtnisses wurden Befunde aus Studien dargelegt, die Korrelationen zwischen dem Arbeitsgedächtnis, der allgemeinen und fluiden Intelligenz, Speed, dem Kurzzeitgedächtnis und der Schulleistung auf Aufgaben- und auf Konstruktebene berichten. Während die Korrelationen zwischen diesen Konstrukten auf der Aufgabenebene zwar in der Regel signifikant, jedoch eher gering ausfallen, finden sich auf der Ebene latenter Faktoren größere Zusammenhänge. Generell bestehen die stärksten Zusammenhänge zwischen dem Arbeitsgedächtnis und der fluiden Intelligenz bzw. der Verarbeitungskapazität sowie der Schulleistung bei Kindern. Die berichteten Zusammenhänge zwischen latenten Faktoren variieren in verschiedenen Studien jedoch beträchtlich. Dies lässt sich zum einen auf deutliche Unterschiede im verwendeten Aufgabenmaterial (z. B. Beschränkung auf komplexe Spannaufgaben oder Einbeziehung auch einfacher Spannaufgaben) und zum anderen auf Besonderheiten der untersuchten Stichproben zurückführen. Im Hinblick auf die Pbn findet sich zum Teil die Verwendung hochselektiver Gruppen mit eingeschränkter Fähigkeitsvarianz, auch das Lebensalter der Pbn kann eine Rolle spielen, falls Entwicklungsprozesse für die beobachteten Zusammenhänge relevant sind. Verschiedene Maße für „Speed“ und „Kurzzeitgedächtnis“ kovariieren ebenfalls mit Arbeitsgedächtnisaufgaben, die Zusammenhänge auf Aufgaben- und Konstruktebene fallen jedoch geringer aus als die Zusammenhänge zwischen Arbeitsgedächtnis und Intelligenz.

## **4. Die Messung der Arbeitsgedächtniskapazität**

Dieses Kapitel beschäftigt sich mit verschiedenen Aspekten der Messung der Arbeitsgedächtniskapazität. Es werden zunächst mit der „Reading Span“ und der „Backward Digit Span“ zwei prototypische Arbeitsgedächtnisaufgaben aus der differentiellen Forschungstradition vorgestellt. Abschnitt 4.2 befasst sich im Anschluss daran mit verschiedenen Methoden der Auswertung der Testergebnisse (Scoringmethoden). Insbesondere werden verschiedene Möglichkeiten der Aggregation von Rohwerten referiert und diskutiert. In bisherigen Studien wurden deutlich voneinander abweichende Methoden verwendet, die jeweils unterschiedliche Vor- und Nachteile aufweisen. In Abschnitt 4.3 werden drei standardisierte Testverfahren beschrieben, in denen bereits Arbeitsgedächtnisaufgaben für die Individualdiagnostik verwendet werden. Ein weiterer Abschnitt (4.4) widmet sich der Messgüte häufig eingesetzter Arbeitsgedächtnisaufgaben. Abschnitt 4.5 fasst die Ausführungen in diesem Kapitel noch einmal zusammen.

### **4.1 Klassische Aufgabentypen zur Erfassung der Arbeitsgedächtniskapazität**

Entsprechend den unterschiedlichen Forschungstraditionen bei der Untersuchung des Arbeitsgedächtnisses (experimentelle Forschung vs. Untersuchung individueller Unterschiede) unterscheiden sich auch die Messmethoden. In der eher experimentell orientierten Arbeitsgedächtnisforschung erfolgt die Messung in der Regel nach dem Dual-Task-Interferenz-Paradigma. Hier wird eine kognitive Aufgabe mit einer sekundären Aufgabe gepaart, von der angenommen wird, dass sie eine bestimmte Komponente des Arbeitsgedächtnisses berührt. Wenn die Leistung in der kognitiven Aufgabe durch die sekundäre Aufgabe – im Vergleich zur Kontrollbedingung ohne sekundäre Aufgabe – beeinträchtigt wird, so wird geschlussfolgert, dass die entsprechende Arbeitsgedächtniskomponente an der Leistung in der primären kognitiven Aufgabe beteiligt ist (Baddeley & Logie, 1999). Im differentiellen Ansatz werden in der Regel komplexe Spannaufgaben eingesetzt, die eine simultane Speicherung und Verarbeitung von Informationen erfordern. Im Folgenden wird die Messung der Arbeitsgedächtniskapazität aus der Forschungstradition der Untersuchung individueller Unterschiede heraus dargestellt.

In der Literatur finden sich Beschreibungen einer Vielzahl von Aufgaben, die die Messung der Kapazität des Arbeitsgedächtnisses anzielen und die die doppelte Anforderung des simultanen Speicherns und Verarbeitens umsetzen. Dabei werden vor allem verbale und numerische, zunehmend auch räumliche Inhalte verwendet. Zwar werden bestimmte Aufgabentypen häufig und von verschiedenen Forschergruppen eingesetzt (z. B. Varianten der Aufgabe „Reading Span“, s. u.), es gibt bisher jedoch keine „Standardtests“ für die Erfassung der Leistungsfähigkeit des Arbeitsgedäch-

nisses (Schweizer & Moosbrugger, 2000). Oberauer (2005) schlägt drei Kriterien für gute Arbeitsgedächtnisaufgaben vor:

- a) Hohe Korrelationen mit anderen Arbeitsgedächtnisaufgaben (Konstruktvalidität);
- b) hohe Korrelationen mit anderen kognitiven Fähigkeiten, von denen angenommen wird, dass sie von der Arbeitsgedächtniskapazität abhängen (konvergente Validität), und keine Zusammenhänge mit Variablen, die unabhängig vom Arbeitsgedächtnis sind (diskriminante Validität);
- c) experimentelle Manipulationen sollten zu Effekten führen, die aufgrund der zugrunde gelegten Arbeitsgedächtnistheorie erwartet werden.

Zwei prototypische und häufig eingesetzte Aufgabentypen, die neben vier weiteren Aufgabentypen auch in der vorliegenden Studie verwendet wurden, werden im Folgenden kurz vorgestellt. Für beide konnte in zahlreichen Studien die Konstruktsowie Kriteriumsvalidität für die Erfassung der Arbeitsgedächtniskapazität nachgewiesen werden (z. B. Oberauer et al., 2000).

#### **4.1.1 Aufgabentyp „Reading Span“**

Die Originalversion des Aufgabentyps „Reading Span“ wurde von Daneman und Carpenter (1980) zur Erklärung von individuellen Unterschieden in der Lesefähigkeit junger Erwachsener entwickelt. Es handelte sich dabei um die erste komplexe Spannaufgabe. Seitdem ist die „Reading Span“ die am häufigsten eingesetzte Arbeitsgedächtnisaufgabe zur Erfassung individueller Unterschiede (Oberauer et al., 2000).

Daneman und Carpenter (1980) führten zunächst folgende Version der „Reading Span“ durch: Der Pb liest nacheinander einzelne Sätze, die ihm auf Karten vom Versuchsleiter präsentiert werden, in seinem eigenen Tempo laut vor. Sobald der Pb den Satz gelesen hat, wird ihm die nächste Karte vorgelegt. Gleichzeitig während des Lesens soll er sich das letzte Wort jedes Satzes merken. Nach mehreren Sätzen soll der Pb die jeweils letzten Wörter der Sätze mündlich in der richtigen Reihenfolge wiedergeben. Die Schwierigkeit der Aufgaben wird variiert über die Anzahl der enthaltenen Sätze (zwei bis sechs Sätze), von jeder Schwierigkeitsstufe werden dem Pb drei Aufgaben vorgelegt. Die Arbeitsgedächtnisspanne des Pb entspricht der maximalen Anzahl von letzten Wörtern, die er sich richtig merken kann. Dabei müssen mindestens zwei der drei Aufgaben eines Schwierigkeitslevels korrekt beantwortet werden. Die einzelnen Items werden nur so lange dargeboten, wie der Pb benötigt, um den Satz zu lesen (ca. fünf Sekunden), und sie folgen ohne Pause direkt aufeinander. Auf diese Weise sollen aktive Strategien wie ein bewusstes Wiederholen der Wörter verhindert werden. Daneman und Carpenter (1980) verwendeten zusätzlich eine auditive Variante der Aufgabe zur Erfassung der „Listening Span“. Damit konnten sie zeigen, dass die gleichen Prozesse auch beim Hörverstehen ablaufen und eine visuelle Verarbeitung der Sätze nicht relevant zu sein scheint. Für die Listening-Span-Variante wurden die Sätze so umgestaltet, dass der Pb jeweils entscheiden muss, ob der präsentierte Satz wahr oder falsch ist. So stellten die Autorinnen sicher, dass die Pbn den ganzen Satz verarbeiten und sich nicht nur auf das jeweils letzte

Wort konzentrieren. Zu Vergleichszwecken konstruierten Daneman und Carpenter zusätzlich eine parallele Variante, bei denen die Pbn die Sätze leise lesen und dann entscheiden sollen, ob der Satz wahr oder falsch ist. Die Reading-Span-Aufgabe, die in der vorliegenden Arbeit verwendet wurde, entspricht einer computergestützten Version dieser letzten Variante. Die Aussagesätze in den Aufgaben von Daneman und Carpenter waren Quizbüchern zu allgemeinem Wissen entnommen und wiesen eine mittlere Schwierigkeit auf.

Die verschiedenen Versionen der Reading-Span-Aufgabe von Daneman und Carpenter (1980) werden heute noch eingesetzt. Einige Studien verwendeten in der Folge computergestützte Aufgabenversionen (z. B. Engle et al., 1992; Oberauer et al., 2000). Die auditive Listening-Span-Aufgabe wird heute bei jüngeren Kindern eingesetzt, bei denen noch nicht von einer ausreichenden Lesefähigkeit ausgegangen werden kann (z. B. Gathercole & Pickering, 2000a; Chiappe, Hasher & Siegel, 2000). In vielen Studien konnte sich die „Reading Span“ psychometrisch als guter Indikator für das Arbeitsgedächtnis bewähren (z. B. Oberauer et al., 2003).

Die meisten Arbeitsgedächtnisaufgaben sind ähnlich aufgebaut wie die klassische Aufgabe „Reading Span“ von Daneman und Carpenter (1980): Die Verarbeitungs- und Speicheranforderungen wechseln sich innerhalb der Aufgabe ab, typischerweise erfolgt dabei zuerst die Verarbeitungsanforderung. Die einzelnen Operationen sind in der Regel relativ einfach gehalten, so dass sie im Prinzip alle Pbn richtig ausführen könnten, wenn sie separat dargeboten werden würden. Bei einigen Aufgabentypen wird dieselbe Information verarbeitet und gespeichert, bei anderen Aufgabentypen handelt es sich um unterschiedliche Informationen.

Es wurden jedoch weitere Aufgabentypen entwickelt und eingesetzt, z. B. Aufgaben nach dem Prinzip von Brown-Peterson-Aufgaben (Stuart-Hamilton, 1996; siehe z. B. Lustig et al., 2001). Hier wird eine Liste mit Items auswendig gelernt, dann werden ablenkende Aufgaben ausgeführt (z. B. Rückwärtszählen, einfache Wahlaufgaben). Anschließend sollen die Items der Liste wiedergegeben werden. Bei diesem Aufgabentyp erfolgt somit zuerst eine Speicheranforderung, dann folgen mehrere Verarbeitungsaufgaben, die die Aufrechterhaltung der zu erinnernden Items stören sollen. In der vorliegenden Arbeit entspricht die Aufgabe „Verbal Span“ diesem Prinzip (vgl. Abschnitt 6.1.2). Weitere Aufgabentypen erfordern ein fortlaufendes „Update“ von bestimmten Informationen. Die Werte von zu erinnernden Items müssen ständig entsprechend bestimmten Transformationsanweisungen verändert werden. Beispielsweise müssen in der Aufgabe „Memory Updating Figural“ in der vorliegenden Arbeit die Positionen von Punkten in einem Feld entsprechend bestimmten Instruktionen mental verschoben werden (vgl. Abschnitt 6.1.5). Eine umfassende Zusammenstellung verschiedener Aufgabentypen zur Erfassung des Arbeitsgedächtnisses findet sich bei Oberauer et al. (2000).

#### 4.1.2 Aufgabentyp „Backward Digit Span“

Aufgaben zur Erfassung der Zahlenspanne („Digit Span“) eines Pbn sind so aufgebaut, dass dem Pb in kurzer Folge entweder auditiv oder visuell (auf einem Computerbildschirm) einzelne Ziffern vorgegeben werden. Anschließend soll der Pb die korrekte Zahlenreihe wiedergeben. Zahlenspannen-Tests werden bereits seit längerem häufig für die Erfassung der Gedächtnisspanne eingesetzt. Interpretiert wurden die Ergebnisse als Ausdruck des auditiven Kurzzeitgedächtnisses (z. B. Torgesen & Houck, 1980) oder der Aufmerksamkeit bzw. Ablenkbarkeit (z. B. Tarver & Hallahan, 1974; Case & Globerson, 1974). Auch als Maß für die Intelligenz wurden und werden Zahlenspannen-Tests vielfach verwendet. Sie waren bereits in den ersten Tests von Binet und Simon (1905) enthalten und werden auch heute noch zur Messung der verbalen Intelligenz eingesetzt (Tewes et al., 2000; Kubinger & Wurst, 2000). Später etablierten sie sich als Indikatoren für die Arbeitsgedächtniskapazität (z. B. Kyllonen & Christal, 1990; Gathercole & Pickering, 2000a). Üblich sind zwei Varianten der Aufgabe, zwischen denen in vielen Studien jedoch nicht explizit unterschieden wird (z. B. de Jong & Das-Smaal, 1995). Bei der *Forward Digit Span (FDS)* soll die Zahlenreihe in der Originalreihenfolge wiedergegeben werden, bei der *Backward Digit Span (BDS)* in der genau umgekehrten Reihenfolge. Beide Varianten werden beispielsweise auch im Hamburg-Wechsler Intelligenztest für Kinder III (HAWIK-III, Tewes et al., 2000) im Untertest „Zahlennachsprechen“ verwendet und gemeinsam interpretiert. Das Adaptive Intelligenz Diagnostikum 2 (AID 2, Kubinger & Wurst, 2000) enthält mit dem Untertest „Unmittelbares Reproduzieren – numerisch“ ebenfalls einen Zahlenspannen-Test. Die Vorwärts- und die Rückwärtsvariante wird hier jedoch separat ausgewertet.

Die kombinierte Verwendung beider Aufgabenvarianten in vielen Untersuchungen vernachlässigt jedoch, dass die Vorwärts- und die Rückwärtsvariante unterschiedliche kognitive Prozesse erfassen. Bei FDS handelt es sich um einen Test zur Messung des Kurzzeitgedächtnisses, da lediglich eine passive Speicherleistung verlangt wird. Im Rahmen des Baddeley-Modells (Baddeley & Hitch, 1974) erfasst FDS die Speicherkapazität der phonologischen Schleife (Gathercole & Pickering, 2000a). Anders bei der BDS: Neben der Speicherung der Ziffern erfordert die Aufgabe eine einfache mentale Transformation, indem eine Neuordnung der Reihenfolge hergestellt werden muss. Somit erfüllt die Aufgabe die Anforderung an eine komplexe Spannaufgabe, da sie sowohl eine Speicher- als auch eine Verarbeitungskomponente enthält. In der Terminologie des Baddeley-Modells handelt es sich um einen Indikator für die Leistungsfähigkeit der zentralen Exekutiven. Das, Kirby und Jarman (1979) interpretieren anhand ihres „Informations-Integrations-Modells“ die Leistungen in der FDS als Ausdruck sequentieller Verarbeitung, die Leistungen in der BDS hingegen als Resultat simultaner Verarbeitung. Bei der Vorhersage komplexerer kognitiver Leistungen zeigte sich auch empirisch, dass FDS und BDS eine unterschiedliche prädiktive Validität aufweisen. Verschiedene Studien konnten höhere Korrelationen zwischen BDS und der Intelligenz als zwischen FDS und der Intelligenz nachweisen (z. B. Jensen & Figueroa, 1975).



In Untersuchungen zur Struktur des Arbeitsgedächtnisses zeigte sich, dass die BDS mit anderen klassischen Arbeitsgedächtnisaufgaben bedeutsam korreliert und in Faktorenanalysen mit ihnen auf einen gemeinsamen Faktor lädt. Oberauer et al. (2000) fanden, dass BDS ein guter Indikator für die Fähigkeit zur simultanen Speicherung und Transformation ist (Oberauer et al., 2003). Die BDS ist – auch im Vergleich zu anderen Arbeitsgedächtnisaufgaben – besonders gut dazu geeignet, zwischen Gruppen unterschiedlicher akademischer Leistungsfähigkeit zu differenzieren (Gathercole et al., 2004). Inzwischen kann die Aufgabe BDS als ein Klassiker unter den Aufgaben zur Erfassung der Arbeitsgedächtniskapazität betrachtet werden.

Möglicherweise ist die Eignung der BDS zur Erfassung der Arbeitsgedächtniskapazität jedoch vom Fähigkeitsniveau der Pbn abhängig. Die Anforderung, die Reihenfolge der Ziffernreihe vor der Wiedergabe umzukehren, könnte bei Pbn mit geringeren Fähigkeiten (z. B. Kindern oder weniger intelligenten Pbn) zentral-exekutive Fähigkeiten erfordern. Für fähigere Pbn könnte die Verarbeitungskomponente hingegen prinzipiell zu einfach sein und daher nicht so stark ins Gewicht fallen, so dass auch die BDS vorwiegend zu einer Speicheraufgabe wird. Schofield und Ashman (1986) untersuchten 323 Schülerinnen und Schüler der fünften und sechsten Klassenstufe mit der FDS und der BDS. Sie fanden eine signifikante Korrelation sowohl zwischen FDS und BDS einerseits und dem HAWIK-IQ andererseits ( $r = .33$  bzw.  $r = .37$ ). Bei einer Teilung der Stichprobe in drei Fähigkeitsgruppen auf Basis des IQ der Pbn (unterdurchschnittlich:  $IQ < 105$ , überdurchschnittlich:  $IQ$  zwischen 105 und 124, deutlich überdurchschnittlich:  $IQ > 124$ ) zeigte sich hingegen ein anderes Bild: In der Gruppe mit unterdurchschnittlicher Intelligenz ( $N = 102$ ) korrelierten – ebenso wie in der Gesamtstichprobe – sowohl FDS als auch BDS signifikant mit dem IQ. In der deutlich überdurchschnittlich intelligenten Gruppe ( $N = 75$ ) hingen jedoch weder FDS noch BDS bedeutsam mit dem IQ zusammen. Lediglich in der mittleren Gruppe ( $N = 146$ ) zeigte sich der erwartete Befund. Es ergab sich ein signifikanter Zusammenhang zwischen IQ und BDS, aber kein signifikanter Zusammenhang zwischen IQ und FDS.

Auch die Ergebnisse von Gathercole et al. (2004) können in diese Richtung interpretiert werden. Sie fanden bei siebenjährigen Kindern, nicht jedoch bei 14-jährigen Jugendlichen signifikante Zusammenhänge zwischen FDS und der Mathematik- und Englischleistung (vgl. Abschnitt 3.4.4). Einen indirekten Anhaltspunkt für diese Annahme liefern auch Engle, Tuholski et al. (1999), die in einer Studie mit Collegestudenten feststellten, dass die verbale, aber strukturell mit BDS vergleichbare Aufgabe „Backward Word Span“ eher einfache Speicherleistungen der Pbn erfasst. Die „Backward Word Span“ lud in einer Faktorenanalyse mit Kurzzeitgedächtnisaufgaben gemeinsam auf einem Faktor (und nicht gemeinsam mit anderen Arbeitsgedächtnisaufgaben). Die Autoren schlussfolgern, dass eine einfache Umkehrung der Reihenfolge aus einer Kurzzeitgedächtnisaufgabe noch keine Arbeitsgedächtnisaufgabe macht, zumindest nicht für Studierendenstichproben. Beim Einsatz in Untersuchungen mit jungen, leistungsfähigen Erwachsenen oder sehr intelligenten Kindern könnte es sich bei der BDS somit eher um eine Aufgabe zur Messung des Kurzzeitgedächtnisses handeln.

## 4.2 Scoring von Arbeitsgedächtnisaufgaben

Die Leistungen von Pbn in Arbeitsgedächtnisaufgaben können auf verschiedene Weise zu Gesamtscores zusammengefasst werden. Obwohl in der Literatur häufig ähnliche Aufgabentypen beschrieben werden, gibt es bisher keine Standardprozedur für die Auswertung und die Aggregation der Ergebnisse. Es werden unterschiedliche Scoringmethoden verwendet, in der Regel ohne dass die verschiedenen Möglichkeiten und ihre jeweiligen Konsequenzen thematisiert werden. Um einen Gesamtscore für einen Aufgabentyp bilden zu können, müssen vorab mehrere Entscheidungen getroffen werden:

- a) Insgesamt erfordern Arbeitsgedächtnisaufgaben zwei verschiedene Leistungen: eine Speicher- und eine Verarbeitungsleistung. Wird nur die Speicherleistung oder auch die Verarbeitungsleistung der Pbn berücksichtigt? (siehe Abschnitt 4.2.1)
- b) Auf welche Weise werden die Leistungen in den einzelnen Aufgaben zu einem Gesamtscore für die Skala zusammengefasst? (siehe Abschnitte 4.2.2 und 4.2.3)
- c) Wird die Einhaltung der richtigen Reihenfolge bei den Antworten der Pbn im Speicherteil einer Aufgabe bewertet? (siehe Abschnitt 4.2.4)

Diese Fragen werden im Folgenden diskutiert und als Fazit wird ein Scoring abgeleitet, das die Anforderungen von Arbeitsgedächtnismaßen möglichst genau abbildet und sich gleichzeitig zur Messung individueller Unterschiede gut eignet.

### 4.2.1 Welche Leistungsdaten werden beim Scoring berücksichtigt?

Bei einigen Aufgabentypen, z. B. bei der Aufgabe „Reading Span“ und bei Aufgaben nach dem Prinzip von Brown-Peterson-Aufgaben, lassen sich die Verarbeitungs- und die Speicherleistung separat erfassen. Andere Aufgabentypen erlauben dies nicht, z. B. die Aufgabe „Backward Digit Span“. Hier wird lediglich die richtige Wiedergabe der Zahlenreihe gefordert. Wenn die Zahlenreihe korrekt wiedergegeben wird, ist sowohl die Speicherung als auch die Verarbeitung der Ziffern und ihrer Reihenfolge gelungen. Üblicherweise besteht jedoch der Skalenscore bei solchen Aufgaben, in denen sich die Leistungen separieren lassen, lediglich aus den Leistungen des Pb in der Speicherkomponente. Bei der Aufgabe „Reading Span“ beispielsweise wird meist nur ausgezählt, wie viele Wörter sich die Person merken konnte. Die Verarbeitungskomponente wird jedoch in der Regel nicht in den Gesamtscore des Pb mit einberechnet (z. B. Kane et al., 2004; Conway et al., 2002; Oberauer et al., 2000; siehe aber auch Klein & Fiss, 1999; Waters & Caplan, 1996, sowie Hacker et al., 1999). Den Pbn wird in der Instruktion deutlich gemacht, dass beide Komponenten gleich wichtig sind. Man hofft so, dass die Instruktion dazu führt, dass die Pbn ihre Aufmerksamkeit auf beide Aufgabenkomponenten in gleichem Maße richten. Im Nachhinein erfolgt über die Daten dann eine grobe Kontrolle, ob die Pbn die Instruktion ausreichend befolgt haben: Geringe Leistungen bzw. viele Fehler eines Pb im Verarbeitungsteil deuten darauf hin, dass sich der Pb vor allem auf die Speicherkomponente der Aufgabe konzentriert hat. Im Extremfall, wenn die Verarbeitungskomponente

völlig ignoriert wurde, misst die Aufgabe nur noch das Kurzzeitgedächtnis, nicht aber das Arbeitsgedächtnis. Die Leistungen in der Speicherkomponente würden dann entsprechend höher ausfallen und die Fähigkeit des Pb überschätzen. Folglich werden Pbn mit Leistungen in der Verarbeitungskomponente unterhalb eines bestimmten Kriteriums (z. B. weniger als 85% richtige Antworten bei Engle, Tuholski et al., 1999) meist aus der Stichprobe ausgeschlossen. Dennoch „belohnt“ dieses Scoring tendenziell solche Pbn, die ihre Aufmerksamkeit stärker auf die Speicherkomponente der Aufgabe richten und die Verarbeitungskomponente eher vernachlässigen – solange ihre Leistung noch oberhalb eines bestimmten kritischen Wertes liegt.

Dass die Ergebnisse aus der Verarbeitungskomponente oft nicht verwendet werden, wird so begründet, dass beide Aufgabenteile in der Regel zwar gering, aber positiv miteinander korrelieren. (Eine negative Korrelation würde hingegen darauf hindeuten, dass eine gute Leistung in einer Aufgabenkomponente zu Lasten der Leistung in der anderen Komponente erreicht wurde.) Auch sind die Verarbeitungsaufgaben für sich genommen sehr einfach gehalten, und es werden nahezu perfekte Lösungen bei allen Pbn angestrebt. Der Verarbeitungsteil allein sollte somit nur wenig zwischen Personen differenzieren können. Beispielsweise sind bei der Aufgabe „Reading Span“ in der Version, wie sie in der vorliegenden Arbeit verwendet wird (vgl. Abschnitt 5.2), die Sätze stets trivial wahr oder falsch, so dass im Prinzip alle Pbn in der Lage sein sollten, die richtige Entscheidung zu treffen. Einzelne auftretende Fehler werden als nicht weiter relevante Konzentrationslücken interpretiert.

#### **4.2.2 Zusammenfassung verschiedener Leistungen zu einem Score**

In der Arbeitsgedächtnisforschung wird der Skalenscore aus den Speicherleistungen auf verschiedene Weise errechnet. Die Verwendung einer bestimmten Methode wird meist nicht näher begründet. Zum Teil werden innerhalb derselben Studie für verschiedene Aufgaben unterschiedliche Methoden verwendet (z. B. Mackintosh & Bennett, 2003). Die verschiedenen Methoden führen jedoch zu unterschiedlichen Scores, die sich z. B. in ihrer Varianz und Verteilung unterscheiden. Im Folgenden werden die verschiedenen Scoringmethoden, über die in der Literatur berichtet wird, zu drei Hauptansätzen zusammengefasst dargestellt.

Für den Vergleich der Methoden ist es zunächst erforderlich, die Ebenen der einzelnen Arbeitsgedächtnisaufgaben zu definieren, da verschiedene Scorings auf unterschiedlichen Ebenen ansetzen. Alle sechs Aufgabentypen, die in der vorliegenden Studie untersucht wurden, lassen sich auf drei verschiedenen Ebenen betrachten: Zunächst besteht jede Aufgabe aus einzelnen Items (z. B. bei der Aufgabe „Reading Span“ ist ein Item ein Satz, für den entschieden werden muss, ob er richtig oder falsch ist und von dem das letzte Wort behalten werden soll). Jede Aufgabe besteht aus mehreren solcher Items; die einfacheren Aufgaben enthalten wenige Items, die schwierigeren Aufgaben mehr Items. Schließlich lassen sich mehrere Aufgaben derselben Schwierigkeitsstufe (das heißt mit jeweils gleich vielen Items) zu so genannten „Sets“ zusammenfassen. Einige Autoren (z. B. Kane et al., 2004) nehmen eine

weitere Unterscheidung vor: Die einzelnen Aufgaben können bei der Aggregation unterschiedlich gewichtet werden, entweder werden alle Aufgaben gleich gewichtet, oder es erfolgt eine Gewichtung nach ihrer Schwierigkeit. Eine Gewichtung nach Schwierigkeit würde hier bedeuten, dass die richtige Lösung einer Aufgabe mit vielen Items – die schwierig ist, da sie hohe Anforderungen an das Gedächtnis stellt – mehr zum Gesamtscore beiträgt als die richtige Lösung einer Aufgabe mit wenigen Items. Im Folgenden werden die verschiedenen Scoringmethoden vorgestellt. Tabelle 4.1 gibt zusätzlich einen vergleichenden Überblick über die verschiedenen Methoden anhand einer Beispiellösung.

*a. Traditionelle Methode: Bestimmung der „Spanne“ auf der Setebene*

Die ersten Arbeitsgedächtnisaufgaben wurden auf der Ebene der Sets ausgewertet (Daneman & Carpenter, 1980), und auch aktuell verwenden viele Autoren diese Methode: Der Score entspricht hier dem Level des höchsten Sets, das der Pb noch lösen konnte. Dabei wird ein Set in der Regel als gelöst betrachtet, wenn mindestens zwei von drei Aufgaben richtig beantwortet wurden. Werden alle drei Aufgaben eines Sets nicht mehr gelöst, wird die Testung abgebrochen. Dieses Scoring ist relativ grob, da der Range möglicher Werte gering ist und einzelne richtige Lösungen höherer Sets nicht berücksichtigt werden. Aus diesen Gründen ergänzten andere Autoren die Methode um die Möglichkeit, zusätzliche Punkte für die teilweise Lösung eines Sets – das heißt einzelne richtig gelöste Aufgaben – zu vergeben („partial credit“). Case, Kurland und Goldberg (1982) vergaben etwa bei einem Set, das aus drei Aufgaben besteht, für jede weitere gelöste Aufgabe 0.33 Punkte. Eine ähnliche Prozedur verwendeten beispielsweise Lustig et al. (2001): Der Score entspricht hier zunächst ebenfalls dem Level des höchsten Sets, das der Pb noch lösen konnte. Dabei wird ein Set als gelöst betrachtet, wenn mindestens drei von fünf Aufgaben richtig beantwortet wurden. Werden in einem Set nur zwei von fünf Aufgaben vollständig richtig gelöst, werden dafür zusätzlich zum bisher erreichten Score 0.5 Punkte vergeben. Es müssen mindestens zwei Aufgaben richtig gelöst werden, damit die Testung mit dem nächst höheren Set fortgesetzt wird, andernfalls wird die Testung abgebrochen.

Tabelle 4.1: Verschiedene Scoringverfahren im Vergleich anhand eines Beispiel-Antwortmusters

| Ebene 1<br>Set | Ebene 2<br>Aufgaben | Ebene 3<br>Items | Antwort-<br>muster<br>(Beispiel) | Scoring<br>auf Set-<br>Ebene | Scoring auf<br>Aufgaben-Ebene |                         | Scoring auf<br>Item-Ebene<br>(partial credit für Items) |                                      | Scoring auf<br>versch.<br>Ebenen<br>Sets, Aufg.,<br>Items <sup>3</sup> |         |   |             |   |           |     |
|----------------|---------------------|------------------|----------------------------------|------------------------------|-------------------------------|-------------------------|---|--------------------------------------|--|---------|---|-------------|---|-----------|-----|
|                |                     |                  |                                  | Trad.<br>Spanne <sup>1</sup> | „weighted<br>items“           | Einheits-<br>gewichtung | Summe richtig<br>gelöster Items                         | Einheits-<br>gewichtung <sup>2</sup> |  |         |   |             |   |           |     |
| Set 1          | Aufg. 1             | Item 1_1         | 1                                | gelöst                       | 2                             | 1                       | 1   | $2/2 = 1$                            | gelöst   |         |   |             |   |           |     |
|                |                     | Item 1_2         | 1                                |                              |                               |                         | 1   |                                      |  |         |   |             |   |           |     |
|                | Aufg. 2             | Item 2_1         | 1                                |                              |                               |                         | 2   |                                      |  | 1       | 1 | $2/2 = 1$   |   |           |     |
|                |                     | Item 2_2         | 1                                |                              |                               |                         | 1   |                                      |  | 1       |   |             |   |           |     |
|                | Aufg. 3             | Item 3_1         | 1                                |                              |                               |                         | 2   |                                      |  | 1       | 1 | $2/2 = 1$   |   |           |     |
|                |                     | Item 3_2         | 1                                |                              |                               |                         | 1   |                                      |  | 1       | 1 |             |   |           |     |
| Set 2          | Aufg. 4             | Item 4_1         | 1                                | 3                            | 3                             | 1                       | 1   | $3/3 = 1$                            | 3  |         |   |             |   |           |     |
|                |                     | Item 4_2         | 1                                |                              |                               |                         | 1   |                                      |  |         |   |             |   |           |     |
|                |                     | Item 4_3         | 1                                |                              |                               |                         | 1   |                                      |  |         |   |             |   |           |     |
|                | Aufg. 5             | Item 5_1         | 1                                |                              |                               |                         | 3   |                                      |  | 1       | 1 | $3/3 = 1$   |   |           |     |
|                |                     | Item 5_2         | 1                                |                              |                               |                         | 1   |                                      |  | 1       | 1 |             |   |           |     |
|                |                     | Item 5_3         | 1                                |                              |                               |                         | 1   |                                      |  | 1       | 1 |             |   |           |     |
|                | Aufg. 6             | Item 6_1         | 1                                |                              |                               |                         | 0   |                                      |  | 0       | 1 | $1/3 = .33$ |   |           |     |
|                |                     | Item 6_2         | 0                                |                              |                               |                         | 0   |                                      |  | 0       | 0 |             |   |           |     |
|                |                     | Item 6_3         | 0                                |                              |                               |                         | 0   |                                      |  | 0       | 0 |             |   |           |     |
|                | Set 3               | Aufg. 7          | Item 7_1                         |                              |                               |                         | 1   |                                      |  | (+ 0.3) | 4 | 1           | 1 | $4/4 = 1$ | 0.5 |
|                |                     |                  | Item 7_2                         |                              |                               |                         | 1   |                                      |  |         |   |             | 1 |           |     |
|                |                     |                  | Item 7_3                         |                              |                               |                         | 1   |                                      |  |         |   |             | 1 |           |     |
| Item 7_4       |                     |                  | 1                                | 1                            |                               |                         |   |                                      |  |         |   |             |   |           |     |
| Aufg. 8        |                     | Item 8_1         | 0                                | 0                            | 0                             | 0                       | $2/4 = .5$  | $2/4 \cdot 0.5 = 0.25$               |  |         |   |             |   |           |     |
|                |                     | Item 8_2         | 0                                | 0                            | 0                             | 0                       |   |                                      |  |         |   |             |   |           |     |
|                |                     | Item 8_3         | 1                                | 1                            | 1                             | 1                       |   |                                      |  |         |   |             |   |           |     |
|                |                     | Item 8_4         | 1                                | 1                            | 1                             | 1                       |   |                                      |  |         |   |             |   |           |     |
| Aufg. 9        |                     | Item 9_1         | 1                                | 0                            | 0                             | 1                       | $3/4 = .75$   | $3/4 \cdot 0.5 = 0.375$              |  |         |   |             |   |           |     |
|                |                     | Item 9_2         | 1                                | 1                            | 1                             | 1                       |   |                                      |  |         |   |             |   |           |     |
|                |                     | Item 9_3         | 1                                | 1                            | 1                             | 1                       |   |                                      |  |         |   |             |   |           |     |
|                |                     | Item 9_4         | 0                                | 0                            | 0                             | 0                       |   |                                      |  |         |   |             |   |           |     |
| Score          |                     |                  |                                  | 3 (3.3)                      | 16                            | 6                       | 22 bzw. $22/9=2.44$                                     | $7.58/9 = .84$                       | 4.125  |         |   |             |   |           |     |
| Max. Range     |                     |                  |                                  | 0, 2-4                       | 0-27                          | 0-9                     | 0-27 bzw. 0-3   | 0-1                                  | 0, 2- 4.25   |         |   |             |   |           |     |

Anmerkung. <sup>1</sup> (mit partial credit, nach Case et al., 1982); <sup>2</sup> Methode nach Kane et al. (2004); <sup>3</sup> Methode nach Towse et al. (1998)

### *b. Scoring auf der Aufgabenebene*

*Methode der „gewichteten Items“ (z. B. Engle, Tuholski et al., 1999; Chiappe et al., 2000; Ackerman et al., 2002)*

In diesem Ansatz entspricht der Score eines Pb der Anzahl richtig gelöster Items aus den Aufgaben, die komplett richtig gelöst wurden. Einzelne richtig gelöste Items innerhalb nicht vollständig gelöster Aufgaben werden ignoriert. Zum Beispiel löst ein Pb in einem Set mit dem Schwierigkeitslevel 2 (= zwei Items pro Aufgabe) drei Aufgaben richtig und in einem Set auf Level 3 (= drei Items pro Aufgabe) eine Aufgabe richtig. Somit erhält er einen Score von  $2 \cdot 3 + 3 \cdot 1 = 9$ . Die höhere Gedächtnisbelastung der Sets auf höherem Level fließt daher als Gewichtung in den Score mit ein. Die richtige Lösung einer Aufgabe auf einem hohen Level trägt überproportional mehr zum Gesamtscore bei als die richtige Lösung eines Items in einer Aufgabe auf niedrigem Level (z. B. Kane et al., 2004; Lustig et al., 2001). Eine Besonderheit dieses Scorings ist somit, dass Leistungen in den schwierigeren Aufgaben belohnt werden. Nachteilig ist jedoch, dass es in der Regel zu positiv (d.h. links-) schiefen Verteilungen führt (Kane et al., 2004).

*Einheitsgewichtung auf der Aufgabenebene (z. B. Swanson, 1996; Gathercole & Pickering, 2000a)*

Hier wird die Anzahl vollständig richtig gelöster Aufgaben aufsummiert, einzelne richtig gelöste Items werden ignoriert. Dieses Scoringverfahren entspricht dem Vorgehen in der klassischen Intelligenzdiagnostik: Für jede richtig gelöste Aufgabe wird – unabhängig von ihrer Schwierigkeit und von einzelnen richtig bearbeiteten Aufgabenkomponenten – ein Punkt vergeben. Beide bisher entwickelten standardisierten Testbatterien zur Erfassung der Arbeitsgedächtniskapazität verwenden dieses Scoring (Swanson, 1996; Pickering & Gathercole, 2001; vgl. Abschnitt 4.3). Eine weitere Scoringmethode auf Aufgabenebene, die auf der Schwellenbestimmung in der Psychophysik basiert, schlagen Hacker et al. (1999) vor.

### *c. Scoring auf der Itemebene*

*Summe richtig gelöster Items (z. B. Mackintosh & Bennett, 2003; Turner & Engle, 1989)*

Hier werden alle richtig gelösten Items aufsummiert. Dabei ist es unerheblich, ob die gesamte Aufgabe richtig gelöst wurde, oder ob nur einzelne Items daraus korrekt beantwortet wurden. Oberauer et al. (2000) berechneten so zunächst die Summe der richtigen Lösungen und bildeten dann den Mittelwert über alle Items einer Skala, den sie schließlich als Skalenwert verwendeten. Vergleicht man dieses Scoring mit der Methode der gewichteten Items, so zeigt sich: Bei vollständig gelösten Aufgaben entsprechen sich die Werte, jedoch werden hier partielle Leistungen zusätzlich gewertet. Dieses Scoring enthält – bei einer Betrachtung auf der Aufgabenebene – eine Gewichtung. Für schwierigere Aufgaben kann der Pb mehr Punkte erhalten, da diese

aus mehr Items bestehen als leichte Aufgaben. Auf der Itemebene handelt es sich jedoch um eine Einheitsgewichtung: Jedes richtig gelöste Item wird mit einem Punkt bewertet, unabhängig davon, ob es zu einem leichten oder einem schwierigen Set gehört.

*Einheitsgewichtete Methode nach Kane et al. (2004):*

Für jede Aufgabe wird berechnet, welcher Anteil der Items korrekt gelöst wurde (z. B. zwei von drei Items einer Aufgabe = 0.67), anschließend erfolgt die Berechnung des Mittelwertes aus diesen Quotienten über alle Aufgaben einer Skala. Kane et al. (2004) argumentieren, dieses Scoring habe im Vergleich zu anderen Methoden zu den am ehesten normal verteilten Werten geführt.

*d. Scoring, das die verschiedenen Ebenen kombiniert*

Towse et al. (1998) vergeben wie in der traditionellen Methode zunächst den Score des höchsten Levels (Sets), das der Pb vollständig gelöst hat. Hat er weitere Aufgaben gelöst, erhält er zusätzliche Punkte nach folgenden Regeln: Wenn eine weitere Aufgabe eines höheren Sets (von jeweils insgesamt drei Aufgaben pro Set) komplett gelöst wurde, werden 0.5 Punkte zum Score hinzuaddiert, denn das Set wird als „halb richtig“ gelöst betrachtet. Wurden weitere einzelne Items korrekt gelöst, werden diese zusätzlich gewertet. Der Anteil der gelösten Items innerhalb einer Aufgabe wird ins Verhältnis gesetzt zu den nicht gelösten Items der gleichen Aufgabe. Dieser Quotient wird mit 0.5 multipliziert und zum bisher errechneten Score addiert.

#### **4.2.3 Vergleich verschiedener Scoringmethoden im Hinblick auf die Aggregation von Aufgabenwerten**

Die verschiedenen Scoringmethoden haben verschiedene Vor- und Nachteile. Ein Vorteil der *traditionellen Scoringmethode* besteht darin, dass sie am eindeutigsten von allen genannten Methoden das abbildet, was mit den Aufgaben gemessen werden soll: Der Gesamtscore ist direkter Ausdruck der Gedächtnisspanne. Bei einem Score von 4 beispielsweise ist eindeutig, dass der Pb mehrmals dazu in der Lage war, sich vier Items in Folge zu merken. Aufgaben mit fünf Items konnte er hingegen gar nicht mehr oder nur manchmal bewältigen. Dieses Vorgehen hat jedoch auch einige Nachteile. Zum einen gibt es kein objektives Kriterium für die mögliche Vergabe eines „partial credits“. Es ist somit willkürlich, ob ein ganzer, halber oder gar kein Punkt für die nicht vollständige Bewältigung eines Levels vergeben wird („zwei von fünf“, „drei von fünf“). Zum anderen differenziert dieses Scoring nur recht grob zwischen Probanden unterschiedlicher Fähigkeiten. Die relativ geringe Streuung der Werte mindert zudem die Korrelationen mit anderen Maßen.

*Scorings auf Aufgabenebene* haben den Vorteil, dass sie direkt abbilden, ob der Pb die gesamte Anforderung der jeweiligen Aufgabe erfüllt hat. Dies entspricht eher als ein Scoring auf Itemebene der Intention von Arbeitsgedächtnisaufgaben: In einer

Aufgabenreihe wird die Aufgabenschwierigkeit sukzessive gesteigert, indem die Gedächtnisbelastung erhöht wird. Es soll auf diese Weise gemessen werden, mit welcher Gedächtnisbelastung der Pb bei gleichzeitiger Verarbeitungsanforderung noch zu Recht kommt. Von Interesse ist, ob der Pb die Anforderung der gesamten Aufgabe bewältigt, sich also alle Items einer Aufgabe merken kann.

Der Vorteil von *Scorings auf Itemebene* besteht hingegen darin, dass die Information aus den Daten maximal ausgeschöpft wird. Teilleistungen werden honoriert und somit feinere Leistungsunterschiede zwischen Pbn abgebildet. Die relativ große Streuung der Scores führt zu tendenziell höheren Korrelationen der Aufgaben untereinander und mit anderen Variablen.

Verschiedene Autoren nahmen auch empirische Vergleiche von Scoringmethoden vor. Turner und Engle (1989) sowie Klein und Fiss (1999) beispielsweise verglichen das traditionelle Spannenmaß mit der Summe über alle Items bei der Aufgabe „Operation Span“. In verschiedenen Durchgängen korrelierten die beiden Summenscores zwischen  $r = .89$  und  $r = .91$  (Klein & Fiss, 1999). Turner und Engle machen keine genauen Angaben, berichten aber: „Beide Spannenmaße führten, mit wenigen Ausnahmen, zu denselben Schlussfolgerungen“ (S. 134; Übersetzung durch die Verfasserin). Auch Waters und Caplan (1996, 2003) verglichen das traditionelle Spannenmaß (mit partial credit) und die Summe über alle Items (bzw. Anteil richtiger Items an der Gesamtzahl der Items). Beide Scorings korrelierten bei verschiedenen Arbeitsgedächtnisaufgaben zwischen  $r = .70$  und  $r = .95$  miteinander. Lustig et al. (2001) verglichen die Methode des einfachen Summenscores mit der traditionellen und der gewichteten Methode. Sie berichten, dass sie keine Unterschiede zwischen den drei Scoringmethoden feststellen konnten. Towse et al. (1998) verglichen die traditionelle Methode (mit partial credit) mit ihrer Scorebildung, die die verschiedenen Ebenen kombiniert (s. o.). Sie fanden in ihrer experimentellen Studie ebenfalls keine bedeutsamen Unterschiede zwischen beiden Scores.

#### **4.2.4 Berücksichtigung der Reihenfolge von Antworten (*item-recall* vs. *serial recall*)**

Die Pbn werden in der Regel so instruiert, dass sie die Antworten zur Speicherkomponente in der exakt richtigen Reihenfolge wiedergeben sollen. Bei der „Reading Span“ beispielsweise soll der Pb alle Wörter in der Reihenfolge ihrer Darbietung wiedergeben. Das Konzept des Arbeitsgedächtnisses erfordert dies nicht zwingend, die Aufgaben werden jedoch üblicherweise so instruiert. Eine Nichtbeachtung der Reihenfolge würde die Aufgaben zwar einfacher machen, sie würden jedoch immer noch die Arbeitsgedächtniskapazität erfassen (siehe z. B. Case et al., 1982). Es sind somit prinzipiell zwei Scoring-Varianten möglich (Richardson, 1996): a. Die richtige Wiedergabe von Items wird nur gewertet, wenn auch die Reihenfolge korrekt ist (*serial recall*), und b. es zählt nur die richtige Nennung der Items, die Reihenfolge der Wiedergabe ist irrelevant (*item recall*). Beide Varianten finden sich in der Literatur, Daneman und Carpenter (1980) beispielsweise verwendeten beide Formen bei verschiedenen Versionen der „Reading Span“.



#### 4.2.5 Fazit zu den verschiedenen Scoringmethoden

Wie oben dargestellt, hat jede der Scoringmethoden bestimmte Vor- und Nachteile. Das Ziel jedes Scorings sollte jedoch sein, die intendierten Anforderungen der Arbeitsgedächtnisaufgaben möglichst genau abzubilden, Leistungsunterschiede zwischen Pbn unterschiedlicher Fähigkeiten darzustellen und der Instruktion, die den Pbn gegeben wird, möglichst genau zu entsprechen.

Hinsichtlich der Berücksichtigung von Leistungen in der Verarbeitungskomponente der Aufgaben (Frage *a*, Abschnitt 4.2.1) scheint es daher sinnvoll, von der häufig verwendeten Praxis abzuweichen, in der nur die Speicherleistung eines Pb gewertet wird. Die Leistung in der Verarbeitungskomponente sollte stärker und auf der Aufgabenebene kontrolliert werden, damit sichergestellt wird, dass die Pbn tatsächlich beide Teilaufgaben kombinieren. Gerade bei der Untersuchung von Kindern ist es wahrscheinlich, dass trotz sehr einfacher Verarbeitungsaufgaben mehr Fehler gemacht werden, als in den Studien mit erwachsenen Pbn berichtet wird. Ein aus Verarbeitungs- und Speicherleistungen kombinierter Score ist daher zu favorisieren.

Auch bei der Frage nach der Aggregation der Werte (Frage *b*, Abschnitte 4.2.2 und 4.2.3) sollte die Eindeutigkeit der Messung im Vordergrund stehen. Diese erscheint am besten gewährleistet durch ein Scoring auf der Aufgabenebene mit Einheitsgewichtung, denn hier wird bewertet, ob ein Pb die Anforderung einer ganzen Aufgabe bewältigen kann. Werden hingegen einzelne Itemleistungen bewertet, so kann ein Pb relativ viele Punkte erhalten, ohne dass er einmal die gesamte Anforderung einer Aufgabe erfüllt hat. Eine Bewertung auf der Itemebene würde auch die Strategie belohnen, sich bei schwierigeren Aufgaben mit vielen Items bewusst nur noch einige Items einzuprägen. Ein Pb, der diese Strategie anwendet, wäre im Vorteil gegenüber einem Pb, der versucht, sich alle Items zu merken – und bei schwierigen Aufgaben dann alle vergisst, wenn sein Arbeitsgedächtnis überlastet ist.

Die Verwendung der traditionellen Spannenmaße würde zwar ebenfalls zu relativ eindeutig interpretierbaren Maßen führen, aufgrund ihrer sehr geringen Varianz eignen sie sich jedoch schlechter als Scorings auf der Aufgabenebene für die Abbildung individueller Unterschiede. Sollen die Aufgaben teststatistisch analysiert werden, so werden zudem Testdaten benötigt, die auf der Aufgabenebene Eins-Null-kodiert sind. Ein Nachteil des Scorings auf der Aufgabenebene ist die geringere Varianz als bei einem Scoring auf Itemebene. Dies führt dazu, dass die Korrelationen der Aufgaben untereinander und mit anderen Maßen etwas geringer ausfallen, auch die innere Konsistenz (Cronbachs Alpha) fällt bei diesem Scoring etwas geringer aus. Dieser Nachteil sollte jedoch, in Anbetracht der Vorteile, die es für die Eindeutigkeit der Messung erbringt, in Kauf genommen werden. Ein Scoring auf der Aufgabenebene entspricht am ehesten dem Vorgehen in der klassischen Intelligenzdiagnostik und weist insgesamt ein mittleres Ausmaß an Informationsgehalt auf. Schließlich erscheint die Verwendung eines *serial recall* in den meisten Fällen sinnvoll. Wenn die Aufgaben im Sinne eines *serial recall* instruiert werden, sollte das Scoring die Reihenfolge der Antworten berücksichtigen (Frage *c*, Abschnitt 4.2.4). Insgesamt resultiert somit ein Scoring, das die Verarbeitungskomponente – wenn es die jeweilige

Aufgabe erlaubt – auf der Aufgabenebene kontrolliert, die richtige Bearbeitung vollständiger Aufgaben bewertet und die richtige Reihenfolge der Itemwiedergabe erfordert.

### 4.3 Individualdiagnostik der Arbeitsgedächtniskapazität mit standardisierten Tests

Nachdem in den vorangegangenen Abschnitten Aufgabentypen und technische Aspekte der Bewertung von Leistungen in Arbeitsgedächtnisaufgaben der differentiellen Arbeitsgedächtnisforschung dargestellt wurden, werden nun drei komplette Testverfahren vorgestellt, die den Einsatz von Arbeitsgedächtnisaufgaben in der Individualdiagnostik erlauben. Es sind bisher drei standardisierte (und zum Teil normierte) Verfahren bekannt, die eine Fähigkeitsdiagnostik bei gesunden Pbn auf der Basis von klassischen und neu entwickelten Arbeitsgedächtnisaufgaben ermöglichen. Diese Tests werden im Folgenden mit ihrem Konzept, der Aufgabenauswahl und Angaben zu den Gütekriterien Validität und Normierung kurz vorgestellt. Fragen der Messgenauigkeit oder Reliabilität der Tests und Arbeitsgedächtnisaufgaben generell werden in Abschnitt 4.4 gesondert behandelt.

Jeder der vorgestellten Tests basiert auf einem unterschiedlichen theoretischen Modell und strebt über die Erfassung der Arbeitsgedächtniskapazität Aussagen zu bestimmten unterschiedlichen Fähigkeiten an. Während etwa der S-CPT (Abschnitt 4.3.3) explizit die allgemeine Intelligenz erfassen will, zielt der SCT (Abschnitt 4.3.1) eher auf die Messung der Aufmerksamkeit. Die WMTB-C hingegen erfasst das Arbeitsgedächtnis entsprechend Baddeleys klassischem Drei-Komponenten-Modell. Allen drei Tests ist jedoch gemein, dass sie teilweise oder ausschließlich solche Aufgaben enthalten, die zur Messung des Arbeitsgedächtnisses im Sinne der simultanen Speicherung und Transformation von Informationen verwendet werden.

#### 4.3.1 Der „Star Counting Test“ (SCT)

Der *Star Counting Test* (SCT) von de Jong und Das-Smaal (1990, 1995) wurde zur Messung der zentralen Exekutiv (im Sinne des Baddeley-Modells) bzw. der Aufmerksamkeit (im Sinne des Aufmerksamkeits-Modells von Norman und Shallice, 1986) entwickelt. Ziel der Testentwicklung war die Bereitstellung eines Tests zur Diagnose von Aufmerksamkeitsdefiziten bei Kindern in Gruppentestungen. Die 44 Aufgaben des SCT erfordern das abwechselnde Vorwärts- und Rückwärtszählen mehrerer Reihen von Sternchen. Vor dem ersten Sternchen steht eine Zahl zwischen 14 und 78, von der ausgehend weitergezählt werden soll. Zwischen den Sternchen stehen in unregelmäßigen Abständen Plus- und Minuszeichen. Im ersten Testteil soll der Pb nach jedem Pluszeichen weiter vorwärts und nach jedem Minuszeichen weiter rückwärts zählen. Im zweiten Testteil wird die Bedeutung der Rechenzeichen umgekehrt: Nach einem Pluszeichen wird dann rückwärts und nach einem Minuszeichen vorwärts gezählt. Der Wechsel der Zählrichtung nach jedem Rechenzeichen erfordert die aktive Unterdrückung eines gerade ablaufenden Prozesses. Der Bedeutungswech-

sel der Rechenzeichen im zweiten Testteil erhöht die Anforderung an die zentrale Exekutive zusätzlich, denn eine gut eingeübte Reaktionsweise muss unterdrückt werden. Der Test wurde im Rahmen einer landesweiten Studie zu Aufmerksamkeitsdefizit-Störungen an 2.222 Grundschulern im Alter von 9;10 Jahren ( $SD = 5.0$  Monate) aus 111 niederländischen Grundschulen erprobt (de Jong & Das-Smaal, 1990, 1995).

Die Reliabilität des SCT erwies sich als gut (siehe Angaben in Abschnitt 4.4). Weiterhin konnte gezeigt werden, dass die Aufgaben des SCT Rasch-Modell konform sind. Die Konstruktvalidität wurde an einer Substichprobe mit 381 Pbn überprüft. Hier bildete der SCT mit drei anderen Arbeitsgedächtnisaufgaben einen gemeinsamen Faktor ( $\chi^2 = 4.11$ ,  $df = 2$ ,  $p = .13$ ). Die Leistung im SCT korreliert signifikant mit anderen als Arbeitsgedächtnisaufgaben klassifizierten Tests ( $r = .29-.53$ ), verschiedenen Speed-Maßen ( $r = .21-.23$ ), der verbalen Intelligenz ( $r = .29$ ), der figuralen Intelligenz ( $r = .34$ ) sowie mit Schulleistungstests zum Leseverstehen ( $r = .29$ ) und Rechnen ( $r = .35$ ).

Der SCT wurde äußerst sorgfältig konstruiert und weist eine hohe Messgenauigkeit auf. Die Möglichkeit der Gruppentestung ist ein großer Vorteil. Etwas unklar bleibt die Konstruktvalidität: Zwar finden sich bei de Jong und Das-Smaal (1995) hohe Zusammenhänge zwischen dem SCT und anderen Aufgaben zur Erfassung des Arbeitsgedächtnisses. Dennoch fanden de Jonge und de Jong (1996) einen separaten SCT-Faktor bei gleichzeitiger Analyse weiterer Arbeitsgedächtnisaufgaben. In der Studie von Oberauer et al. (2000) wurde die dort verwendete computergestützte SCT-Variante im hypothetischen Zwei-Facetten-Modell als Supervisions-Aufgabe eingeordnet (vgl. Abschnitt 3.3).

Trotz der gezeigten Passung auf das Rasch-Modell, das eine Eindimensionalität des SCT impliziert, erscheint es problematisch, dass die SCT-Aufgaben über zwei verschiedene Strategien gelöst werden können: Eine mögliche Lösungsstrategie ist das Vorwärts- und Rückwärtszählen auf die von den Autoren beschriebene Weise. Eine alternative Strategie ist, die Aufgaben als einfache Rechenaufgaben zu verstehen: Zunächst werden hierzu die Sternchen zwischen den Rechenzeichen gezählt, und diese Summen werden dann addiert bzw. subtrahiert. Beide Strategien entsprechen zwar der Definition der simultanen Speicherung und Verarbeitung von Informationen, berühren jedoch möglicherweise unterschiedliche kognitive Prozesse.

#### 4.3.2 Die „Working Memory Test Battery for Children“ (WMTB-C)

Die *Working Memory Test Battery for Children* (WMTB-C) von Pickering und Gathercole (2001) basiert auf dem klassischen Baddeley-Modell und wurde für die Individualdiagnostik bei Kindern zwischen 4;7 und 15;9 Jahren entwickelt. Die Testung erfolgt individuell, Gruppentestungen sind nicht möglich, da in einigen der Aufgaben die Items mündlich vom Testleiter vorgesprochen werden. Mit insgesamt neun Aufgabentypen werden die phonologische Schleife (PhS), die zentrale Exekutive (ZE) und der visuell-räumliche Skizzenblock (VRS) erfasst.

Zur Erfassung der *phonologischen Schleife (PhS)* werden vier verschiedene verbale und numerische Kurzzeitgedächtnisaufgaben verwendet. In der Aufgabe „Digit Recall“ werden dem Pb einzelne Ziffern vom Testleiter vorgesprochen, und der Pb soll sie in derselben Reihenfolge wiederholen. Zusätzlich zu dieser numerischen Aufgabe werden zwei verbale Aufgaben, die ebenfalls die Wiedergabe einer bestimmten Liste von Items erfordern, eingesetzt („Word List Recall“ und „Nonword List Recall“). Weiterhin wird mit „Word List Matching“ ein weiterer verbaler Subtest zur Messung der PhS verwendet, der keine Wiederholung der Wörter erfordert und somit weniger durch das Langzeitgedächtnis beeinflusst wird und auch für Kinder mit Sprachschwierigkeiten geeignet ist. Hier werden dem Pb zwei Listen mit denselben Wörtern präsentiert, und er soll angeben, ob die Reihenfolge der Wörter in beiden Listen identisch ist.

Zur Erfassung des *visuell-räumlichen Skizzenblocks (VRS)* dienen zwei Aufgabentypen. Der Aufgabentyp „Block Recall“ wird in der Neuropsychologie häufig als Maß für das räumliche Kurzzeitgedächtnis verwendet. Hier zeigt der Testleiter in einer bestimmten Reihenfolge nacheinander auf verschiedene, in dreidimensionaler Form angeordnete „Blocks“. Anschließend soll der Pb auf dieselben Blocks in genau der gleichen Reihenfolge zeigen. Im Aufgabentyp „Mazes Memory“ (Pickering, Gathercole, Hall & Lloyd, 2001) werden dem Pb gezeichnete Labyrinth vorgelegt, in denen der Weg vom Mittelpunkt zu einem von zwei möglichen Ausgängen eingezeichnet ist. Der Testleiter zeichnet den Weg einmal mit dem Finger nach, anschließend soll das Kind den gezeigten Weg in dasselbe Labyrinth, in dem jedoch der Weg fehlt, einzeichnen.

Die *zentrale Exekutive (ZE)* wird über drei verbale und numerische klassische Aufgabentypen erfasst, die die gleichzeitige Speicherung und Manipulation von Informationen erfordern. In der Aufgabe „Listening Recall“ wird dem Kind eine Reihe kurzer Sätze vorgelesen. Nach jedem Satz soll das Kind entscheiden, ob der Satz wahr oder falsch ist. Anschließend soll jeweils das letzte Wort jeden Satzes in der richtigen Reihenfolge aufgesagt werden. Die Aufgabe „Counting Recall“ erfordert das Zählen von Punkten, die dem Pb auf verschiedenen, nacheinander vorgelegten Karten dargeboten werden. Die Anzahl der Punkte auf den einzelnen Karten soll der Pb anschließend in der richtigen Reihenfolge nennen (ursprüngliche Version von Case et al., 1982). In der Aufgabe „Backward Digit Recall“ soll eine vorgesprochenen Zahlenreihe in der genau umgekehrten Reihenfolge wiedergegeben werden.

Die WMTB-C wurde an einer Stichprobe von insgesamt 729 Pbn normiert (zwischen 59 und 179 Kindern pro Jahrgang). Die Normierung deckt die Altersgruppen 4;7 bis 11;8 Jahre, 12;1 bis 13;1 Jahre sowie 14;9 bis 15;9 Jahre ab. Die Retestreliabilitäten der Untertests fallen eher gering aus und werden in Abschnitt 4.4 dargestellt.

Die angenommene Faktorstruktur im Sinne des Baddeley-Modells konnte in Vorläuferstudien zur Entwicklung der WMTB-C nur teilweise bestätigt werden (Gathercole & Pickering, 2000a). Die Faktorladungen der visuell-räumlichen Aufgaben waren nicht eindeutig zu interpretieren, so dass sie für weitere Analysen außer Acht gelassen wurden. Bei einer Analyse der PhS- und ZE-Aufgaben ergab sich die erwartete

Zwei-Faktoren-Struktur mit korrelierten Faktoren ( $r = .55$ ). Für die Endversion der Testbatterie resultierte jedoch eine gute Passung des erwarteten Drei-Faktor-Modells auf die Daten. Die Zusammenhänge der Faktoren untereinander schwankten in verschiedenen Altersgruppen jedoch deutlich. So korrelierte die ZE mit der PhS je nach Altersgruppe zwischen  $r = .08$  und  $r = .90$  und mit dem VRS zu zwischen  $r = .57$  und  $r = .82$ .

Die Kriteriumsvalidität wurde an sechs- und siebenjährigen Kindern überprüft, indem ihre Testleistungen in Aufgaben der PhS und der ZE mit den Ergebnissen in verschiedenen standardisierten Leistungstests zu Wortschatz, Lesefähigkeit und Rechenfähigkeit zu zwei Zeitpunkten korreliert wurden (Gathercole & Pickering, 2000a). Die ZE-Leistung korrelierte signifikant mit allen Leistungsmaßen, sowohl im Alter von sieben als auch im Alter von acht Jahren ( $r = .49$ -.61). Die PhS-Leistung korrelierte zunächst auch signifikant mit sämtlichen Schulleistungsmaßen im Alter von sieben Jahren und der Wortschatzleistung im Alter von acht Jahren. Wenn jedoch die ZE-Leistung und das Alter statistisch kontrolliert wurden, wies das PhS-Maß nur noch mit dem Wortschatz einen starken und über ein Jahr andauernden Zusammenhang auf ( $r = .41$ ), mit der Lese- und Rechenfähigkeit ergab sich hingegen weder im Alter von sieben noch im Alter von acht Jahren ein signifikanter Zusammenhang. Die externe Validität der VRS-Aufgaben wird nicht berichtet.

Somit verfügen die ZE-Aufgaben über bedeutsame Zusammenhänge mit den schulischen Leistungen, sowohl im Jahr der Testung als auch ein Jahr später. Die PhS-Aufgaben sind extern valide vor allem im Hinblick auf den Wortschatz, eine Vorhersage der anderen schulischen Leistungen über den Vorhersagebeitrag der ZE-Aufgaben hinaus erfolgt jedoch nicht.

Zusätzlich ergaben Gruppenvergleiche, dass Pbn mit unterdurchschnittlichen Leistungen in den Schulleistungstests insbesondere in den ZE-Aufgaben Defizite aufwiesen. Bei den PhS-Aufgaben waren die Ergebnisse weniger eindeutig, nur in einigen der Aufgaben schnitten schulleistungsschwache Kinder schlechter ab (Gathercole & Pickering, 2000b). Insgesamt kann festgestellt werden, dass Aufgaben zur Erfassung der zentralen Exekutive gute Prädiktoren für Schulleistungen darstellen, Aufgaben zur Erfassung der PhS jedoch eher nicht (Gathercole & Pickering, 2000a, 2000b).

Als Fazit kann festgehalten werden, dass es sich bei der WMTB-C um eine umfassende Testbatterie handelt, die die Erfassung des Arbeitsgedächtnisses im Sinne des Baddeley-Modells erlaubt. Die WMTB-C hat jedoch einige Schwachpunkte: Durch die individuelle Präsentation der Items durch einen Testleiter ist die Durchführungsobjektivität eingeschränkt. Eine Gruppentestung mit der Testbatterie ist nicht möglich. Die ZE wird lediglich über drei verbale und numerische Aufgaben erfasst, figurale Aufgaben fehlen. Die Aufgaben „Counting Recall“ und „Listening Recall“ weisen für die älteren Pbn eine zu geringe Retestrelabilität auf, die Aufgabe „Backward Digit Recall“ für die jüngeren Pbn (siehe Abschnitt 4.4). Im Hinblick auf die Vorhersage schulischer Leistungen erweisen sich die ZE-Aufgaben als valide, der zusätzliche Einsatz von PhS-Aufgaben dient lediglich einer weiteren Aufklärung des

Wortschatzes. Die PhS-Aufgaben tragen über den Beitrag der ZE-Aufgaben hinaus nichts zur Aufklärung der Lese- und Rechenleistung bei.

### 4.3.3 Der „Swanson Cognitive Processing Test“ (S-CPT)

Der *Swanson Cognitive Processing Test* (S-CPT) von Swanson (1996) erfasst über elf Untertests mit unterschiedlichen Arbeitsgedächtnisaufgaben die Leistungsfähigkeit des Arbeitsgedächtnisses. Angestrebt werden Aussagen über verschiedene Aspekte intellektueller Fähigkeiten, insbesondere Spearman's  $g$ , und das Potential zur Informationsverarbeitung bei Personen im Alter von 4;5 – 78;6 Jahren.

Die zentrale Eigenschaft aller Subtests des S-CPT (siehe Tabelle 4.2) ist die Anforderung, bestimmte Informationen zu speichern, während andere Informationen verarbeitet werden. Abgefragt wird jeweils sowohl die Speicher- als auch die Verarbeitungsleistung (durch Verständnisfragen zum gespeicherten Material). Es handelt sich somit um komplexe Spannenaufgaben, die der Vorstellung des Arbeitsgedächtnisses entspricht, die in der differentiellen Tradition vorherrscht (vgl. Abschnitt 3.2). Der S-CPT enthält sowohl solche Aufgabentypen, die eine verbale Repräsentation von Informationen erfordern (auditive Vorgabe der Items), als auch solche Aufgabentypen, die eine nicht-verbale Repräsentation erfordern (visuelle Vorgabe der Items). Auch wenn die Existenz entsprechender Subsysteme im Arbeitsgedächtnis umstritten ist, wird hier auf Befunde verwiesen, dass es bedeutsame Fähigkeitsunterschiede zwischen Individuen gibt, sich verbal oder nicht-verbal kodierte Informationen zu merken.

Zusätzlich basiert der Test auf einem Konzept, das Beziehungen zwischen dem Arbeitsgedächtnis und dem Langzeitgedächtnis betont, da davon ausgegangen wird, dass ein aktives Gedächtnissystem fortlaufend die Anforderungen der aktuellen Situation mit früher gespeicherten Informationen integrieren muss. Im S-CPT werden zwei Langzeitgedächtniskomponenten voneinander unterschieden: eine *episodische* und eine *semantische* Gedächtniskomponente (nach Tulving, 1972, 1986). Das episodische Gedächtnis spielt beispielsweise in den Untertests 5 und 6 der Testbatterie eine Rolle, das semantische Gedächtnis in den Untertests 7, 9 und 10. Eine weitere Unterteilung des Gedächtnissystems, die der S-CPT berücksichtigt, bezieht sich auf die Einteilung in ein *retrospektives* und ein *prospektives* Gedächtnis. Retrospektive Gedächtnisleistungen werden in den Untertests 1, 2, 5, 6 und 9 gefordert, prospektive Leistungen erfassen die Untertests 3, 4, 7, 8, 10 und 11.

Weiterhin verfolgt der Test einen Ansatz dynamischen Testens, der als „*testing the limits*“ (nach Carlson & Wiedl, 1990) bezeichnet wird. Im S-CPT wird die Leistung daher stets zweimal gemessen: Einmal direkt nach der Instruktion und ein zweites Mal, nachdem der Testleiter bestimmte Hinweise und Hilfestellungen gegeben hat. Es soll das Potential des Pb für Leistungsveränderungen erfasst werden, um die spätere Leistungsentwicklung besser prognostizieren zu können. Nach bestimmten Vorgaben im Testmanual werden dem Pb Teilantworten gegeben, um die Reproduktion der noch fehlenden Items zu erleichtern. Eine Durchführung als reiner Statustest

ohne Hilfestellungen ist jedoch ebenfalls möglich. Um zusätzlich die Stabilität der Leistung zu messen, wird die letzte jeweils – mit Hilfestellungen – noch richtig gelöste Aufgabe eines Untertests nach der Durchführung mehrerer anderer Untertests noch einmal bearbeitet. Es wird überprüft, ob die einmal erreichte Leistung beibehalten werden kann, ohne dass erneut Hilfestellungen erfolgen.

*Tabelle 4.2: Bezeichnungen und gemessene Fähigkeiten der Untertests des S-CPT*

| <b>Untertest</b>                 | <b>Gemessene Fähigkeit</b>   |
|----------------------------------|--|
| <b>Retrospektives Gedächtnis</b> |  |
| <b>Auditiv</b>                   |  |
| 1. Reimende Wörter               | Auditiver Abruf akustisch ähnlicher Wörter   |
| 5. Geschichte nacherzählen       | Auditiver sequentieller Abruf von Text   |
| 9. Semantische Assoziationen     | Auditiver Abruf von Wörtern nach Anordnung in entsprechende übergeordnete Konzepte   |
| <b>Visuell</b>                   |  |
| 2. Visuelle Matrix               | Räumlicher organisierter Abruf von Punktmustern                                      |
| 6. Bilderfolge                   | Räumlicher organisierter Abruf von Sets mit komplexen abstrakten Bildern             |
| <b>Prospektives Gedächtnis</b>   |  |
| <b>Auditiv</b>                   |  |
| 3. Auditive Zahlenfolge          | Geplanter Abruf von Ziffern in einer Adresse   |
| 7. Satzfolge                     | Geplanter sequentieller Abruf von voneinander unabhängigen Sätzen                    |
| 10. Semantische Kategorisierung  | Geplanter Abruf von über- und untergeordneten Klassifikationen von Wörtern           |
| <b>Visuell</b>                   |  |
| 4. Richtungen erinnern           | Geplanter räumlicher organisierter Abruf von verbundenen Punkten auf einem Stadtplan |
| 8. Räumliche Organisation        | Geplanter räumlicher geordneter Abruf von bedeutungslosen Formen                     |
| 11. Nonverbale Folge             | Geplanter über- und untergeordneter organisierter Abruf idiosynkratischer Sequenzen  |

Für jeden Untertest werden vier bis fünf Scores vergeben, die jeweils zu Gesamtwerten für den ganzen Test verrechnet werden können. Der „Initial Score“ stellt die Anzahl spontan richtig wiedergegebener Items dar und entspricht daher dem Wert, den der Pb ohne Hilfe des Testleiters erreicht. Er entspricht klassischen, über Statustests erfassten IQ-Werten, repräsentiert allgemeine Verarbeitungsfertigkeiten über verschiedene akademische Domänen hinweg und spiegelt die Fähigkeit des Pb wider, Informationen aufzunehmen. Swanson (1996, S. 83) interpretiert diesen Wert auch als ein Maß für Spearman's  $g$ . Die weiteren Scores stellen Maße für verschiedene

Aspekte des Lernpotentials des Pb dar. Der „Gain Score“ ist der maximale Wert, den der Pb nach den Hilfestellungen des Testleiters erzielt. Die Anzahl insgesamt erforderlicher Hinweise wird im „Probe Score“ ausgedrückt. Leistungen in den wiederholten Aufgaben zur Messung der Leistungsstabilität werden mit dem „Maintenance Score“ erfasst. Für die Untertests, die prospektive Gedächtnisleistungen erfassen, wird zusätzlich ein „Strategy Efficiency Score“ berechnet, der die vom Pb gewählten Strategien bewertet.

Es finden sich recht hohe Korrelationen zwischen verschiedenen Intelligenztests und dem Gesamtwert aus den Initial Scores, die jedoch nur an kleinen Stichproben ermittelt wurden (z. B. K-ABC:  $r = .82-84$ ,  $N = 41$ ; Raven-Test:  $r = .64$ ,  $N = 30$ ; Verbalteil des WISC-R:  $r = .62$ , Handlungsteil des WISC-R:  $r = .88$ ,  $N = 30$ ). Die anderen Scores des S-CPT weisen deutlich geringere Zusammenhänge mit Intelligenztests auf; dies wird so erklärt, dass diese Maße Aspekte kognitiver Fähigkeiten erfassen, die klassische Statusmessungen nicht abbilden können. (Zur Reliabilität des S-CPT siehe Abschnitt 4.4.) Der Test wurde in den Jahren 1987-1994 an einer repräsentativen Stichprobe von 1.611 Pbn in den USA und in Kanada normiert.

Zusammenfassend lässt sich festhalten, dass die Aufgaben des S-CPT eine valide und reliable Erfassung der Arbeitsgedächtniskapazität erlauben, zumindest dann, wenn die Initial Scores herangezogen werden. An hoch begabten Pbn wurde der Test nicht erprobt, es ist zu vermuten, dass sich der S-CPT bei der Untersuchung von hoch Begabten als zu einfach erweisen würde. Die Testdurchführung des S-CPT muss individuell erfolgen, Gruppentestungen sind nicht möglich. Die Vorgabe der Items durch einen Versuchsleiter kann zudem zu einer eingeschränkten Durchführungsobjektivität führen.

#### **4.4 Messgenauigkeit von Arbeitsgedächtnisaufgaben**

Im Rahmen der Untersuchung individueller Unterschiede ist die Messgüte der verwendeten Arbeitsgedächtnisaufgaben von Interesse. Dies trifft gleichermaßen auf klassische Aufgabentypen, die in wissenschaftlichen Studien verwendet werden, und auf die Skalen der oben beschriebenen Individualtests zu. Ob die Arbeitsgedächtniskapazität eines Pb durch Leistungen in bestimmten Aufgaben angemessen abgebildet wird, hängt vor allem von der Objektivität, Reliabilität und Validität der verwendeten Aufgaben ab. Die Objektivität der Messung ist in der Regel durch eine standardisierte Durchführung der Testung gegeben. Bei der Itemvorgabe durch einen Testleiter (im Gegensatz zu einer Darbietung am Computer) kann die Durchführungsobjektivität vermindert sein, da die Darbietungsdauer der zu merkenden Items nicht exakt konstant gehalten werden kann. Während die Konstrukt- und die Kriteriumsvalidität von Arbeitsgedächtnisaufgaben Gegenstand vieler Studien sind, wird die Reliabilität oder Messgenauigkeit der Skalen häufig nicht untersucht.

Die Studien, die Angaben zur Messgenauigkeit machen, sprechen überwiegend für eine ausreichend hohe innere Konsistenz der Skalen. Die innere Konsistenz (Cronbachs Alpha) der einzelnen Subtests des S-CPT (vgl. Abschnitt 4.3.3) etwa liegt für



die Initial Scores zwischen  $\alpha = .72$  und  $\alpha = .86$ , für den Gesamttest (über alle Scores hinweg) bei  $\alpha = .96$ . Verschiedene Versionen des Star Counting Test (vgl. Abschnitt 4.3.1) erbrachten in mehreren Substichproben Werte der inneren Konsistenz zwischen  $\alpha = .83$  und  $\alpha = .88$  ( $N = 82-727$ ).

Die meisten Angaben zur Reliabilität liegen in verschiedenen Studien insgesamt für die Aufgabe „Reading Span“ vor; Daneman und Merikle (1996) berichten über mehrere Studien von Engle und Kollegen mit der „Reading Span“ ( $k = 4$  Studien;  $n = 473$  Pbn) eine mittlere, nach Stichprobengröße gewichtete Reliabilität von  $\alpha = .80$ . Neuere Studien fanden Koeffizienten in vergleichbarer Höhe (Conway et al., 2002:  $\alpha = .70$ ; Kane et al., 2004:  $\alpha = .78$ ). Swanson (1996) berichtet über eine innere Konsistenz von  $\alpha = .92$  für die auditiv vorgegebene „Listening Span“, die mit der „Reading Span“ vergleichbar ist. Die Split-Half-Koeffizienten liegen in verschiedenen Studien zwischen  $r = .78$  und  $r = .95$  (Waters & Caplan, 1996; Park, Lautenschlager & Hedden, 2002; Tirre & Pena, 1992). Ebenfalls ausreichend hohe Werte erreicht die Aufgabe „Operation Span“ (OSPAN, Turner & Engle, 1989). Klein und Fiss (1999) berichten hier über Werte der inneren Konsistenz zwischen .77 und .83.

Nur in wenigen Fällen wurde eine Retest-Reliabilität von Arbeitsgedächtnisaufgaben bestimmt, die berichteten Werte fallen hier häufig gering aus. Bei einer Retestung mit der „Reading Span“ nach einem Intervall von ein bis sechs Monaten ermittelten Waters und Caplan (1996) eine Reliabilität von lediglich  $r_{tt} = .41$ . Die Retest-Reliabilität der OSPAN betrug nach drei Wochen  $r_{tt} = .73$  und nach neun bis zehn Wochen  $r_{tt} = .67$ , die messfehlerbereinigte Retest-Reliabilität (Verfahren nach Heise, 1969) betrug  $r_{tt} = .88$ . Hacker et al. (1999) fanden eine Retest-Reliabilität von nur  $r_{tt} = .55$  für die Aufgabe „Rechenspanne“ (Drei-Wochen-Intervall), führen den geringen Koeffizienten jedoch auf ihre relativ homogene Studierendenstichprobe zurück.

Für die Aufgaben der oben beschriebenen Individualtests werden sehr unterschiedlich hohe Retest-Reliabilitäten berichtet. Während die Retest-Reliabilität des Star Counting Tests (siehe Abschnitt 4.3.1) mit  $r_{tt} = .77$  ( $N = 82$ , Drei-Wochen-Intervall) als ausreichend zu betrachten ist, berichten Pickering und Gathercole (2001) für die meisten Subtests der „Working Memory Test Battery for Children“ (WMTB-C, siehe Abschnitt 4.3.2) über geringe Retest-Reliabilitäten. Keiner der Untertests zur Messung der zentralen Exekutiv erreichte hier ausreichend hohe Werte.

Für 50 Kinder der ersten und zweiten Klasse fanden sie nach einem Intervall von zwei Wochen Werte zwischen  $r_{tt} = .45$  und  $r_{tt} = .81$ , für 49 Kinder der fünften und sechsten Klassen lagen die Werte zwischen  $r_{tt} = .38$  und  $r_{tt} = .82$ . Die Retest-Reliabilität der „Listening Recall“ betrug bei den jüngeren Pbn zwar  $r_{tt} = .83$ , bei den älteren Pbn hingegen nur  $r_{tt} = .38$ . Die Aufgabe „Backward Digit Recall“ (entspricht der Aufgabe „Backward Digit Span“) erbrachte in beiden Altersgruppen ebenfalls keine sehr stabile Messungen (jüngere Pbn:  $r_{tt} = .58$ ; ältere Pbn:  $r_{tt} = .71$ ). Für die Aufgabe „Backward Digit Span“ liegen zusätzlich Angaben zur Retest-Reliabilität für Kinder und Jugendliche aus dem AID 2 (Kubinger & Wurst, 2000) vor. Nach

einem Intervall von vier Wochen betrug die Retest-Reliabilität hier  $r_{tt} = .67$  bzw.  $r_{tt} = .74$ , nach einem Jahr jedoch nur noch  $r_{tt} = .53$ .

In einer Vorläuferstudie zur Entwicklung der WMTB-C untersuchten Gathercole und Pickering (2000a) an einer Stichprobe von 27 Kindern die Retest-Reliabilität verschiedener Aufgabentypen nach einem Intervall von drei Wochen. Es fand sich eine Reliabilität von  $r_{tt} = .62$  für die Aufgabe „Listening Recall“; die Aufgabe „Counting Recall“ erreichte jedoch nur einen Wert von  $r_{tt} = .15$ . Für den S-CPT (siehe Abschnitt 4.3.3) wurde keine Retest-Reliabilität bestimmt.

Insgesamt sind die Befunde zur Retest-Reliabilität von Arbeitsgedächtnisaufgaben nicht als zufrieden stellend zu betrachten. Offenbar schwanken die Leistungen in den genannten Arbeitsgedächtnisaufgaben über verschiedene Messzeitpunkte relativ stark. Die Retest-Reliabilität wurde bisher jedoch zu selten erhoben, um Arbeitsgedächtnisaufgaben insgesamt eine geringe Reliabilität zu bescheinigen. Zu vermuten ist, dass die computergestützte Aufgabenvorgabe, die eine größtmögliche Standardisierung der Durchführung ermöglicht, zu einer besseren Übereinstimmung der Messungen zu verschiedenen Zeitpunkten führt. Sollte sich der Befund der geringen Retest-Reliabilität jedoch in späteren Untersuchungen erhärten, so ist davon auszugehen, dass die Leistung des Arbeitsgedächtnisses tatsächlich nicht sehr stabil über die Zeit ist.

Arbeitsgedächtnisaufgaben wurden bisher nicht auf ihre Passung auf Modelle der Item Response Theorie (vgl. Abschnitt 2.3.1) geprüft. Einzige Ausnahme bildet hier der „Star Counting Test“ (SCT, de Jong & Das-Smaal, 1990, vgl. Abschnitt 4.3.1) zur Messung der zentralen Exekutive des Arbeitsgedächtnisses bei Kindern. Für die 44 Aufgaben des SCT konnte gezeigt werden, dass sie dem Rasch-Modell entsprechen.

#### 4.5 Zusammenfassung

Es existieren bereits verschiedene Aufgabentypen und ganze Testverfahren zur Erfassung der Arbeitsgedächtniskapazität, deren prinzipielle psychometrische Eignung mehrfach untersucht wurde. Lediglich die Retest-Reliabilität ist bislang fraglich. Für einige Aufgabentypen wird jedoch vermutet, dass sie die intendierten Fähigkeiten bei bestimmten, sehr fähigen Probandengruppen nicht messen. So wurde diskutiert, ob die „Backward Digit Span“ auch bei sehr fähigen Pbn noch den Charakter einer Arbeitsgedächtnisaufgabe hat, oder ob sie nur noch die reine Speicherleistung erfasst.

Ein bisher häufig vernachlässigtes Thema ist weiterhin die Frage des Scorings von Arbeitsgedächtnisaufgaben. Die verschiedenen Scoringmethoden, über die in der Literatur berichtet wird, wurden dargestellt und diskutiert. Als relativ beste Methode wurde ein Scoring auf Aufgabenebene mit Einheitsgewichtung abgeleitet, das die Leistungen sowohl in der Verarbeitungs- als auch in der Speicherkomponente berücksichtigt und die richtige Reihenfolge der Antworten erfordert. Die drei bisher existierenden standardisierten Arbeitsgedächtnistests unterscheiden sich deutlich in theoretischem Konzept, in der Breite der Messung und in der Aufgabenauswahl.

## 5. Aufgabenauswahl und Fragestellungen

Der Aufbau dieses Kapitels gliedert sich wie folgt: Zunächst wird in Abschnitt 5.1 dargelegt, inwiefern Arbeitsgedächtnisaufgaben für einen Test zur Messung hoher intellektueller Begabung geeignet scheinen. Auf der Basis des in den Kapiteln 3 und 4 referierten Forschungsstands zum Arbeitsgedächtnis wird begründet, warum Arbeitsgedächtnisaufgaben in der vorliegenden Arbeit als Basis eines neuen Testkonzepts verwendet werden. Anschließend wird die Auswahl der hier verwendeten *Typen* von Arbeitsgedächtnisaufgaben beschrieben (Abschnitt 5.2). Im darauf folgenden Abschnitt 5.3 werden die Fragestellungen dargelegt, die sich aus dem Forschungsstand zum Arbeitsgedächtnis ergeben und die in der vorliegenden Arbeit untersucht werden. Zunächst werden Fragestellungen zum Konstrukt und zur Faktorstruktur des Arbeitsgedächtnisses formuliert. Im letzten Teil dieses Kapitels geht es um Fragestellungen, die im Hinblick auf die Entwicklung eines Tests für die Begabungsdiagnostik, der auf der Erfassung der Arbeitsgedächtniskapazität basiert, untersucht werden sollen.

### 5.1 Verwendung von Arbeitsgedächtnisaufgaben für die Diagnostik hoher intellektueller Fähigkeiten

Die Frage nach der Eignung von Arbeitsgedächtnisaufgaben für die Messung hoher intellektueller Fähigkeiten wurde bereits in Abschnitt 2.6 kurz thematisiert. Nachdem in den Kapiteln 3 und 4 die theoretischen Grundlagen von Arbeitsgedächtnisaufgaben und zahlreiche empirische Befunde aus differentiellen Studien dargelegt wurden, soll nun noch einmal umfassend begründet werden, warum Arbeitsgedächtnisaufgaben für die hier angestrebte Grundlegung einer Testentwicklung verwendet werden.

#### *Theoretische Fundierung der Aufgaben*

Die Darstellung und Diskussion der Befunde aus der differentiellen Arbeitsgedächtnisforschung hat bereits deutlich gemacht, dass es sich beim Arbeitsgedächtnis um ein theoretisch und empirisch fundiertes Konstrukt handelt, das sich als Grundlage für einen neu zu entwickelnden Begabungstest sehr gut eignen könnte. Arbeitsgedächtnisaufgaben stehen nicht in der direkten Tradition klassischer Intelligenztestaufgaben, auch wenn einzelne Aufgabentypen bereits in Intelligenztests Verwendung fanden (z. B. die „Backward Digit Span“ in den Wechsler-Tests, vgl. Abschnitt 4.1.2). Stattdessen stammen Arbeitsgedächtnisaufgaben aus der experimentellen und differentiellen Kognitionspsychologie und wurden auf der Basis expliziter Modellvorstellungen und zahlreicher experimenteller Befunde entwickelt und erprobt. Die Aufgaben erfassen – anders als klassische Intelligenztestaufgaben – keine höheren mentalen Fähigkeiten, sondern eher grundlegende kognitive Prozesse. Dennoch sind sie in hohem Maße in der Lage, verschiedene intellektuelle Leistungen vorherzusa-

gen. Noch nicht hinreichend geklärt ist bisher jedoch die Binnenstruktur des Konstrukts Arbeitsgedächtnis: Handelt es sich um einen globalen Faktor, oder besteht das Arbeitsgedächtnis aus verschiedenen, domänenspezifischen Subsystemen? Mit dieser Frage verknüpft ist die Frage, ob verschiedene Arbeitsgedächtnisaufgaben möglicherweise unterschiedliche Fähigkeiten erfassen.

#### *Erfassung fluider Fähigkeiten mit verbalem, numerischem und figuralem Aufgabematerial*

Kyllonen (1996) konnte die zentrale Rolle des Arbeitsgedächtnisfaktors für unterschiedlichste kognitive Aktivitäten aufzeigen und kommt zu der Schlussfolgerung „dass die Arbeitsgedächtniskapazität tatsächlich im Wesentlichen Spearman's  $g$  sein könnte“ (S. 73; Übersetzung durch die Verfasserin). Auch wenn sich nicht alle Forscher dieser Schlussfolgerung anschließen, so besteht doch weitgehender Konsens über die ausgeprägte Verwandtschaft von Arbeitsgedächtnis und Intelligenz. Verschiedene Studien konnten hohe bis sehr hohe Zusammenhänge der Konstrukte „Arbeitsgedächtnis“ und „Verarbeitungskapazität“ bzw. „Reasoning“ zeigen. Als Unterschiede zwischen beiden Konstrukten zeigte sich, dass Reasoning tendenziell stärker mit Wissen, das Arbeitsgedächtnis hingegen stärker mit der Geschwindigkeit der Informationsverarbeitung verknüpft ist (Kyllonen & Christal, 1990). Arbeitsgedächtnisaufgaben erfassen weitgehend fluide Prozesse, während in vielen klassischen Intelligenztestaufgaben Wissensaspekte und kristalline Fähigkeiten eine Rolle spielen. Ein Vorteil gegenüber stärker wissensabhängigen Testaufgaben besteht darin, dass mit Arbeitsgedächtnisaufgaben eher das bildungsunabhängige intellektuelle Potential erfasst wird. Typische Arbeitsgedächtnisaufgaben erfordern kein (Schul-)wissen, das über rudimentäre Lese- und Rechenfähigkeiten hinausgeht.

Es existieren bereits vielfältige Arbeitsgedächtnisaufgaben mit verbalem, numerischem und figural-räumlichem Inhalt. Bisher ist jedoch umstritten, ob sich im Arbeitsgedächtnis alle drei Inhaltsaspekte differenzieren lassen. Die derzeitige Befundlage stützt am ehesten die Annahme zweier separater inhaltlicher Faktoren (verbal-numerisch und räumlich), wengleich auch Befunde sowohl für ein domänengenerelles als auch für ein dreifaktorielles Arbeitsgedächtnis vorliegen.

#### *Psychometrische Bewährung*

Wie in Abschnitt 4.3 dargestellt, gibt es bereits einige standardisierte Tests für Kinder und Erwachsene, die Arbeitsgedächtnisaufgaben für die Individualdiagnostik verwenden. In verschiedenen Studien werden ebenfalls Angaben zur psychometrischen Qualität der eingesetzten Aufgabentypen gemacht. Insgesamt konnte hier für verschiedene Arbeitsgedächtnisaufgaben überwiegend eine gute Messgüte gezeigt werden, insbesondere für die häufig untersuchte „Reading Span“. Die Messgüte wird jedoch durch eine möglicherweise relativ geringe Retest-Reliabilität eingeschränkt (vgl. Abschnitt 4.4). Eine weitere mögliche Einschränkung besteht hinsichtlich der Validität einzelner Aufgabentypen beim Einsatz bei Pbn mit deutlich überdurch-

schnittlichen Fähigkeiten (vgl. Ausführungen zum Aufgabentyp „Backward Digit Span“ in Abschnitt 4.1.2).

#### *Einfache und systematische Modifizierbarkeit der Schwierigkeit*

Arbeitsgedächtnisaufgaben sind in der Regel sehr einfach strukturiert, ihre Schwierigkeit bestimmt sich über die Kombination aus einer relativ simplen Verarbeitungsanforderung und einer Speicheraufgabe, die das Behalten unterschiedlich vieler Informationen erfordert. Tendenziell gilt daher: Je mehr Informationen gespeichert werden müssen, desto schwieriger wird die Aufgabe. Arbeitsgedächtnisaufgaben können daher sehr einfach in ihrer Schwierigkeit manipuliert werden, oftmals auch ohne dass empirische Lösungshäufigkeiten bekannt sind (Kyllonen, 1996). Theoretisch gibt es bei Arbeitsgedächtnisaufgaben keine Obergrenze für die Schwierigkeit, die Konstruktion sehr schwieriger Aufgaben ist ohne weiteres möglich, in der Regel indem die Speicheranforderung stark erhöht wird.

#### *Eignung für Kinder zwischen etwa acht und zwölf Jahren*

Verschiedene Studien konnten bereits zeigen, dass Kinder die doppelte Anforderung des Speicherns und gleichzeitigen Verarbeitens von Informationen, wie sie in den Arbeitsgedächtnisaufgaben gefordert wird, bereits im Grundschulalter gut beherrschen (z. B. Leather & Henry, 1994; Gathercole & Pickering, 2000a). Auch die standardisierten Tests, die in Abschnitt 4.3 vorgestellt wurden, belegen die Eignung von Arbeitsgedächtnisaufgaben für die Diagnostik bei Kindern.

#### *Computergestützte Durchführung und Gruppentestung*

Klassischerweise werden Arbeitsgedächtnisaufgaben – ebenso wie viele Intelligenztestaufgaben – in Form eines „standardisierten Dialogs“ (Tewes, 1991) zwischen Testleiter und Pb vorgegeben. In einigen Studien wurden Arbeitsgedächtnisaufgaben bei Erwachsenen jedoch bereits erfolgreich computergestützt eingesetzt (z. B. Oberauer et al., 2000). Die Durchführungsobjektivität ist bei einer computergestützten Darbietung der Testaufgaben deutlich erhöht. Viele Arbeitsgedächtnisaufgaben eignen sich zudem für eine Gruppentestung, wenn auf auditive Vorgaben und mündliche Antworten verzichtet wird. De Jong und Das-Smaal (1990) weisen darauf hin, dass eine Gruppentestung nicht nur ökonomischer ist, sondern unter Umständen auch ökologisch valider sein kann. Dies ist dann der Fall, wenn die schulischen Leistungen eines Kindes prognostiziert werden sollen, die in der Regel im Klassensetting erbracht werden müssen. Eine Testung in der Gruppe entspricht dann eher den Bedingungen, in denen Schülerinnen und Schüler ihre intellektuellen Leistungen zeigen müssen.

Die Verwendung von Arbeitsgedächtnisaufgaben scheint also für eine Individualdiagnostik vielversprechend. Ob sie sich tatsächlich bei der Messung sehr hoher intellektueller Fähigkeiten bewähren, ist Gegenstand der vorliegenden Untersuchung.

## 5.2 Auswahl von Typen von Arbeitsgedächtnisaufgaben

Da in der Literatur mindestens 50 verschiedene Aufgabentypen zur Erfassung der Arbeitsgedächtniskapazität beschrieben sind, ist es nicht erforderlich, für die vorliegende Arbeit ganz neuartige Aufgaben zu konzipieren. Stattdessen kann auf Erfahrungen und Befunde aus anderen Studien zurückgegriffen werden. Die systematische und theoretisch geleitete Aufgabenauswahl von Oberauer et al. (2000) lässt es sinnvoll erscheinen, die Ergebnisse dieser Studie als Grundlage für eine Aufgabenauswahl zu verwenden.

Oberauer und Kollegen klassifizierten insgesamt 23 Aufgabentypen nach ihren Anforderungen und konnten diese a priori vorgenommene Zuordnung überwiegend auch empirisch belegen (vgl. Abschnitt 3.3). Verschiedene Aufgabentypen bewährten sich in dieser Studie psychometrisch besonders gut, so dass Oberauer et al. (2003) diese später als „Marker-Tests“ einsetzten, anhand derer sie die Güte neu entwickelter Aufgaben bewerteten. Ein weiterer Vorteil dieser Testbatterie besteht zudem darin, dass alle Aufgaben bereits computergestützt und in Gruppentestungen eingesetzt wurden, was auch dem hier angestrebten Vorgehen entspricht. Jedoch ist es nicht möglich, in Studien mit Kindern eine ähnlich umfangreiche Testbatterie einzusetzen, weil dies die Belastungsgrenze der Pbn deutlich überschreiten würde. Da die Befundlage für die Arbeitsgedächtnisfunktion „Simultane Speicherung und Transformation“ am besten gesichert ist und diese Funktion einen gemeinsamen Nenner für die meisten Forscher darstellt, werden für die vorliegende Untersuchung ausschließlich solche Aufgaben verwendet, die bisherigen Studien zufolge diesen Bereich erfassen.

Bei der Auswahl geeigneter Aufgabentypen aus der Batterie von Oberauer et al. (2000) wurde weiterhin darauf geachtet, dass die Aufgaben bereits ausreichend hohe Koeffizienten für die interne Konsistenz ( $\alpha > .80$ ) und relativ hohe Zusammenhänge mit der allgemeinen Intelligenz (BIS-g;  $r > .40$ ) sowie mit der BIS-Skala „Verarbeitungskapazität“ ( $r > .40$ ) gezeigt hatten. Die Aufgabentypen sollten zudem nur solche Routinen erfordern, die auch Kinder schon sicher beherrschen; nicht vorausgesetzt werden können etwa die „schlafwandlerische“ Kenntnis des Alphabets oder das Kopfrechnen mit mehrstelligen Zahlen. Es ist möglich, dass bestimmte Verarbeitungsanforderungen, die für Erwachsene trivial sind, für Kinder eine hohe kognitive Anforderung darstellen. Besonders beachtet wurden solche Aufgabentypen, die eigens für Kinder entwickelt und bereits an diesen erprobt wurden. Die Aufgabe „Reading Span“ beispielsweise wurde in Form einer „Listening Span“ bereits mehrfach mit Kindern durchgeführt (z. B. Siegel, 1994). Auch die Aufgabe „Backward Digit Span“ hat sich bereits für den Einsatz bei Kindern bewährt (z. B. Gathercole et al., 2004). Inhaltlich wurden alle einzelnen Aufgaben für die vorliegende Arbeit neu gestaltet, lediglich die Aufgabenprinzipien wurden aus der Literatur übernommen. Die Darbietung aller Arbeitsgedächtnisaufgaben erfolgt computergestützt.

Für die vorliegende Arbeit wurden Arbeitsgedächtnisaufgaben mit jeweils einem eindeutig verbalen, numerischen und figuralen Inhalt ausgewählt, um die Frage der Domänengeneralität vs. -spezifität prüfen zu können (vgl. Abschnitt 5.3.1). Jeder

Inhaltsbereich wird über zwei verschiedene Arbeitsgedächtnisaufgaben erfasst. Insgesamt werden schließlich sechs Aufgabentypen verwendet: „Reading Span“ (verbal), „Verbal Span“ (verbal), „Computation Span“ (numerisch), „Backward Digit Span“ (numerisch), „Memory Updating Figural“ (figural) und „Spatial Working Memory“ (figural). Die genauen Beschreibungen der hier verwendeten Aufgabenversionen folgen in Abschnitt 6.1. Bei der Konstruktion der einzelnen Aufgaben wurde zudem stets das Ziel verfolgt, Deckeneffekte und Motivationsverluste zu vermeiden, die durch zu wenig schwierige und zu viele einfache Aufgaben verursacht werden (vgl. Abschnitt 2.4.2).

### 5.3 Fragestellungen

Ein wesentliches Ziel der vorliegenden Arbeit besteht in der Entwicklung der Grundlagen für einen Begabungstest auf der Basis von Arbeitsgedächtnisaufgaben. Die Arbeitsgedächtniskapazität wurde bereits ausführlich untersucht, wesentliche Befunde wurden in den Kapiteln 3 und 4 dargestellt und diskutiert. Jedoch sind bislang einige Aspekte des Konstrukts „Arbeitsgedächtnis“ nicht hinreichend geklärt. So deuten zwar inzwischen viele Befunde darauf hin, dass sich das Arbeitsgedächtnis in zwei Komponenten – eine verbal-numerische und eine räumliche – untergliedern lässt, jedoch gibt es auch Befunde, die entweder für einen generellen Arbeitsgedächtnisfaktor oder für ein Drei-Faktor-Modell (mit separaten verbalen und numerischen Faktoren) sprechen. Bisher ist auch unklar, ob dieselbe Faktorstruktur für Kinder unterschiedlichen Alters und unterschiedlicher intellektueller Begabung gilt.

Der Stellenwert des Arbeitsgedächtnisses im Netzwerk zentraler kognitiver Fähigkeiten ist ebenfalls noch nicht abschließend geklärt. Die Höhe des Zusammenhangs mit fluider Intelligenz variiert in bisherigen Untersuchungen relativ stark. Es kann davon ausgegangen werden, dass mittlere positive Zusammenhänge zwischen der Arbeitsgedächtniskapazität auf der einen Seite und Speed, Kurzzeitgedächtnis und Schulleistung auf der anderen Seite bestehen. Für Kinder unterschiedlicher Altersgruppen sowie für Kinder mit deutlich überdurchschnittlich hohen intellektuellen Fähigkeiten fehlen hier bislang jedoch detaillierte Befunde. Eine weitere Exploration des Konstrukts „Arbeitsgedächtnis“ für die hier untersuchte Zielgruppe der durchschnittlich und hoch begabten Kinder ist somit eine wichtige Voraussetzung für die Erprobung eines Testkonzepts auf der Basis von Arbeitsgedächtnisaufgaben.

Daraus ergeben sich zwei Komplexe von Fragestellungen für die vorliegende Arbeit. Der erste Komplex bezieht sich auf das Konstrukt des Arbeitsgedächtnisses, der zweite auf die Eignung von Arbeitsgedächtnisaufgaben für die Messung hoher intellektueller Begabungen bei Kindern und die Entwicklung der Grundlagen für ein Testverfahren. Technisch erfordern diese beiden Fragenkomplexe Datenanalysen auf zwei unterschiedlichen Ebenen:

Bei der Untersuchung des Konstrukts „Arbeitsgedächtnis“ sind die Analyseeinheiten die Aufgabenscores der sechs Gruppen von Arbeitsgedächtnisaufgaben „Reading Span“ (RS), „Verbal Span“ (VS), „Computation Span“ (CS), „Backward Digit Span“

(BDS), „Memory Updating Figural“ (MUF) und „Spatial Working Memory“ (SWM). Für jede Aufgabengruppe wird die Summe über alle richtig gelösten Aufgaben gebildet (Scoring auf Aufgabenebene), die Verwendung dieses Scorings wurde in Abschnitt 4.2 begründet. Es werden zu Vergleichszwecken zusätzlich die deskriptiven Statistiken auf der Basis einer alternativen, häufig verwendeten Scoringmethode (Scoring auf Itemebene) berechnet. Über empirische Vergleiche verschiedener Scoringmethoden wird in vereinzelt Studien berichtet, diese fanden überwiegend keine bedeutsamen Unterschiede zwischen den Verfahren.

Bei der Untersuchung der Frage, ob sich aus Arbeitsgedächtnisaufgaben ein Begabungstest zusammenstellen lässt, erfolgt eine Analyse auf Basis der einzelnen *Testaufgaben*, die zu Testskalen zusammengefasst werden. Diese werden nach den Maßgaben der Klassischen Testtheorie und der Item Response Theorie hinsichtlich ihrer Messqualität untersucht. Die einzelnen Fragestellungen für die vorliegende Arbeit zu beiden Fragenkomplexen werden im Folgenden beschrieben.

### 5.3.1 Fragestellungen zum Konstrukt „Arbeitsgedächtnis“

Eine wesentliche Frage, die in der aktuellen Literatur zum Arbeitsgedächtnis bei Kindern und Erwachsenen diskutiert wird, betrifft die Domänengeneralität oder -spezifität des Arbeitsgedächtnisses (vgl. Kapitel 3). Die bisherige Befundlage hierzu ist uneindeutig: Während die Ergebnisse verschiedener neuerer Studien auf eine Domänengeneralität hinweisen (z. B. Ackerman et al., 2002), liegen auch vielfältige Befunde vor, die eine Domänenspezifität stützen (z. B. Shah & Miyake, 1996). Die Frage der Domänengeneralität kann, wie in Kapitel 3 ausgeführt, über verschiedene Wege untersucht werden. Eine Möglichkeit besteht darin, die Korrelationen verschiedener Arbeitsgedächtnisaufgaben miteinander zu betrachten. Eine Grundvoraussetzung für das Vorliegen von Domänengeneralität wäre, dass Arbeitsgedächtnisaufgaben verschiedener Domänen positiv und substantiell miteinander korrelieren. In der Regel werden positive Zusammenhänge zwischen Arbeitsgedächtnisaufgaben verschiedener Domänen gefunden, es gibt jedoch auch entgegengesetzte Befunde (z. B. Hacker et al., 1999). In vielen Studien wurden bisher nicht Aufgaben aller drei Domänen – verbal, numerisch, figural – verwendet, oder es wurden zusätzlich zum Inhaltsbereich über die Aufgaben hinweg auch operative Aspekte der Aufgaben variiert (z. B. einfache vs. komplexe Spannaufgaben), so dass häufig nur eingeschränkte Aussagen möglich sind. Das Aufzeigen positiver Korrelationen reicht jedoch nicht aus, um über Domänengeneralität vs. -spezifität zu entscheiden. Ein Problem besteht darin, dass je nach dem, ob positive Korrelationen als Belege für die Konstrukt- oder Kriteriumsvalidität verstanden werden, sie im Sinne der Generalitäts- oder der Spezifitätshypothese interpretiert werden können (Oberauer, 2005; vgl. Abschnitt 3.3.2).

Ein weiterer Untersuchungsansatz besteht darin, die Zusammenhänge der Arbeitsgedächtnisaufgaben zu anderen kognitiven Tests, die entweder dieselbe oder eine andere Domäne erfassen, zu analysieren (vgl. z. B. Shah & Miyake, 1996). Sollte das Arbeitsgedächtnis tatsächlich in hohem Maße domänenspezifisch sein, so wird



erwartet, dass Arbeitsgedächtnis- und andere Aufgaben deutlich höher miteinander korrelieren, wenn sie dieselbe Domäne erfassen als wenn sie unterschiedliche Domänen erfassen.

Da bei beiden genannten Ansätzen jedoch stets lediglich einzelne Maße miteinander korreliert werden, die jeweils auch testspezifische Varianz aufweisen, sind entsprechende Ergebnisse nur eingeschränkt aussagefähig. Besser gelingt eine Untersuchung auf der Ebene kognitiver Konstrukte, wenn jede Domäne über mehrere, mindestens jedoch zwei einzelne Maße erfasst wird. Daher soll auch in der vorliegenden Arbeit die Frage nach der Domänengeneralität zusätzlich über exploratorische und konfirmatorische Faktorenanalysen über verschiedene Arbeitsgedächtnisaufgaben aus unterschiedlichen Domänen geprüft werden.

Insgesamt ist die Befundlage in den bisherigen Studien, die unterschiedliche Methoden verwendeten, sehr uneinheitlich. Mögliche Ursachen für die stark abweichenden Befunde werden – neben Unterschieden in den methodischen Ansätzen – vor allem in der Aufgabenauswahl und in den jeweils untersuchten Stichproben gesehen (vgl. Abschnitt 3.3.4). Eine eingeschränkte Fähigkeitsvarianz führt generell eher zu einer Bestätigung der Spezifitätshypothese (Shah & Miyake, 1996). Die ausschließliche Untersuchung von Pbn mit hohen Fähigkeiten führt ebenfalls eher zu geringen Korrelationen der Aufgaben untereinander und somit eher zu einer Betonung spezifischer Aspekte des Arbeitsgedächtnisses (Mackintosh & Bennett, 2003).

In der vorliegenden Arbeit wird daher eine Stichprobe mit relativ großer Fähigkeitsvarianz untersucht, es werden Kinder aller Fähigkeitsbereiche – von unterdurchschnittlich bis weit überdurchschnittlich – in die Studie einbezogen. Die Auswahl der Arbeitsgedächtnisaufgaben orientiert sich an einem relativ eindeutigen Kriterium: Es werden nur solche Aufgaben verwendet, die eine simultane Speicherung und Transformation von Informationen erfordern. Es handelt sich somit ausschließlich um komplexe Spannaufgaben; einfache Spannaufgaben, die lediglich eine passive Informationsspeicherung erfordern, werden nicht eingesetzt.

Die Zuordnung der verwendeten Aufgaben zur Operation „Simultane Speicherung und Verarbeitung“ konnte bereits – in Abgrenzung zu Arbeitsgedächtnisaufgaben, die andere Operationen erfordern – empirisch gezeigt werden (Oberauer et al., 2000). Zudem wurde die Messgüte der verwendeten Aufgabentypen bereits in anderen Studien mit erwachsenen Pbn gezeigt (Oberauer et al., 2000; Süß et al., 2002). Verschiedene Vortests, in denen die einzelnen Aufgaben bereits geprüft und sukzessive für den Einsatz bei Kindern optimiert wurden, lassen auch für die hier eingesetzten Aufgabengruppen eine ausreichende Messqualität erwarten. Schließlich beinhaltet die Zusammenstellung der verwendeten Aufgaben eine Ausgewogenheit in den Inhalten, es werden je zwei Aufgaben mit verbalen, numerischen und figuralen Inhalten eingesetzt. Insgesamt werden somit eine möglichst repräsentative Stichprobe von Pbn und eine klar definierte, ausgewogene Stichprobe von Aufgabentypen verwendet.

Die meisten der bisherigen Studien, die latente Faktoren verwendeten, wurden mit erwachsenen Pbn durchgeführt. Es ist jedoch möglich, dass sich die Binnenstruktur

des Arbeitsgedächtnisses im Laufe der Kindheit und Jugend verändert. Die Faktorstruktur könnte sich mit zunehmendem Lebensalter entweder von einer eher generellen Fähigkeit weiter ausdifferenzieren oder ausgehend von einer mehrdimensionalen Struktur zunehmend dedifferenzieren. Zu dieser Fragestellung liegen in der Arbeitsgedächtnisforschung bisher kaum Befunde vor. Für den relativ kurzen, aber für die kognitive Entwicklung wichtigen Lebensabschnitt zwischen 8 und 13 Jahren, der in der vorliegenden Arbeit untersucht wird, soll die Invarianz der Faktorstruktur daher geprüft werden. Weiterhin interessieren das Vorhandensein und das Ausmaß von mittleren Leistungsunterschieden: Aus anderen Studien ist bereits bekannt, dass die Leistungsfähigkeit des Arbeitsgedächtnisses mit steigendem Alter zunimmt (z. B. Swanson, 1999). Dies soll auch für die vorliegende Aufgabenauswahl überprüft werden.

Bisher gar nicht untersucht wurde die Binnenstruktur des Arbeitsgedächtnisses bei hoch begabten Kindern. Auch hier wäre es möglich, dass sich die Faktorstrukturen für durchschnittlich und hoch begabte Kinder unterscheiden. In der Intelligenzstrukturforschung ist bisher nicht eindeutig geklärt, ob sich die Intelligenzstruktur hoch Begabter von der durchschnittlich Begabter unterscheidet, neuere Befunde sprechen – zumindest für Kinder ab zwölf Jahren – eher dagegen (Jäger et al., 2005; siehe Kapitel 2). Die Invarianz der Faktorstruktur für durchschnittlich und hoch Begabte soll daher in dieser Arbeit untersucht werden. Auch werden die mittleren Leistungsunterschiede zwischen durchschnittlich und hoch begabten Kindern betrachtet. Trifft es zu, dass hoch begabte Kinder in allen Aufgabengruppen im Mittel höhere Leistungen zeigen als durchschnittlich begabte Kinder? Wenn das der Fall ist, wäre dies ein weiterer Beleg für die zentrale Rolle der Arbeitsgedächtniskapazität im Bereich der kognitiven Fähigkeiten.

Auch die Zusammenhänge zwischen den Arbeitsgedächtnisaufgaben und anderen kognitiven Fähigkeiten und damit die Rolle, die das Arbeitsgedächtnis im nomologischen Netzwerk kognitiver Fähigkeiten spielt, ist bisher nicht eindeutig geklärt. Leistungen in Arbeitsgedächtnisaufgaben korrelieren in aller Regel positiv mit den Leistungen in Tests, die fluide Intelligenz erfassen (z. B. Kyllonen & Christal, 1990; Engle et al., 1999). Auch mit kristallinen Aspekten der Intelligenz finden sich positive Zusammenhänge (z. B. Swanson, 1996). Ebenso wurden bereits verschiedentlich bedeutsame Zusammenhänge zwischen dem Arbeitsgedächtnis und dem Kurzzeitgedächtnis sowie Speed berichtet (z. B. Kane et al., 2004; Ackerman et al., 2002). Das Ausmaß der Zusammenhänge variiert in bisherigen Studien jedoch deutlich (siehe Kapitel 3). Während einige Autoren über nahezu perfekte Zusammenhänge zwischen dem Arbeitsgedächtnis und einzelnen Intelligenzkomponenten berichten (z. B. Kyllonen, 1996), fallen die Zusammenhänge in anderen Studien geringer aus (z. B. Ackerman et al., 2002; Oberauer et al., 2000).

Hier scheinen ebenfalls Fragen der Aufgabenauswahl und Eigenschaften der jeweils untersuchten Stichproben zentral. Kyllonen (1996; Kyllonen & Christal, 1990) etwa verwendete Arbeitsgedächtnis- und Reasoningaufgaben, die sich inhaltlich stark ähnelten, so dass sehr hohe Zusammenhänge eher zu erwarten waren. In der Studie von de Jong und Das-Smaal (1995), die hohe Zusammenhänge zwischen Arbeitsgedäch-

nis, Intelligenz und Schulleistung bei neunjährigen Kindern zeigen konnte, waren die vier eingesetzten Arbeitsgedächtnisaufgaben inhaltlich recht heterogen. Neben Aufgaben, die eine simultane Speicherung und Verarbeitung erfordern, setzten sie mit der „Forward Digit Span“ auch eine einfache Spannenaufgabe ein. Viele bisherige Studien wurden mit Studierendenstichproben durchgeführt (z. B. Ackerman et al., 2002; Oberauer et al., 2000). Es handelte sich somit um Probandengruppen, deren Fähigkeitsvarianz eingeschränkt ist, so dass die Zusammenhänge zwischen dem Arbeitsgedächtnis und anderen Konstrukten hier eher unterschätzt werden.

Wie oben bereits dargelegt, wird in der vorliegenden Arbeit eine Stichprobe mit großer Fähigkeitsvarianz untersucht und die verwendeten Arbeitsgedächtnisaufgaben entsprechen einer klaren Definition. Zur Messung der anderen kognitiven Konstrukte werden Maße eingesetzt, die häufig in der Intelligenzdiagnostik Verwendung finden und die sich inhaltlich und strukturell deutlich von den Arbeitsgedächtnisaufgaben unterscheiden. Zusammenhänge zwischen Arbeitsgedächtnis- und Intelligenzfaktoren sollten daher in der vorliegenden Untersuchung nicht auf zu große Ähnlichkeiten zwischen den Aufgabentypen, die unterschiedliche Konstrukte messen sollen, zurückzuführen sein.

Die einzelnen Fragestellungen und Hypothesen, die in der vorliegenden Studie untersucht werden, werden im Folgenden dargelegt.

### *1. Faktorstruktur der Aufgaben*

Wie in den Abschnitten 3.3 und 3.4 ausgeführt wurde, sind sowohl die Binnenstruktur als auch das Ausmaß der Zusammenhänge des Arbeitsgedächtnisses mit anderen kognitiven Konstrukten – insbesondere für Kinder – umstritten. Zur weiteren Klärung dieser Aspekte des Arbeitsgedächtnisses soll zunächst die Binnenstruktur des Arbeitsgedächtnisses geprüft werden.

### *Korrelationen*

Es wird zunächst geprüft, ob die Scores aller hier verwendeten sechs Arbeitsgedächtnisaufgaben bei Kindern positiv untereinander korrelieren. Das Korrelationsmuster kann als ein erster (schwacher) Hinweis auf Domänengeneralität vs. -spezifität interpretiert werden. Positive Korrelationen auch über verschiedene Domänen hinweg wären ein erster Beleg für eine mögliche Domänengeneralität. Höhere Zusammenhänge der Aufgaben mit Aufgaben derselben Domäne als mit Aufgaben aus einer anderen Domäne deuten dagegen auf domänenspezifische Aspekte hin.

### *Faktorenanalysen*

Die Faktorenstruktur des Arbeitsgedächtnisses wird über exploratorische und konfirmatorische Faktorenanalysen untersucht. Drei mögliche Modelle erscheinen aufgrund des bisherigen Forschungsstands plausibel und werden mittels konfirmatorischer Faktorenanalyse geprüft:

- *Modell 1*: ein domänengenerelles Ein-Faktor-Modell, in dem alle Arbeitsgedächtnisaufgaben auf einen globalen Faktor laden;
- *Modell 2*: ein domänenspezifisches Zwei-Faktoren-Modell, in dem ein verbal-numerischer und ein figural-räumlicher Faktor spezifiziert werden, die miteinander korrelieren;
- *Modell 3*: ein domänenspezifisches Drei-Faktoren-Modell, in dem ein verbaler, ein numerischer und ein figural-räumlicher Faktor spezifiziert werden, die miteinander korrelieren.

Es wird aufgrund des in Kapitel 3 referierten Forschungsstands erwartet, dass Modell 2 die Daten am besten erklären kann. Zwar gibt es Hinweise auf einen separaten numerischen Faktor, es wird jedoch vermutet, dass verbale und numerische Arbeitsgedächtnisleistungen zumindest enger kovariieren als jeweils verbale oder numerische Leistungen mit Leistungen in figuralen Aufgaben. Zur Faktorstruktur werden daher die folgenden beiden Hypothesen formuliert:

- a. Modell 2 erklärt die Kovarianzstruktur der Daten besser als Modell 1.
- b. Modell 3 erklärt die Kovarianzstruktur nicht besser als Modell 2.

Ferner wird erwartet, dass die inhaltsspezifischen Faktoren positiv korreliert sind, so dass ein übergeordneter allgemeiner Arbeitsgedächtnisfaktor angenommen werden kann.

#### *Prüfung der Invarianz der Faktorstruktur für verschiedene Gruppen*

Nach der Untersuchung, welches Modell die Kovarianzstruktur in den Daten aus der Gesamtstichprobe am besten erklärt, soll weiterhin die Invarianz der Faktorstruktur für verschiedene Subgruppen untersucht werden. Es ist bisher nicht geklärt, ob sich die Struktur des Arbeitsgedächtnisses in bestimmten Gruppen systematisch unterscheidet.

#### *Invarianz der Faktorstruktur über Altersgruppen*

Bisher ist nur wenig bekannt über mögliche Veränderungen der Faktorstruktur des Arbeitsgedächtnisses im Kindesalter. Aktuelle Studien aus der Intelligenzstrukturforschung (z. B. Carroll, 1993; Bickley et al., 1995) sprechen eher für eine Invarianz der Faktorstruktur der Intelligenz im Kindesalter. Erste Studien mit Arbeitsgedächtnisaufgaben deuten darauf hin, dass sich die Struktur im hier betrachteten Altersbereich nicht verändert (de Jonge & de Jong, 1996). Es wird daher erwartet, dass die für die Gesamtstichprobe gefundene Faktorstruktur für Kinder verschiedener Altersgruppen invariant ist, das heißt, dass sich die geschätzten Modelle für verschiedene Altersgruppen weder in ihren Faktorkovarianzen noch in ihren Faktorladungen signifikant unterscheiden. Als Hypothese kann hier formuliert werden:

- Die Faktorstruktur des Arbeitsgedächtnisses unterscheidet sich für Kinder verschiedener Altersgruppen nicht.

### *Invarianz der Faktorstruktur über Begabungsgruppen*

Die Faktorstruktur des Arbeitsgedächtnisses wurde bisher nicht für Kinder verschiedener Begabungsgruppen untersucht. Auch in der Intelligenzstrukturforschung sind die Befunde zur Invarianz der Faktorstruktur für durchschnittlich und hoch begabte Kinder bisher rar und zudem widersprüchlich. Eine aktuelle und umfassende Studie mit dem BIS-HB konnte jedoch zeigen, dass das BIS-Modell für durchschnittlich und hoch begabte Kinder gleichermaßen gilt (Jäger et al., 2005).

Da die Befunde in der Intelligenzstrukturforschung eher auf eine Invarianz der Intelligenzstruktur schließen lassen und das Arbeitsgedächtnis und die fluide Intelligenz als eng zusammenhängende Konstrukte angesehen werden können, wird erwartet, dass die für die Gesamtstichprobe gefundene Faktorstruktur für Kinder verschiedener Begabungsgruppen invariant ist. Das heißt, dass sich die geschätzten Modelle für Kinder verschiedener Begabungsgruppen und für hoch begabte Kinder weder in ihren Faktorkovarianzen noch in ihren Faktorladungen signifikant unterscheiden. Die zu prüfende Hypothese lautet daher:

- Die Faktorstruktur des Arbeitsgedächtnisses unterscheidet sich für hoch begabte Kinder und für eine hinsichtlich kognitiver Fähigkeiten repräsentative Gruppe von Kindern nicht.

### *Invarianz der Faktorstruktur bei Mädchen und Jungen*

Zur Invarianz der Faktorstruktur des Arbeitsgedächtnisses bei Mädchen und Jungen liegen bisher keine Befunde vor. Da in der bisherigen Literatur nicht über Unterschiede in der Faktorstruktur berichtet wurde, wird hier angenommen, dass sich die Faktorstruktur des Arbeitsgedächtnisses von Mädchen und Jungen nicht unterscheidet. Als Hypothese wird daher formuliert:

- Die Faktorstruktur des Arbeitsgedächtnisses unterscheidet sich für Mädchen und Jungen nicht.

## *2. Leistungsunterschiede zwischen verschiedenen Gruppen*

Es wird angenommen, dass sich die Leistungsfähigkeit des Arbeitsgedächtnisses in verschiedenen Gruppen unterscheidet. So wird erwartet, dass ältere Kinder in Arbeitsgedächtnisaufgaben bessere Leistungen erbringen als jüngere Kinder und hoch intelligente Kinder bessere Leistungen erbringen als nicht hoch intelligente Kinder. Weiter wird vermutet, dass sich Mädchen und Jungen in ihren mittleren Leistungen generell nicht unterscheiden, dass jedoch die Jungen den Mädchen in Aufgaben mit Anforderungen an räumliche Vorstellungsfähigkeiten überlegen sind und daher möglicherweise in der Aufgabe „Spatial Working Memory“ besser abschneiden. Diese Annahmen werden im Folgenden näher erläutert. Die Annahmen werden auf der Basis der sechs Aufgabengruppen überprüft (und nicht auf der Basis der später erstellten Testskalen), da entsprechende Unterschiede möglicherweise nur in einzelnen Aufgabengruppen auftreten könnten.

*Leistungsunterschiede in den Arbeitsgedächtnisaufgaben zwischen jüngeren und älteren Kindern*

Die Leistungsfähigkeit des Arbeitsgedächtnisses steigt in der Kindheit mit zunehmendem Alter (z. B. Fry & Hale, 2000; Swanson, 1999; Siegel, 1994). Bisherige Studien berichten von einer Leistungszunahme über die gesamte Kindheit und darüber hinaus (Swanson, 1999). In der vorliegenden Studie soll daher geprüft werden, ob sich dieser klare Alterstrend auch für die hier vorgenommene Aufgabenauswahl bestätigen lässt. Es wird daher geprüft, ob sich jüngere und ältere Kinder in ihrer mittleren Leistung im Arbeitsgedächtnis unterscheiden. Als Hypothese wird formuliert:

- Ältere Kinder zeigen im Mittel in allen sechs Arbeitsgedächtnisaufgaben bessere Leistungen als jüngere Kinder.

*Leistungsunterschiede in den Arbeitsgedächtnisaufgaben zwischen hoch intelligenten Pbn und nicht hoch intelligenten Pbn*

Bisher sind keine Befunde zu Unterschieden in der Leistungsfähigkeit des Arbeitsgedächtnisses von durchschnittlich und hoch intelligenten Kindern bekannt. Da jedoch vielfältige Befunde zum Zusammenhang von Intelligenzmaßen und der Arbeitsgedächtniskapazität vorliegen (s. o.), sind hier im Mittel bessere Leistungen der hoch Intelligenten zu erwarten. Als Hypothese wird formuliert:

- Hoch intelligente Pbn erzielen im Mittel höhere Leistungen in den sechs Arbeitsgedächtnisaufgaben als nicht hoch intelligente Pbn.

*Leistungsunterschiede in den Arbeitsgedächtnisaufgaben zwischen Mädchen und Jungen*

Generell wird davon ausgegangen, dass die Leistungsfähigkeit des Arbeitsgedächtnisses bei Mädchen und Jungen nicht unterschiedlich ausgeprägt ist. Jedoch könnten die verschiedenen Inhalte der verwendeten Arbeitsgedächtnisaufgaben dazu führen, dass entweder Jungen oder Mädchen bestimmte Aufgaben leichter fallen.

Zwar erreichen Mädchen in verbalen Testaufgaben in vielen Studien höhere Werte als Jungen (z. B. Loehlin, 2000; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991), häufig sind die gefundenen Effekte jedoch eher klein (z. B. für die Verbalskala im BIS-HB; Jäger et al., 2005). Eine Metaanalyse, die die Daten von 120 Studien zu Geschlechterunterschieden in verbalen Testaufgaben zusammenfasste, konnte keine bedeutsam höheren Leistungen von Mädchen im Schulalter gegenüber gleichaltrigen Jungen belegen (Hyde & Linn, 1988). Werden Geschlechterunterschiede für verschiedene verbale Aufgabentypen separat untersucht, so zeigt sich, dass weibliche Pbn bessere Leistungen im sprachlichen Ausdruck und beim Finden von Synonymen zeigen, männliche Pbn schneiden hingegen beim Lösen verbaler Analogien etwas besser ab (z. B. Hines, 1990). Die hier verwendeten Arbeitsgedächtnisaufgaben stellen jedoch nur sehr geringe Anforderungen an die eigentlichen sprachlichen Fähig-

keiten, daher sollten in den Arbeitsgedächtnisaufgaben mit verbalen Inhalten keine Geschlechterunterschiede auftreten.

- Es wird erwartet, dass Mädchen in den verbalen Arbeitsgedächtnisaufgaben nicht besser abschneiden als Jungen.

Es konnte bereits mehrfach belegt werden, dass männliche Pbn in Aufgaben mit Anforderungen an die räumliche Vorstellungsfähigkeit besser abschneiden als weibliche (z. B. Linn & Petersen, 1985, 1986). Besonders stark ausgeprägt ist dieser Unterschied bei solchen Aufgaben, die eine mentale Rotation erfordern. Diese Geschlechterunterschiede finden sich für alle Altersgruppen, bereits 10-jährige Jungen können Aufgaben, die mentale Rotationen erfordern, besser lösen als gleichaltrige Mädchen. Die in der vorliegenden Untersuchung verwendete Aufgabe „Spatial Working Memory“ (SWM) erfordert eine mentale Rotation einfacher, schachbrettartiger Muster. Für die andere hier verwendete figurale Aufgabe „Memory Updating Figural“ (MUF) wird hingegen kein Geschlechterunterschied erwartet.

- Es wird erwartet, dass Jungen im Mittel höhere Werte im Aufgabentyp SWM erreichen als Mädchen. Die mittleren Leistungen im Aufgabentyp MUF unterscheiden sich bei Mädchen und Jungen nicht.

Im numerischen Bereich unterscheiden sich die mittleren Fähigkeiten von Jungen und Mädchen tendenziell nicht. Eine Metaanalyse, in der 100 Studien zu mathematischen Fähigkeiten zusammengefasst wurden, konnte für unausgelesene Stichproben über alle Altersgruppen hinweg keine Unterschiede in mathematischen Fähigkeiten zwischen den Geschlechtern feststellen (Hyde, Fennema & Lamon, 1990). Daher wird hier die folgende Annahme formuliert:

- Es werden keine Mittelwertsunterschiede zwischen Mädchen und Jungen in der Leistung in den numerischen Arbeitsgedächtnisaufgaben erwartet.

### *3. Zusammenhänge des Konstrukts „Arbeitsgedächtnis“ mit anderen kognitiven Konstrukten*

Auf der Basis des am besten passenden Faktorenmodells für das Arbeitsgedächtnis werden anschließend die Zusammenhänge mit den anderen untersuchten kognitiven Konstrukten auf der Ebene der latenten Faktoren in einem gemeinsamen Modell geschätzt. Hierfür muss für die Konstrukte „Intelligenz“ und „Speed“ vorab ebenfalls die Faktorstruktur, die dem Gesamtmodell zugrunde gelegt werden soll, geklärt werden.

#### *Fluide und kristalline Intelligenz*

Leistungen in Arbeitsgedächtnisaufgaben korrelieren in aller Regel positiv mit den Leistungen in Tests, die die fluide Intelligenz erfassen (z. B. Kyllonen & Christal, 1990; Engle et al., 1999). Das Ausmaß der Zusammenhänge variiert in bisherigen Studien jedoch beträchtlich (siehe Kapitel 3). In Studien mit latenten Faktoren liegen die berichteten Werte zwischen  $r = .66$  (de Jong & Das-Smaal, 1995) und  $r = .96$

(Kyllonen, 1996). De Jong und Das-Smaal (1995) verwendeten jedoch nicht nur komplexe Spannenmaße, sondern mit der „Forward Digit Span“ auch ein einfaches Spannenmaß. Da einfache Spannenaufgaben weniger gut dazu geeignet sind, die fluide Intelligenz vorherzusagen, kann vermutet werden, dass bei dem Einsatz von ausschließlich komplexen Spannenmaßen noch etwas höhere Zusammenhänge gefunden werden. Ackerman et al. (2002) verwendeten für ihre Untersuchung an erwachsenen Pbn sieben verschiedene komplexe Spannenaufgaben mit verbalen, numerischen und figuralen Inhalten und berichten über eine Korrelation von  $r = .70$ . Bei der Verwendung der oben genannten sechs Arbeitsgedächtnisaufgaben in der vorliegenden Arbeit wird daher erwartet, Zusammenhänge zwischen der fluiden Intelligenz und dem Arbeitsgedächtnis von (mindestens) etwa  $r = .70$  zu finden.

Auch mit eher kristallinen Aspekten der Intelligenz finden sich positive Zusammenhänge des Arbeitsgedächtnisses (z. B. Swanson, 1996; Engle, Tuholski et al., 1999), die jedoch in der Regel etwas geringer ausfallen als die Zusammenhänge mit der fluiden Intelligenz. Kristalline Fähigkeiten, wie sie beispielsweise in Wortschatztests gefordert sind, werden in stärkerem Maße als fluide Fähigkeiten durch Bildungs- und Wissensaspekte sowie durch das Langzeitgedächtnis beeinflusst.

- Es wird erwartet, dass das Arbeitsgedächtnis positiv mit der fluiden und der kristallinen Intelligenz korreliert.
- Weiter wird erwartet, dass die Zusammenhänge zwischen Arbeitsgedächtnis und fluider Intelligenz stärker ausfallen als zwischen Arbeitsgedächtnis und kristalliner Intelligenz.

Die Analyse der Zusammenhänge erfolgt generell auf der Konstruktebene, das heißt über latente Faktoren. Dennoch werden hier zusätzlich die Zusammenhänge auf der Aufgabenebene betrachtet, da für jeden einzelnen Typ von Arbeitsgedächtnisaufgaben geprüft werden soll, ob er sich prinzipiell für die Prognose von Intelligenztestleistungen eignet. Dieses Vorgehen ist deshalb sinnvoll, da sich einige Aufgabentypen in anderen Studien bei sehr fähigen Pbn nicht als gute Indikatoren des Arbeitsgedächtnisses bewähren konnten, und sich bei überdurchschnittlich intelligenten Pbn keine Zusammenhänge mehr zwischen dem Arbeitsgedächtnis und der allgemeinen Intelligenz fanden. Für die klassische Arbeitsgedächtnisaufgabe „Backward Digit Span“ (vgl. Abschnitt 4.1.2), die auch in der vorliegenden Arbeit verwendet wird, wurde beispielsweise festgestellt, dass sie bei sehr intelligenten Kindern nicht dazu geeignet ist, Intelligenztestleistungen vorherzusagen; bei durchschnittlich und unterdurchschnittlich intelligenten Pbn gelang dies jedoch (Schofield & Ashman, 1986). Eine mögliche Erklärung dafür ist, dass die Verarbeitungsanforderung von Rückwärtsspannen-Aufgaben für sehr fähige Pbn so gering ist, dass die komplexe Spannenaufgabe de facto zu einer einfachen Spannenaufgabe wird (Engle, Tuholski et al., 1999). Prinzipiell könnte dies jedoch auch für andere Typen von Arbeitsgedächtnisaufgaben zutreffen. Es ist daher zu prüfen, ob die in dieser Arbeit vorgenommene Aufgabenauswahl dazu geeignet ist, auch bei hoch begabten Kindern Leistungen in kognitiven Tests vorherzusagen.



*Geschwindigkeit der Informationsverarbeitung*

Die Zusammenhänge zwischen dem Arbeitsgedächtnis und der Geschwindigkeit der Informationsverarbeitung oder Speed wurden auf der Konstruktebene bereits vielfach untersucht. Die gefundenen Zusammenhänge fallen jedoch geringer aus als die Zusammenhänge zwischen Arbeitsgedächtnis und fluider Intelligenz. Insgesamt variiert auch hier die Höhe der Zusammenhänge in den verschiedenen Studien beträchtlich, in Abschnitt 3.4.2 wurden Korrelationen auf der Konstruktebene in verschiedenen Studien im Bereich zwischen  $r = .35$  (Kyllonen & Christal, 1990) und  $r = .60$  (de Jong & Das-Smaal, 1995) referiert.

Eine Ursache für die unterschiedlich starken Zusammenhänge in verschiedenen Studien ist die Zusammenstellung der Speed-Aufgaben. Es scheint plausibel (und konnte auch empirisch bestätigt werden, siehe z. B. Babcock, 1994; Ackerman et al., 2002), dass die Höhe des Zusammenhangs unter anderem davon abhängt, in welchem Ausmaß die Speed-Aufgaben Anforderungen an eine kognitive Verarbeitung stellen. Speed-Aufgaben mit höheren Anforderungen an die kognitive Verarbeitung sollten höher mit Arbeitsgedächtnisleistungen korrelieren als Speed-Aufgaben mit geringen Anforderungen an die kognitive Verarbeitung. Es werden daher in der vorliegenden Arbeit sowohl Speed-Aufgaben mit komplexeren als auch Speed-Aufgaben mit geringeren kognitiven Anforderungen eingesetzt. Entsprechend wird untersucht, ob die eher elementaren und die komplexeren Speed-Aufgaben in einer gemeinsamen Faktorenanalyse zwei separate Faktoren bilden, die jeweils unterschiedlich stark mit dem Arbeitsgedächtnis assoziiert sind.

- Es wird erwartet, dass die Speed-Aufgaben positiv mit den Arbeitsgedächtnisaufgaben korrelieren. Speed-Aufgaben mit höheren kognitiven Anforderungen korrelieren höher mit dem Arbeitsgedächtnis als Speed-Aufgaben mit geringeren kognitiven Anforderungen.

Sollten sich faktorenanalytisch zwei Speed-Faktoren auffinden lassen, die als Speed mit unterschiedlichem Ausmaß an kognitiver Komplexität interpretiert werden können, so wird erwartet, dass der Speed-Faktor, der Speed-Prozesse mit stärkerer Beteiligung höherer kognitiver Prozesse widerspiegelt, im Gesamtmodell (s. u.) stärker mit dem Arbeitsgedächtnis korreliert.

Speed-Aufgaben mit eher komplexen Anforderungen finden sich in verschiedenen Intelligenztests. In der vorliegenden Arbeit wird der Zahlen-Verbindungs-Test (ZVT) von Oswald und Roth (1987) eingesetzt, der in der Intelligenzdiagnostik auch als „Kurzintelligenztest“ verwendet wird. Zusätzlich werden zwei Untertests aus dem HAWIK III, der „Zahlen-Symbol-Test“ und die „Symbolsuche“, eingesetzt. Sowohl der ZVT als auch der Zahlen-Symbol-Test sind klassische Verfahren zur Messung der Informationsverarbeitungsgeschwindigkeit in der psychometrischen Tradition. Der Zahlen-Symbol-Test wurde bereits mehrfach als eher komplexes Speed-Maß klassifiziert (siehe Babcock, 1994; Ackerman et al., 2002). Der ZVT und die Speed-Tests aus dem HAWIK III wurden zudem bereits mehrfach in Studien mit hoch begabten Kindern eingesetzt (z. B. Jäger et al., 2005; Rost, 2000).

Die Operationalisierung der Geschwindigkeit der Informationsverarbeitung mit eher geringer kognitiver Anforderung erfolgt in der vorliegenden Studie über Aufgaben nach dem Hick-Paradigma im Paper-Pencil-Format. Mit dem Einsatz dieses Aufgabentyps wird zusätzlich das Ziel verfolgt, zu prüfen, ob sich eine Paper-Pencil-Version der klassischen Hick-Aufgaben prinzipiell für die Untersuchung elementarer Informationsverarbeitungsgeschwindigkeit bei Kindern mit durchschnittlicher und hoher Intelligenz eignet. Die sehr ökonomische Paper-Pencil-Version erlaubt eher als der klassische Versuchsaufbau die Verwendung der Hick-Aufgaben für die psychologische Diagnostik.

### *Gesamtmodell kognitiver Fähigkeiten*

Zusammenfassend wird für die Analysen auf der Konstruktebene erwartet, dass alle kognitiven Konstrukte positiv miteinander korrelieren. Aufgrund des Forschungsstands kann weiterhin erwartet werden, dass die höchsten Zusammenhänge zwischen dem Arbeitsgedächtnis und der fluiden Intelligenz bestehen. Weiterhin wird angenommen, dass der Zusammenhang zwischen Arbeitsgedächtnis und kristalliner Intelligenz, die stärker wissens- und gedächtnisabhängig ist als die fluide Intelligenz, ebenfalls hoch ist, aber etwas geringer ausfällt. Sollten sich empirisch elementare und komplexe Speed-Aufgaben differenzieren lassen, so werden weiterhin höhere Zusammenhänge des Arbeitsgedächtnisses mit eher komplexen als mit eher elementaren Prozessen der Informationsverarbeitungsgeschwindigkeit erwartet.

- Es wird erwartet, dass alle untersuchten kognitiven Konstrukte positiv korrelieren.
- Es wird erwartet, dass der höchste Zusammenhang im Gesamtmodell zwischen dem Arbeitsgedächtnis und der fluiden Intelligenz besteht. Ein etwas geringerer Zusammenhang wird zwischen Arbeitsgedächtnis und kristalliner Intelligenz erwartet.

### *Vorhersage von Schulleistungen durch Arbeitsgedächtnisaufgaben*

Verschiedene Studien konnten zeigen, dass die Leistungen in Arbeitsgedächtnisaufgaben mit den schulischen Leistungen zusammenhängen (de Jong & Das-Smaal, 1995; Lehto, 1995). Gathercole et al. (2004) untersuchten die Zusammenhänge zwischen den Aufgaben „Listening Recall“ und „Backward Digit Recall“ und Schulleistungen (erfasst über standardisierte Leistungstests) bei britischen Schülerinnen und Schülern der Klassenstufen 3 und 10. Bei den jüngeren Kindern korrelierten diese Aufgaben zwischen  $r = .41$  und  $r = .53$  mit der Englisch- und der Mathematikleistung. Bei den älteren Pbn korrelierte die Englischleistung nur zu  $r = .32$  bzw.  $r = .35$  mit den Arbeitsgedächtnisaufgaben, die Mathematikleistung hingegen zu  $r = .54$  bzw.  $r = .45$ . Für die älteren Schülerinnen und Schüler wurde zusätzlich der Zusammenhang der Aufgaben mit der Leistung in den Naturwissenschaften bestimmt, dieser lag für die Aufgabe „Listening Recall“ bei  $r = .59$  und für die Aufgabe „Backward Digit Recall“ bei  $r = .39$ . Lehto (1995) berichtet für die Aufgaben „Reading Span“ und „Operation Span“ über Korrelationen zwischen  $r = .39$  und  $r = .56$

mit der Schulleistung in sprachlichen Fächern bei 15- bis 16-jährigen Jugendlichen. Die Mathematikleistung korrelierte bei den Jugendlichen zu  $r = .33$  mit der „Reading Span“ und zu  $r = .55$  mit der „Operation Span“. Für das deutsche Schulsystem sind keine entsprechenden Studien bekannt. Verschiedene Autoren haben darauf hingewiesen, dass die Zusammenhänge zwischen kognitiven Fähigkeiten und den Schulnoten je nach Klassenstufe unterschiedlich ausgeprägt sind; in den unteren Schuljahrgängen finden sich in der Regel stärkere Zusammenhänge als in den höheren Klassen (siehe z. B. Jensen, 1980, 1998). Die Höhe der Zusammenhänge zwischen kognitiven Maßen und Schulnoten ist insgesamt vor allem dadurch begrenzt, dass Schulnoten nur über eine eingeschränkte Objektivität, Reliabilität und Validität verfügen.

Für die vorliegende Arbeit wird daher erwartet, dass die sechs ausgewählten Arbeitsgedächtnisaufgaben negativ mit den Leistungen in den Kernfächern korrelieren, dabei werden aufgrund der oben referierten Befunde Korrelationen in einer Höhe von mindestens ca.  $r = -.30$  erwartet. (Es werden negative Zusammenhänge angenommen, da Schulnoten im deutschen Schulsystem von „sehr gut“ = 1 bis „ungenügend“ = 6 skaliert sind.) Dabei wird zusätzlich erwartet, dass die Korrelationen in der Primarstufe höher ausfallen als in der Sekundarstufe.

- Es wird erwartet, dass die Arbeitsgedächtnisaufgaben mit den Zeugnisnoten in den Kernfächern (Deutsch, Fremdsprachen, Mathematik und Naturwissenschaften bzw. Sachkunde) negativ korrelieren.
- Weiter wird erwartet, dass die Korrelationen in der Primarstufe höher ausfallen als in der Sekundarstufe.

Möglicherweise finden sich zudem jeweils höhere Zusammenhänge zwischen Schulfächern und Arbeitsgedächtnisaufgaben aus derselben Domäne als zwischen Aufgaben und Schulfächern aus unterschiedlichen Domänen. Dies wäre z. B. der Fall, wenn verbale Arbeitsgedächtnisaufgaben höher mit den Leistungen in sprachlichen Fächern korrelieren als mit den Leistungen in Mathematik. Ein entsprechender Befund könnte als ein (schwacher) Hinweis auf eine Domänenspezifität des Arbeitsgedächtnisses gewertet werden (vgl. Abschnitt 3.3).

### 5.3.2 Fragestellungen zur Testentwicklung

Diese Arbeit verfolgt das Ziel, zu prüfen, inwieweit sich computergestützte Arbeitsgedächtnisaufgaben für die Diagnostik hoher und sehr hoher intellektueller Fähigkeiten bei Kindern zwischen 8 und 13 Jahren eignen. Eine erste Grundlegung für ein entsprechendes Testverfahren soll daher bereitgestellt werden. Im Folgenden wird dargelegt, welche Fragestellungen im Hinblick auf die Testentwicklung in der vorliegenden Arbeit geprüft werden sollen.

Die oben ausgeführten Fragestellungen zum Konstrukt erfordern Analysen auf der Ebene der einzelnen Aufgabengruppen, hierzu werden die Summenscores über die einzelnen Aufgabengruppen verwendet. Für die Fragestellungen zur Testentwicklung

erfolgen die statistischen Analysen hingegen auf der Ebene der einzelnen *Testaufgaben*, die dichotom kodiert sind.

### *Dimensionalität der Arbeitsgedächtnisaufgaben*

Als erster Schritt wird faktorenanalytisch untersucht, zu welchen Skalen sich die einzelnen Aufgaben aller sechs Aufgabengruppen zusammenfassen lassen. Je nach gefundener Faktorstruktur auf der Ebene der Aufgaben-Scores (s. o.) kann hier ein entsprechendes Ergebnis erwartet werden, das die Bildung von einer Gesamtskala oder von zwei oder drei inhaltsspezifischen Skalen nahe legt.

### *Bestimmung der Aufgaben- und Teststatistiken nach der Klassischen Testtheorie und der Item Response Theorie*

Zur Prüfung der Messeigenschaften der Skalen werden die psychometrischen Eigenschaften der Testaufgaben und der Skalen anhand von Teststatistiken entsprechend der Klassischen Testtheorie untersucht. Aufgrund der bekannten Schwächen klassischer Testmodelle (siehe auch Abschnitt 2.4.1) werden in der vorliegenden Arbeit zusätzlich Modelle der Item Response Theorie (IRT) herangezogen. Die hier verwendeten Testmodelle werden ausführlich z. B. bei Rost (2004) oder Hambleton et al. (1991) dargestellt.

In der vorliegenden Arbeit werden drei häufig verwendete IRT-Modelle auf ihre Datenpassung geprüft. Das erste dieser Modelle ist das *1-Parameter-logistische Modell* (1PL-Modell, auch: Rasch-Modell; Rasch, 1960), das die Grundlage aller IRT-Modelle darstellt. Im 1PL-Modell wird angenommen, dass nur ein Parameter, nämlich die Schwierigkeit der Aufgaben, die Leistungen der Pbn beeinflusst.

Ein zweites häufig verwendetes Modell ist das etwas weniger restriktive *2-Parameter-logistische Modell* (2PL-Modell, auch: Birnbaum-Modell; Birnbaum, 1968). Zusätzlich zur Schwierigkeit wird hier als weiterer Parameter die Trennschärfe der Testaufgaben in Form des Trennschärfe- oder Diskriminationsparameters betrachtet. Die Testaufgaben eines Tests können somit bei Geltung des 2PL-Modells sowohl in ihrer Schwierigkeit als auch in ihrer Trennschärfe variieren.

Ein weiterer Parameter – die Ratewahrscheinlichkeit – wird im *3-Parameter-logistischen Modell* (3PL-Modell) spezifiziert. Mit dem Rateparameter (auch: „pseudo-chance-level parameter“; Hambleton et al., 1991) wird die Wahrscheinlichkeit beschrieben, allein durch Raten auf die richtige Lösung zu kommen. Das 3PL-Modell ist insbesondere bei Testaufgaben im Multiple-Choice-Format relevant. Dieser dritte Parameter spezifiziert die untere Asymptote der ICC und wird entweder ebenfalls aus den Daten als Modellparameter geschätzt oder direkt aus dem Antwortformat der Items abgeleitet (z. B.  $p = .25$  bei vier Antwortmöglichkeiten).

Es wird angestrebt, dass sich der Zusammenhang zwischen der Personenfähigkeit und den Antworten der Pbn in den Arbeitsgedächtnisaufgaben über das 1PL- oder das 2PL-Modell beschreiben lässt. Da das Finden der richtigen Lösungen durch Raten in den hier untersuchten Arbeitsgedächtnisaufgaben sehr unwahrscheinlich ist,

wird nicht erwartet, dass das 3PL-Modell die Testdaten aus den Arbeitsgedächtnisaufgaben besser erklärt als das 1PL- oder 2PL-Modell. Sollten verschiedene Modellgeltungstests (vgl. Beschreibung des Vorgehens bei der testtheoretischen Analyse in Abschnitt 7.5.1) für eine prinzipielle Passung der Testdaten auf eines der Modelle sprechen, so werden einzelne Aufgaben, die von den Vorgaben des jeweiligen Modells abweichen, eliminiert, um modellkonforme Skalen zu erhalten.

Für diese Skalen können dann die Reliabilität und die externe Validität (im Sinne einer Übereinstimmungsvalidität mit anderen Testverfahren und schulischen Leistungen) entsprechend den Vorgaben der Klassischen Testtheorie bestimmt werden. Es werden positive Zusammenhänge der Skalen mit anderen Intelligenztests erwartet. Hierbei ist zu berücksichtigen, dass auch für relativ konstruktnahe Intelligenztests oft nur Korrelationen in mittlerer Höhe erwartet werden können. So berichtet Weiß (1998) über eine Korrelation des CFT 20 mit den Coloured Progressive Matrices (CPM) von  $r = .58$  bei Grundschulern ( $N = 600$ ). CFT 20 und CPM sind konstruktnahe Verfahren, sie erfassen beide die fluide Intelligenz über figurale Testaufgaben. Das Zusatzmodul „Wortschatztest“ des CFT 20 korreliert im Mittel zu  $r = .65$  mit verschiedenen konstruktnahen und zu  $r = .15$  (n.s.) mit verschiedenen konstruktfernen Intelligenztests (Weiß, 1998). Für das Zusatzmodul „Zahlenfolgentest“ werden mittlere Korrelationen von  $r = .57$  mit konstruktnahen und von  $r = .34$  mit konstruktfernen Tests berichtet (Weiß, 1998). Sollten sich Korrelationen zwischen den Arbeitsgedächtnisskalen und den Intelligenztests finden lassen, die den Korrelationen von konstruktnahen Intelligenztests entsprechen, wäre das ein guter Beleg für die Validität.

Die Bestimmung der Validität eines Intelligenztests über Korrelationen mit Schulnoten ist übliche Praxis, da davon ausgegangen wird, dass ein entsprechender Test dazu in der Lage sein sollte, schulische Leistungen in gewissem Maße zu prognostizieren. Die nur relativ geringe Objektivität, Reliabilität und Validität von Schulnoten begrenzt jedoch von vornherein die Höhe möglicher Zusammenhänge. Auf der Grundlage der Ergebnisse zur Reliabilität und Validität der Testskalen kann eine erste Einschätzung erfolgen, ob sich die Arbeitsgedächtnisskalen auch im Vergleich mit klassischen Intelligenztests zur Messung von intellektuellen Fähigkeiten eignen.

Wenn sich die Arbeitsgedächtnisskalen nach dem hier skizzierten Vorgehen psychometrisch bewähren, wären zwei weitere Schritte erforderlich – die jedoch im Rahmen dieser Arbeit nicht mehr geleistet werden können –, damit die Skalen für die Individualdiagnostik einsetzbar sind. Zunächst müsste die Stabilität der Messungen im Rahmen einer Retestung untersucht werden. Die Ergebnisse verschiedener Studien ergaben für verschiedene Arbeitsgedächtnisaufgaben nur relativ geringe Retest-Reliabilitäten (z. B. Waters & Caplan, 1996; Hacker et al., 1999). Für die vorliegenden Skalen wäre zu prüfen, ob die Messergebnisse ausreichend stabil sind, um sie diagnostisch verwerten zu können. Es wird jedoch vermutet, dass die in bisherigen Studien aufgefundenen geringen Reliabilitätswerte zumindest zum Teil auch auf verschiedene Bedingungen zurückzuführen sind, die den Messfehler erhöhen, z. B. eine geringere Durchführungsobjektivität aufgrund der mündlichen Vorgabe der Testaufgaben durch einen Testleiter. Für die hier verwendeten Skalen wird eine höhere

Retest-Reliabilität erwartet, da die Aufgabenvorgabe durch den Computer erfolgt und die Durchführung in hohem Maße standardisiert ist. Auch die Zusammenstellung von Skalen aus jeweils mehreren verschiedenen Aufgabentypen könnte dazu beitragen, die Übereinstimmung zwischen verschiedenen Messungen zu verbessern. Wenn die Retest-Reliabilität bestimmt ist, müssten die Arbeitsgedächtnisskalen anschließend an einer Stichprobe durchschnittlich und hoch intelligenter Kinder normiert werden.

Im folgenden Kapitel wird die Untersuchung mit allen eingesetzten Materialien und den untersuchten Stichproben beschrieben. Die Darstellung der Ergebnisse folgt in Kapitel 7.

## 6. Untersuchung

In diesem Kapitel werden zunächst die verwendeten Aufgabentypen zur Erfassung der Arbeitsgedächtniskapazität beschrieben (Abschnitt 6.1). Abschnitt 6.2 stellt die zusätzlich verwendeten Materialien (Fragebogen und verschiedene kognitive Tests) dar. Im darauf folgenden Abschnitt (6.3) wird die untersuchte Stichprobe vorgestellt, die sich aus unausgelesenen Pbn und Kindern aus spezialisierten Einrichtungen und Organisationen für hoch Begabte zusammensetzt. Im letzten Abschnitt (6.4) wird die Durchführung der Untersuchung beschrieben.

### 6.1 Arbeitsgedächtnisaufgaben

In der hier beschriebenen Untersuchung werden sechs verschiedene Aufgabentypen zur Erfassung der Arbeitsgedächtniskapazität eingesetzt. Diese waren in vier Vorstudien mit durchschnittlich und hoch begabten Kindern ( $N = 22-118$ ) erprobt und optimiert worden. Im Folgenden wird die Testendversion beschrieben, wie sie in der Hauptuntersuchung eingesetzt wurde.

Für die computergestützte Darbietung der Arbeitsgedächtnisaufgaben wurde das Programm „WMC 2000“ (Oberauer & Süß, 1996) verwendet. Das Programm erlaubt den Pbn eine teilweise Selbststeuerung. Auf Anweisung des Versuchsleiters starten die Pbn einen neuen Aufgabentyp durch Tastendruck. Innerhalb eines Aufgabentyps, das heißt zwischen den einzelnen Aufgaben, können die Pbn selbst entscheiden, wann sie mit der nächsten Aufgabe beginnen möchten. Dies erlaubt den Pbn kurze individuelle Pausen und vermeidet Leistungseinbußen durch kurzzeitige Konzentrationslücken.

Alle Pbn bearbeiten die Aufgaben in jeweils derselben festgelegten Reihenfolge. Stets wird jeder Aufgabentyp vorab anhand von Beispielen erklärt. Anschließend bearbeitet jeder Pb mehrere Übungsaufgaben unterschiedlicher Schwierigkeit am Computer. Nach jeder Übungsaufgabe erfolgt ein automatisches Feedback, in dem der Pb über seine Leistung in der Aufgabe informiert wird. Wenn der Pb eine Übungsaufgabe nicht korrekt löst, wird dieselbe Aufgabe so lange wiederholt, bis der Pb sie einmal vollständig richtig beantwortet. Dieses Verfahren soll zum einen sicherstellen, dass jeder Pb die Instruktion richtig verstanden hat und es soll zum anderen den Pbn die Gelegenheit geben, mit dem Ablauf der Aufgabe vertraut zu werden. Es kann so davon ausgegangen werden, dass zu Beginn der Testaufgaben alle Pbn ausreichend mit dem Prinzip des jeweiligen Aufgabentyps vertraut sind. Alle Testungen in dieser Untersuchung wurden von mindestens zwei Testleitern durchgeführt, so dass es möglich war, individuelle Fragen der Pbn zu beantworten, das Verhalten der Kinder zu beobachten und das Befolgen der Instruktionen zu kontrollieren.

### 6.1.1 Aufgabe „Reading Span“ (RS)

Die hier verwendete Variante der „Reading Span“ (RS) unterscheidet sich etwas von der klassischen Version, die in Abschnitt 4.1.1 vorgestellt wurde, und orientiert sich an der Version von Oberauer et al. (2000). Die Schwierigkeit der Aufgaben wird über die Anzahl von Sätzen bestimmt, die nacheinander gelesen werden müssen. Nach dem kurzen Erscheinen des Hinweises „Achtung“ werden nacheinander einzelne Sätze auf dem Bildschirm gezeigt (siehe Abbildung 6.1). So lange der Pb noch keine Antwort eingegeben hat, ist zusätzlich ein Fragezeichen sichtbar. Die einzelnen Sätze werden 5 s lang präsentiert. Während jeder Satzdarbietung soll der Pb den Satz so schnell wie möglich als wahr oder falsch bewerten. Die Bewertung erfolgt über die Tastatur; zwei Tasten sind mit farbigen Aufklebern und der Beschriftung „ja“ (,‘-Taste) bzw. „nein“ (,>/<-Taste) versehen. Sobald der Pb eine der beiden Tasten gedrückt hat, verschwindet das Fragezeichen auf dem Bildschirm. Der Satz wird jedoch konstant 5 s präsentiert, so dass alle Pbn dieselbe Zeit zum Einprägen des Satzes zur Verfügung haben – unabhängig davon, wie schnell sie eine Entscheidung treffen. Wenn der Pb innerhalb der vorgegebenen Zeitspanne keine Entscheidung trifft, wird das Item als falsch gelöst bewertet, und das nächste Item erscheint.

Nach einer Reihe von Sätzen erfolgt über den Bildschirm die Aufforderung: „Bitte die Wörter aufschreiben!“. Nun soll der Pb jeweils das letzte Wort jedes Satzes in der Reihenfolge des Erscheinens auf einem Antwortbogen in vorgegebenen Kästchen notieren. Für das Notieren der Wörter liegt keine Zeitbegrenzung vor. Um die Pbn dazu anzuhalten, die richtige Reihenfolge bei der Wiedergabe der Wörter zu beachten, werden sie instruiert, mit einem Strich im entsprechenden Kästchen auf dem Antwortbogen zu markieren, dass sie sich an das entsprechende Wort nicht erinnern können.

Bei der Formulierung der Items wurden das Lesenniveau und das Vorwissen von Kindern berücksichtigt. Alle Sätze sind kurz und einfach gehalten, stets trivial wahr oder falsch und das letzte Wort jedes Satzes bildet ein bekanntes Substantiv, bestehend aus maximal drei Silben. Zusätzlich wurde darauf geachtet, dass das letzte Wort einfach zu schreiben ist. Auch sollten sich die letzten Worte innerhalb einer Aufgabe möglichst nicht semantisch ähnlich sein, da eine semantische Ähnlichkeit das Einprägen der Worte erleichtern würde.

Der Pb erhält für jede komplett gelöste Aufgabe einen Punkt. Dies beinhaltet das Notieren der richtigen Wörter an der richtigen Position auf dem Antwortbogen und das Treffen der richtigen Entscheidungen („Satz ist wahr“ versus „Satz ist falsch“). Auf der niedrigsten Schwierigkeitsstufe (Level) enthalten die RS-Aufgaben drei Sätze, auf der höchsten Stufe sechs Sätze in Folge. Die Skala enthält insgesamt vier Schwierigkeitsstufen mit insgesamt elf Aufgaben (siehe Tabelle 6.1). Vorab bearbeiten die Pbn drei Übungsaufgaben.



**Beispielaufgabe „Reading Span“**

1. Bildschirm

Ein Schneemann ist aus Schnee.  
?

*Richtige Antwort:  
„Ja“*

2. Bildschirm

Ich sehe mit den Ohren.  
?

*Richtige Antwort:  
„Nein“*

---

Korrekte Antwort auf dem Antwortbogen:

|        |
|--------|
| Schnee |
| Ohren  |

Abbildung 6.1: Übungsaufgabe 1 der Aufgabengruppe „Reading Span“ mit richtiger Lösung

Tabelle 6.1: Struktur der Skala „Reading Span“

| Aufgabe         | Level  | Anzahl Sätze |
|-----------------|--------|--------------|
| Übungsaufgabe 1 | RS 0   | 2            |
| Übungsaufgabe 2 | RS I   | 3            |
| Übungsaufgabe 3 | RS II  | 4            |
| Aufgabe 1       | RS I   | 3            |
| Aufgabe 2       |        | 3            |
| Aufgabe 3       | RS II  | 4            |
| Aufgabe 4       |        | 4            |
| Aufgabe 5       |        | 4            |
| Aufgabe 6       | RS III | 5            |
| Aufgabe 7       |        | 5            |
| Aufgabe 8       |        | 5            |
| Aufgabe 9       | RS IV  | 6            |
| Aufgabe 10      |        | 6            |
| Aufgabe 11      |        | 6            |

Anmerkung. Level: RS 0 = 2 Sätze; RS I = 3 Sätze; RS II = 4 Sätze; RS III = 5 Sätze; RS IV = 6 Sätze.

### 6.1.2 Aufgabe „Verbal Span“ (VS)

Der Aufgabentyp „Verbal Span“ (VS) wurde von Oberauer et al. (2000) übernommen, jedoch wurde bei den einzelnen Aufgaben ein Vokabular verwendet, das dem Wortschatz von Kindern entspricht. Die Arbeitsgedächtnisanforderung wird in dieser Aufgabe über die Kombination zweier voneinander unabhängiger Teilaufgaben realisiert. Im ersten Teil der Aufgabe liest der Pb eine Liste kurzer Wörter, die für 6 s auf dem Bildschirm dargeboten wird. Es folgen dann zwei bis drei verbale Wahlaufgaben, bei denen der Pb jeweils ein Wort aus vier vorgegebenen Wörtern auswählen muss, das zu einem bestimmten Oberbegriff gehört. Der Oberbegriff ist in der Mitte des Bildschirms platziert, die vier Auswahlwörter sind nummeriert und befinden sich in den vier Ecken des Bildschirms (siehe Abbildung 6.2). Der Pb gibt seine Wahl über die Zahlentasten 1-4 in den Computer ein. Sobald eine Eingabe erfolgt ist, erscheint die nächste Wahlaufgabe. Die Wahlaufgaben unterliegen keiner Zeitbegrenzung, die Pbn werden jedoch instruiert, so schnell wie möglich zu antworten. Nach zwei bis drei Wahlaufgaben erscheint auf dem Bildschirm die Aufforderung „Bitte die Wörter aufschreiben!“, und der Pb soll die Wörterliste in der richtigen Reihenfolge auf einem Antwortbogen notieren. Für das Notieren der Wörter besteht ebenfalls keine Zeitbegrenzung. Die Schwierigkeit der Aufgaben ergibt sich aus der Anzahl von Wörtern und der Anzahl von nachfolgenden Wahlaufgaben.

Der Pb erhält für eine Aufgabe einen Punkt, wenn er jedes Listen-Item in der richtigen Position auf dem Antwortbogen notiert hat. Auf der niedrigsten Schwierigkeitsstufe (Level) enthalten die zu merkenden Wortlisten der VS-Aufgaben drei Wörter, und es sind zwei Wahlaufgaben zu lösen. Auf der höchsten Stufe sind sechs Wörter einzuprägen und zwei Wahlentscheidungen zu treffen. Die VS-Skala enthält vier Schwierigkeitsstufen mit insgesamt zehn Aufgaben (siehe Tabelle 6.2). Vorab bearbeiten die Pbn zwei Übungsaufgaben der Schwierigkeitsstufen 1 und 2.

**Beispielaufgabe „Verbal Span“**

1. Bildschirm: Wortliste

Blume  
Berg  
Mantel

2. Bildschirm: 1. Wahlaufgabe

1. Bild      2. Amsel  
Vogel  
3. Kette    4. Sofa

*Richtige Antwort: „2“*

3. Bildschirm: 2. Wahlaufgabe

1. Hammer    2. Spatz  
Werkzeug  
3. Bauch      4. Haus

*Richtige Antwort: „1“*

---

Korrekte Antwort auf dem Antwortbogen:

|        |
|--------|
| Blume  |
| Berg   |
| Mantel |

Abbildung 6.2: Übungsaufgabe 1 der Aufgabengruppe „Verbal Span“ mit richtiger Lösung

Tabelle 6.2: Struktur der Skala „Verbal Span“

| Aufgabe         | Level   | Anzahl Wörter | Anz. Entscheidungen |
|-----------------|---------|---------------|---------------------|
| Übungsaufgabe 1 | VS I    | 3             | 2                   |
| Übungsaufgabe 2 | VS IIb  | 4             | 3                   |
| Aufgabe 1       | VS I    | 3             | 2                   |
| Aufgabe 2       | VS IIa  | 4             | 2                   |
| Aufgabe 3       |         | 4             | 2                   |
| Aufgabe 4       | VS IIb  | 4             | 3                   |
| Aufgabe 5       |         | 4             | 3                   |
| Aufgabe 6       | VS IIIa | 5             | 2                   |
| Aufgabe 7       |         | 5             | 2                   |
| Aufgabe 8       | VS IIIb | 5             | 3                   |
| Aufgabe 9       |         | 5             | 3                   |
| Aufgabe 10      | VS IV   | 6             | 2                   |

Anmerkung. Level: VS I = 3 Wörter, 2 Entscheidungen; VS IIA = 4 Wörter, 2 Entscheidungen; VS IIb = 4 Wörter, 3 Entscheidungen; VS IIIa = 5 Wörter, 2 Entscheidungen; VS IIIb = 5 Wörter, 3 Entscheidungen; VS IV = 6 Wörter, 2 Entscheidungen.

### 6.1.3 Aufgabe „Computation Span“ (CS)

Die Aufgabe „Computation Span“ (CS; nach Daneman & Carpenter, 1980; siehe auch Turner & Engle, 1989) ist nach dem gleichen Prinzip aufgebaut wie die Aufgabe „Reading Span“. Hier werden jedoch an Stelle von Sätzen einfache Gleichungen dargeboten, die die Addition oder Subtraktion einstelliger Zahlen beinhalten, z. B.  $2 + 4 = 6$ . Die Gleichungen sind zum Teil richtig und zum Teil falsch. Die Darbietungszeit der Gleichungen beträgt 5 s. Innerhalb dieser Zeitspanne muss eine Bewertung richtig/falsch mittels Tastendruck erfolgen. Nach einer Reihe von Gleichungen sollen alle Ergebnisse der Gleichungen in der richtigen Reihenfolge über die Tastatur in den Computer eingegeben werden. Die Pbn werden instruiert, eine Null einzugeben, wenn sie sich an das richtige Ergebnis der Gleichung (stets zwischen 1 und 9) nicht erinnern können. Die Schwierigkeit der Aufgaben wird über die Anzahl von Gleichungen bestimmt, die nacheinander kontrolliert werden müssen.

Der Pb erhält für jede Aufgabe, die er komplett richtig gelöst hat, einen Punkt. Dies beinhaltet das Eingeben der richtigen Ziffern in der richtigen Reihenfolge und das Treffen der richtigen Entscheidungen („Gleichung ist richtig“ versus „Gleichung ist falsch“). Auf der niedrigsten Schwierigkeitsstufe bestehen die Aufgaben aus drei Gleichungen, auf der höchsten Stufe aus sechs Gleichungen in Folge. Die Skala enthält fünf Schwierigkeitsstufen (Level) mit insgesamt zehn Aufgaben (siehe Tabelle

6.3). Vorab bearbeiten die Pbn drei Übungsaufgaben der Schwierigkeitsstufen 0, 1 und 2.

Tabelle 6.3: Struktur der Skala „Computation Span“

| Aufgabe         | Level  | Anzahl Gleichungen |
|-----------------|--------|--------------------|
| Übungsaufgabe 1 | CS 0   | 2                  |
| Übungsaufgabe 2 | CS I   | 3                  |
| Übungsaufgabe 3 | CS II  | 4                  |
| Aufgabe 1       | CS I   | 3                  |
| Aufgabe 2       |        | 3                  |
| Aufgabe 3       | CS II  | 4                  |
| Aufgabe 4       |        | 4                  |
| Aufgabe 5       | CS III | 5                  |
| Aufgabe 6       |        | 5                  |
| Aufgabe 7       | CS IV  | 6                  |
| Aufgabe 8       |        | 6                  |
| Aufgabe 9       | CS V   | 7                  |
| Aufgabe 10      |        | 7                  |

Anmerkung. Level: CS 0 = 2 Gleichungen; CS I = 3 Gleichungen; CS II = 4 Gleichungen; CS III = 5 Gleichungen; CS IV = 6 Gleichungen; CS V = 7 Gleichungen.

#### 6.1.4 Aufgabe „Backward Digit Span“ (BDS)

Die verwendete Version der „Backward Digit Span“ (BDS) entspricht der in Abschnitt 4.1.2 vorgestellten klassischen Variante. Nach dem kurzen Hinweis „Achtung!“ erscheinen auf dem Bildschirm nacheinander einstellige Zahlen. Jede Zahl wird für 1.5 s in der Mitte des Bildschirms präsentiert. Der Pb soll sich die Zahlen merken und sie im Anschluss in genau umgekehrter Reihenfolge in den Computer eingeben. Die Ziffernfolge „1 – 2 – 3 – 4“ beispielsweise lautet in genau umgekehrter Reihenfolge „4 – 3 – 2 – 1“. Die Eingabe ist zeitlich nicht begrenzt. Um die Raterwahrscheinlichkeit zu minimieren, werden die Pbn instruiert, eine Null einzugeben, wenn sie sich an die richtige Ziffer (stets zwischen 1 und 9) nicht erinnern können.

Der Pb erhält für jede komplett richtig bearbeitete Aufgabe einen Punkt. Auf der einfachsten Schwierigkeitsstufe (Level) der BDS-Aufgaben werden vier Ziffern, auf der höchsten Stufe sieben Ziffern in Folge dargeboten (siehe Tabelle 6.4). Vorab bearbeiten die Pbn zwei Übungsaufgaben der Schwierigkeitsstufe 1.

Tabelle 6.4: Struktur der Skala „Backward Digit Span“

| Aufgabe         | Level   | Anzahl Ziffern |
|-----------------|---------|----------------|
| Übungsaufgabe 1 | BDS I   | 4              |
| Übungsaufgabe 2 |         | 4              |
| Aufgabe 1       | BDS I   | 4              |
| Aufgabe 2       |         | 4              |
| Aufgabe 3       | BDS II  | 5              |
| Aufgabe 4       |         | 5              |
| Aufgabe 5       | BDS III | 6              |
| Aufgabe 6       |         | 6              |
| Aufgabe 7       |         | 6              |
| Aufgabe 8       |         | 6              |
| Aufgabe 9       | BDS IV  | 7              |
| Aufgabe 10      |         | 7              |
| Aufgabe 11      |         | 7              |
| Aufgabe 12      |         | 7              |

Anmerkung. Level: BDS I = 4 Ziffern; BDS II = 5 Ziffern; BDS III = 6 Ziffern; BDS IV = 7 Ziffern.

### 6.1.5 Aufgabe „Memory Updating Figural“ (MUF)

Diese Version der Aufgabe „Memory Updating Figural“ (MUF) ist eine abgewandelte und deutlich vereinfachte Version des Aufgabentyps von Oberauer et al. (2000; ursprünglich von Salthouse, Babcock & Shaw, 1991). In ein bis maximal vier aktiven Zellen einer Matrix erscheint jeweils für 1.5 s sukzessive ein Punkt in einer der vier Ecken jeder Zelle (jeweils vier mögliche Positionen, siehe Abbildung 6.3). Anschließend erscheinen in den Zellen nacheinander Pfeile, die in verschiedene Richtungen zeigen können (waagrecht und senkrecht: links, rechts, oben, unten; diagonal: rechts oben, rechts unten, links oben, links unten, siehe Abbildung 6.3).

Der Pb soll nun den vorher erschienenen Punkt innerhalb der Zelle mental in die Position verschieben, in die der Pfeil zeigt (siehe Abbildung 6.4). Anschließend gibt der Pb die neue Position des Punktes in jeder Zelle über den Nummernblock der Tastatur ein. Alle neun Tasten des Nummernblocks wurden mit kleinen weißen Aufklebern beklebt. Die Zifferntasten 1, 3, 7 und 9 (Eckpositionen) wurden zusätzlich mit einem kleinen schwarzen Punkt versehen, diese stellen die möglichen Positionen der Punkte dar. Die Schwierigkeit der Aufgaben wird über die Anzahl von Punkten und über die Anzahl an Operationen (Pfeilen) variiert, die insgesamt in der Aufgabe ausgeführt werden müssen. Der Pb erhält für jede komplett richtig gelöste Aufgabe einen Punkt.

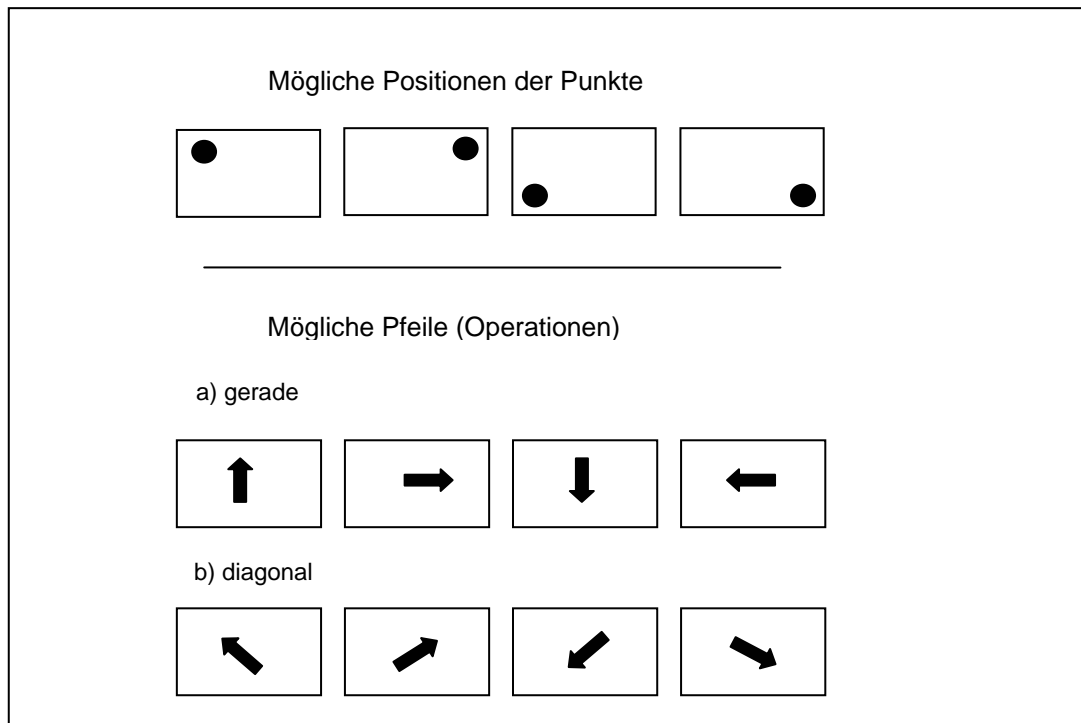


Abbildung 6.3: Mögliche Positionen der Punkte und mögliche Pfeilrichtungen in der Aufgabe „Memory Updating Figural“

Die MUF-Skala enthält drei Schwierigkeitsstufen (Level) mit insgesamt 14 Aufgaben (siehe Tabelle 6.5). Auf der Schwierigkeitsstufe 1 muss die Position zweier Punkte in zwei Zellen behalten werden, und es müssen drei bis vier mentale Operationen vorgenommen werden. Die schwierigste Aufgabe erfordert das Behalten der Position von vier Punkten und die Durchführung von vier mentalen Operationen. Vorab bearbeiten die Pbn sechs Übungsaufgaben. Dieser Aufgabentyp wird ausführlicher geübt, da er etwas komplizierter strukturiert und durch den figuralen Charakter weniger gut verbal vermittelbar ist. Auch sollen die Pbn beim Üben bereits alle Darbietungsformate kennen lernen, insbesondere weil die Darstellung der vier Zellen in Aufgabe 13 und 14 von der der anderen Aufgaben abweicht. Da Aufgaben mit vier Punkten jedoch sehr schwierig sind, enthält die Übungsaufgabe 6 zwar vier Zellen, es erscheinen jedoch nur in zwei dieser Zellen Punkte.

Tabelle 6.5: Struktur der Skala „Memory Updating Figural“

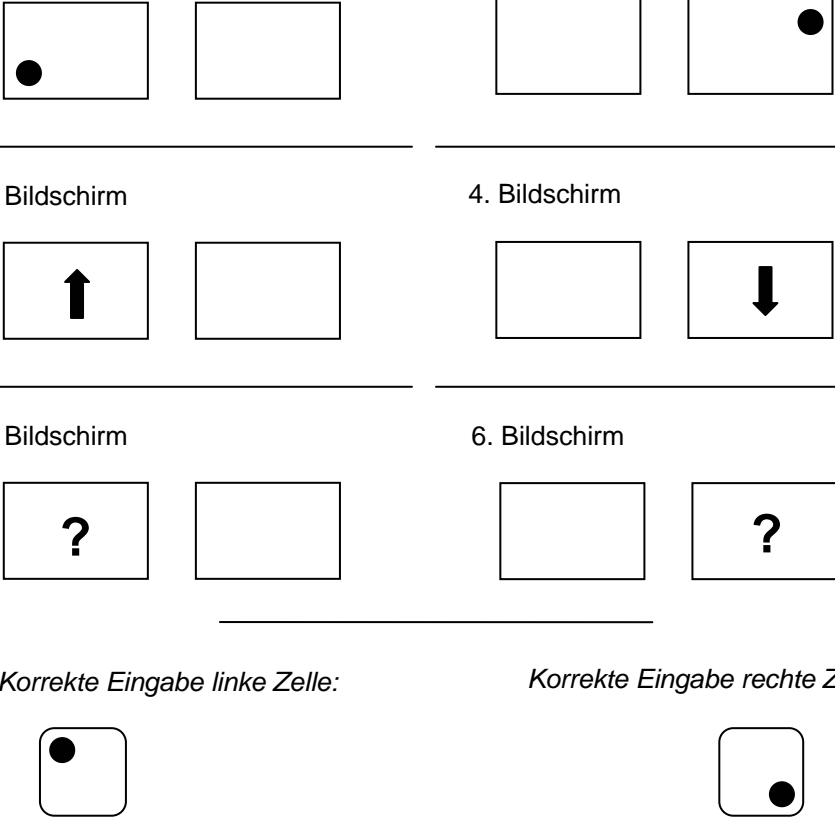
| Aufgabe           | Level    | Anzahl Punkte | Anzahl Pfeile |
|-------------------|----------|---------------|---------------|
| Übungsaufgabe 1*  | MUF 0    | 1             | 1             |
| Übungsaufgabe 2*  | MUF 0    | 1             | 1             |
| Übungsaufgabe 3   | MUF 0    | 2             | 2             |
| Übungsaufgabe 4   | MUF 0    | 2             | 2             |
| Übungsaufgabe 5   | MUF 0    | 3             | 1             |
| Übungsaufgabe 6** | MUF 0    | 2             | 2             |
| Aufgabe 1         | MUF Ia   | 2             | 3             |
| Aufgabe 2         |          | 2             | 3             |
| Aufgabe 3         | MUF Ib   | 2             | 4             |
| Aufgabe 4         |          | 2             | 4             |
| Aufgabe 5         | MUF IIa  | 3             | 2             |
| Aufgabe 6         |          | 3             | 2             |
| Aufgabe 7         | MUF IIb  | 3             | 3             |
| Aufgabe 8         | MUF IIc  | 3             | 4             |
| Aufgabe 9         |          | 3             | 4             |
| Aufgabe 10        |          | 3             | 4             |
| Aufgabe 11        |          | 3             | 4             |
| Aufgabe 12        |          | 3             | 4             |
| Aufgabe 13        | MUF IIIa | 4             | 2             |
| Aufgabe 14        | MUF IIIb | 4             | 4             |

Anmerkung. Level: MUF Ia = 2 Zellen, 3 Operationen; MUF Ib = 2 Zellen, 4 Operationen; MUF IIa = 3 Zellen, 2 Operationen; MUF IIb = 3 Zellen, 3 Operationen; MUF IIc = 3 Zellen, 4 Operationen; MUF IIIa = 4 Zellen, 2 Operationen; MUF IIIb = 4 Zellen, 4 Operationen. \* Darstellung: zwei Zellen, nur in einer davon erscheint ein Punkt; \*\* Darstellung: vier Zellen, nur in zwei davon erscheint ein Punkt.



**Beispielaufgabe „Memory Updating Figural“**

1. Bildschirm 2. Bildschirm



3. Bildschirm 4. Bildschirm

5. Bildschirm 6. Bildschirm

Korrekte Eingabe linke Zelle: Korrekte Eingabe rechte Zelle:

Abbildung 6.4: Übungsaufgabe 3 der Aufgabengruppe „Memory Updating Figural“ mit richtiger Lösung

### 6.1.6 Aufgabe „Spatial Working Memory“ (SWM)

In dieser Aufgabe von Oberauer et al. (2000) werden dem Pb einfache schachbrettartige Muster dargeboten (siehe Abbildung 6.5). Diese Muster soll der Pb je nach Anweisung mental um 90° nach links oder nach rechts rotieren lassen und anschließend das rotierte Muster in den Antwortbogen einzeichnen. Nach einigen Durchgängen mit nur einem Muster werden dann zwei bis maximal vier Muster kurz hintereinander gezeigt. Diese sollen zunächst im Kopf behalten und entsprechend der Anweisung mental rotiert werden. Der Ablauf einer Aufgabe ist folgender: Zunächst erscheint der kurze Hinweis „Achtung!“ in der Mitte des Bildschirms. Danach wird für 3 s eine verbale und graphisch über einen Pfeil verdeutlichte Rotationsanweisung

um 90° nach links oder rechts gezeigt. Anschließend wird das (erste) Muster für 4 s dargeboten. Gegebenenfalls folgen weitere Muster direkt im Anschluss. Erst nach der Aufforderung „Bitte die Ergebnisse eintragen!“ dürfen die Muster dann in die leeren Matrizen auf dem Antwortbogen eingezeichnet werden. Dabei sollen die Muster in derselben Reihenfolge eingezeichnet werden, wie sie dargeboten wurden. Relevant ist dabei die richtige Position auf dem Antwortbogen, nicht zu kontrollieren ist, ob die Pbn die Muster tatsächlich der Reihe nach von links nach rechts eintragen.

Der Pb erhält für jedes an der richtigen Position auf dem Antwortbogen eingezeichnete Muster einen Punkt. Die SWM-Skala enthält vier Schwierigkeitsstufen (Level) mit insgesamt 13 Aufgaben (siehe Tabelle 6.6). Auf der Schwierigkeitsstufe 1 muss lediglich ein Muster mental gedreht werden. Die höchste Schwierigkeitsstufe erfordert das mentale Rotieren und Behalten von vier Mustern. Vorab bearbeiten die Pbn vier Übungsaufgaben der Schwierigkeitsstufen 1 und 2.

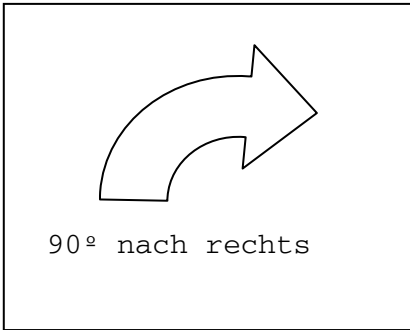
Tabelle 6.6: Struktur der Skala „Spatial Working Memory“

| Aufgabe         | Level   | Anzahl Muster |
|-----------------|---------|---------------|
| Übungsaufgabe 1 | SWM I   | 1             |
| Übungsaufgabe 2 |         | 1             |
| Übungsaufgabe 3 | SWM II  | 2             |
| Übungsaufgabe 4 |         | 2             |
| Aufgabe 1       | SWM I   | 1             |
| Aufgabe 2       |         | 1             |
| Aufgabe 3       |         | 1             |
| Aufgabe 4       | SWM II  | 2             |
| Aufgabe 5       |         | 2             |
| Aufgabe 6       |         | 2             |
| Aufgabe 7       |         | 2             |
| Aufgabe 8       | SWM III | 3             |
| Aufgabe 9       |         | 3             |
| Aufgabe 10      |         | 3             |
| Aufgabe 11      |         | 3             |
| Aufgabe 12      | SWM IV  | 4             |
| Aufgabe 13      |         | 4             |

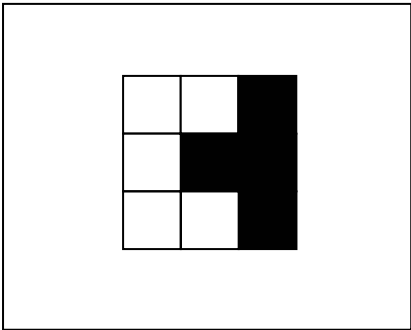
Anmerkung. Level: SWM I = 1 Muster; SWM II = 2 Muster; SWM III = 3 Muster; SWM IV = 4 Muster.

**Beispielaufgabe „Spatial Working Memory“**

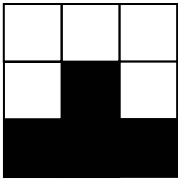
1. Bildschirm: Rotationsanweisung



2. Bildschirm: Präsentation des Musters



*Korrektter Eintrag auf dem Antwortbogen:*



*Abbildung 6.5: Übungsaufgabe 1 der Aufgabengruppe „Spatial Working Memory“ mit richtiger Lösung*

## 6.2 Weitere Materialien

### 6.2.1 Fragebogen

Allen Pbn wurde zu Beginn der Untersuchung ein kurzer Fragebogen vorgelegt (siehe Anhang A). Abgefragt wurden das Geburtsdatum, die Klassenstufe, die Schulform, das Geschlecht, die Muttersprache, die Händigkeit und die Berufe der Eltern. Die Händigkeit wurde erfasst, um kontrollieren zu können, ob die Leistung in den Arbeitsgedächtnisaufgaben durch Links- oder Rechtshändigkeit beeinflusst wird. Die Berufe der Eltern wurden erfragt, um den sozialen Status der Stichproben abschätzen zu können. Jedoch konnten viele Kinder keine hinreichend genauen Angaben zum Beruf ihrer Eltern machen, so dass die Ergebnisse nicht ausgewertet werden. Zusätzlich wurde über vier kurze Fragen die Vorerfahrung der Pbn im Umgang mit Computern erfragt. Weiterhin wurden die Schulnoten des letzten Schulzeugnisses erfasst. In der Sekundarstufe wurde dabei nach allen Noten gefragt, in der Primarstufe lediglich nach den Noten in den Kernfächern Deutsch, Mathematik und Sachkunde.

## 6.2.2 Kognitive Tests

Es wurden mehrere Tests und Aufgaben zur Messung verschiedener kognitiver Bereiche eingesetzt, so dass jeweils über verschiedene Aufgaben oder Skalen die Konstrukte „Intelligenz“, „Geschwindigkeit der Informationsverarbeitung“ und „Kurzzeitgedächtnis“ (bzw. „Merkfähigkeit“) erfasst wurden. Im BIS-Modell von Jäger (1982) entspricht dies den Operationen „Verarbeitungskapazität“, „Bearbeitungsgeschwindigkeit“ und „Merkfähigkeit“. Die Tests dienen zum einen der Messung verschiedener kognitiver Konstrukte, deren Zusammenhänge mit dem Arbeitsgedächtnis untersucht werden sollen. Zum anderen werden sie für die Validierung der Arbeitsgedächtnisaufgaben und -skalen benötigt.

### *Intelligenztests*

Zur Erfassung der fluiden Intelligenz der Pbn wird mit dem *Grundintelligenztest – Skala 2* (CFT 20, Kurzform; Weiß, 1998) die deutsche Version des *Culture Fair Tests* von Cattell (1973) verwendet. Der CFT 20 misst über die vier sprachfreien, figuralen Subtests „Series“, „Classifications“, „Matrices“ und „Topologies“ die fluide Intelligenz. In faktorenanalytischen Studien konnte gezeigt werden, dass der Culture Fair Test hoch auf dem g-Faktor der Intelligenz lädt (Carroll, 1993).

Zusätzlich werden die Zusatzmodule „Zahlenfolgentest“ (CFT-ZF) und „Wortschatztest“ (CFT-WS) des CFT 20 eingesetzt. Die beiden Zusatzmodule zielen auf die Messung kristalliner Aspekte der Intelligenz. Der Wortschatztest erfasst anhand von 30 verbalen Mehrfachwahlaufgaben den über den Grundwortschatz hinausgehenden Wortschatz. Das Ergebnis kann als Ausdruck der Allgemeinbildung eines Pb verstanden werden. Der Zahlenfolgentest enthält 21 Mehrfachwahlaufgaben, die die korrekte Fortsetzung einer gegebenen Zahlenfolge erfordern. Gemessen wird die Fähigkeit, Regeln und Gesetzmäßigkeiten bei einfachen und komplexen numerischen Aufgaben zu erkennen. In der Taxonomie des BIS-Modells (Jäger, 1982; siehe Abschnitt 2.2) erfasst der CFT 20 die figurale Verarbeitungskapazität, der Wortschatztest die verbale, und der Zahlenfolgentest die numerische Verarbeitungskapazität (Weiß, 1998). Es ist anzumerken, dass sich bei der Verwendung dieser Tests zur separaten Messung fluider und kristalliner Intelligenz eine Konfundierung von Domänen bzw. Inhalten und Operationen nicht vermeiden lässt: Die Aufgaben zur Messung der fluiden Fähigkeiten haben alle einen figuralen Inhalt, während die Aufgaben zur Messung kristalliner Fähigkeiten einen verbalen (Wortschatztest) und numerischen (Zahlenfolgentest) Inhalt aufweisen.

### *Aufgaben zur Erfassung der Geschwindigkeit der Informationsverarbeitung*

Zur Erfassung der Geschwindigkeit der Informationsverarbeitung werden insgesamt vier verschiedene Verfahren eingesetzt. Für die Messung solcher Prozesse, die eher komplexe kognitive Anforderungen an die Informationsverarbeitungsgeschwindigkeit stellen, werden (Unter-)Tests verwendet, die in der Intelligenzdiagnostik häufig zur Erfassung von Speed-Komponenten eingesetzt werden. Als standardisierte und

normierte Verfahren werden der Zahlen-Verbindungs-Test (ZVT, Oswald & Roth, 1987) sowie zwei Untertests des HAWIK III, der „Zahlen-Symbol-Test“ und die „Symbolsuche“, verwendet (Tewes et al., 2000). Die beiden HAWIK-Untertests lassen sich zu einem Index für die Arbeitsgeschwindigkeit zusammenfassen (AG-Index). Der ZVT misst die Geschwindigkeit der Informationsverarbeitung über vier Aufgaben, in denen die Zahlen von 1 bis 90, die auf einem Testbogen angeordnet sind, so schnell wie möglich mit einer Linie verbunden werden sollen. Rost und Hanes (1993) konnten zeigen, dass der ZVT auch für Kinder im Grundschulalter ausreichend reliabel ist ( $N = 6008$ ;  $r_{tt} = .88$ ;  $\alpha = .91$ ).

Zur Messung der eher elementaren Informationsverarbeitungsgeschwindigkeit, die ohne die Beteiligung komplexer kognitiver Prozesse erfolgt, werden Papier-und-Bleistift-Aufgaben nach dem Hick-Paradigma eingesetzt. Es werden vier Hick-Aufgaben aus der PAST-Batterie von Roberts, Pallier und Stankov (1999) verwendet, die Instruktionen wurden von der Verfasserin ins Deutsche übersetzt. In der einfachsten Hick-Aufgabe (Hick-0, 0 Bits) sind die Pbn aufgefordert, lediglich einen Pfeil, dessen Spitze nach rechts zeigt, in jedes Kästchen zu zeichnen (über jedem Kästchen steht in dieser einfachsten Version eine 1, siehe Abbildung 6.6).

**Aufgabe Hick-0 (0 Bit):** 1 = →

**Aufgabe Hick-1 (1 Bit):** 1 = → 2 = ←

**Aufgabe Hick-2 (2 Bit):** 1 = ↑ 2 = → 3 = ↓ 4 = ←

**Aufgabe Hick-3 (3 Bit):** 1 = ↑ 2 = ↗ 3 = → 4 = ↘  
5 = ↓ 6 = ↙ 7 = ← 8 = ↖

**Beispielzeile für Hick-0:**

|   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| → | → | → | → | → |   |   |   |   |   |   |   |   |

**Beispielzeile für Hick-3:**

|   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 6 | 8 | 5 | 4 | 3 | 2 | 1 | 5 | 6 | 7 | 4 | 3 |
| ← | ↙ | ↖ | ↓ | ↘ |   |   |   |   |   |   |   |   |

Abbildung 6.6: Hick-Aufgaben 0-3. Darstellung der Stimuli (Zahlen) und geforderten Antworten (Pfeile), Beispielzeilen aus Hick-0 und Hick-3

Die zweite Hick-Aufgabe (Hick-1, 1 Bit) erfordert eine Wahlreaktion: Unter jede 1 ist ein Pfeil nach rechts und unter jede 2 ein Pfeil nach links zu zeichnen. Die Aufgaben Hick-2 und Hick-3 erfordern mit zwei bzw. drei Bits ebenfalls Wahlreaktionen. Jede Hick-Aufgabe wird vorab in einer Zeile geübt, für die eigentliche Aufgabe haben die Pbn 60 s zur Verfügung. Das Testblatt ist so gestaltet, dass nicht alle Kästchen in der vorgegebenen Zeit bearbeitet werden können, Deckeneffekte sind somit auszuschließen. Jeder richtig eingetragene Pfeil wird mit einem Punkt bewertet, falsche Lösungen werden ignoriert.

#### *Aufgaben zur Erfassung des Kurzzeitgedächtnisses bzw. der Merkfähigkeit*

Zur Erfassung der Merkfähigkeit werden drei Aufgaben aus dem *Berliner Intelligenzstruktur-Test für Jugendliche – Begabungs- und Hochbegabungsdiagnostik* (BIS-HB; Jäger et al., 2005) verwendet. Merkfähigkeit wird im BIS-HB als Fähigkeit zum aktiven Einprägen und kurzfristigen Wiedererkennen oder Reproduzieren von verschiedenartigem Material verstanden. Bestimmte Informationen müssen bei diesen Aufgaben auswendig gelernt und unmittelbar im Anschluss an das Lernen als Leistung des Kurzzeitgedächtnisses wiedergegeben werden. Es handelt sich somit um Aufgaben, die lediglich eine passive Speicherung von Informationen erfordern. Der BIS-HB enthält insgesamt neun Merkfähigkeitsaufgaben, je drei mit numerischem, verbalem und figuralem Material. Für die vorliegende Untersuchung wurde je eine Aufgabe pro Inhaltsbereich ausgewählt. Da der BIS-HB lediglich für den Altersbereich von 12;6 bis 16;5 Jahren normiert ist, können für die vorliegende Stichprobe keine standardisierten Werte ermittelt werden. Die Aufgaben werden – bis auf eine Abweichung in der Zeitbegrenzung bei der Aufgabe „Wege erinnern“ (s. u.) – entsprechend den Angaben im BIS-HB durchgeführt.

*Aufgabe „Worte merken“ (WM):* Die Pbn erhalten eine Liste mit 19 allgemein bekannten kurzen Substantiven und haben 40 s Zeit, um sich die Wörter so gut wie möglich einzuprägen. Direkt im Anschluss sollen die Wörter auf der nächsten Seite im Testheft aufgeschrieben werden, die Reihenfolge der Wörter muss dabei nicht beachtet werden. Zum Aufschreiben haben die Pbn 90 s Zeit. Für jedes richtig erinnerte Wort erhält der Pb einen Punkt, falsche Lösungen werden ignoriert.

*Aufgabe „Zweistellige Zahlen“ (ZZ):* Den Pbn wird eine Liste mit 14 zweistelligen Zahlen vorgelegt. Innerhalb von 60 s sind diese einzuprägen, anschließend haben die Pbn 50 s Zeit, die gemerkten Zahlen im Testheft zu notieren. Die Reihenfolge der Zahlen spielt dabei keine Rolle. Jede richtig erinnerte Zahl wird mit einem Punkt bewertet, falsche Lösungen werden ignoriert.

*Aufgabe „Wege erinnern“ (WE):* Den Pbn wird ein stilisierter Stadtplanausschnitt vorgelegt. Dieser enthält die schematische Darstellung von Gebäuden und Straßen (ohne Beschriftungen). Eines der Gebäude ist als „Wohnung“, ein zweites als „Schule“ gekennzeichnet. Zwischen diesen beiden Gebäuden ist ein Weg eingezeichnet, den der Schulbus fährt. Die Pbn haben 30 s Zeit, sich diesen Weg einzuprägen. Anschließend erhalten sie denselben Stadtplanausschnitt, in den sie den eingepprägten

Weg einzeichnen sollen. Für das Einzeichnen sind im BIS-HB 40 s vorgesehen, aufgrund des allgemein langsameren Arbeitstempos von jüngeren Pbn wurde diese Zeit in der vorliegenden Untersuchung auf 60 s erhöht. Die Auswertung erfolgt mittels einer Schablone, es wird jeweils ein Punkt vergeben, wenn der eingezeichnete Weg des Pb bestimmte Markierungen auf der Schablone kreuzt. So erhält der Pb umso mehr Punkte, je öfter er an den richtigen Stellen im Plan richtig „abgebogen“ ist (maximale Punktzahl: 31). Falsche Wege werden ignoriert.

### 6.3 Stichprobe

Mit der Endversion der Testbatterie wurden insgesamt 382 Pbn untersucht. Bei drei Pbn gingen die Testdaten aufgrund von Computerproblemen verloren, ein Schüler hat die Bearbeitung der Aufgaben abgebrochen und vier Pbn nahmen nur an einem der beiden Untersuchungstage teil. Insgesamt konnten somit 374 Fälle verwendet werden.

Es wurde angestrebt, dass sich ein Teil der Stichprobe aus weitgehend unausgelesenen Kindern zusammensetzt, so dass diese Substichprobe möglichst repräsentativ für die Gesamtpopulation der Kinder der Klassenstufen 4 bis 6 ist. Der zweite Teil der Stichprobe sollte möglichst viele hoch begabte Kinder enthalten. Da es definitionsgemäß nur sehr wenig hoch begabte Kinder in der Bevölkerung gibt, wurden für die Rekrutierung hoch begabter Pbn verschiedene Organisationen und Einrichtungen kontaktiert. Es konnten so Kinder aus verschiedenen, auf die Förderung hoch begabter Kinder spezialisierten Einrichtungen und Vereinen untersucht werden. Dabei musste damit gerechnet werden, dass sich nur ein Teil der dort organisierten Kinder als tatsächlich besonders intelligent erweisen würde, weil die Einrichtungen und Vereine häufig keine Intelligenzdiagnostik vornehmen, sondern auf eine Selbstselektion der Kinder bzw. ihrer Familien setzen. Zusätzlich wurde mit der „Mathematik-Schülerakademie“ (s. u.) eine Gruppe in die Studie einbezogen, die zwar keine explizite Hochbegabtengruppe ist, bei der aber eine hohe Quote besonders intelligenter Teilnehmer erwartet werden konnte. Ein weiterer Vorteil der Schülerakademie besteht darin, dass dort 50% der Teilnehmer Mädchen sind. Mädchen sind in den meisten Einrichtungen für hoch Begabte deutlich unterrepräsentiert. Da die untersuchten Gruppen aus Einrichtungen für hoch Begabte vorwiegend auf einer Selbstselektion der Mitglieder basieren und zum Teil auch spezielle Interessen und Begabungen eine Rolle spielen, sind die *einzelnen* Gruppen sicherlich nicht repräsentativ für die Population aller hoch Begabten.

Die Zusammensetzung der einzelnen Substichproben aus relativ unterschiedlichen „Hochbegabtengruppen“ zu einer größeren „Hochbegabtenstichprobe“ ermöglicht jedoch eine Annäherung an eine repräsentative Abbildung der Hochbegabtenpopulation. Diese Teilstichprobe, die sich aus Pbn mehrerer Hochbegabtengruppen zusammensetzt, wird im Folgenden der Einfachheit halber als „Hochbegabtenstichprobe“ bezeichnet, auch wenn nur ein Teil der Pbn die Kriterien für eine Hochbegabung tatsächlich erfüllt. Einen Überblick über die Probandenzahlen der einzelnen

Substichproben der unausgelesenen Stichprobe sowie der Hochbegabtenstichprobe gibt Tabelle 6.7 in Abschnitt 6.3.3.

### 6.3.1 Stichprobe von Schülerinnen und Schülern aus regulären Schulen

Es wurden Schülerinnen und Schüler aus zwei vierten Klassen einer Grundschule in Steinfurt (bei Münster, Nordrhein-Westfalen) und Schülerinnen und Schüler der Klassenstufen 5 und 6 einer Orientierungsstufe in Belm (bei Osnabrück, Niedersachsen) getestet. Die Orientierungsstufe war zum Untersuchungszeitpunkt für die Schülerinnen und Schüler der Klassenstufen 5 und 6 in Niedersachsen die Regelschule. Es wurde für die Untersuchung deshalb eine Orientierungsstufe ausgewählt, weil hier – anders als in anderen Bundesländern – in den Klassenstufen 5 und 6 noch keine Separierung nach Leistungen in verschiedene Schultypen erfolgt. Es wurde daher erwartet, dass es sich um eine Schülergruppe handelt, deren intellektuelle Fähigkeiten und schulische Leistungen breit verteilt sind. Insgesamt wurden die Schülerinnen und Schüler aus drei fünften und drei sechsten Klassen getestet. Bis auf einzelne fehlende Kinder nahmen in der Grundschule und in der Orientierungsstufe vollständige Klassen an der Untersuchung teil (Gründe für eine Nichtteilnahme waren Krankheit, keine ausreichenden Kenntnisse der deutschen Sprache und keine Einwilligung durch die Eltern).

Für die drei untersuchten sechsten Klassen liegen so genannte Trendaussagen der Schule vor, in denen nach dem ersten Halbjahr aufgrund der bis dahin gezeigten Leistungen der Schülerinnen und Schüler prognostiziert wird, wie vielen Kindern nach dem Ende der sechsten Klasse der Besuch einer Hauptschule, einer Realschule oder eines Gymnasiums empfohlen wird. Von den insgesamt 67 Schülerinnen und Schülern der drei untersuchten sechsten Klassen erhielten 43% eine Prognose für die Hauptschule, 40% für die Realschule und 16% für das Gymnasium. Nach der Einschätzung verschiedener Lehrkräfte deuten diese Trendaussagen darauf hin, dass es sich um eine leistungsmäßig relativ typische Schülerschaft einer Orientierungsstufe handelt. Diese Quoten weichen zwar deutlich von dem Anteil von Schülerinnen und Schülern ab, die in der siebten Klasse tatsächlich in den jeweiligen Schulformen zu finden sind (es besuchen anteilig mehr Schülerinnen und Schüler das Gymnasium und weniger die Hauptschule). Dies liegt jedoch daran, dass die Empfehlungen der Schule für die Entscheidung der Eltern nicht bindend sind.

### 6.3.2 Stichprobe von Schülerinnen und Schülern aus Einrichtungen und Organisationen für hoch Begabte

*Mitglieder des Vereins „Deutsche Gesellschaft für das hochbegabte Kind“ (DGhK):* Die DGhK ist ein bundesweit vertretener Verein, in dem vor allem Eltern hoch begabter Kinder organisiert sind. Die Mitgliedschaft in der DGhK ist nicht an eine nachgewiesene hohe Begabung des Kindes gebunden. Die Mitglieder wurden über Ankündigungen in regionalen Rundschreiben des Vereins auf die Studie aufmerksam



gemacht. Insgesamt nahmen 14 Kinder des Regionalverbands Hannover und 21 Kinder aus Regionalverbänden in NRW an der Studie teil.

*Mitglieder des Vereins „Hochbegabtenförderung e.V.“ (HBF e.V.):* Die HBF e.V. ist ein Verein für die Eltern hoch begabter Kinder, der in mehreren Städten Deutschlands vertreten ist und dort Kurse für hoch begabte Kinder anbietet. Der Verein setzt für eine Mitgliedschaft bzw. eine Teilnahme des Kindes am Kursprogramm einen IQ von mindestens 120 voraus, der über ein testpsychologisches Gutachten nachzuweisen ist. Die Mitglieder wurden über ein Rundschreiben des Vereins auf die Studie aufmerksam gemacht. Insgesamt nahmen 26 Kinder des Vereins an der Untersuchung teil.

*„Mathetreff“ der Universität Münster:* Die Pbn waren Teilnehmer/innen am „Mathetreff“ der Universität Münster, einem wöchentlichen Kursprogramm für begabte und interessierte Kinder im Grundschulalter. Die Kurse werden im Rahmen der Erforschung von Denkprozessen besonders begabter Kinder vom Institut für Mathematikdidaktik der Universität Münster angeboten. Einige der Kinder nahmen aufgrund von Empfehlungen ihrer Lehrer teil, für manche der Kinder lagen bereits Hinweise auf eine hohe Begabung durch eine psychologische Diagnostik vor. Die übrigen Kinder wurden aufgrund der Einschätzungen der Eignung für eine Teilnahme durch die Eltern angemeldet. Eine nachgewiesene hohe Begabung ist keine Voraussetzung für eine Teilnahme an einem der Kurse. Aus diesen Kursen nahmen insgesamt 14 Kinder an der Untersuchung teil.

*Mathematik-Schülerakademie der Bezirksregierung Münster:* Zur Mathematik-Schülerakademie der Bezirksregierung werden in jedem Jahr die besten Schülerinnen und Schüler in Mathematik der Klassen 5 und 6 des Regierungsbezirks Münster eingeladen. Die Schulen nominieren je eine Teilnehmerin und einen Teilnehmer. Bei den Schülerakademien der Jahre 2002 und 2003 erhielten die Teilnehmer/innen eine Einladung für die Studie. Die Teilnehmer/innen des Jahres 2002 wurden im Frühjahr 2003 getestet, zu diesem Zeitpunkt besuchte ein Teil dieser Gruppe bereits die siebte Klasse. Die Teilnehmerinnen und Teilnehmer des Jahres 2003 wurden im Herbst 2003 untersucht. Insgesamt nahmen 34 Teilnehmerinnen und Teilnehmer der Schülerakademie an der Studie teil. Zusätzlich wurden insgesamt vier jüngere Geschwister von Teilnehmerinnen und Teilnehmern untersucht, da die Eltern bei diesen einen Verdacht auf eine ebenfalls sehr hohe intellektuelle Begabung äußerten. Diese vier Pbn besuchten die vierte Klassenstufe.

*Schülerinnen und Schüler der Christophorusschule des Christlichen Jugenddorfwerks in Braunschweig (CJD Braunschweig):* Das CJD Braunschweig unterhält ein privates Gymnasium und eine private Grundschule, die auf die Förderung intellektuell hoch begabter Schülerinnen und Schüler spezialisiert sind. Während die hoch Begabten in den höheren Klassen des Gymnasiums separat unterrichtet werden, findet die Förderung in der Grundschule und in der Unterstufe des Gymnasiums in integrierten Klassen statt. Da für diese Klassen keine systematische Auswahl der Schülerinnen und Schüler über die Intelligenz erfolgt, ist nicht bekannt, welche und wie viele der Schülerinnen und Schüler intellektuell hoch begabt sind. Durch die Ausrichtung der Schule auf die Förderung hoch begabter Kinder konnte erwartet

werden, unter den Schülerinnen und Schülern der Schule eine relativ hohe Quote besonders intelligenter Kinder zu finden. Daher wurden ganze Klassen des CJD Braunschweig (zwei vierte, eine fünfte und zwei sechste Klassen) in die Studie mit einbezogen. Anders als in den Schulen in Steinfurt und Belm gab es hier – insbesondere in den beiden sechsten Klassen – einzelne Schülerinnen und Schüler, die nicht an der Untersuchung teilnehmen wollten. Dies ist vermutlich darauf zurückzuführen, dass die Themen Intelligenz und Begabung am CJD Braunschweig präsenter sind als an anderen Schulen und einige Schülerinnen und Schüler keine weitere Auseinandersetzung mit den Themen wünschten bzw. bereits eine bewusstere Auseinandersetzung stattgefunden und zu einer Ablehnung von Intelligenzmessungen geführt hatte.

### 6.3.3 Überblick über die untersuchten Gruppen

Die Tabellen 6.7 und 6.8 geben einen Überblick über die Verteilung der einzelnen Teilstichproben und der Gesamtstichprobe auf verschiedene Altersgruppen, Klassenstufen und den Anteil von Mädchen und Jungen. In den Tabellen 6.9 und 6.10 sind die Mittelwerte und Standardabweichungen der einzelnen Gruppen in den verschiedenen eingesetzten kognitiven Tests wiedergegeben.

Das Alter der Pbn streut von 8;6 bis 13;11 Jahren und liegt im Mittel bei 11.4 Jahren ( $SD = 1.12$ ). Die Rekrutierung der Pbn aus regulären Schulen erfolgte über die jeweiligen Schulen und die Kinder nahmen klassenweise an der Untersuchung teil. Es wurden Schülerinnen und Schüler der Klassenstufen 4 bis 6 untersucht. Da die Testungen am Ende des Schuljahres stattfanden, waren in den vierten Klassen die meisten Kinder bereits mindestens zehn Jahre alt. Insgesamt ist daher nur ein relativ geringer Teil der Stichprobe jünger als zehn Jahre; der größte Teil der Stichprobe setzt sich aus Kindern zwischen zehn und zwölf Jahren zusammen (siehe Tabelle 6.7). In der Gesamtstichprobe sind mehr Jungen als Mädchen enthalten. Dies ist darauf zurückzuführen, dass in den meisten der in die Untersuchung einbezogenen Einrichtungen für hoch Begabte Jungen deutlich überrepräsentiert sind.

Tabelle 6.7: Anzahl der Pbn pro Altersgruppe (8;6 bis 13;11 Jahre) und der Anteil von Mädchen und Jungen in den Substichproben sowie in der Gesamtstichprobe

| Gruppen                                       | 8;6 –<br>9;11  | 10;0 –<br>10;11 | 11;0 –<br>11;11 | 12;0 –<br>12;11 | 13;0 –<br>13;11 | ♀               | ♂               |
|---|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| <i>Pbn aus regulären Schulen</i>              |                |                 |                 |                 |                 |                 |                 |
| Grundschule                                   | 4              | 43              | 3               | -               | -               | 26              | 24              |
| Orientierungsstufe                            | -              | 11              | 46              | 52              | 13              | 58              | 64              |
| Regul. Schul. ges.                            | 4              | 54              | 49              | 52              | 13              | 84              | 88              |
| <i>Pbn aus Einrichtungen für hoch Begabte</i> |                |                 |                 |                 |                 |                 |                 |
| DGhK  | 11             | 8               | 6               | 8               | 3               | 16              | 20              |
| HBf e.V.                                      | -              | 4               | 11              | 6               | 5               | 9               | 17              |
| Mathetreff                                    | 5              | 9               | -               | -               | -               | 5               | 9               |
| Schülerakademie                               | 3 <sup>1</sup> | 1 <sup>1</sup>  | 10              | 12              | 12              | 16              | 22              |
| CJD Braunschweig                              | 11             | 28              | 28              | 19              | 2               | 38 <sup>2</sup> | 49 <sup>2</sup> |
| Hoch Begabte ges.                             | 30             | 59              | 49              | 46              | 18              | 84              | 117             |
| Gesamtstichprobe                              | 34             | 113             | 98              | 98              | 31              | 168             | 205             |

Anmerkung. <sup>1</sup> = jüngere Geschwister der Akademieteilnehmer/innen; <sup>2</sup>: eine fehlende Angabe beim Geschlecht.

Tabelle 6.8: Anzahl der Pbn pro Klassenstufe nach Substichproben

| Gruppen                                       | 3. Kl. | 4. Kl.         | 5. Kl. | 6. Kl. | 7. Kl. | 8. Kl. | Σ   |
|---|--------|----------------|--------|--------|--------|--------|-----|
| <i>Pbn aus regulären Schulen</i>              |        |                |        |        |        |        |     |
| Grundschule                                   | -      | 50             | -      | -      | -      | -      | 50  |
| Orientierungsstufe                            | -      | -              | 62     | 60     | -      | -      | 122 |
| Regul. Schul. ges.                            | -      | 50             | 62     | 60     | -      | -      | 172 |
| <i>Pbn aus Einrichtungen für hoch Begabte</i> |        |                |        |        |        |        |     |
| DGhK  | 5      | 5              | 12     | 4      | 7      | 3      | 36  |
| HBf e.V.                                      | -      | -              | 4      | 11     | 6      | 5      | 26  |
| Mathetreff                                    | 1      | 13             | -      | -      | -      | -      | 14  |
| Schülerakademie                               | -      | 4 <sup>1</sup> | -      | 18     | 16     | -      | 38  |
| CJD Braunschweig                              | -      | 36             | 22     | 30     | -      | -      | 88  |
| Hoch Begabte ges.                             | 6      | 58             | 38     | 63     | 29     | 8      | 202 |
| Gesamtstichprobe                              | 6      | 108            | 100    | 123    | 29     | 8      | 374 |

Anmerkung. <sup>1</sup> = jüngere Geschwister der Akademieteilnehmer/innen.

Tabelle 6.9: Mittelwerte und Standardabweichungen der IQ-Werte der Intelligenztests für die Substichproben und die Gesamtstichprobe

| Gruppen                                       | CFT 20 IQ |           | CFT-WS IQ |           | CFT-ZF IQ |           |
|---|-----------|-----------|-----------|-----------|-----------|-----------|
|   | <i>M</i>  | <i>SD</i> | <i>M</i>  | <i>SD</i> | <i>M</i>  | <i>SD</i> |
| <i>Pbn aus regulären Schulen</i>              |           |           |           |           |           |           |
| Grundschule                                   | 110.20    | 13.80     | 104.88    | 14.84     | 107.16    | 14.51     |
| Orientierungsstufe                            | 109.57    | 12.71     | 98.92     | 12.15     | 104.45    | 12.33     |
| Regul. Schul. ges.                            | 109.76    | 13.00     | 100.65    | 13.23     | 105.24    | 13.01     |
| <i>Pbn aus Einrichtungen für hoch Begabte</i> |           |           |           |           |           |           |
| DGhK  | 125.94    | 15.21     | 118.25    | 13.09     | 120.58    | 16.40     |
| HBF e.V.                                      | 132.46    | 9.96      | 118.54    | 9.28      | 124.19    | 10.10     |
| Mathetreff                                    | 122.43    | 14.46     | 114.86    | 8.66      | 124.64    | 16.04     |
| Schülerakademie <sup>1</sup>                  | 124.94    | 12.16     | 118.62    | 9.86      | 122.62    | 10.20     |
| CJD Braunschweig                              | 118.76    | 12.85     | 113.36    | 12.12     | 113.74    | 14.25     |
| Hoch Begabte ges.                             | 123.04    | 13.69     | 115.97    | 11.65     | 118.62    | 14.31     |
| Gesamtstichprobe                              | 116.93    | 14.91     | 108.91    | 14.56     | 112.45    | 15.25     |

Anmerkung. CFT 20: Grundintelligenztest Skala 2; CFT-WS: Modul „Wortschatztest“ des CFT 20; CFT-ZF: Modul „Zahlenfolgen“ des CFT 20. <sup>1</sup> = Teilnehmer/innen an der Akademie ohne jüngere Geschwister ( $N = 34$ ).

Tabelle 6.10: Mittelwerte und Standardabweichungen der IQ-Werte des AG-Index und des ZVT für die Substichproben und die Gesamtstichprobe

| Gruppen                                       | AG-Index IQ |           | ZVT IQ   |           |
|---|-------------|-----------|----------|-----------|
|   | <i>M</i>    | <i>SD</i> | <i>M</i> | <i>SD</i> |
| <i>Pbn aus regulären Schulen</i>              |             |           |          |           |
| Grundschule                                   | 99.48       | 13.14     | 104.36   | 12.80     |
| Orientierungsstufe                            | 103.11      | 13.30     | 109.60   | 13.51     |
| Regul. Schul. ges.                            | 102.07      | 13.32     | 108.08   | 13.48     |
| <i>Pbn aus Einrichtungen für hoch Begabte</i> |             |           |          |           |
| DGhK  | 115.91      | 17.60     | 118.64   | 15.00     |
| HBF e.V.                                      | 123.38      | 15.94     | 119.69   | 13.64     |
| Mathetreff                                    | 113.21      | 12.94     | 117.00   | 7.98      |
| Schülerakademie <sup>1</sup>                  | 118.24      | 13.53     | 121.74   | 11.63     |
| CJD Braunschweig                              | 112.02      | 15.22     | 120.61   | 13.03     |
| Hoch Begabte ges.                             | 115.07      | 15.75     | 119.91   | 12.90     |
| Gesamtstichprobe                              | 109.13      | 16.04     | 114.47   | 14.42     |

Anmerkung. AG-Index: Index für die Arbeitsgeschwindigkeit aus dem HAWIK III; ZVT: Zahlen-Verbindungs-Test. <sup>1</sup> = Teilnehmer/innen an der Akademie ohne jüngere Geschwister ( $N = 34$ ).

*Bildung einer „Hochintelligenten-Stichprobe“*

Es zeigt sich, dass nur ein Teil der Pbn aus den „Hochbegabtengruppen“ auch definitionsgemäß hoch begabt ist, wenn die Messung mit dem CFT 20 zugrunde gelegt wird (vgl. Tabellen 6.9 und 6.10). Klassischerweise wird zur Definition intellektueller Hochbegabung ein Kriterium festgelegt, das einen IQ von mindestens 130 voraussetzt (siehe Abschnitt 2.1). Intelligenz ist jedoch ein kontinuierlich verteiltes Merkmal, so dass das Kriterium eines IQ von 130 lediglich eine willkürliche Setzung darstellt.

37% der Pbn der „Hochbegabtenstichprobe“ erreichen nur einen  $IQ < 120$ . 35% erzielen einen IQ im Bereich 120-129 und 29% einen  $IQ \geq 130$ . Dabei tendiert der CFT 20 vermutlich eher zu einer leichten Überschätzung der Intelligenz, was bei der Betrachtung der Mittelwerte des IQ im CFT 20 in den unausgelesenen Gruppen deutlich wird: Sowohl in der Grundschule als auch in der Orientierungsstufe liegt die mittlere Intelligenz im CFT 20 deutlich über 100. Es ist jedoch unwahrscheinlich, dass es sich bei den beiden Gruppen um überdurchschnittlich intelligente Pbn handelt (siehe die Trendergebnisse für die Schülerinnen und Schüler der sechsten Klassen). Andererseits enthält auch die Gruppe mit den Pbn aus regulären Schulen einige Schülerinnen und Schüler, die über einen hohen IQ verfügen. 23% dieser Pbn haben einen IQ im CFT 20 von mindestens 120 und 6% haben einen IQ von mindestens 130.

Somit sind in der Hochbegabtenstichprobe einige Pbn enthalten, deren mit dem CFT 20 gemessene Intelligenz im durchschnittlichen Bereich liegt und die mit hoher Wahrscheinlichkeit nicht intellektuell hoch begabt sind. Auf der anderen Seite finden sich auch in den regulären Schulen erwartungsgemäß einige intellektuell sehr begabte Kinder. Es scheint somit nicht sinnvoll, die Unterschiede zwischen allen Pbn aus Hochbegabtengruppen und solchen Pbn aus regulären Schulen zu analysieren, da dies tatsächlich vorhandene Unterschiede in fähigkeitsbezogenen Maßen wie den Arbeitsgedächtnismaßen unterschätzen würde. Eine Erstellung einer weiteren Substichprobe nach den Ergebnissen im CFT 20 ist daher erforderlich.

Das Festhalten an dem strengen Kriterium eines IQ von mindestens 130 würde in der vorliegenden Untersuchung zu einer zu kleinen Stichprobengröße führen, daher wird das Intelligenzkriterium etwas gelockert. Es wird eine Substichprobe erstellt, in der alle Kinder enthalten sind, die im CFT 20 einen IQ von mindestens 120 aufweisen und somit nachgewiesenermaßen deutlich überdurchschnittlich intelligent sind. Dieser Wert liegt zwar unter dem klassischen Kriterium für intellektuelle Hochbegabung von 130, es ist jedoch davon auszugehen, dass es sich um eine zumindest deutlich überdurchschnittlich intelligente Gruppe handelt. Diese Substichprobe wird im Folgenden als Gruppe der „hoch Intelligenten“ (HI) bezeichnet. Die Kinder, die dieses Intelligenzkriterium nicht erreichen, bilden entsprechend die Gruppe der „nicht hoch Intelligenten“ (NHI).

Für einige der Analysen ist zusätzlich eine Betrachtung der Substichprobe mit allen Kindern aus den regulären Schulen sinnvoll. Es handelt sich hierbei um eine Gruppe weitgehend unausgelesener Kinder mit breit gestreuten intellektuellen Fähigkeiten.

Insbesondere für die testtheoretischen Analysen ist es sinnvoll, die Eignung der Testaufgaben – zum Beispiel im Hinblick auf die Schwierigkeit und Trennschärfe – nicht nur für eine stark selektierte, sehr intelligente Gruppe, sondern auch für eine relativ repräsentative Gruppe von Schülerinnen und Schülern aller Begabungsstufen zu betrachten. Diese Stichprobe wird im Folgenden auch als die Gruppe der „Unausgelesenen“ (UA) bezeichnet.

Insgesamt werden in den statistischen Analysen in Kapitel 7 je nach Fragestellung Ergebnisse für die folgenden Teilstichproben berichtet:

- Hoch Intelligente (HI; Pbn der Gesamtstichprobe mit einem IQ im CFT 20 von mindestens 120;  $N = 170$ );
- nicht hoch Intelligente (NHI; Pbn der Gesamtstichprobe mit einem IQ im CFT 20 von  $< 120$ ;  $N = 204$ );
- Unausgelesene (UA; Pbn von regulären Schulen,  $N = 172$ ).

### *Muttersprache*

37 Pbn (10%) geben eine andere Sprache als Deutsch als Muttersprache an. Die Frage, seit wie vielen Jahren sie bereits Deutsch sprechen, beantworten insgesamt 35 Pbn. Im Mittel sprachen die Pbn mit anderer Muttersprache zum Testzeitpunkt bereits seit 7.75 Jahren Deutsch ( $SD = 3.17$ ). 15 dieser Pbn geben ihr Lebensalter bzw. ihr Lebensalter abzüglich ein bis zwei Jahre an. (Diese Angabe machen auch einige Kinder mit Deutsch als Muttersprache, um auszudrücken, dass sie „schon immer“ Deutsch sprechen.) Bei diesen 15 Pbn ist somit davon auszugehen, dass sie bereits zweisprachig – mit Deutsch und ihrer Muttersprache – aufgewachsen sind.

## **6.4 Durchführung der Untersuchung**

Die Testungen an den Schulen, der Pbn der DGhK, der Teilnehmerinnen und Teilnehmer an der Mathe-Schülerakademie 2002 und des Mathetreffs wurden zunächst im Zeitraum von den Oster- bis zu den Sommerferien 2003, das heißt von März bis Juli 2003, durchgeführt. Bei einer ersten Sichtung der Daten stellte sich heraus, dass in der bis dahin untersuchten Stichprobe deutlich weniger hoch intelligente Pbn waren als erwartet. Daher wurden im Herbst 2003 zusätzlich Kinder des Vereins Hochbegabtenförderung e.V. und die Teilnehmerinnen und Teilnehmer der Mathe-Schülerakademie 2003 getestet.

Alle Untersuchungen wurden vormittags durchgeführt. Die gesamte Testung wurde auf zwei Testtermine à ca. drei Stunden verteilt, am ersten Termin wurde der Fragebogen ausgefüllt und es wurden die Paper-Pencil-Tests bearbeitet. Am zweiten Termin wurden die Arbeitsgedächtnisaufgaben durchgeführt. Die Retestung der Hick-Aufgaben mit 140 Pbn fand am zweiten Untersuchungstermin im Anschluss an die Arbeitsgedächtnis-Testung statt. Die Testsitzungen wurden jeweils durch kurze Pausen à 10-20 min unterbrochen. Beide Untersuchungstermine fanden im Abstand von

ein bis sieben Tagen statt. Die Reihenfolge der Aufgaben an beiden Testterminen ist Abbildung 6.7 zu entnehmen.

| <b>1. Testtermin:</b><br><b>Kognitive Tests</b><br>(Paper-Pencil-Format) | <b>2. Testtermin:</b><br><b>Arbeitsgedächtnis</b><br>(computergestützt) |
|--|---|
| Fragebogen<br>CFT 20<br>CFT-ZF   | Reading Span<br>Computation Span  |
| <i>Pause</i>   | <i>Pause</i>  |
| CFT-WS<br>AG-Aufgaben (HAWIK III)<br>Hick-Aufgaben                       | Memory Updating Figural<br>Verbal Span                                  |
| ZVT<br>Merkaufgaben  | <i>Pause</i>  |
|  | Backward Digit Span<br>Spatial Working Memory                           |
|  | (ggf. <i>Pause</i> + Hick-Retest)                                       |

Abbildung 6.7: Ablauf der Untersuchung

Die Testungen in den drei Schulen (Grundschule, Orientierungsstufe und CJD Braunschweig) fanden während der Unterrichtszeit statt. Der erste Teil der Untersuchung wurde im Klassenraum durchgeführt, der zweite Teil der Untersuchung im Computerraum der Schule. Für den zweiten Teil wurden die Grundschulklassen in zwei Hälften (à 10-14 Pbn) geteilt und an zwei verschiedenen Terminen getestet, um die Gruppengrößen bei der Testung des Arbeitsgedächtnisses überschaubarer zu halten. In der Orientierungsstufe und im CJD Braunschweig konnte jeweils eine ganze Klasse auf zwei benachbarte Räume aufgeteilt werden, so dass hier alle Schülerinnen und Schüler einer Klasse zeitgleich getestet werden konnten. Die Computerräume wurden so umgeräumt, dass es den Pbn nicht möglich war, von ihrem Sitzplatz aus auf den Monitor eines anderen Pbn zu sehen. Die Testungen der Pbn aus den beiden Hochbegabtenvereinen, dem Mathetreff und der Schülerakademie fanden vormittags an Wochenenden und in den Schulferien statt. Bis auf die Testung der Pbn der DGhK Hannover, die in den Räumlichkeiten des CJD Hannover durchgeführt wurde, fanden die Untersuchungen in der Universität Münster statt. Die Pbn aus den Hochbegabtengruppen wurden in Gruppen getestet, die in der Regel 8-15 Kinder umfassten.

In den Schulen waren bei den Paper-Pencil-Testungen stets zwei Versuchsleiter, bei den Arbeitsgedächtnis-Testungen drei Versuchsleiter pro Testung beteiligt. Bei den weiteren Untersuchungen waren – je nach Gruppengröße – ebenfalls stets zwei bis drei Versuchsleiter anwesend. Der erste Versuchsleiter gab die Instruktionen zu den einzelnen Aufgaben und beantwortete Verständnisfragen. Während der Aufgabenbearbeitung war es die Aufgabe aller Versuchsleiter, individuelle Fragen der Pbn zu beantworten und eine instruktionsgemäße Bearbeitung der Aufgaben zu kontrollieren.

Die Eltern der Pbn wurden mit einem Informationstext über die Studie aufgeklärt und mussten schriftlich ihr Einverständnis erklären, dass die Kinder an der Untersuchung teilnehmen durften. Alle Pbn konnten im Fragebogen angeben, ob sie eine Rückmeldung über ihre Ergebnisse wünschten. Dieses Angebot wurde von der Mehrzahl der Pbn in Anspruch genommen. Die Pbn erhielten eine schriftliche Rückmeldung darüber, wie viele Aufgaben sie jeweils in den einzelnen Arbeitsgedächtnis-Aufgabengruppen richtig gelöst haben. Eine Rückmeldung standardisierter Ergebnisse aus den Intelligenztests an die Pbn in den Schulen erfolgte nicht. Die Eltern der Pbn aus den Hochbegabtengruppen erhielten jedoch auf Wunsch die standardisierten Ergebnisse ihres Kindes in den Intelligenztests.



## 7. Ergebnisse

Die Darstellung der Ergebnisse gliedert sich wie folgt: Zunächst werden die deskriptiven Ergebnisse aus dem Fragebogen, den verschiedenen kognitiven Tests sowie den sechs Arbeitsgedächtnisaufgaben dargestellt (Abschnitt 7.1). Zwei verschiedene Scoringmethoden für die vorliegenden Testdaten werden im Hinblick auf deskriptive Daten miteinander verglichen. Der Abschnitt 7.2 widmet sich verschiedenen Strukturanalysen. Zunächst wird die Binnenstruktur des Arbeitsgedächtnisses mit den vorliegenden sechs Aufgabentypen exploriert und überprüft. Anschließend werden Zusammenhänge zwischen den Arbeitsgedächtnisfaktoren und anderen kognitiven Fähigkeiten berichtet (Abschnitt 7.3). Im darauf folgenden Abschnitt 7.4 werden die Zusammenhänge zwischen den Leistungen in den Arbeitsgedächtnisaufgaben und den Schulleistungen in der Primar- und Sekundarstufe dargestellt.

In Abschnitt 7.5 folgen die testtheoretischen Analysen. Zunächst wird die Dimensionalität der insgesamt 70 Arbeitsgedächtnisaufgaben aller sechs Aufgabengruppen geprüft. Es werden dann die Itemstatistiken entsprechend der Klassischen Testtheorie und der Item Response Theorie (IRT) für alle Aufgaben bestimmt, und es wird über verschiedene Modellgeltungstests berichtet. Im Anschluss daran werden Befunde zur Reliabilität und Validität der modellkonformen Testskalen dargelegt. Abschließend werden die testtheoretischen Analysen noch einmal zusammengefasst.

### 7.1 Deskriptive Ergebnisse und Vergleich von Scorings

Trotz großer Sorgfalt bei der Datenerhebung konnte das Auftreten einzelner fehlender Werte nicht verhindert werden. Es fehlen jedoch maximal 3.5% der Fragebogendaten (= 13 Fälle) zu einer Variablen, 1.3% bei den Intelligenz- und Speed-Test-Daten (= 5 Fälle) und 0.8% (= 3 Fälle) einzelner Daten aus den Arbeitsgedächtnisaufgaben. Diese Raten liegen deutlich unter dem Richtwert von maximal 10% fehlender Werte, wie sie etwa für Strukturgleichungsmodelle angegeben werden (Kline, 1998). Für die ganz überwiegende Zahl der Variablen liegen vollständige Daten vor. Fehlende Werte resultieren zum einen aus unvollständigen Angaben der Pbn im Fragebogen (z. B. konnten sich einige Pbn nicht an alle ihre letzten Schulnoten erinnern; einige besuchten Schulen, in denen keine Noten vergeben wurden) und zum anderen aus einzelnen nicht bearbeiteten Testaufgaben oder Items. In einigen Fällen gingen die Daten einzelner Testaufgaben durch Computerprobleme verloren.

### 7.1.1 Fragebogen

#### *Vorerfahrung im Umgang mit Computern und Händigkeit*

Im Fragebogen wurde die Erfahrung der Pbn im Umgang mit Computern über vier Fragen abgefragt (mit ja/nein-Antwortmöglichkeit). Ziel dieser Befragung war es zu prüfen, ob die Pbn ein Mindestmaß an Vorerfahrung mit Computern aufweisen, um solche Effekte auf die Leistungen auszuschließen, die darauf zurückgehen, dass die Bedienung eines PC für die Kinder völlig neu ist. Die Ergebnisse zeigen, dass die meisten Pbn zumindest über erste Erfahrungen am PC verfügen. 96% gaben an, manchmal Computerspiele zu spielen, 50% antworteten mit „Ja“ auf die Frage, bereits einmal eine E-Mail geschrieben zu haben und 90% berichteten, schon einmal einen Text an einem PC geschrieben zu haben. Insgesamt 84% gaben an, bereits einmal mit der Klasse im Computerraum der Schule gewesen zu sein. 82% der Pbn bejahten drei oder vier der Fragen. Lediglich zwei Pbn (0.5%) verneinten alle vier Fragen zur PC-Erfahrung.

Die 36 Linkshänder der Stichprobe erreichen in allen sechs Aufgabengruppen im Mittel höhere Werte als die Rechtshänder. Für die Arbeitsgedächtnisaufgaben „Reading Span“, „Verbal Span“, „Backward Digit Span“ und „Spatial Working Memory“ werden die Mittelwertsunterschiede im t-Test (mit zweiseitiger Testung) signifikant ( $p < .05$ ). Eine Benachteiligung der Linkshänder in den Arbeitsgedächtnisaufgaben kann somit ausgeschlossen werden.

#### *Schulnoten*

Die Pbn wurden im Fragebogen gebeten, ihre Schulnoten des letzten Zeugnisses anzugeben. Bei den Schülerinnen und Schülern der Primarstufe wurde nur nach den Kernfächern Deutsch, Mathematik und Sachkunde gefragt, bei den Schülerinnen und Schülern der Sekundarstufe nach allen Noten. Berichtet werden hier die Noten in den sprachlichen und mathematisch-naturwissenschaftlichen Fächern (siehe Tabelle 7.1).

### 7.1.2 Kognitive Tests

Die Mittelwerte der IQ der standardisierten kognitiven Tests liegen für die Gesamtstichprobe erwartungsgemäß über dem Normmittelwert von 100 (siehe Tabelle 7.2). Bis auf den CFT 20 weichen die Verteilungen der Testscores signifikant von einer Normalverteilung ab. Dies ist auf die Zusammensetzung der Gesamtstichprobe zurückzuführen, die einen deutlich überproportionalen Anteil überdurchschnittlich intelligenter Pbn enthält.

Die Erfassung des Kurzzeitgedächtnisses erfolgte über drei Merkfähigkeitsaufgaben aus dem BIS-HB. Da der BIS-HB für Schülerinnen und Schüler ab zwölf Jahren entwickelt und normiert wurde, können für die vorliegende Stichprobe keine standardisierten Werte ermittelt werden. In Tabelle 7.2 werden daher die Rohwerte (RW) berichtet.

Tabelle 7.1: Mittelwerte und Standardabweichungen der Schulnoten  
(Primar- und Sekundarstufe)

| <b>Noten</b>         | <b>N</b> | <b>M</b> | <b>SD</b> |
|----------------------|----------|----------|-----------|
| <b>Primarstufe</b>   |          |          |           |
| Deutsch              | 113      | 2.30     | 0.88      |
| Mathematik           | 113      | 2.09     | 0.80      |
| Sachkunde            | 111      | 2.08     | 0.85      |
| <b>Sekundarstufe</b> |          |          |           |
| Deutsch              | 256      | 2.63     | 0.98      |
| Englisch             | 244      | 2.60     | 1.02      |
| Französisch          | 16       | 2.06     | 0.68      |
| Latein               | 25       | 1.68     | 0.85      |
| Mathematik           | 257      | 2.40     | 0.96      |
| Physik               | 144      | 2.78     | 0.92      |
| Chemie               | 28       | 2.25     | 0.89      |
| Biologie             | 197      | 2.44     | 0.88      |

Für das Modul „Zahlenfolgen“ des CFT zeigt sich – bei Betrachtung der Rohwerte (RW) – ein starker Deckeneffekt. 56 Pbn (15%) der Gesamtstichprobe haben den Test komplett gelöst (RW = 21), hieraus folgt eine deutliche Einschränkung der IQ-Varianz, insbesondere für die älteren Pbn. Während Pbn im Altersbereich 10;7-11;6 bei kompletter Lösung des Zahlenfolgentests noch einen IQ von 139 erhalten, können die Pbn zwischen 11;7 und 12;6 Jahren nur noch maximal einen IQ von 130 erreichen, bei Pbn im Alter von 12;7-13;6 liegt der maximale IQ nur noch bei 126.

Auch bei der Betrachtung der Gruppe der unausgelesenen Pbn zeigt sich, dass die Mittelwerte des IQ im CFT 20 ( $M = 109.76$ ) und im ZVT ( $M = 108.08$ ) über dem mittleren Normwert liegen. Die anderen eingesetzten Intelligenztests überschreiten den mittleren Normwert nur leicht. Die IQ-Werte des Zahlenfolgentests, des ZVT sowie die Rohwerte der verbalen und der numerischen Merkaufgabe weichen signifikant von einer Normalverteilung ab.

Der mittlere IQ im CFT 20 in der Gruppe der hoch Intelligenten liegt mit 130.08 zwei Standardabweichungen über dem Normmittelwert. Die Verteilung der IQ-Werte im CFT 20 weicht hier signifikant von einer Normalverteilung ab, da die Stichprobe anhand dieses Tests (IQ im CFT 20  $\geq 120$ ) zusammengestellt wurde und der untere Teil der Verteilung somit abgeschnitten ist. Die IQ-Werte aus dem Zahlenfolgentest und dem Wortschatztest sowie die Rohwerte der verbalen und der numerischen Merkaufgabe weichen ebenfalls signifikant von einer Normalverteilung ab.

Tabelle 7.2: Mittelwerte, Standardabweichungen und Verteilungen der kognitiven Tests für die Gesamtstichprobe und die Teilstichproben

| Kognitive Tests                            | <i>M</i> | <i>SD</i> | Schiefe | Exzess | K-S-Test<br><i>Z</i> |
|--|----------|-----------|---------|--------|----------------------|
| <i>Gesamtstichprobe (N = 374)</i>          |          |           |         |        |                      |
| CFT 20 (IQ)                                | 116.93   | 14.91     | -.02    | -.38   | .05                  |
| CFT-WS (IQ) <sup>1</sup>                   | 108.91   | 14.56     | -.12    | -.09   | .06**                |
| CFT-ZF (IQ) <sup>1</sup>                   | 112.45   | 15.25     | -.11    | -.47   | .07**                |
| AG-Index (IQ) <sup>2</sup>                 | 109.13   | 16.04     | .12     | -.24   | .06**                |
| ZVT (IQ)                                   | 114.47   | 14.42     | -.20    | .24    | .05*                 |
| BIS-WM (RW)                                | 6.72     | 2.37      | .37     | .42    | .11**                |
| BIS-ZZ (RW)                                | 5.64     | 1.89      | .30     | .44    | .12**                |
| BIS-WE (RW)                                | 14.92    | 5.45      | .07     | -.48   | .05*                 |
| <i>Unausgelesene Pbn (UA; N = 172)</i>     |          |           |         |        |                      |
| CFT 20 (IQ)                                | 109.76   | 13.00     | .07     | -.24   | .06                  |
| CFT-WS (IQ)                                | 100.65   | 13.23     | .10     | .25    | .05                  |
| CFT-ZF (IQ)                                | 105.24   | 13.01     | .30     | .32    | .09**                |
| AG-Index (IQ) <sup>3</sup>                 | 102.07   | 13.32     | -.05    | .01    | .06                  |
| ZVT (IQ)                                   | 108.08   | 13.48     | -.44    | .17    | .07*                 |
| BIS-WM (RW)                                | 5.95     | 2.08      | .30     | .81    | .11**                |
| BIS-ZZ (RW)                                | 5.41     | 1.78      | .01     | -.15   | .13**                |
| BIS-WE (RW)                                | 14.36    | 5.30      | -.10    | -.75   | .06                  |
| <i>Hoch intelligente Pbn (HI; N = 170)</i> |          |           |         |        |                      |
| CFT 20 (IQ)                                | 130.08   | 8.42      | .68     | -.58   | .15**                |
| CFT-WS (IQ) <sup>4</sup>                   | 115.80   | 11.73     | .06     | -.17   | .08*                 |
| CFT-ZF (IQ) <sup>4</sup>                   | 121.09   | 13.64     | -.17    | -.30   | .07*                 |
| AG-Index (IQ) <sup>5</sup>                 | 115.84   | 14.52     | .09     | -.55   | .06                  |
| ZVT (IQ)                                   | 119.67   | 12.52     | .02     | -.49   | .05                  |
| BIS-WM (RW)                                | 7.18     | 2.45      | .31     | .64    | .13**                |
| BIS-ZZ (RW)                                | 6.08     | 1.95      | .34     | .80    | .12**                |
| BIS-WE (RW)                                | 15.82    | 5.74      | -.01    | -.56   | .05                  |

Anmerkung. CFT-WS = Wortschatztest des CFT 20, CFT-ZF = Zahlenfolgentest des CFT 20, AG-Index = Arbeitsgeschwindigkeitsindex aus dem HAWIK III, ZVT = Zahlen-Verbindungs-Test, BIS-WM = Merkfähigkeitsaufgabe „Wörtermerken“ aus dem BIS-HB, BIS-ZZ = Merkfähigkeitsaufgabe „Zweistellige Zahlen“ aus dem BIS-HB, BIS-WE = Merkfähigkeitsaufgabe „Wege erinnern“ aus dem BIS-HB; 1:  $N = 373$ , 2:  $N = 368$ , 3:  $N = 168$ , 4:  $N = 169$ , 5:  $N = 167$ .

K-S-Test: Kolmogorov-Smirnov-Test auf Normalverteilung, Signifikanzkorrektur nach Lilliefors. \* $p < .05$ . \*\* $p < .01$ .

### 7.1.3 Arbeitsgedächtnisaufgaben

Tabelle 7.3 zeigt die Mittelwerte, Standardabweichungen, den Range sowie Angaben zu den Verteilungen der Scores der sechs Aufgabentypen für die gesamte Stichprobe ( $N = 374$ ). Die Werte für die innere Konsistenz der Aufgabengruppen (Cronbachs Alpha) liegen mit Werten zwischen  $\alpha = .73$  und  $\alpha = .81$  in einem akzeptablen Bereich (vgl. auch Abschnitt 7.1.4).

Tabelle 7.3: Mittelwerte, Standardabweichungen, Verteilungen und Reliabilität der Skalenscores der Arbeitsgedächtnisaufgaben in der Gesamtstichprobe und den Teilstichproben

| Arbeitsgedächtnis<br>Aufgabentyp           | <i>M</i> | <i>SD</i> | Range<br>(theoret.<br>Max.) | Schiefe | Exzess | K-S-<br>Test<br>Z | $\alpha$ |
|--|----------|-----------|-----------------------------|---------|--------|-------------------|----------|
| <i>Gesamtstichprobe (N = 374)</i>          |          |           |                             |         |        |                   |          |
| Reading Span                               | 3.80     | 2.56      | 0-10 (11)                   | .24     | -.79   | .11               | .79      |
| Verbal Span                                | 4.02     | 2.65      | 0-10 (10)                   | .17     | -.87   | .10               | .78      |
| Computation Span                           | 4.54     | 2.32      | 0-10 (10)                   | -.13    | -.58   | .11               | .73      |
| Backw. Digit Span                          | 3.91     | 2.81      | 0-12 (12)                   | .69     | -.26   | .15               | .81      |
| Memory Upd. Fig.                           | 5.50     | 2.97      | 0-14 (14)                   | .04     | -.74   | .09               | .74      |
| Spatial WM                                 | 4.86     | 2.79      | 0-12 (13)                   | .29     | -.56   | .12               | .76      |
| <i>Unausgelesene Pbn (UA; N = 172)</i>     |          |           |                             |         |        |                   |          |
| Reading Span                               | 2.70     | 2.15      | 0-10 (11)                   | .67     | .01    | .15               | .73      |
| Verbal Span                                | 3.12     | 2.48      | 0-9 (10)                    | .44     | -.86   | .15               | .76      |
| Computation Span                           | 3.90     | 2.27      | 0-10 (10)                   | .04     | -.54   | .11               | .72      |
| Backw. Digit Span                          | 2.91     | 2.29      | 0-10 (12)                   | .88     | .26    | .18               | .75      |
| Memory Upd. Fig.                           | 4.55     | 2.78      | 0-11 (14)                   | .25     | -.84   | .11               | .72      |
| Spatial WM                                 | 3.75     | 2.37      | 0-12 (13)                   | .56     | .28    | .13               | .69      |
| <i>Hoch intelligente Pbn (HI; N = 170)</i> |          |           |                             |         |        |                   |          |
| Reading Span                               | 5.06     | 2.26      | 0-10 (11)                   | -.16    | -.39   | .12               | .70      |
| Verbal Span                                | 5.25     | 2.28      | 0-10 (10)                   | -.03    | -.49   | .09               | .67      |
| Computation Span                           | 5.39     | 2.20      | 0-10 (10)                   | -.41    | -.02   | .12               | .69      |
| Backw. Digit Span                          | 5.19     | 2.78      | 0-12 (12)                   | .35     | -.69   | .11               | .77      |
| Memory Upd. Fig.                           | 6.68     | 2.75      | 0-14 (14)                   | -.29    | -.27   | .14               | .67      |
| Spatial WM                                 | 5.98     | 2.63      | 0-12 (13)                   | -.04    | -.63   | .11               | .73      |

Anmerkung. K-S-Test: Kolmogorov-Smirnov-Test auf Normalverteilung, Signifikanzkorrektur nach Lilliefors, alle Aufgaben:  $p < .01$ .

Zur Darstellung der Verteilung der Scores sind die Schiefe, der Exzess sowie der Kolmogorov-Smirnov-Test auf Normalverteilung (mit Signifikanzkorrektur nach Lilliefors) angegeben. Die Verteilungen aller Scores – mit Ausnahme des Scores für „Computation Span“ – sind leicht rechtsschief (positive Schiefe), das heißt ein

größerer Anteil der Pbn hat geringere Scores erzielt, als bei einer Normalverteilung der Daten zu erwarten wäre. Die Verteilungen der Arbeitsgedächtnis-Scores weichen laut Kolmogorov-Smirnov-Test signifikant von der Normalverteilung ab. In vier der sechs Aufgabengruppen wurde von einigen Pbn die jeweils maximale Punktzahl erreicht, das heißt, dass sie alle Aufgaben richtig lösen konnten. In den Aufgabengruppen „Reading Span“ und „Spatial Working Memory“ hingegen gelang es keinem Pbn, alle Aufgaben richtig zu lösen.

Da die vorliegende Gesamtstichprobe im Hinblick auf die intellektuellen Fähigkeiten, bedingt durch die Alters- und Intelligenzstreuung, sehr heterogen ist, werden zusätzlich die Verteilungen für die zwei Substichproben „Unausgelesene Pbn“ (UA) und „hoch intelligente Pbn“ (HI) untersucht. Es zeigt sich, dass die Arbeitsgedächtnis-Scores auch in beiden Substichproben nach dem Kolmogorov-Smirnov-Test (mit Signifikanzkorrektur nach Lilliefors) signifikant von einer Normalverteilung abweichen. In der UA-Gruppe sind die Scores deutlich rechtsschief verteilt. Bei den Pbn der Gruppe HI sind bis auf die Aufgabe „Backward Digit Span“ alle Scores leicht linksschief verteilt, hier haben somit mehr Pbn hohe Scores erzielt, als es bei normal verteilten Scores der Fall wäre.

#### **7.1.4 Empirischer Vergleich der Scorings der Arbeitsgedächtnisaufgaben auf Aufgaben- und auf Itemebene**

Wie in Abschnitt 4.2 ausgeführt wurde, gibt es verschiedene Möglichkeiten, die Leistungen in Arbeitsgedächtnisaufgaben zu bewerten. Aufgrund theoretischer Überlegungen wird in der vorliegenden Arbeit eine Scoringmethode verwendet, die die Verarbeitungskomponente mit einbezieht, eine Einheitsgewichtung auf der Aufgabenebene vornimmt, und einen *serial recall* erfordert. Im Folgenden wird dieses Scoring mit einigen Alternativen auf Basis der vorliegenden Testdaten verglichen.

##### *Berücksichtigung der Verarbeitungskomponente?*

In den meisten Studien wird die Leistung in Arbeitsgedächtnisaufgaben nur anhand der Speicherleistung der Pbn bewertet (vgl. Abschnitt 4.2.1). Über die Instruktion wird versucht sicherzustellen, dass die Pbn dennoch beide Komponenten der Aufgabe (Speicherung und Verarbeitung) beachten und nicht etwa schlechte Leistungen in der Verarbeitungskomponente in Kauf nehmen, um bessere Speicherleistungen zu erzielen. Auch in der vorliegenden Studie wurden die Pbn so instruiert, dass beide Komponenten der Aufgaben gleich wichtig sind. Die Verarbeitungskomponenten der Aufgaben wurden zudem so einfach gestaltet, dass sie für Kinder der untersuchten Altersgruppen einfach lösbar sind. Dennoch traten in dieser Studie in den Verarbeitungskomponenten der Aufgaben „Reading Span“ und „Computation Span“ mehr Fehler auf als bei Studien mit erwachsenen Pbn, über die in der Literatur berichtet wird. In der Aufgabe „Reading Span“ fanden sich bei 26% der Pbn > 10% falsche Entscheidungen und bei 10% der Pbn > 20% falsche Entscheidungen. 40% der Pbn trafen > 10% falsche Entscheidungen und 20% der Pbn > 20% falsche Entscheidungen.

gen in der Aufgabe „Computation Span“. In der vorliegenden Arbeit wurde daher der Score für die Aufgaben „Reading Span“ und „Computation Span“ so gebildet, dass sowohl die Antwort als auch die Entscheidung korrekt sein muss, damit der Pbn für ein Item einen Punkt erhält. Verschiedene Autoren argumentieren, dass eine solche kombinierte Scorebildung exakter ist und die Leistungen der Pbn genauer abbildet (z. B. Klein & Fiss, 1999; Hacker et al., 1999).

Der im Folgenden verwendete kombinierte Score korreliert in dieser Untersuchung für den Aufgabentyp „Reading Span“ sehr hoch mit dem Score, der nur aus der Speicherkomponente gebildet wird ( $r = .95, p < .01$ ). Da ein relativ großer Teil der Pbn viele Fehler in der Aufgabengruppe „Computation Span“ gemacht hat, fällt der entsprechende Zusammenhang hier etwas geringer aus ( $r = .87, p < .01$ ). Insbesondere für die „Computation Span“ scheint daher ein kombinierter Score für die vorliegende Stichprobe angemessener.

#### *Art der Datenaggregation*

Mit dem bisher verwendeten Scoring (Einheitsgewichtung auf der *Aufgabenebene*) wurde die Scorebildung nach dem Prinzip der Einheitsgewichtung auf der *Itemebene* verglichen (vgl. Abschnitt 4.2.2 sowie Tabelle 4.1). Beide Scorebildungen führen zu Scores, die hoch miteinander korrelieren (siehe Tabelle 7.4). Die Scorebildung auf Itemebene weist jedoch bestimmte günstige Eigenschaften auf.

*Tabelle 7.4: Kennwerte für die Aufgabenscores, die auf Itemebene mit Einheitsgewichtung gebildet wurden (Gesamtstichprobe, N = 374)*

| <b>Arbeitsgedächtnis Aufgabentyp</b> | <b>M</b> | <b>SD</b> | <b>Range (theoret. Max.)</b> | <b>K-S-Test Z</b> | <b><math>\alpha</math></b> | <b>Korr. mit Aufg.score</b> |
|--------------------------------------|----------|-----------|------------------------------|-------------------|----------------------------|-----------------------------|
| Reading Span                         | 29.86    | 10.93     | 2-49 (51)                    | .07**             | .89                        | .89**                       |
| Verbal Span                          | 28.52    | 10.31     | 1-45 (45)                    | .11**             | .87                        | .91**                       |
| Computation Span                     | 32.90    | 10.26     | 1-50 (50)                    | .09**             | .84                        | .90**                       |
| Backw. Digit Sp.                     | 39.02    | 13.19     | 3-70 (70)                    | .06**             | .83                        | .88**                       |
| Memory Upd. Fig.                     | 24.76    | 6.68      | 6-40 (40)                    | .09**             | .78                        | .92**                       |
| Spatial WM                           | 15.25    | 6.55      | 0-30 (30)                    | .04               | .82                        | .92**                       |

Anmerkung. K-S-Test: Kolmogorov-Smirnov-Test auf Normalverteilung, Signifikanzkorrektur nach Lilliefors; \*\* $p < .01$ .

Aufgrund der insgesamt größeren Varianz der Scores fallen die Korrelationen mit anderen Variablen in der Regel etwas größer aus. Entsprechend sind auch die auf Basis der Klassischen Testtheorie berechneten Trennschärfen höher, wenn die Leistungen der Pbn auf Itemebene bewertet werden. Auch die Kennwerte für die innere Konsistenz der Aufgabengruppen fallen höher aus als bei einem Scoring auf Aufgabenebene. Bis auf den Wert für „Memory Updating Figural“ überschreiten alle Kennwerte einen Wert von  $\alpha = .80$ , die beiden verbalen Aufgaben erreichen sogar Werte von  $\alpha > .85$ . Die Werte auf Basis der einzelnen Items weichen nicht für alle

Skalen signifikant von einer Normalverteilung ab, wie es bei dem Scoring auf Aufgabenebene der Fall ist. Die Scores für die Aufgabengruppen „Spatial Working Memory“ können hier als normalverteilt angesehen werden.

### *Serial recall vs. item recall*

Im Folgenden werden zusätzlich zu den Scores, die bisher betrachtet wurden und die einen *serial recall* der Items erfordern, solche Scores berechnet, die lediglich einen *item recall* erfordern. Während ein Item beim serial recall nur als richtig gewertet wird, wenn auch die Reihenfolge der Antworten korrekt ist, reicht beim item recall die reine Nennung des Items (vgl. Abschnitt 4.2.4). Diese Unterscheidung im Scoring lässt sich bei den verbalen Aufgabengruppen sowie bei „Computation Span“ und bei „Spatial Working Memory“ vornehmen (bei der „Backward Digit Span“ ist die Einhaltung einer bestimmten Reihenfolge unerlässlich, bei „Memory Updating Figural“ wird die korrekte Reihenfolge durch den Computer vorgegeben). Für alle betrachteten Skalen korrelieren beide Scorevarianten sehr hoch, insbesondere bei einem Scoring auf Aufgabenebene, so dass davon auszugehen ist, dass es für weitere Analysen nicht bedeutsam ist, welcher Wert verwendet wird. Die Korrelationen der Scores auf Aufgaben- bzw. auf Itemebene fallen folgendermaßen aus:  $r_{RSAufg.} = .99$  bzw.  $r_{RSItem} = .96$ ;  $r_{VSAufg.} = .99$  bzw.  $r_{VSItem} = .98$ ;  $r_{CSAufg.} = .95$  bzw.  $r_{CSItem} = .92$ ;  $r_{SWMAufg.} = 1.0$  bzw.  $r_{SWMItem} = .99$ . Auch die Betrachtung der Verteilungen, Schwierigkeiten, Trennschärfen, der internen Konsistenz sowie der Zusammenhänge mit anderen Variablen wie Schulnoten und kognitiven Tests zeigt nur sehr geringe Unterschiede zwischen beiden Scoringmethoden.

## **7.2 Untersuchung der Binnenstruktur des Arbeitsgedächtnisses**

Im Folgenden werden die Befunde zur Faktorstruktur des Arbeitsgedächtnisses dargestellt. Hierzu werden zunächst die bivariaten und Partialkorrelationen der Arbeitsgedächtnisaufgaben untereinander betrachtet. Anschließend wird eine exploratorische Faktorenanalyse durchgeführt. Im nächsten Schritt werden die drei Faktorenmodelle entsprechend den Hypothesen (Abschnitt 5.3.1) mittels konfirmatorischer Faktorenanalysen überprüft. Es wird ferner untersucht, ob die Faktorenstruktur jeweils für verschiedene Altersgruppen, für die Gruppen der hoch Intelligenten und der nicht hoch Intelligenten sowie für Jungen und Mädchen invariant ist. Anschließend werden die Mittelwertsunterschiede in den Leistungen in den Arbeitsgedächtnisaufgaben für verschiedene Gruppen berichtet.

### **7.2.1 Bivariate und Partialkorrelationen der Arbeitsgedächtnisaufgaben**

Tabelle 7.5 zeigt die Korrelationen der Arbeitsgedächtnis-Scores. Alle Arbeitsgedächtnisaufgaben korrelieren positiv und hochsignifikant miteinander, das heißt, es korrelieren auch Arbeitsgedächtnisaufgaben verschiedener Domänen positiv miteinander. Angegeben sind oberhalb der Diagonale der Tabelle zusätzlich die Korrelatio-



nen, bei denen der Alterseffekt auspartialisiert wurde. Die Partialkorrelationen fallen etwas geringer aus als die bivariaten Korrelationen. Die erste Grundvoraussetzung für die Annahme einer Domänengeneralität des Arbeitsgedächtnisses ist damit erfüllt, denn die positiven Zusammenhänge finden sich nicht nur innerhalb einer Domäne, sondern auch über die Domänen hinweg.

*Tabelle 7.5: Korrelationen der Gesamtscores aus den Arbeitsgedächtnisaufgaben (Korrelationen und Partialkorrelationen, Alter kontrolliert)*

| <b>Arbeitsgedächtnisaufgaben</b> | <b>RS</b> | <b>VS</b> | <b>CS</b> | <b>BDS</b> | <b>MUF</b> | <b>SWM</b> |
|----------------------------------|-----------|-----------|-----------|------------|------------|------------|
| Reading Span (RS)                | 1         | .64       | .56       | .58        | .41        | .35        |
| Verbal Span (VS)                 | .65       | 1         | .53       | .62        | .37        | .34        |
| Computation Span (CS)            | .59       | .56       | 1         | .56        | .34        | .29        |
| Backward Digit Span (BDS)        | .60       | .63       | .58       | 1          | .47        | .40        |
| Memory Upd. Figural (MUF)        | .44       | .40       | .38       | .50        | 1          | .47        |
| Spatial WM (SWM)                 | .38       | .35       | .31       | .42        | .49        | 1          |

Anmerkung. Unterhalb der Diagonale: Korrelationskoeffizienten; oberhalb der Diagonale: Partialkorrelationen, Alter kontrolliert. Alle Korrelationen  $p < .01$ .

Die beiden verbalen Aufgabentypen korrelieren jeweils höher untereinander als mit den anderen Aufgabentypen. Die Korrelationen der verbalen Aufgaben mit den beiden numerischen Aufgaben fallen etwas geringer aus, liegen jedoch noch in vergleichbarer Höhe. Beide numerische Aufgaben korrelieren jeweils am stärksten mit den verbalen Aufgaben und etwas schwächer untereinander. Die Korrelationen der figuralen Aufgaben untereinander und mit den anderen Aufgaben fallen insgesamt am schwächsten aus. Die stärksten Zusammenhänge weisen sie untereinander und mit der numerischen Aufgabe „Backward Digit Span“ auf. Die figuralen Aufgaben hingegen teilen insgesamt weniger gemeinsame Varianz untereinander und mit den anderen Aufgabentypen. Somit deuten die leicht höheren Zusammenhänge der verbalen und numerischen Aufgaben untereinander als mit den figuralen Aufgaben auch auf domänenspezifische Aspekte des Arbeitsgedächtnisses hin.

### 7.2.2 Exploratorische und konfirmatorische Faktorenanalysen

Wie in Abschnitt 7.1 bereits dargestellt, konnte das Auftreten vereinzelter fehlender Werte im Datensatz nicht verhindert werden. Insgesamt ist der Anteil fehlender Werte jedoch sehr gering. Der MCAR-Test nach Little zeigt zudem, dass nicht von einem systematischen Auftreten fehlender Werte ausgegangen werden muss ( $\chi^2 = 8951.99$ ,  $df = 23577$ ,  $p > .99$ ).

Vor der Berechnung von Faktorenanalysen werden die fehlenden Werte über den EM-Algorithmus in SPSS ersetzt, so dass vollständige Datensätze für sämtliche 374 Pbn vorliegen. Es wird hier der EM-Ansatz gewählt, da dieser versucht, über Maximum-Likelihood-Schätzungen die Informationen aus den vorhandenen Daten

optimal für die Ergänzung der fehlenden Werte auszunutzen. Der EM-Ansatz erbringt in der Regel die genauesten Schätzungen für die fehlenden Werte (vgl. Hair, Anderson, Tatham & Black, 1998, S. 55). Weiterhin werden die Daten auf univariate und multivariate Ausreißer geprüft. Es liegen keine univariaten Extremwerte vor, die  $> 3 SD$  vom Mittelwert abweichen, ebenso ergibt eine Prüfung auf multivariate Ausreißer mittels der Mahalanobis Distanz keine Auffälligkeiten. Zwar weichen die Scores laut Kolmogorov-Smirnov-Test signifikant von einer Normalverteilung ab, jedoch überschreiten die Werte für Schiefe und Exzess nicht die Grenzen, für die aus Simulationsstudien bekannt ist, dass sie die Güte der Schätzung von Strukturmodellen beeinträchtigen (Schiefe  $> 3.0$ , Kurtosis  $> 10.0$ ; Kline, 1998). Eine Prüfung auf Abweichung von einer multivariaten Normalverteilung ergibt ebenfalls keinen signifikanten Befund (multivariate Kurtosis =  $-1.19$ , critical ratio  $< 1.96$ ).

Zuerst wird die Struktur der Arbeitsgedächtnisaufgaben mit einer exploratorischen Faktorenanalyse (Extraktion: Hauptachsenanalyse) untersucht. Nach dem Screeplot (siehe Abbildung 7.1) empfiehlt sich die Extraktion von ein bis zwei Faktoren, jedoch weist nur der erste Faktor einen Eigenwert  $> 1$  auf (erster Faktor: 3.45, zweiter Faktor: 0.87). Bei der Extraktion eines Faktors zeigen die verbalen und die numerischen Aufgaben die höchsten Faktorladungen und die beiden figuralen Aufgaben etwas geringere Ladungen. Insgesamt ergibt sich das folgende Ladungsmuster: Reading Span:  $r = .78$ , Verbal Span:  $r = .77$ , Computation Span:  $r = .70$ , Backward Digit Span:  $r = .81$ , Memory Updating Figural:  $r = .60$ , Spatial Working Memory:  $r = .52$ .

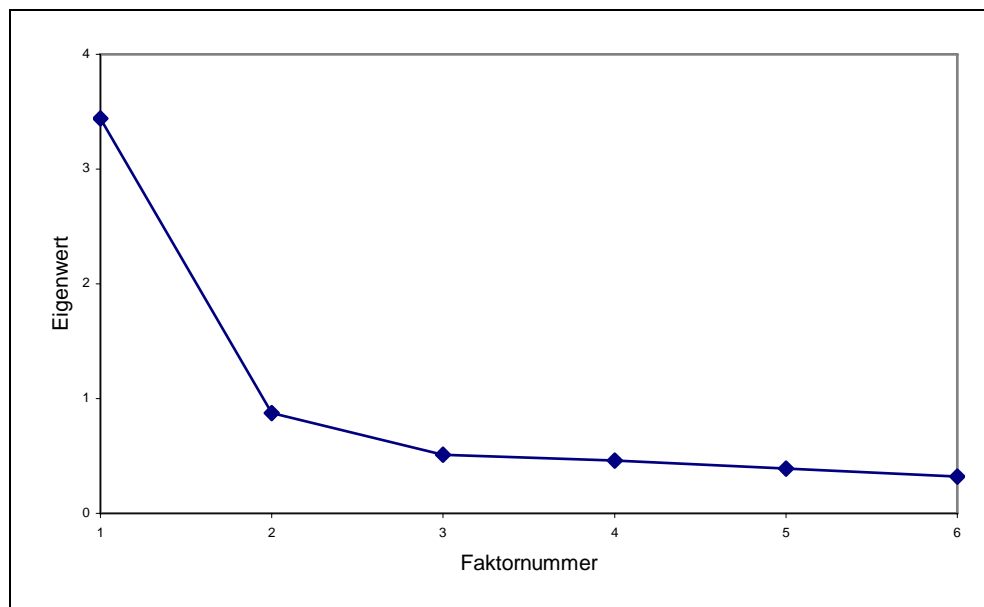


Abbildung 7.1: Screeplot über die sechs Arbeitsgedächtnis-Scores

Zur Prüfung der Passung verschiedener Faktorstrukturen werden im Folgenden konfirmatorische Faktorenanalysen durchgeführt. Die Modelle werden mit dem Programm AMOS 4.0 (Arbuckle, 1999) berechnet, die Schätzung der Parameter er-

folgt anhand der Maximum-Likelihood-Methode. Der Fit der konfirmatorischen Modelle wird anhand der folgenden Maße bewertet:

Das Fit-Maß *Root Mean Square Error of Approximation* (RMSEA) basiert auf der geschätzten Diskrepanz zwischen dem Modell mit optimal gewählten Parametern und der Kovarianzmatrix in der Population und sollte möglichst  $<.05$  betragen. RMSEA-Werte bis  $.08$  gelten als Indikatoren für einen akzeptablen Fit, Werte zwischen  $.08$  und  $.10$  als mittelmäßig und Werte  $>.10$  deuten auf eine schlechte Passung hin (Browne & Cudeck, 1993). Das relative Fit-Maß *Comparative Fit Index* (CFI) bewertet die Passung des untersuchten Modells in Relation zum Independence-Modell. Der CFI sollte über  $.95$  liegen (Byrne, 2001). Zusätzlich wird der *Adjusted Goodness of Fit Index* (AGFI) berichtet. Dieser Index ist ein Maß für den relativen Anteil von Varianz und Kovarianz in der Stichprobe, der durch das angenommene Modell erklärt wird. Die Verwendung von mehr Parametern in einem Modell wird durch dieses Maß „bestraft“, da es an der Anzahl der Freiheitsgrade relativiert wird. Der AGFI kann zwischen 1 und 0 liegen und fällt umso größer aus, je besser das Modell die Daten erklärt. Darüber hinaus werden die informationstheoretischen Maße *Akaike's Information Criterion* (AIC) und der am Stichprobenumfang relativierte *Consistent AIC* (CAIC) berichtet. Beide Maße dienen insbesondere dem Vergleich verschiedener Modelle, wobei ein kleinerer Wert für eine bessere Modellanpassung spricht.

Zur Prüfung der Hypothesen, dass ein Zwei-Faktoren-Modell die Kovarianzstruktur der Daten besser erklärt als ein Ein-Faktor-Modell, ein Drei-Faktoren-Modell jedoch keine bessere Erklärung der Daten erlaubt, werden im Folgenden drei Faktorenmodelle spezifiziert und mit konfirmatorischen Faktorenanalysen untersucht. Zunächst wird – auch entsprechend dem Ergebnis der exploratorischen Faktorenanalyse – ein Modell mit einem gemeinsamen Faktor spezifiziert, auf den die Scores aller sechs Arbeitsgedächtnisaufgaben laden (Modell 1, siehe Abbildung 7.2). Die Fit-Maße zeigen, dass der Fit des Ein-Faktor-Modells nicht optimal ist. Zwar fallen die Werte CFI und AGFI befriedigend aus, der  $\chi^2$ -Test wird jedoch signifikant und auch der RMSEA deutet auf eine eher schlechte Passung des Modells hin.

In einem zweiten Schritt wird daher ein Modell mit zwei Arbeitsgedächtnisfaktoren spezifiziert (Modell 2, siehe Abbildung 7.3). Die Scores aus den verbalen Aufgaben RS und VS sowie aus den numerischen Aufgaben CS und BDS laden in diesem Modell auf den ersten Faktor, die figuralen Aufgaben MUF und SWM laden auf den zweiten Faktor. Zwischen dem verbal-numerischen und dem figuralen Faktor wird eine Korrelation zugelassen. Modell 2 passt deutlich besser als Modell 1: Der  $\chi^2$ -Test wird nicht signifikant, und auch alle anderen Maße sprechen dafür, dass das Modell gut dazu geeignet ist, die Daten zu erklären. AIC und CAIC fallen in diesem Modell zudem geringer aus als in Modell 1, was ebenfalls darauf hindeutet, dass Modell 2 das angemessenere Modell ist.

Zur Prüfung der zweiten Hypothese, dass ein Drei-Faktoren-Modell die Kovarianzstruktur der Testdaten nicht besser erklären kann als ein Zwei-Faktoren-Modell, wird nun Modell 3 spezifiziert, und es werden die Parameter des Modells geschätzt.

Abbildung 7.4 zeigt eine sehr gute Passung für Modell 3. Nach dem AIC passt Modell 3 besser als Modell 2 auf die Daten, der CAIC favorisiert jedoch Modell 2. Bei der Betrachtung der Faktorenkorrelationen fällt auf, dass der verbale und der numerische Faktor extrem hoch miteinander korrelieren ( $r = .95$ ) und daher 90% ihrer Varianz teilen. Die Spezifizierung zweier separater Faktoren scheint damit überflüssig.

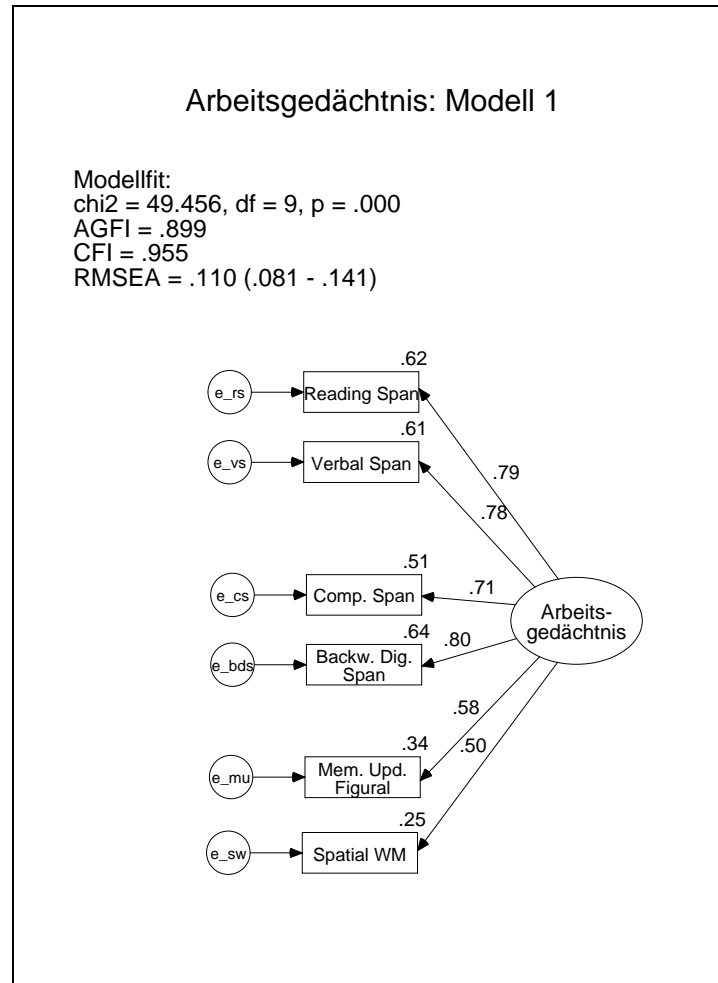


Abbildung 7.2: Modell 1. Ein-Faktor-Modell des Arbeitsgedächtnisses (CFA)

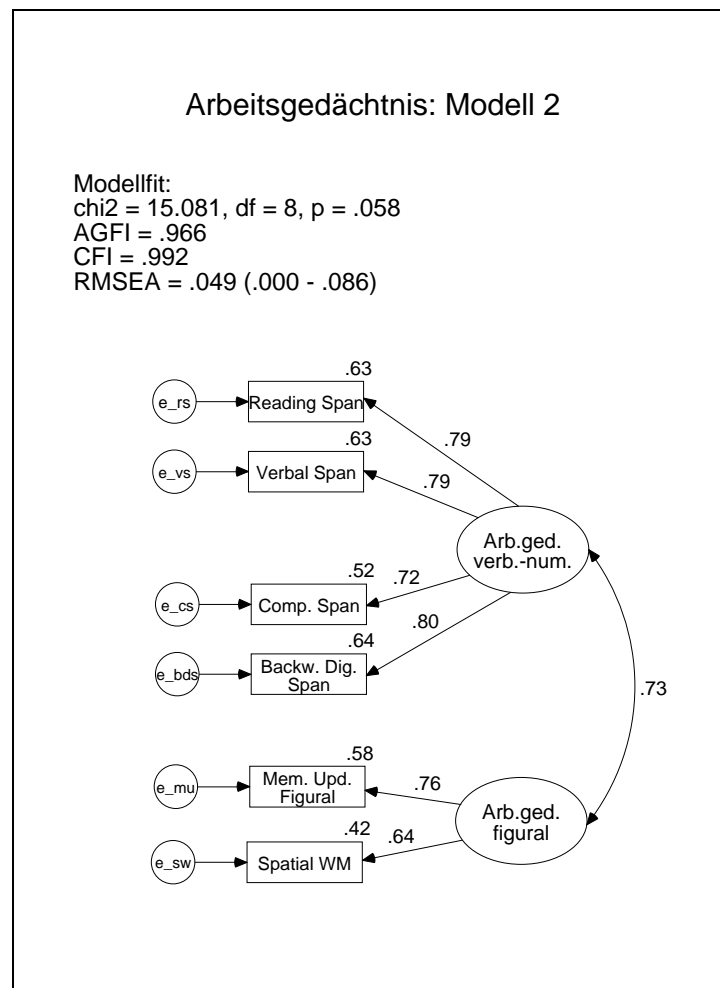


Abbildung 7.3: Modell 2. Zwei-Faktoren-Modell des Arbeitsgedächtnisses (CFA)

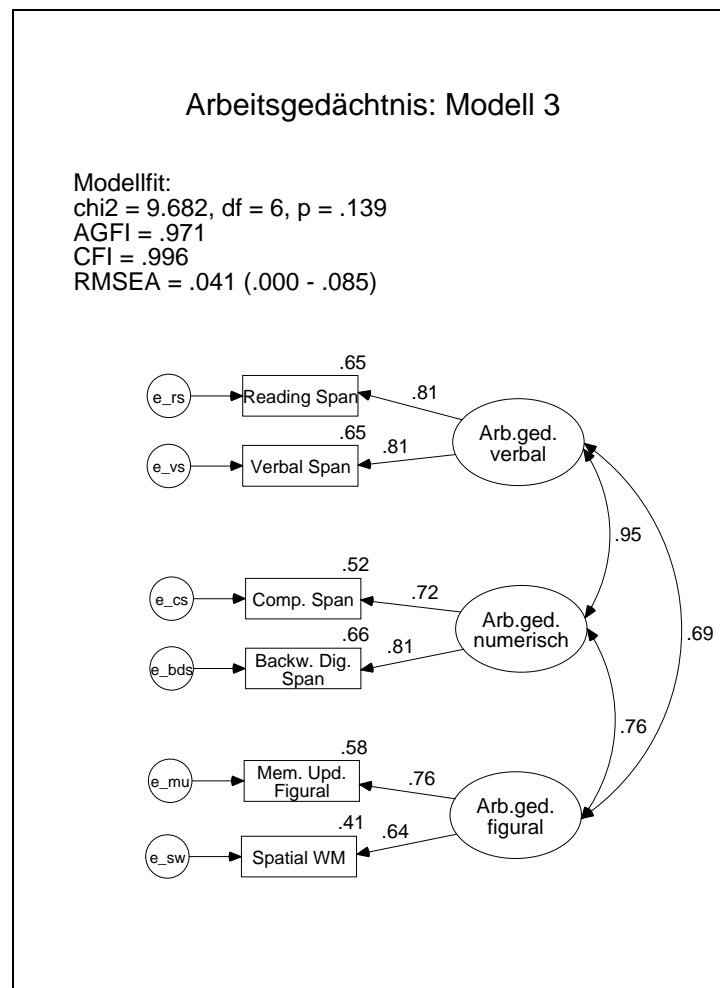


Abbildung 7.4: Modell 3. Drei-Faktoren-Modell des Arbeitsgedächtnisses (CFA)

Tabelle 7.6: Modellpassung eines Ein-, Zwei- und Drei-Faktor-Modells des Arbeitsgedächtnisses und Modellvergleich

| Modell                 | $\chi^2$ | df | p   | CFI  | RMSEA<br>(Konfidenz) | AGFI | AIC   | CAIC   |
|------------------------|----------|----|-----|------|----------------------|------|-------|--------|
| 1                      | 49.46    | 9  | .00 | .96  | .11 (.08-.14)        | .90  | 73.46 | 132.55 |
| 2                      | 15.08    | 8  | .06 | .99  | .05 (.00-.09)        | .97  | 41.08 | 105.10 |
| 3                      | 9.68     | 6  | .14 | 1.00 | .04 (.00-.09)        | .97  | 39.68 | 113.55 |
| <i>Modellvergleich</i> |          |    |     |      |                      |      |       |        |
| 1 vs. 2                | 34.38    | 1  | .00 |      |                      |      | 2 < 1 | 2 < 1  |
| 2 vs. 3                | 5.40     | 2  | .07 |      |                      |      | 3 < 2 | 2 < 3  |

Anmerkung. Modell 1 = Ein-Faktor-Modell; Modell 2 = Zwei-Faktoren-Modell (ein verbal-numerischer und ein figuraler Faktor); Modell 3 = Drei-Faktoren-Modell (ein verbaler, ein numerischer und ein figuraler Faktor).

Aus diesen Befunden wird bereits deutlich, dass Modell 2 relativ am besten dazu in der Lage ist, die Daten zu erklären. Eine weitere Aufteilung der Faktoren scheint nicht gewinnbringend, da der verbale und der numerische Faktor hoch korreliert sind. Für eine Entscheidung über das relativ am besten passende Modell ist jedoch zusätzlich ein direkter Vergleich der drei Modelle notwendig (Loehlin, 1998). Dieser erfolgt erneut über einen  $\chi^2$ -Test. Der Modellvergleich in Tabelle 7.6 zeigt, dass Modell 2 signifikant besser passt als Modell 1. Die weitere Aufspaltung des verbal-numerischen Faktors in zwei einzelne Faktoren in Modell 3 führt jedoch nicht zu einer signifikant besseren Datenpassung. Aufgrund der hohen Faktorkorrelation kann davon ausgegangen werden, dass durch eine Fusion des verbalen und des numerischen Faktors kein bedeutender Informationsverlust erfolgt.

Es können somit beide Hypothesen bestätigt werden: Modell 2, das eine Differenzierung in verbal-numerische versus figurale Komponenten des Arbeitsgedächtnisses vornimmt, ist signifikant besser dazu in der Lage, die Kovarianzstruktur der Daten zu erklären als Modell 1, das einen domänengenerellen Faktor annimmt. Dennoch hängen auch verbal-numerische und figurale Arbeitsgedächtnisfähigkeiten bedeutsam miteinander zusammen, beide Faktoren korrelieren zu  $r = .75$ . Eine weitere Aufgliederung des verbal-numerischen Faktors in Modell 3 erbringt jedoch keine bedeutsame Verbesserung des Modell-Fits.

### **7.2.3 Prüfung der Invarianz der Faktorstruktur für verschiedene Alters- und Begabungsgruppen sowie für Mädchen und Jungen**

Im Folgenden wird geprüft, ob die gefundene Faktorstruktur für verschiedene Alters- und Begabungsgruppen sowie für Mädchen und Jungen gleichermaßen gut dazu geeignet ist, die Kovarianzstruktur der Daten zu erklären. Es wird dabei so verfahren, dass zunächst die Invarianz der Kovarianzstruktur simultan für die interessierenden Teilstichproben überprüft wird. Die  $\chi^2$ -Statistik für die simultane Prüfung aller Gruppen ist die Summe der  $\chi^2$ -Statistiken für jede einzelne Gruppe, ebenso ist die Zahl der Freiheitsgrade die Summe der Freiheitsgrade jeder gruppenspezifischen Testung. Es wird die Nullhypothese getestet, dass die Kovarianzstruktur in allen Gruppen gleich ist. Muss die Nullhypothese aufgrund des  $\chi^2$ -Tests verworfen werden, so kann dieses Ergebnis als Nichtäquivalenz der Gruppen interpretiert werden. Wenn die Nullhypothese hingegen nicht verworfen werden kann, so kann davon ausgegangen werden, dass die Kovarianzstruktur in den betrachteten Gruppen invariant ist. Die Frage der Invarianz sollte jedoch nicht nur auf der Basis eines globalen Tests im Sinne einer simultanen Prüfung beantwortet werden (Byrne, 2001). Zusätzlich werden daher systematisch einzelne Parameter des Modells (Faktorladungen und/oder Faktorkovarianzen) für die jeweils interessierenden Gruppen gleich gesetzt. Die  $\chi^2$ -Statistik eines solchen Modells mit gleichgesetzten Parametern kann zur  $\chi^2$ -Statistik des Baseline-Modells, in dem die Parameter für alle Gruppen frei geschätzt wurden, in Beziehung gesetzt werden. Es wird geprüft, ob Modelle, die in allen Gruppen dieselbe Ausprägung bestimmter festgelegter Parameter aufweisen, gut dazu geeignet sind, die Kovarianzstruktur der Daten zu erklären. Die Signifikanz der Differenz der  $\chi^2$ -

Werte im Baseline-Modell und im Modell mit fixierten Parametern ( $\Delta\chi^2$ ) ist dann ausschlaggebend für die Beurteilung, ob die gleichgesetzten Parameter in allen Gruppen invariant sind.

### *Altersgruppen*

Zunächst werden die drei hypothetischen Modelle für die jüngeren ( $\leq 11.5$  Jahre,  $N=195$ ) und die älteren Pbn ( $> 11.5$  Jahre,  $N=178$ ) der Stichprobe separat geschätzt (siehe Tabelle 7.7, die Abbildungen der einzelnen Modelle finden sich in Anhang B). Eine Aufspaltung der Stichprobe nach einzelnen Jahrgängen wird nicht vorgenommen, da die Stichprobengrößen dann zu gering ausfallen würden, um zu stabilen Schätzungen zu gelangen.

*Tabelle 7.7: Passung verschiedener Faktormodelle des Arbeitsgedächtnisses für jüngere und ältere Pbn*

| <b>Altersgruppen</b>  | $\chi^2$      | <i>df</i> | <i>p</i>   | <b>CFI</b> | <b>RMSEA<br/>(Konfidenz)</b> | <b>AGFI</b> |
|---|---------------|-----------|------------|------------|------------------------------|-------------|
| <i>Modell 1</i>   |               |           |            |            |                              |             |
| ≤ 11.5 Jahre  | 25.12         | 9         | .00        | .96        | .10 (.05-.14)                | .91         |
| > 11.5 Jahre  | 34.87         | 9         | .00        | .94        | .13 (.09-.17)                | .85         |
| <i>Modell 2</i>   |               |           |            |            |                              |             |
| ≤ 11.5 Jahre  | 6.19          | 8         | .63        | 1.00       | .00 (.00-.07)                | .97         |
| > 11.5 Jahre  | 19.03         | 8         | .02        | .97        | .09 (.04-.14)                | .91         |
| <i>Modell 3</i>   |               |           |            |            |                              |             |
| ≤ 11.5 Jahre  | 5.81          | 6         | .44        | 1.00       | .00 (.00-.09)                | .97         |
| > 11.5 Jahre  | 12.85         | 6         | .05        | .98        | .08 (.01-.14)                | .92         |
| <i>Simultane Schätzung für Modell 2 für beide Altersgruppen</i> |               |           |            |            |                              |             |
| Baseline-Modell<br>(Parameter frei)                             | 25.22         | 16        | .07        | .99        | .04 (.00 - .07)              | .95         |
| Kovarianz fix<br>Vergleich ( $\Delta\chi^2$ )                   | 26.50<br>1.28 | 17<br>1   | .07<br>.26 | .99        | .04 (.00 - .07)              | .95         |
| Ladungen fix<br>Vergleich ( $\Delta\chi^2$ )                    | 27.83<br>2.61 | 20<br>4   | .11<br>.63 | .99        | .03 (.00 - .06)              | .95         |
| Ladungen u. Kov. fix<br>Vergleich ( $\Delta\chi^2$ )            | 29.36<br>4.14 | 21<br>5   | .11<br>.53 | .99        | .03 (.00 - .06)              | .95         |

Anmerkung. Modell 1 = Ein-Faktor-Modell; Modell 2 = Zwei-Faktoren-Modell (ein verbal-numerischer und ein figuraler Faktor); Modell 3 = Drei-Faktoren-Modell (ein verbaler, ein numerischer und ein figuraler Faktor).



Die Fit-Maße zeigen, dass Modell 1 für beide Altersgruppen nicht gut passt. Modell 2 hingegen passt für die jüngeren Pbn sehr gut, für die älteren Pbn ergibt sich ein nur mäßiger Fit. Gute Fit-Maße weist Modell 3 für beide Altersgruppen auf, jedoch passt Modell 3 für beide Altersgruppen nicht signifikant besser als Modell 2 (jüngere Pbn:  $\Delta\chi^2 = 0.37$ ,  $\Delta df = 2$ ,  $p = .83$ ; ältere Pbn:  $\Delta\chi^2 = 6.18$ ,  $\Delta df = 2$ ,  $p = .05$ ). In Modell 3 korrelieren der verbale und der numerische Faktor bei den jüngeren Pbn extrem hoch ( $r = .99$ ), bei den älteren Pbn etwas geringer ( $r = .91$ ). Erneut ist nicht von einem bedeutsamen Informationsverlust auszugehen, wenn der verbale und der numerische Faktor zusammengefasst werden.

Im Folgenden wird geprüft, ob die Faktorstruktur für beide Gruppen invariant ist. Das Baseline-Modell für beide Altersgruppen zeigt eine gute Datenpassung. Die Nullhypothese, dass beide Gruppen eine äquivalente Kovarianzstruktur aufweisen, muss daher nicht verworfen werden. Die Invarianz beider Gruppen bestätigt sich auch bei der strengeren Prüfung durch die sukzessive Fixierung von Parametern, da der  $\chi^2$ -Test für die verschiedenen  $\Delta\chi^2$  nicht signifikant wird. Es kann somit davon ausgegangen werden, dass die Ladungsmuster und die Faktorkovarianz in beiden Altersgruppen invariant sind.

### *Intelligenzgruppen*

Ferner wird ein Vergleich der Modellpassung für die beiden untersuchten Begabungsgruppen, die Gruppe der hoch Intelligenten (HI,  $N = 170$ ) und die Gruppe der nicht hoch Intelligenten (NHI,  $N = 204$ ), vorgenommen. Zunächst werden hierzu erneut separate Modelle für beide Gruppen geschätzt. Tabelle 7.8 zeigt, dass für die hoch Intelligenten sowohl Modell 2 als auch Modell 3 eine gute Passung aufweist. In Modell 3 korrelieren jedoch der verbale und der numerische Faktor erneut recht hoch ( $r = .91$ ), so dass Modell 2 als eine gute Beschreibung der Struktur angesehen werden kann. Für die Gruppe der nicht hoch Intelligenten findet sich hingegen keine optimale Modellpassung. Die Fitmaße fallen für die Modelle 2 und 3 zwar besser aus als für Modell 1, jedoch erreichen sie keine zufrieden stellenden Werte.

Das bisher am besten passende Modell 2 wird dennoch zusätzlich einer Invarianzprüfung unterzogen. Bei simultaner Schätzung und Fixierung der Kovarianzen kann noch von einer hinreichenden Invarianz ausgegangen werden, bei den strengeren Prüfungen (Fixierung der Ladungen bzw. der Ladungen und Kovarianzen) kann hingegen keine ausreichende Passung mehr festgestellt werden. Es ist somit festzuhalten, dass sich die Struktur des Arbeitsgedächtnisses für die Gruppe der hoch Intelligenten mit Modell 2 gut abbilden lässt. Für die weniger intelligenten Pbn findet sich hingegen keine ausreichende Modellpassung.

Tabelle 7.8: Passung verschiedener Faktormodelle des Arbeitsgedächtnisses für die Teilstichproben „Hoch Intelligente“ ( $N = 170$ ) und „Nicht hoch Intelligente“ ( $N = 204$ )

| Begabungsgruppen /<br>Modell   | $\chi^2$ | <i>df</i> | <i>p</i> | CFI  | RMSEA<br>(Konfidenz) | AGFI |
|--|----------|-----------|----------|------|----------------------|------|
| <i>Modell 1</i>  |          |           |          |      |                      |      |
| Hoch Intelligente  | 24.82    | 9         | .00      | .94  | .10 (.06-.15)        | .89  |
| Nicht hoch Intelligente  | 37.99    | 9         | .00      | .93  | .13 (.09-.17)        | .85  |
| <i>Modell 2</i>  |          |           |          |      |                      |      |
| Hoch Intelligente  | 6.09     | 8         | .64      | 1.00 | .00 (.00-.08)        | .97  |
| Nicht hoch Intelligente  | 22.72    | 8         | .00      | .96  | .10 (.05-.14)        | .91  |
| <i>Modell 3</i>  |          |           |          |      |                      |      |
| Hoch Intelligente  | 3.36     | 6         | .76      | 1.00 | .00 (.00-.07)        | .98  |
| Nicht hoch Intelligente  | 19.29    | 6         | .00      | .97  | .10 (.05-.16)        | .90  |
| <i>Simultane Schätzung für Modell 2 für beide Intelligenzgruppen</i> |          |           |          |      |                      |      |
| Baseline-Modell<br>(Parameter frei)                                  | 28.80    | 16        | .03      | .98  | .05 (.02-.07)        | .94  |
| Kovarianz fix  | 29.15    | 17        | .03      | .98  | .04 (.01-.07)        | .94  |
| Vergleich ( $\Delta\chi^2$ )   | 0.35     | 1         | .55      |      |                      |      |
| Ladungen fix   | 41.24    | 20        | .00      | .97  | .05 (.03-.08)        | .93  |
| Vergleich ( $\Delta\chi^2$ )   | 12.44    | 4         | .01      |      |                      |      |
| Ladungen u. Kov. fix   | 42.11    | 21        | .00      | .97  | .05 (.03-.08)        | .93  |
| Vergleich ( $\Delta\chi^2$ )   | 13.31    | 5         | .02      |      |                      |      |

### *Mädchen und Jungen*

Zusätzlich wird im Folgenden die Hypothese geprüft, dass sich die Faktorstruktur für Jungen und Mädchen nicht unterscheidet. Die Ergebnisse in Tabelle 7.9 zeigen, dass die simultane Schätzung von Modell 2 für Mädchen und Jungen eine gute Datenpassung aufweist (die Abbildungen der einzelnen Modelle finden sich in Anhang B). Die Fixierung der Parameter führt zudem nicht zu einer signifikant schlechteren Modellpassung. Die Hypothese kann somit bestätigt werden.

Tabelle 7.9: Passung des Zwei-Faktoren-Modells des Arbeitsgedächtnisses für Jungen ( $N = 205$ ) und Mädchen ( $N = 168$ )

| Geschlecht   | $\chi^2$ | $df$ | $p$ | CFI  | RMSEA (Konfidenz) | AGFI |
|--|----------|------|-----|------|-------------------|------|
| Jungen   | 13.68    | 8    | .09 | .99  | .06 (.00 - .11)   | .94  |
| Mädchen  | 6.38     | 8    | .61 | 1.00 | .00 (.00 - .08)   | .97  |
| <i>Simultane Schätzung für Modell 2 für beide Geschlechter</i> |          |      |     |      |                   |      |
| Baseline-Modell (Parameter frei)                               | 20.05    | 16   | .22 | 1.00 | .03 (.00 - .06)   | .95  |
| Kovarianz fix  | 20.57    | 17   | .25 | 1.00 | .02 (.00 - .06)   | .96  |
| Vergleich ( $\Delta\chi^2$ , $\Delta df$ )                     | 0.52     | 1    | .47 |      |                   |      |
| Ladungen fix   | 25.90    | 20   | .17 | .99  | .03 (.00 - .06)   | .95  |
| Vergleich ( $\Delta\chi^2$ , $\Delta df$ )                     | 5.85     | 4    | .21 |      |                   |      |
| Ladungen u. Kov. fix   | 25.96    | 21   | .21 | .99  | .03 (.00 - .05)   | .96  |
| Vergleich ( $\Delta\chi^2$ , $\Delta df$ )                     | 5.91     | 5    | .32 |      |                   |      |

#### Fazit zur Binnenstruktur des Arbeitsgedächtnisses

Insgesamt kann festgehalten werden, dass in zwei der drei vorgenommenen Unterteilungen der Stichprobe das Zwei-Faktoren-Modell gut dazu in der Lage ist, die Kovarianzstruktur der Daten zu erklären. Bei der Aufgliederung der Stichprobe nach ihrer Intelligenz fand sich zwar kein guter Fit für Modell 2, jedoch war das domänengenerelle Modell 1 noch weniger gut dazu in der Lage, die Daten zu beschreiben. Auf der Basis dieser Befunde kann somit ein domänengenerelles Modell nicht bestätigt werden. Stattdessen ist es für die hier untersuchte Altersgruppe erforderlich, zwei Subsysteme im Arbeitsgedächtnis anzunehmen, die jedoch nicht unabhängig voneinander sind. Damit weichen die Befunde von den Ergebnissen anderer Studien ab, in denen ebenfalls konfirmatorische Faktorenanalysen verwendet wurden und die ein domänengenerelles Modell stützen. Die vorliegenden Ergebnisse können insbesondere mit den Befunden von Ackerman et al. (2002), Kane et al. (2004), Oberauer et al. (2000) sowie Süß et al. (2002) verglichen werden, die ebenfalls ausgewogene Zusammenstellungen verbaler, numerischer und figuraler Aufgaben verwendeten (siehe Abschnitt 3.3).

Während die Analysen von Ackerman et al. eindeutig ein Ein-Faktor-Modell präferieren, ergab die Untersuchung von Kane et al. (2004) neben einer guten Passung für ein Ein-Faktor-Modell auch einen sehr guten Fit für ein Zwei-Faktor-Modell. Jedoch fand sich hier eine deutlich höhere Korrelation zwischen beiden Faktoren als in der vorliegenden Untersuchung ( $r = .93$  bzw.  $r = .84$ ), so dass die Autoren das Ein-Faktor-Modell als die bessere Beschreibung der Daten akzeptierten. Die Analysen von Oberauer et al. (2000) und Süß et al. (2002) ergaben ebenfalls eine Differenzierung zwischen den Faktoren „verbal-numerisch“ und „räumlich“. Auch hier fanden

sich jedoch sehr hohe Zusammenhänge zwischen den beiden Faktoren ( $r = .82$  bzw.  $r = .80$ ).

In Abschnitt 3.3.4 wurde diskutiert, dass die Verwendung von Stichproben, deren Fähigkeitsvarianz stark eingeschränkt ist, eher zu einer Bestätigung der Spezifitätshypothese führt. Während die oben zitierten Studien relativ fähigkeitshomogene Stichproben untersuchten (überwiegend Studierende), wurde in der vorliegenden Untersuchung jedoch eine Gruppe von Kindern mit breiter Fähigkeitsstreuung untersucht. Die unterschiedlichen Befunde lassen sich daher über die Fähigkeitsvarianz der Stichproben nicht erklären.

Ein weiterer Unterschied der oben genannten Studien zu der vorliegenden Arbeit besteht darin, dass erwachsene Pbn untersucht wurden. Die bisherigen Studien mit Kindern lassen sich jedoch nur bedingt mit den vorliegenden Ergebnissen vergleichen, da sehr unterschiedliche Aufgabenzusammenstellungen verwendet wurden (siehe Abschnitt 3.3.2). So stützen die Befunde von de Jonge und de Jong (1996) zwar ein Ein-Faktor-Modell, jedoch wurden hier keine figuralen Aufgaben verwendet. Die Studie von Jarvis und Gathercole (2003) mit 11- und 14-jährigen Pbn zielte auf die Erfassung des Arbeitsgedächtnisses im Sinne des klassischen Baddeley-Modells, so dass neben komplexen Spannaufgaben vorwiegend einfache Spannaufgaben verwendet wurden. Die Befunde von Jarvis und Gathercole lassen sich – trotz dieser Einschränkung – auch im Hinblick auf eine Unterteilung des Arbeitsgedächtnisses in eine verbale und eine räumliche Komponente deuten (wie es auch das Baddeley-Modell annimmt). Möglicherweise sind die verbal-numerische und die figurale Komponente des Arbeitsgedächtnisses bei Kindern noch weniger stark miteinander verknüpft als bei Erwachsenen. Zur Prüfung dieser Hypothese müssten jedoch in einer größer angelegten Untersuchung verschiedene Altersgruppen mit demselben Aufgabenmaterial systematisch untersucht werden.

Auf der Grundlage der Analysen zur Binnenstruktur des Arbeitsgedächtnisses lässt sich somit ein domänengenerelles Arbeitsgedächtnissystem nicht bestätigen. In Abschnitt 7.3 werden die Befunde zum Zusammenhang der beiden Arbeitsgedächtnisfaktoren mit den anderen kognitiven Maßen berichtet. Diese Analysen geben zusätzlichen Aufschluss über die Frage der Domänenspezifität vs. -generalität. Eine weitere Bestätigung der Spezifitätshypothese wäre gegeben, wenn kognitive Aufgaben mit verbal-numerischem Inhalt stärker mit den verbal-numerischen Arbeitsgedächtnisaufgaben korrespondieren, und kognitive Aufgaben mit figuralen Inhalten eher mit figuralen Arbeitsgedächtnisaufgaben.

#### **7.2.4 Untersuchung von mittleren Leistungsunterschieden zwischen Gruppen**

Im Folgenden wird untersucht, ob sich verschiedene Teilgruppen den Erwartungen entsprechend (vgl. Abschnitt 5.3.1) in ihren mittleren Leistungen in den Arbeitsgedächtnisaufgaben unterscheiden. Die Unterschiede in den Leistungen in den Arbeitsgedächtnisaufgaben der verschiedenen Gruppen werden über t-Tests für unabhängige Stichproben geprüft. Zur Abschätzung der praktischen Bedeutsamkeit der Unter-

schiede werden zusätzlich die Effektstärken ( $d$ ) angegeben. Bei einer Effektstärke ab 0.2 kann von einem kleinen, bei einer Effektstärke ab 0.5 von einem mittleren und bei Werten ab 0.8 von einem großen Effekt ausgegangen werden (Bortz & Döring, 2002).

*Leistungsunterschiede in den Arbeitsgedächtnisaufgaben zwischen jüngeren und älteren Pbn*

Es wird die Hypothese geprüft, dass ältere Pbn im Mittel bessere Leistungen in den Arbeitsgedächtnisaufgaben zeigen als jüngere Pbn. Die Ergebnisse des t-Tests in Tabelle 7.10 bestätigen die Hypothese und zeigen, dass die älteren Pbn ( $> 11.5$  Jahre) im Mittel in allen Aufgaben signifikant besser abschneiden als die jüngeren Pbn ( $\leq 11.5$  Jahre). Die Effektstärken machen deutlich, dass die Unterschiede bei dieser Aufteilung der Stichprobe in zwei Altersgruppen jedoch relativ gering sind.

*Tabelle 7.10: Vergleich der Mittelwerte der Arbeitsgedächtnisaufgaben von jüngeren ( $N = 196$ ) und älteren Pbn ( $N = 177$ )*

| <b>Arbeitsgedächtnis<br/>Aufgabentyp</b> | <b><math>\leq 11.5</math> Jahre<br/><math>M (SD)</math></b> | <b><math>&gt; 11.5</math> Jahre<br/><math>M (SD)</math></b> | <b>t-Test<br/><math>t (df)</math></b> | <b>Effekt-<br/>stärke<br/><math>d</math></b> |
|--|---|---|---------------------------------------|--|
| Reading Span                             | 3.23 (2.43)   | 4.45 (2.55)   | -4.70 (370)                           | 0.49   |
| Verbal Span                              | 3.53 (2.63)   | 4.57 (2.56)   | -3.86 (371)                           | 0.40   |
| Comp. Span                               | 3.98 (2.33)   | 5.16 (2.17)   | -5.05 (371)                           | 0.52   |
| Backw. Digit Span                        | 3.30 (2.56)   | 4.60 (2.91)   | -4.59 (371)                           | 0.47   |
| Memory Upd. Fig.                         | 4.78 (2.74)   | 6.31 (3.03)   | -5.08 (370)                           | 0.53   |
| Spatial WM                               | 4.52 (2.57)   | 5.26 (2.97)   | -2.55 (349.88) <sup>1</sup>           | 0.27   |

Anmerkung. t-Test: einseitige Testung, alle  $p < .01$ ; <sup>1</sup> = Levene-Test auf Varianzgleichheit:  $p < .05$ .

*Leistungsunterschiede in den Arbeitsgedächtnisaufgaben zwischen hoch intelligenten und nicht hoch intelligenten Pbn*

Als nächstes wird die Hypothese untersucht, dass hoch intelligente Pbn im Mittel höhere Leistungen in den Arbeitsgedächtnisaufgaben zeigen als nicht hoch intelligente Pbn. Die Betrachtung der Mittelwerte beider Gruppen in Tabelle 7.11 zeigt, dass die hoch intelligenten Pbn (HI) in allen sechs Arbeitsgedächtnisaufgaben im Mittel höhere Werte erzielen als die restlichen Pbn (NHI). Beide Gruppen unterscheiden sich nicht signifikant in ihrem mittleren Alter ( $M_{HI} = 11.38$ ;  $M_{NHI} = 11.49$ ;  $t = .96$ ,  $df = 372$ ,  $p > .05$ ; zweiseitige Testung), so dass eine zusätzliche Differenzierung nach Altersgruppen an dieser Stelle nicht notwendig ist.

Die hoch Intelligenten schneiden nach dem t-Test in allen sechs Aufgabengruppen signifikant besser ab als die weniger Intelligenten, die Hypothese kann somit bestätigt werden. Die Effektstärken zeigen, dass die Leistungsunterschiede zwischen bei-

den Gruppen überwiegend als groß einzuschätzen sind. Obwohl die Grundlage der Gruppenbildung der figurale Intelligenztest CFT 20 ist, finden sich hier die größten Unterschiede zwischen weniger und hoch intelligenten Pbn für die beiden verbalen Aufgabengruppen.

Tabelle 7.11: Vergleich der Mittelwerte der Arbeitsgedächtnisaufgaben der Teilstichproben „hoch Intelligente“ (HI;  $N = 170$ ) und „nicht hoch Intelligente“ (NHI;  $N = 204$ )

| Arbeitsgedächtnis<br>Aufgabentyp | HI<br>$M (SD)$ | NHI<br>$M (SD)$ | t-Test<br>$t (df)$          | Effekt-<br>stärke<br>$d$ |
|----------------------------------|----------------|-----------------|-----------------------------|--------------------------|
| Reading Span                     | 5.06 (2.26)    | 2.74 (2.31)     | -9.77 (371)                 | 1.02                     |
| Verbal Span                      | 5.25 (2.28)    | 2.99 (2.49)     | -9.11 (372)                 | 0.95                     |
| Comp. Span                       | 5.39 (2.20)    | 3.83 (2.19)     | -6.88 (372)                 | 0.71                     |
| Backw. Digit Span                | 5.19 (2.78)    | 2.84 (2.35)     | -8.71 (331.51) <sup>1</sup> | 0.91                     |
| Memory Upd. Fig.                 | 6.68 (2.75)    | 4.52 (2.80)     | -7.46 (371)                 | 0.78                     |
| Spatial WM                       | 5.98 (2.64)    | 3.94 (2.56)     | -7.58 (371)                 | 0.79                     |

Anmerkung. t-Test: einseitige Testung, alle  $p < .01$ ; <sup>1</sup> = Levene-Test auf Varianzgleichheit:  $p < .05$ .

#### Leistungsunterschiede zwischen Mädchen und Jungen

Abschließend werden nun die Hypothesen zu Leistungsunterschieden zwischen den Geschlechtern geprüft. Es wird erwartet, dass Jungen in den „Spatial Working Memory“-Aufgaben besser abschneiden als Mädchen, da diese Aufgaben eine mentale Rotation erfordern und die Überlegenheit von männlichen Pbn bei dieser Anforderung bereits häufig belegt wurde. Bei den anderen fünf Aufgabentypen sollten sich Mädchen und Jungen im Mittel nicht unterscheiden. Zwar wird häufig über eine mögliche Überlegenheit weiblicher Pbn in verbalen Aufgaben berichtet, jedoch bestätigen neuere Metaanalysen zu verbalen Fähigkeiten diese Befunde nicht (siehe Abschnitt 5.3.2).

Die Tabelle 7.12 zeigt, dass Jungen etwas höhere Werte als Mädchen in den figuralen Aufgaben erzielen. Entgegen den Erwartungen erreichen Mädchen im Mittel etwas höhere Werte in den verbalen Aufgaben als Jungen. Die Mittelwerte beider Geschlechter in den Arbeitsgedächtnisaufgaben unterscheiden sich nach dem t-Test jedoch nicht signifikant.

Die Gesamtstichprobe enthält mehr Jungen als Mädchen aus Hochbegabteinrichtungen (117 Jungen und 84 Mädchen). Daher könnte die mittlere Leistung der Jungen in den Arbeitsgedächtnisaufgaben besser ausfallen. Aus diesem Grund wird zur Prüfung von Geschlechterunterschieden zusätzlich die Gruppe der Unausgelesenen betrachtet, bei der davon ausgegangen werden kann, dass sich die Geschlechter nicht in ihren allgemeinen kognitiven Fähigkeiten unterscheiden (siehe Tabelle 7.13). Der mittlere Leistungsunterschied zwischen Mädchen und Jungen in der Aufgabe „Verbal Span“ zeigt sich hier deutlicher: Mädchen erzielen im Mittel einen Score von

3.65 ( $SD = 2.63$ ), Jungen einen Score von 2.61 ( $SD = 2.23$ ). Dieser Unterschied ist hochsignifikant, die Bedeutsamkeit ist jedoch eher als gering einzuschätzen, wie die Effektstärke von  $d = 0.43$  zeigt. In allen anderen Aufgabengruppen unterscheidet sich auch die mittlere Leistung der Jungen und Mädchen aus der unausgelesenen Gruppe nicht signifikant. Somit kann die Annahme, dass Jungen in den „Spatial Working Memory“-Aufgaben besser abschneiden als Mädchen, nicht bestätigt werden.

Tabelle 7.12: Vergleich der Mittelwerte in den Arbeitsgedächtnis-Scores von Mädchen ( $N = 168$ ) und Jungen ( $N = 205$ ) in der Gesamtstichprobe

| Arbeitsgedächtnis<br>Aufgabentyp | Mädchen<br>$M (SD)$ | Jungen<br>$M (SD)$ | t-Test<br>$t (df)$ | Effektstärke<br>$d$ |
|----------------------------------|---------------------|--------------------|--------------------|---------------------|
| Reading Span                     | 3.88 (2.58)         | 3.74 (2.55)        | 0.55 (370)         | 0.06                |
| Verbal Span                      | 4.31 (2.69)         | 3.78 (2.55)        | 1.94 (371)         | 0.20                |
| Computation Span                 | 4.57 (2.50)         | 4.52 (2.18)        | 0.20 (334)         | 0.02                |
| Backw. Digit Span                | 4.05 (2.80)         | 3.80 (2.82)        | 0.86 (370)         | 0.09                |
| Memory Upd. Fig.                 | 5.30 (3.08)         | 5.67 (2.89)        | -1.19 (371)        | -0.12               |
| Spatial WM                       | 4.61 (2.73)         | 5.05 (2.84)        | -1.54 (370)        | -0.16               |

Anmerkung. t-Test: einseitige Testung, alle  $p > .05$ . Levene-Test auf Varianzgleichheit: alle  $p > .05$ .

Tabelle 7.13: Vergleich der Mittelwerte in den Arbeitsgedächtnis-Scores von Mädchen ( $N = 84$ ) und Jungen ( $N = 88$ ) in der Teilstichprobe „Unausgelesene Pbn“

| Arbeitsgedächtnis<br>Aufgabentyp | Mädchen<br>$M (SD)$ | Jungen<br>$M (SD)$ | t-Test<br>$t (df)$           | Effektstärke<br>$d$ |
|----------------------------------|---------------------|--------------------|------------------------------|---------------------|
| Reading Span                     | 2.99 (2.35)         | 2.43 (1.91)        | 1.70 (159.74)                | 0.26                |
| Verbal Span                      | 3.65 (2.63)         | 2.61 (2.23)        | 2.79 (162.81) <sup>1**</sup> | 0.43                |
| Computation Span                 | 3.90 (2.52)         | 3.89 (2.02)        | 0.05 (159.10) <sup>1</sup>   | 0.00                |
| Backw. Digit Span                | 3.21 (2.50)         | 2.63 (2.04)        | 1.70 (170)                   | 0.26                |
| Memory Upd. Fig.                 | 4.39 (3.01)         | 4.70 (2.56)        | -0.73 (162.93) <sup>1</sup>  | -0.11               |
| Spatial WM                       | 3.57 (2.35)         | 3.92 (2.38)        | -0.97 (170)                  | -0.15               |

Anmerkung. t-Test: einseitige Testung, \*\* =  $p < .01$ ; <sup>1</sup> = Levene-Test auf Varianzgleichheit:  $p < .05$ .

### 7.3 Prüfung der Zusammenhänge der Arbeitsgedächtnisfaktoren mit kognitiven Fähigkeiten

Die Konstrukte „Intelligenz“, „Geschwindigkeit der Informationsverarbeitung“ („Speed“) und „Kurzzeitgedächtnis“ werden jeweils über mehrere Indikatoren erfasst. Es kann daher der Zusammenhang dieser Konstrukte mit dem Arbeitsgedächtnis über latente Faktoren betrachtet werden. Bei den Konstrukten „Intelligenz“ und „Speed“ ist es hierzu zunächst erforderlich, die jeweilige Binnenstruktur zu explorieren, da die Aufgabenauswahl aufgrund theoretischer Überlegungen jeweils ein Zwei-Faktoren-Modell nahe legt, das jedoch an den vorliegenden Daten geprüft

werden soll. Das Kurzzeitgedächtnis wird lediglich über drei reine Merkfähigkeitsaufgaben erfasst, von denen angenommen wird, dass sie einen gemeinsamen Faktor bilden.

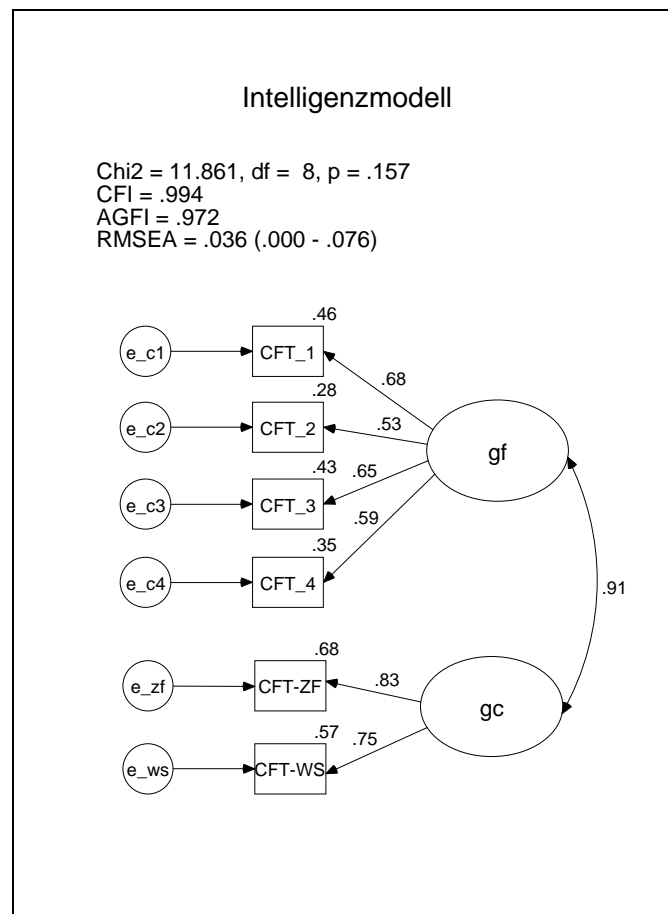
### 7.3.1 Fluide und kristalline Intelligenz

Die Intelligenz der Pbn wird insgesamt über sechs Indikatoren erfasst: die vier figuralen Subtests des CFT 20, den Wortschatztest des CFT 20 (CFT-WS) und den Zahlenfolgen-Test des CFT 20 (CFT-ZF). Aufgrund der theoretischen Konzeption des CFT 20 und seiner beiden Zusatzmodule wird zunächst entsprechend Cattells Intelligenzmodell von einem Zwei-Faktoren-Intelligenzmodell ausgegangen. Die vier Subtests des CFT 20 sollten diesem Modell zufolge hoch auf einem fluiden Intelligenzfaktor ( $g_f$ ) laden, die beiden Zusatzmodule auf einem kristallinen Intelligenzfaktor ( $g_c$ ). Eine alternative Interpretation, die vom Testautor vorgeschlagen wird (Weiß, 1998), versteht den CFT 20 und seine beiden Zusatzmodule als Maße für die Operation „Verarbeitungskapazität“ des BIS-Modells (vgl. Abschnitt 2.2). Der CFT 20 erfasst nach dieser Auslegung die figurale, der Wortschatztest die verbale und der Zahlenfolgentest die numerische Verarbeitungskapazität. Da die beiden Zusatzmodule jedoch jeweils nur aus einem Untertest bestehen, lässt sich dieses Modell hier nicht überprüfen.

Ein entsprechendes Modell ergibt einen guten Fit, die Faktorenkorrelation fällt mit  $r = .91$  erwartungsgemäß hoch aus (siehe Abbildung 7.5). Zum Vergleich wird zusätzlich ein Ein-Faktor-Modell gerechnet, in dem alle Indikatoren auf einen gemeinsamen Intelligenzfaktor laden. Dieses Modell erbringt einen signifikant schlechteren Fit als das Zwei-Faktoren-Modell ( $\chi^2 = 19.33$ ,  $df = 9$ ,  $p = .02$ ; CFI = .98; AGFI = .96; RMSEA = .06, RMSEA Konfidenz = .02-.09;  $\Delta\chi^2 = 7.47$ ,  $\Delta df = 1$ ,  $p = .01$ ).

Es kann somit angenommen werden, dass die verschiedenen eingesetzten Intelligenztests zwei Intelligenzkomponenten erfassen, die als fluide und als kristalline Komponente interpretiert werden können. Im Rahmen des BIS-Modells lassen sich die Faktoren des Modells auch als verbal-numerische Verarbeitungskapazität und als figurale Verarbeitungskapazität verstehen.





Anmerkung. CFT\_1 = Subtest „Series“, CFT\_2 = Subtest „Classifications“, CFT\_3 = Subtest „Matrices“, CFT\_4 = Subtest „Topologies“, CFT-ZF = Zahlenfolgentest, CFT-WS = Wortschatztest.

Abbildung 7.5: Zwei-Faktorenmodell der Intelligenz

### *Zusammenhänge der Arbeitsgedächtnisaufgaben mit den Intelligenztests*

Neben der Untersuchung des Zusammenhangs von Arbeitsgedächtnis und Intelligenz auf der Ebene der Konstrukte (siehe Abschnitt 7.3.3) ist weiterhin der Zusammenhang der einzelnen Aufgabentypen mit Maßen für die fluide und die kristalline Intelligenz von Interesse. Verschiedene Studien weisen darauf hin, dass sich bestimmte Arbeitsgedächtnisaufgaben – insbesondere wurden hier Rückwärtsspannen-Aufgaben wie die „Backward Digit Span“ diskutiert – bei überdurchschnittlich fähigen Pbn nicht mehr zur Vorhersage von höheren kognitiven Leistungen eignen (z. B. Schofield & Ashman, 1986). Es wurde vermutet, dass die Verarbeitungsanforderung bestimmter Arbeitsgedächtnisaufgaben für sehr fähige Pbn nur minimal ist, so dass im Wesentlichen nur noch die Speicherkapazität erfasst wird (Engle, Tuholski et al., 1999; vgl. Abschnitt 4.1.2).

Bereits die Mittelwertsunterschiede in den Arbeitsgedächtnisaufgaben zwischen Pbn mit hoher und mit geringerer fluider Intelligenz (s. o.) deuteten darauf hin, dass alle Arbeitsgedächtnisaufgaben mit der im CFT 20 gemessenen Intelligenz korrespondieren. An dieser Stelle werden zusätzlich die Korrelationen zwischen den Arbeitsgedächtnisaufgaben und allen drei verwendeten Intelligenztests sowohl für die Gesamtstichprobe als auch für die Gruppe der „hoch Intelligenten“ (HI) berichtet. Es werden hier die Rohwerte aus den standardisierten Tests verwendet, da sie – ebenso wie die Arbeitsgedächtnis-Scores – noch keine Alterskorrektur enthalten. Bei den Korrelationen zwischen Arbeitsgedächtnis- und Testscores wird daher das Alter auspartialisiert.

Für die Gesamtstichprobe zeigt sich, dass alle Scores der Arbeitsgedächtnisaufgaben hochsignifikant mit den drei standardisierten Intelligenztests korrelieren (siehe Tabelle 7.14). Die Zusammenhänge liegen in mittlerer Höhe, die höchsten Zusammenhänge finden sich insgesamt für den Score der Aufgabe „Reading Span“. Die Partialkorrelationen in der Gruppe der hoch Intelligenten fallen durchweg etwas geringer aus, was auf die eingeschränkte Fähigkeitsvarianz in dieser Teilstichprobe zurückgeführt werden kann.

*Tabelle 7.14: Partialkorrelationen der Arbeitsgedächtnisaufgaben mit den Intelligenztests (Rohwerte, Alter kontrolliert) für die Gesamtstichprobe (N = 374) und die Teilstichprobe „hoch Intelligente“ (N = 170)*

| <b>Arbeitsgedächtnis<br/>Aufgabentyp</b> | <b>CFT 20</b> | <b>CFT-ZF</b> | <b>CFT-WS</b> |
|--|---------------|---------------|---------------|
| <i>Gesamtstichprobe (N = 374)</i>        |               |               |               |
| Reading Span                             | .52**         | .53**         | .63**         |
| Verbal Span                              | .49**         | .44**         | .51**         |
| Computation Span                         | .43**         | .51**         | .40**         |
| Backward Digit Span                      | .49**         | .47**         | .43**         |
| Memory Updating Figural                  | .51**         | .51**         | .40**         |
| Spatial Working Memory                   | .54**         | .46**         | .36**         |
| <i>Hoch intelligente Pbn (N = 170)</i>   |               |               |               |
| Reading Span                             | .35**         | .39**         | .49**         |
| Verbal Span                              | .25**         | .28**         | .32**         |
| Computation Span                         | .23**         | .40**         | .20**         |
| Backward Digit Span                      | .31**         | .32**         | .25**         |
| Memory Updating Figural                  | .25**         | .39**         | .26**         |
| Spatial Working Memory                   | .24**         | .28**         | .10           |

Anmerkung. CFT-WS = Wortschatztest des CFT 20, CFT-ZF = Zahlenfolgentest des CFT 20, ZVT = Zahlen-Verbindungs-Test. \*\*  $p < .01$ .

Die Zusammenhänge zwischen den verbalen und den numerischen Arbeitsgedächtnisaufgaben einerseits und der Leistung in den Tests CFT 20, CFT-ZF und CFT-WS

andererseits sind auch in dieser Gruppe alle signifikant. Somit erweist sich auch die Aufgabe „Backward Digit Span“ – ebenso wie die anderen fünf Arbeitsgedächtnisaufgaben – als geeignet, zur Vorhersage höherer kognitiver Leistungen beizutragen. Die Ergebnisse von Schofield und Ashman (1986) können auf der Basis dieser Befunde nicht bestätigt werden (vgl. Abschnitt 4.1.2).

In der Gruppe der hoch Intelligenten findet sich lediglich für die figurale Aufgabe „Spatial Working Memory“ kein signifikanter Zusammenhang mit dem Wortschatztest. Dies kann als ein weiterer Hinweis auf die Unterteilung des Arbeitsgedächtnisses in eine verbal-numerische und eine figurale Komponente interpretiert werden – wie sie auch die Analysen zur Binnenstruktur in Abschnitt 7.2 zeigen konnten. Jedoch ist zu berücksichtigen, dass diese Aufgabe in der Gesamtstichprobe, die eine größere Fähigkeitsvarianz beinhaltet, auch signifikant mit dem Wortschatztest korreliert. Diese Ergebnisse sind insgesamt als ein Indiz für die Validität der sechs Aufgaben auch für den sehr fähigen Teil der Stichprobe zu deuten.

### 7.3.2 Geschwindigkeit der Informationsverarbeitung

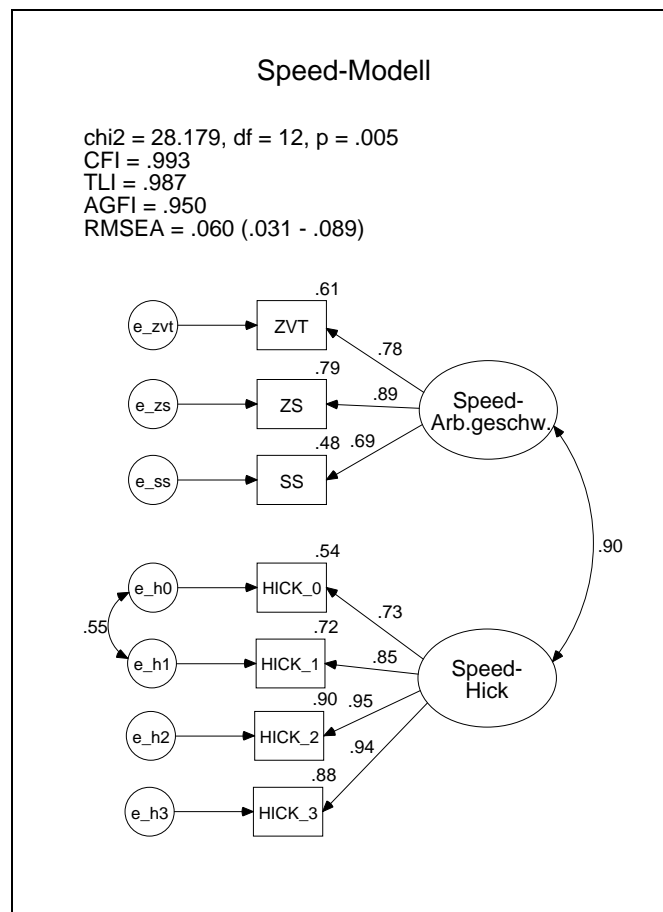
Hinsichtlich der Geschwindigkeit der Informationsverarbeitung sind entsprechend den Fragestellungen in Abschnitt 5.3.1 die folgenden vier Fragen zu beantworten:

- 1) Wie hoch ist der Zusammenhang von Arbeitsgedächtnis und Speed bei Kindern der hier untersuchten Altersgruppe?
- 2) Ist der Zusammenhang zwischen Arbeitsgedächtnis und Speed geringer als zwischen dem Arbeitsgedächtnis und der fluiden und der kristallinen Intelligenz?
- 3) Hängen die verschiedenen Speed-Maße, von denen angenommen wird, dass sie unterschiedlich komplexe kognitive Prozesse erfordern, unterschiedlich stark mit dem Arbeitsgedächtnis zusammen?
- 4) Sind die Hick-Aufgaben im Paper-Pencil-Format dazu geeignet, bei durchschnittlich und hoch begabten Kindern zwischen 8 und 13 Jahren relativ elementare Prozesse der Informationsverarbeitungsgeschwindigkeit reliabel und valide zu erfassen?

Die Beantwortung der ersten beiden Fragen erfolgt in Abschnitt 7.3.3, in dem die Zusammenhänge der verschiedenen kognitiven Konstrukte mit dem Arbeitsgedächtnis betrachtet werden. An dieser Stelle ist hierfür zunächst erforderlich, die Faktorstruktur der verwendeten Speed-Maße zu überprüfen. Insgesamt werden drei standardisierte Tests zur Erfassung der Geschwindigkeit der Informationsverarbeitung (die Untertests „Zahlen-Symbol-Test“ und „Symbolsuche“ aus dem HAWIK III sowie der Zahlen-Verbindungs-Test ZVT) und die Aufgaben nach dem Hick-Paradigma (Hick-Aufgaben Hick-0 bis Hick-3) eingesetzt.

Es wird vermutet, dass die drei standardisierten Tests, die auch für die Intelligenzdiagnostik eingesetzt werden, etwas komplexere geistige Prozesse erfordern als die relativ elementaren Hick-Aufgaben. Aus diesem Grund wird eine gute Passung für ein obliques Zwei-Faktoren-Modell erwartet, in dem die Hick-Aufgaben auf den einen und die standardisierten Tests auf den anderen Faktor laden.

Zunächst wird jedoch mit einem Ein-Faktor-Modell, in dem alle Indikatoren auf einen gemeinsamen Speed-Faktor laden, das sparsamste Modell mittels einer konfirmatorischen Faktorenanalyse überprüft. Für dieses Modell resultiert ein mangelhafter Fit ( $\chi^2 = 208.20$ ,  $df = 14$ ,  $p = .00$ ; CFI = .91; AGFI = .71; RMSEA = .19, RMSEA Konfidenz = .17-.22). Entsprechend der Ausgangshypothese, dass die Hick-Aufgaben und die anderen Speed-Aufgaben leicht unterschiedlich komplexe Prozesse erfordern, wird daraufhin ein Zwei-Faktoren-Modell mit den korrelierten Faktoren „Speed-Arbeitsgeschwindigkeit“, auf den der „Zahlen-Symbol-Test“, die „Symbol-suche“ und der ZVT laden, und „Speed-Hick“, auf den die vier Hick-Aufgaben laden, spezifiziert (siehe Abbildung 7.6). Für dieses Modell ergibt sich eine immer noch nur mäßige, aber signifikant bessere Passung ( $\chi^2 = 144.92$ ,  $df = 13$ ,  $p = .00$ ; CFI = .94; AGFI = .76; RMSEA = .17, RMSEA Konfidenz = .14-.19;  $\Delta\chi^2 = 63.28$ ,  $\Delta df = 1$ ,  $p = .00$ ).



Anmerkung. ZVT = Zahlen-Verbindungs-Test, ZS = Zahlen-Symbol-Test (HAWIK III), SS = Symbolsuche (HAWIK III).

Abbildung 7.6: Zwei-Faktoren-Modell für Speed

Aufgrund der beträchtlichen standardisierten Residualkovarianz ( $z = 2.43$ ) von Hick-0 und Hick-1 und der großen Ähnlichkeit beider Aufgaben (beide stellen mit nur Null bzw. einem Bit äußerst geringe Anforderungen an eine kognitive Verarbeitung; vgl. Faktorenanalyse von Roberts et al., 1999; siehe Abschnitt 5.3.1) wird eine erneute Modellschätzung vorgenommen, bei der eine Korrelation der Residuen von Hick-0 und Hick-1 zugelassen wird ( $r = .55$ ). Diese Modellvariante erbringt eine deutliche Verbesserung des Fits ( $\chi^2 = 28.12$ ,  $df = 12$ ,  $p = .01$ ; CFI = .99; AGFI = .95; RMSEA = .06, RMSEA Konfidenz = .03-.09). Daher wird dieses Speed-Modell als Grundlage für das gemeinsame Modell latenter Faktoren (s. u.) verwendet, anhand dessen die ersten beiden der oben aufgeführten Fragen beantwortet werden können.

Die dritte Frage wird anhand der Korrelationen auf der Ebene der Testaufgaben beantwortet. Der Vorteil der Hick-Aufgaben besteht darin, dass genau bekannt ist, welches Ausmaß von Informationen in jeder der Aufgaben verarbeitet werden muss. Während bei der Aufgabe Hick-0 keine Wahlreaktion erfolgen muss (es kommt nur auf ein möglichst schnelles Eintragen von Pfeilen in die Kästchen an), müssen in den Aufgaben Hick-1 bis Hick-3 bei der Bearbeitung der Items zwischen ein und drei Bit(s) verarbeitet werden. Es wird nun vermutet, dass die Zusammenhänge der Arbeitsgedächtnisaufgaben mit den Hick-Aufgaben umso größer ausfallen, je mehr Bits in den Items der Hick-Aufgaben verarbeitet werden müssen.

Tatsächlich zeigt sich (siehe Tabelle 7.15), dass der Zusammenhang mit zunehmender Komplexität der Hick-Aufgaben steigt. Die Zusammenhänge zwischen den Arbeitsgedächtnis-Scores und den Hick-Aufgaben 1 bis 3 liegen insgesamt in niedriger bis mittlerer Höhe und sind hochsignifikant. Die Aufgabe Hick-0 korreliert nur schwach mit den verbalen Aufgaben und der Aufgabe „Backward Digit Span“, jedoch nicht mit den anderen Arbeitsgedächtnis-Scores. Für Hick-1 finden sich bereits geringe signifikante Korrelationen mit allen Arbeitsgedächtnis-Scores. Die komplexeren Aufgaben Hick-2 und Hick-3 korrelieren beide in mittlerer Höhe mit allen Arbeitsgedächtnisaufgaben. Die Annahme, dass die Höhe des Zusammenhangs zwischen den Arbeitsgedächtnisaufgaben und den Speed-Aufgaben von der Komplexität der Anforderungen der Speed-Aufgaben abhängt, lässt sich somit bestätigen.

*Tabelle 7.15: Partialkorrelationen der Arbeitsgedächtnisaufgaben mit Speed-Aufgaben (Alter kontrolliert)*

| <b>Arbeitsgedächtnis-aufgaben</b> | <b>Hick-0</b> | <b>Hick-1</b> | <b>Hick-2</b> | <b>Hick-3</b> |
|-----------------------------------|---------------|---------------|---------------|---------------|
| Reading Span                      | .20           | .36           | .43           | .41           |
| Verbal Span                       | .06 n.s.      | .21           | .37           | .35           |
| Computation Span                  | .08 n.s.      | .25           | .33           | .27           |
| Backward Digit Span               | .21           | .34           | .39           | .38           |
| Memory Updating Figural           | .19           | .31           | .39           | .37           |
| Spatial Working Memory            | .01 n.s.      | .16           | .22           | .22           |

Anmerkung. Alle Korrelationen  $p < .01$ ; n.s.:  $p > .05$ .

*Eignen sich Hick-Aufgaben für die Erfassung elementarer Informationsverarbeitungsgeschwindigkeit bei durchschnittlich und hoch begabten Kindern?*

Zur Beantwortung der oben genannten vierten Frage werden in Tabelle 7.16 zunächst die Mittelwerte, Standardabweichungen, der Range der Rohwerte sowie Angaben zu den Verteilungen der vier Hick-Aufgaben dargestellt. Die mittlere Leistung in den Hick-Aufgaben sinkt erwartungsgemäß von der einfachsten zur komplexesten Hick-Aufgabe: In stets demselben Zeitintervall von 60 s bewältigen die Pbn in den Aufgaben mit mehr zu verarbeitenden Bits deutlich weniger Items als in den Aufgaben, in denen weniger Bits verarbeitet werden müssen. Der Kolmogorov-Smirnov-Test zeigt, dass die Verteilungen der Aufgabenscores bei Hick-0 bis Hick-2 nicht signifikant von einer Normalverteilung abweichen; die Prüfstatistik für die Verteilung bei Hick-3 überschreitet hingegen die Signifikanzgrenze. Alle Verteilungen sind leicht rechtsschief (positive Schiefe). Da für die hier verwendete Aufgabenversion keine Daten zur Messgüte vorliegen, wird die Reliabilität an der eigenen Stichprobe überprüft. Mit 140 Pbn (aus verschiedenen Teilstichproben, im Mittel 11.19 Jahre alt,  $SD = 1.18$ ) wurde eine Retestung durchgeführt. Im ersten Teil der Gesamtuntersuchung erfolgte die Ersttestung mit den vier Hick-Aufgaben, im zweiten Teil der Untersuchung (nach ein bis sieben Tagen) die Retestung mit denselben Aufgaben. Alle vier Hick-Aufgaben erweisen sich dabei als ausreichend stabil.

Die Hick-Aufgaben korrelieren hoch untereinander (siehe Tabelle 7.17), und die innere Konsistenz über alle vier Hick-Aufgaben erweist sich mit  $\alpha = .93$  (Cronbachs Alpha) als hoch. Es zeigen sich bedeutsame Zusammenhänge mit den anderen in dieser Studie verwendeten Tests zur Erfassung der Geschwindigkeit der Informationsverarbeitung. Mit zunehmender kognitiver Anforderung von der einfachen Aufgabe Hick-0, die noch keine Wahlreaktion erfordert, bis zur kognitiv schon anspruchsvolleren Aufgabe Hick-3 steigen die Zusammenhänge mit den anderen Speed-Tests (siehe Tabelle 7.17). Die Hick-Aufgaben korrelieren am höchsten mit dem Zahlen-Symbol-Test aus dem HAWIK III, der den Hick-Aufgaben strukturell ähnlich ist. (Im Zahlen-Symbol-Test müssen den Zahlen von 1 bis 9 verschiedene einfache Symbole zugeordnet werden.)

*Tabelle 7.16: Mittelwerte, Standardabweichungen, Range der Rohwerte, Verteilungen und Retest-Reliabilitäten der Aufgaben Hick-0 bis Hick-3*

| Aufgabe             | <i>M</i> | <i>SD</i> | Range | Schiefe | Exzess | K-S-Test<br>Z | $r_{tt}$<br>( <i>N</i> = 140) |
|---------------------|----------|-----------|-------|---------|--------|---------------|-------------------------------|
| Hick-0 <sup>1</sup> | 49.66    | 11.66     | 15-91 | .22     | .18    | .04           | .84                           |
| Hick-1 <sup>2</sup> | 44.54    | 12.21     | 10-80 | .11     | -.03   | .04           | .90                           |
| Hick-2 <sup>3</sup> | 35.69    | 10.16     | 13-62 | .20     | -.34   | .05           | .91                           |
| Hick-3 <sup>1</sup> | 29.88    | 8.39      | 8-57  | .22     | -.25   | .06**         | .90                           |

Anmerkung. K-S-Test: Kolmogorov-Smirnov-Test auf Normalverteilung, Signifikanzkorrektur nach Lilliefors, \*\*  $p < .01$ . <sup>1</sup>:  $N = 373$ ; <sup>2</sup>:  $N = 370$ ; <sup>3</sup>:  $N = 369$ . Alle Korrelationen:  $p < .01$ .

Tabelle 7.17: (Partial-)Korrelationen der Hick-Aufgaben (0 bis 3) untereinander und Partiale Korrelationen der Hick-Aufgaben mit den standardisierten Speed-Tests (Rohwerte, Alter kontrolliert)

| Aufgabe             | Hick-0 | Hick-1 | Hick-2 | Hick-3 | ZVT* | Symb.-Suche | Zahlen-Symb.Test |
|---------------------|--------|--------|--------|--------|------|-------------|------------------|
| Hick-0 <sup>1</sup> | 1.00   | .82    | .71    | .68    | .33  | .27         | .49              |
| Hick-1 <sup>2</sup> | .77    | 1.00   | .81    | .79    | .48  | .42         | .64              |
| Hick-2 <sup>3</sup> | .63    | .76    | 1.00   | .88    | .60  | .48         | .71              |
| Hick-3 <sup>1</sup> | .61    | .74    | .86    | 1.00   | .61  | .49         | .68              |

Anmerkung. Korrelationen der Hick-Aufgaben untereinander: oberhalb der Diagonale: bivariate Korrelationen, unterhalb der Diagonale: Partiale Korrelationen, Alter kontrolliert.

<sup>1</sup>:  $N = 373$ ; <sup>2</sup>:  $N = 370$ ; <sup>3</sup>:  $N = 369$ . \* Mittelwert der bit/sec-Rohwerte der vier ZVT-Matrizen. Alle Korrelationen  $p < .01$ .

Schließlich wird untersucht, ob die Gruppe der hoch Intelligenten (HI) in den Hick-Aufgaben im Mittel bessere Leistungen erbringt als die Gruppe der nicht hoch Intelligenten (NHI). Tabelle 7.18 zeigt die Mittelwerte der beiden Gruppen für alle vier Hick-Aufgaben und die Ergebnisse des t-Tests. Es zeigt sich, dass die hoch intelligenten Pbn in den Aufgaben Hick-1 bis Hick-3 im Mittel signifikant mehr Items richtig bearbeiteten als die nicht hoch intelligenten Pbn. Für die Aufgabe Hick-0 findet sich kein bedeutsamer Mittelwertsunterschied ( $p > .05$ ). Die beiden Gruppen unterscheiden sich somit nicht in ihrer Geschwindigkeit beim Zeichnen von Pfeilen. Sobald jedoch eine einfache Entscheidung zwischen zwei Reaktionsmöglichkeiten hinzu kommt (Hick-1), sind die hoch Intelligenten im Vorteil. Wie die Effektstärken zeigen, handelt es sich jedoch um einen eher kleinen Unterschied.

Tabelle 7.18: Vergleich der Mittelwerte der Hick-Aufgaben der Teilstichproben „hoch Intelligente“ (HI;  $N = 170$ ) und „nicht hoch Intelligente“ (NHI;  $N = 203$ )

| Hick-Aufgaben | HI<br>$M (SD)$ | NHI<br>$M (SD)$ | t-Test<br>$t (df)$ | Effektstärke<br>$d$ |
|---------------|----------------|-----------------|--------------------|---------------------|
| Hick-0        | 49.95 (11.24)  | 49.41 (12.02)   | -0.45 (371)        | 0.05                |
| Hick-1        | 46.95 (12.20)  | 42.57 (11.88)   | -3.49 (368)**      | 0.36                |
| Hick-2        | 37.62 (10.25)  | 34.04 (9.82)    | -3.43 (367)**      | 0.36                |
| Hick-3        | 31.62 (8.19)   | 28.43 (8.31)    | -3.71 (371)**      | 0.39                |

Anmerkung. t-Test: einseitige Testung; \*\*  $p < .01$ . Levene-Test auf Varianzgleichheit: alle  $p > .05$ .

Insgesamt kann festgehalten werden, dass sich die Hick-Aufgaben im Paper-Pencil-Format für die hier untersuchte Stichprobe als reliable und extern valide Maße für eine basale Form der Informationsverarbeitungsgeschwindigkeit erweisen.

### 7.3.3 Gesamtmodell: Arbeitsgedächtnis und andere kognitive Faktoren

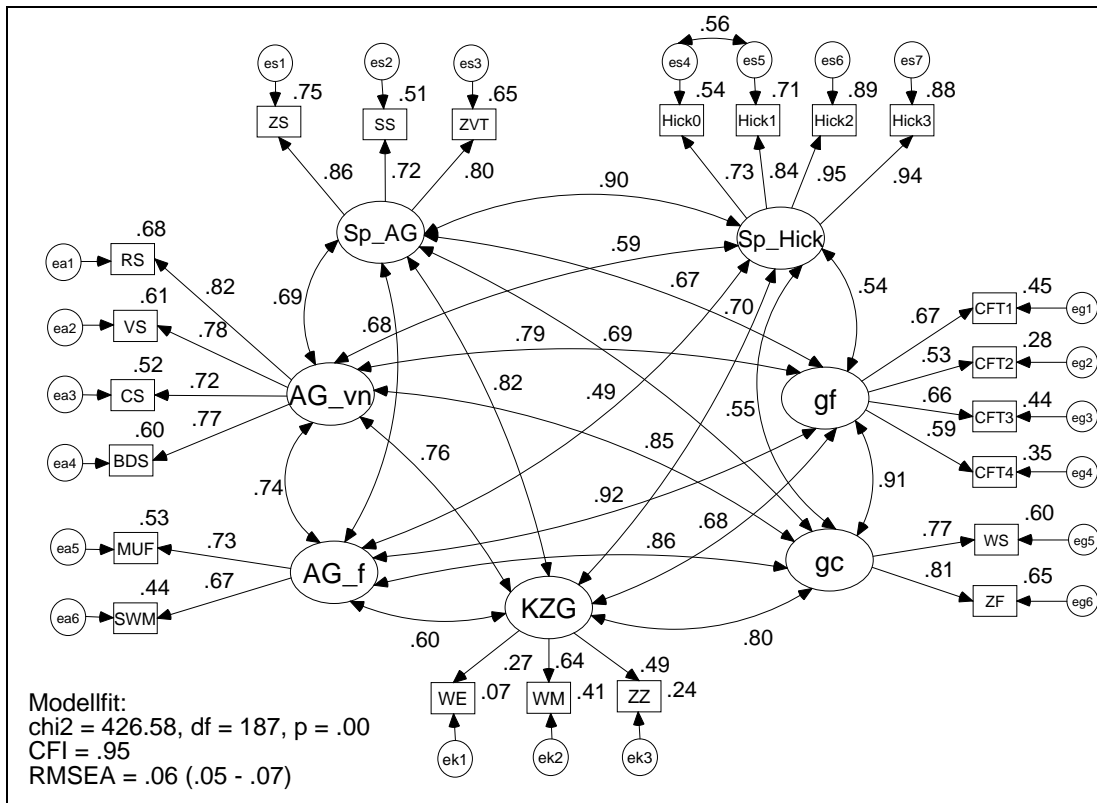
Nachdem die Faktorenstrukturen des Arbeitsgedächtnisses, der Intelligenz und der Informationsverarbeitungsgeschwindigkeit bestimmt wurden, können nun die Zusammenhänge aller hier untersuchten kognitiven Konstrukte in einem gemeinsamen Modell geschätzt werden. Es werden hierzu Korrelationen zwischen den beiden Arbeitsgedächtnisfaktoren „verbal-numerisches Arbeitsgedächtnis“ (AG\_vn) und „figurales Arbeitsgedächtnis“ (AG\_f) und den anderen kognitiven Konstrukten, das heißt den Faktoren „fluide Intelligenz“ (gf), „kristalline Intelligenz“ (gc), „Kurzzeitgedächtnis“ (KZG), „Speed-Arbeitsgeschwindigkeit“ (Sp\_AG) und „Speed-Hick“ (Sp\_Hick) angenommen.

Für dieses Modell resultiert ein befriedigender Fit (siehe Abbildung 7.7; Modell-Fit ohne Korrelation der Residuen von Hick-0 und Hick-1:  $\chi^2 = 544.50$ ,  $df = 188$ ,  $p = .00$ ; CFI = .93, RMSEA = .07, Konfidenz: RMSEA = .07-.08). Aus Gründen der Übersichtlichkeit werden die Korrelationen der Faktoren zusätzlich zu den Angaben in Abbildung 7.7 noch einmal in Tabelle 7.19 dargestellt.

Erwartungsgemäß finden sich hohe Korrelationen jeweils zwischen den beiden Faktoren, die näher miteinander verwandte Konstrukte abbilden; so korrelieren beide Arbeitsgedächtnisfaktoren zu  $r = .74$ , beide Intelligenzfaktoren zu  $r = .91$  und beide Speed-Faktoren zu  $r = .90$ . Beide Arbeitsgedächtnisfaktoren weisen hohe Zusammenhänge mit der fluiden und der kristallinen Intelligenz auf. Die Erwartung, hier Zusammenhänge in einer Größenordnung von mindestens  $r = .70$  zu finden, kann somit bestätigt werden. Die Zusammenhänge mit dem Kurzzeitgedächtnis und dem Faktor „Speed-Arbeitsgeschwindigkeit“ fallen geringer, aber immer noch beträchtlich aus. Die relativ geringsten Korrelationen finden sich zwischen den Arbeitsgedächtnisfaktoren und dem Faktor „Speed-Hick“. Insgesamt fällt somit der Zusammenhang der Arbeitsgedächtnisfaktoren mit den Speed-Faktoren geringer aus als mit den Intelligenzfaktoren (vgl. Frage 2 in Abschnitt 7.3.2).

Es wurde weiterhin vermutet, dass die Zusammenhänge zwischen Arbeitsgedächtnis und den anderen kognitiven Konstrukten einer bestimmten Rangreihe entsprechen. So wurde angenommen, dass der stärkste Zusammenhang zwischen dem Arbeitsgedächtnis und der fluiden Intelligenz besteht und ein etwas schwächerer zwischen dem Arbeitsgedächtnis und der kristallinen Intelligenz. Der Hintergrund dieser Annahme war, dass die kristalline Intelligenz in stärkerem Maße als die fluide Intelligenz durch die Bildung und das erworbene Wissen einer Person sowie durch Gedächtnisaspekte beeinflusst ist. Die geschätzten Parameter bestätigen diese Annahme zunächst nur zum Teil. Die beiden Arbeitsgedächtnisfaktoren weisen ihre höchsten Zusammenhänge mit den beiden Intelligenzfaktoren auf. Zwar korreliert der figurale Arbeitsgedächtnisfaktor stärker mit der fluiden als mit der kristallinen Intelligenz, jedoch ist die Beziehung des verbal-numerischen Arbeitsgedächtnisfaktors zu den Intelligenzfaktoren genau umgekehrt: Hier besteht der stärkere Zusammenhang zur kristallinen Intelligenz. Die Unterschiede fallen jedoch recht gering aus, daher wird vor einer Interpretation dieses Befunds geprüft, ob sich die Werte tatsächlich bedeutsam voneinander unterscheiden.





Anmerkung. CFT\_1 = Subtest „Series“, CFT\_2 = Subtest „Classifications“, CFT\_3 = Subtest „Matrices“, CFT\_4 = Subtest „Topologies“, CFT-ZF = Zahlenfolgentest, CFT-WS = Wortschatztest; ZS = Zahlen-Symbol-Test (HAWIK III), SS = Symbolsuche (HAWIK III), ZVT = Zahlen-Verbindungs-Test, WE = Wege erinnern (BIS-HB), WM = Wörter merken (BIS-HB), ZZ = Zweistellige Zahlen (BIS-HB).

Abbildung 7.7: Modell latenter Faktoren

Tabelle 7.19: Korrelationen der Faktoren des Gesamtmodells

| Faktoren             | AG vn | AG f | g <sub>f</sub> | g <sub>c</sub> | KZG  | Sp AG | Sp Hick |
|----------------------|-------|------|----------------|----------------|------|-------|---------|
| <b>AG_vn</b>         | 1.00  | .74  | .79            | .85            | .76  | .69   | .59     |
| <b>AG_f</b>          | .74   | 1.00 | .92            | .86            | .61  | .68   | .49     |
| <b>g<sub>f</sub></b> | .79   | .92  | 1.00           | .91            | .68  | .67   | .54     |
| <b>g<sub>c</sub></b> | .85   | .86  | .91            | 1.00           | .80  | .69   | .55     |
| <b>KZG</b>           | .76   | .61  | .68            | .80            | 1.00 | .82   | .70     |
| <b>Sp_AG</b>         | .69   | .68  | .67            | .69            | .82  | 1.00  | .90     |
| <b>Sp_Hick</b>       | .59   | .49  | .54            | .55            | .70  | .90   | 1.00    |

Anmerkung. AG\_vn = verbal-numerisches Arbeitsgedächtnis, AG\_f = figurales Arbeitsgedächtnis, g<sub>f</sub> = fluide Intelligenz, g<sub>c</sub> = kristalline Intelligenz, KZG = Kurzzeitgedächtnis, Sp\_AG = Speed-Arbeitsgeschwindigkeit, Sp\_Hick = Speed-Hick.

Diese Prüfung wird für die beiden Arbeitsgedächtnisfaktoren separat vorgenommen. Es werden zunächst die Kovarianzen zwischen AG\_vn und  $g_f$  sowie zwischen AG\_vn und  $g_c$  gleichgesetzt. Sollte dieses Modell keinen signifikant schlechteren Fit aufweisen als das Ausgangsmodell, so kann angenommen werden, dass sich die Beziehungen des verbal-numerischen Arbeitsgedächtnisses zur fluiden und zur kristallinen Intelligenz nicht voneinander unterscheiden. Die Gleichsetzung der Parameter führt jedoch zu einer signifikant schlechteren Modellpassung, so dass davon ausgegangen werden muss, dass tatsächlich unterschiedlich hoch ausgeprägte Zusammenhänge bestehen ( $\chi^2 = 547.09$ ,  $df = 188$ ,  $p = .00$ ;  $\Delta\chi^2 = 120.51$ ,  $\Delta df = 1$ ,  $p = .00$ ). Für den figuralen Arbeitsgedächtnisfaktor wird in gleicher Weise verfahren, auch hier werden die Kovarianzen des Faktors mit den Faktoren  $g_f$  und mit  $g_c$  gleichgesetzt. Erneut führt die Gleichsetzung der Parameter zu einer signifikanten Verschlechterung des Modell-Fits ( $\chi^2 = 527.35$ ,  $df = 188$ ,  $p = .00$ ;  $\Delta\chi^2 = 100.77$ ,  $\Delta df = 1$ ,  $p = .00$ ). Für den figuralen Faktor kann somit angenommen werden, dass er eine stärkere Beziehung zur fluiden Intelligenz aufweist als zur kristallinen Intelligenz.

Eine weitere Annahme betrifft die Stärke des Zusammenhangs zwischen den Arbeitsgedächtnisfaktoren und den Speed-Faktoren: Es wurde vermutet – und die Befunde auf der Ebene der Testwerte deuten bereits in diese Richtung (siehe Abschnitt 7.3.2) – dass das Arbeitsgedächtnis stärker mit Speed zusammenhängt, wenn Speed über komplexere Speed-Aufgaben erfasst wird, wie sie auch in der Intelligenzdiagnostik verwendet werden. Die geschätzten Zusammenhänge fallen für beide Arbeitsgedächtnisfaktoren für den Faktor „Speed-Arbeitsgeschwindigkeit“ höher aus als für den Faktor „Speed-Hick“. Erneut wird durch Gleichsetzung der Parameter überprüft, ob diese Unterschiede bedeutsam sind. Für beide Arbeitsgedächtnisfaktoren ergibt sich bei der Fixierung der Parameter ein signifikant schlechterer Modell-Fit (AG\_vn:  $\chi^2 = 443.00$ ,  $df = 188$ ,  $p = .00$ ;  $\Delta\chi^2 = 16.04$ ,  $\Delta df = 1$ ,  $p = .00$ ; AG\_f:  $\chi^2 = 455.61$ ,  $df = 188$ ,  $p = .00$ ;  $\Delta\chi^2 = 29.03$ ,  $\Delta df = 1$ ,  $p = .00$ ). Die Annahme, dass Speed stärker mit dem Arbeitsgedächtnis korreliert, wenn es über Aufgaben mit komplexeren kognitiven Anforderungen erfasst wird als mit weniger komplexen, kann somit auf der Basis dieser Analysen erneut bestätigt werden.

#### **7.3.4 Zusammenfassung und Diskussion der Analysen der Zusammenhänge zwischen Arbeitsgedächtnis und anderen kognitiven Fähigkeiten**

Es wurden die Zusammenhänge zwischen dem Arbeitsgedächtnis und weiteren kognitiven Konstrukten untersucht. Zunächst wurden hierzu die verwendeten Intelligenztests daraufhin untersucht, ob sie tatsächlich – wie es intendiert war – die beiden Konstrukte „fluide Intelligenz“ und „kristalline Intelligenz“ erfassen. Es fand sich wie erwartet eine gute Passung für ein entsprechendes Zwei-Faktoren-Modell. Im nächsten Schritt wurde die Faktorstruktur der verwendeten Speed-Maße untersucht. Es war beabsichtigt, zum einen relativ elementare Prozesse der Informationsverarbeitungsgeschwindigkeit zu erfassen, und zum anderen solche Speed-Prozesse zu messen, die ein höheres Ausmaß an kognitiver Verarbeitung erfordern. Der Hintergrund des Einsatzes dieser verschiedenen Maße bestand in dem Versuch, das Speed-Konstrukt etwas breiter als oft üblich zu operationalisieren, da es Belege dafür gibt,

dass unterschiedliche Speed-Komponenten in unterschiedlichem Maße mit anderen kognitiven Fähigkeiten zusammenhängen. Die hier verwendeten Speed-Tests erwiesen sich als für diesen Zweck geeignet. Die eher elementaren Hick-Aufgaben bildeten in einem Zwei-Faktoren-Modell einen gemeinsamen Faktor und die kognitiv eher komplexen Speed-Aufgaben aus Intelligenztests einen zweiten Faktor.

Anschließend wurden die Zusammenhänge der Intelligenz- und der Speed-Faktoren, einem Kurzzeitgedächtnisfaktor (gebildet aus drei Merkfähigkeitsaufgaben) und den beiden Arbeitsgedächtnisfaktoren in einem gemeinsamen Modell geschätzt. Es ergaben sich hohe Zusammenhänge zwischen den Arbeitsgedächtnis- und den Intelligenzfaktoren ( $r = .79$  bis  $r = .92$ ), die in etwa vergleichbar sind mit den Befunden von Kyllonen und Christal (1990; Korrelationen zwischen Reasoning und Arbeitsgedächtnis:  $r = .80$ -.88), jedoch höher liegen als in den bisherigen Studien mit Kindern (de Jong und Das-Smaal, 1995; Korrelation zwischen fluider Intelligenz und Arbeitsgedächtnis:  $r = .66$ ) und verschiedenen anderen Studien mit Erwachsenen, die über etwas Zusammenhänge im Bereich von  $r = .70$  berichten (z. B. Ackerman et al., 2002; Colom et al., 2003).

Die hohen Zusammenhänge können hier jedoch nicht auf große Ähnlichkeiten zwischen den Aufgabentypen zurückgeführt werden, die einerseits das Arbeitsgedächtnis und andererseits Intelligenz erfassen sollen, wie es etwa zu den Studien von Kyllonen und Christal (1990) angemerkt wurde (z. B. von Süß, 2001). Die in dieser Studie verwendeten Indikatoren unterscheiden sich nicht nur in der Darbietungsweise (computergestützte Messung des Arbeitsgedächtnisses vs. Messung aller anderen Konstrukte über Paper-Pencil-Tests), sondern auch deutlich in den Anforderungen. So erfordern die Arbeitsgedächtnisaufgaben durchweg relativ simple Informationsverarbeitungsprozesse wie das Überprüfen sehr einfacher Gleichungen oder trivialer Aussagesätze, während die Intelligenztestaufgaben beispielsweise logisches Schließen (z. B. in den Subtests des CFT 20) oder das Abrufen von Wissen (im Wortschatztest) erfordern. Auch wurden das Arbeitsgedächtnis und die Intelligenz an zwei verschiedenen Untersuchungsterminen getestet; dieses Vorgehen lässt eher geringere Zusammenhänge erwarten als wenn beide Konstrukte in einer Sitzung erfasst werden (vgl. Kane et al., 2004). Schließlich führt auch die hier verwendete Scoringmethode aufgrund geringerer Varianz eher zu kleineren Korrelationen als die üblicherweise verwendeten Methoden (vgl. Abschnitte 4.2 und 7.1.4).

Die Stichprobe, die in dieser Studie verwendet wird, weist eine relativ große Fähigkeitsvarianz auf, da sie Kinder unterschiedlicher intellektueller Begabungsstufen und verschiedener Altersgruppen beinhaltet. Diese größere Varianz kann eine Ursache für die höheren Zusammenhänge sein. Die gefundenen Korrelationen stellen somit Schätzungen der Zusammenhänge in einer Gruppe mit einem breiten Begabungsspektrum dar. Zudem sind die Erkenntnisse zu diesen Zusammenhängen bei Kindern bisher rar. Es ist keine Studie bekannt, in der zentral-exekutive Prozesse im Sinne der simultanen Speicherung und Verarbeitung mit Aufgaben aus allen drei Domänen bei Kindern untersucht wurden und in der die Struktur der Daten mit konfirmatorischen Faktorenanalysen untersucht wurde. Neben den Unterschieden in der Fähigkeitsvarianz der Stichproben könnte es daher auch sein, dass die Zusammenhänge zwischen

Arbeitsgedächtnis- und Intelligenzfaktoren bei Kindern dieser Altersgruppe generell höher ausfallen als bei Erwachsenen. Diese Annahme müsste jedoch in weiteren Studien mit Erwachsenen und Kindern systematisch untersucht werden.

Es wurde erwartet, dass die Arbeitsgedächtnisfaktoren stärker mit der fluiden als mit der kristallinen Intelligenz korrespondieren. Jedoch zeigte sich, dass der figurale Arbeitsgedächtnisfaktor stärker mit der fluiden Intelligenz korreliert und der verbal-numerische Arbeitsgedächtnisfaktor stärker mit der kristallinen Intelligenz. Die naheliegendste Erklärung für diesen Befund ist darin zu sehen, dass die Aufgaben, die auf den  $g_f$ -Faktor laden, sämtlich figural sind, während die kristallinen Aufgaben verbale und numerische Inhalte aufweisen. Werden die Ergebnisse im Rahmen des BIS-Modells interpretiert (siehe Abschnitt 7.3.1), so korrespondiert das verbal-numerische Arbeitsgedächtnis stärker mit der verbal-numerischen Verarbeitungskapazität und das figurale Arbeitsgedächtnis stärker mit der figuralen Verarbeitungskapazität. Die gleiche Domäne der Arbeitsgedächtnis- und Intelligenzaufgaben überlagert hier offenbar mögliche Unterschiede hinsichtlich eher fluider oder kristalliner Prozesse. Dieses Ergebnis kann erneut als Beleg für domänenspezifische Aspekte des Arbeitsgedächtnisses angesehen werden.

Aus bisherigen Studien ist nur wenig über den Zusammenhang zwischen Arbeitsgedächtnis und Speed bei Kindern und mögliche Veränderungen dieses Zusammenhangs im Lauf der kognitiven Entwicklung bekannt (Fry & Hale, 2000). Zwar fanden de Jong und Das-Smaal (1995) in ihrer Studie mit Kindern einen Zusammenhang in Höhe von  $r = .60$ , jedoch war hier das Arbeitsgedächtnis nicht optimal operationalisiert (siehe Abschnitt 3.4.1). Mit den in der vorliegenden Studie verwendeten Aufgaben wurden Zusammenhänge im Bereich zwischen  $r = .49$  und  $r = .69$  gefunden, die in etwa in der Größenordnung liegen, wie sie auch in Studien zu Arbeitsgedächtnis und Speed mit erwachsenen Pbn berichtet wurden (z. B.  $r = .55$  bei Ackerman et al., 2002;  $r = .47$  bei Kyllonen & Christal, 1990, Studie 2).

Es wurde zusätzlich angenommen, dass sich schwächere Zusammenhänge der Arbeitsgedächtnisfaktoren mit dem Faktor „Speed-Hick“ finden als mit dem komplexeren Faktor „Speed-Arbeitsgeschwindigkeit“. Tatsächlich fanden sich entsprechend unterschiedlich starke Zusammenhänge. Speed-Arbeitsgeschwindigkeit korrelierte zu  $r = .68$  bzw.  $r = .69$  mit den Arbeitsgedächtnisfaktoren, Speed-Hick hingegen lediglich zu  $r = .49$  bzw.  $r = .59$ . Die Analysen auf der Ebene der Testaufgaben bestätigen diese Annahme zusätzlich: Hier ergaben sich höhere Korrelationen der Arbeitsgedächtnisaufgaben mit den komplexeren Hick-Aufgaben, deren Items die Verarbeitung von zwei oder drei Bits erfordert, als mit den simpleren Hick-Aufgaben, bei deren Items nur ein Bit verarbeitet werden muss. Die Aufgabe Hick-0, die keine Wahlreaktion verlangt, korrelierte nur gering bis gar nicht mit den Arbeitsgedächtnisaufgaben. Unterschiedliche Speed-Aufgaben teilen somit ein unterschiedliches Ausmaß an Varianz mit Leistungen des Arbeitsgedächtnisses.

Diese Ergebnisse entsprechen tendenziell den Befunden von Babcock (1994), der ebenfalls Unterschiede zwischen Speed-Aufgaben mit niedrigen und mit hohen kognitiven Anforderungen fand (Korrelation zwischen Arbeitsgedächtnis und Speed

von  $r = .29$  bei niedrigen und  $r = .59$  bei hohen Anforderungen). Auch Ackerman et al. (2002) berichten über höhere Zusammenhänge des Arbeitsgedächtnisses mit komplexeren Speed-Faktoren als mit Speed-Faktoren, deren Aufgaben geringere Anforderungen an die kognitive Verarbeitung stellen ( $r = .48$  und  $r = .46$  vs.  $r = .23$  und  $r = .39$ ). Es scheint somit nötig, bei der Analyse der Rolle, die Speed im Kontext von Arbeitsgedächtnis und verschiedenen anderen Intelligenzkomponenten spielt, der Operationalisierung des Speed-Konstrukts größere Beachtung zu schenken.

Schließlich fanden sich im Gesamtmodell erwartungsgemäß auch positive Zusammenhänge der Arbeitsgedächtnisfaktoren mit dem Kurzzeitgedächtnis. Der Kurzzeitgedächtnisfaktor wurde über je eine verbale, eine numerische und eine figurale Merkfähigkeitsaufgabe gebildet, so fanden sich Zusammenhänge sowohl mit dem verbal-numerischen als auch mit dem figuralen Arbeitsgedächtnisfaktor.

#### **7.4 Vorhersage von Schulleistungen anhand von Leistungen in Arbeitsgedächtnisaufgaben**

Aufgrund des bisherigen Forschungsstands werden Korrelationen zwischen den Arbeitsgedächtnisleistungen und den Zeugnisnoten in den Kernfächern in der Höhe von mindestens ca.  $r = -.30$  erwartet. Aus vielen Studien ist zudem bekannt, dass die Zusammenhänge zwischen kognitiven Fähigkeiten und den Schulnoten je nach Klassenstufe unterschiedlich ausgeprägt sind; in den unteren Schuljahrgängen finden sich in der Regel stärkere Zusammenhänge als in den höheren Klassen (siehe z. B. Jensen, 1980). Die Berechnungen der Zusammenhänge zwischen den Schulnoten und den Leistungen in den Arbeitsgedächtnisaufgaben erfolgen daher separat für die Primar- und die Sekundarstufe. Es wird angenommen, dass die Zusammenhänge in der Primarstufe etwas höher ausfallen als in der Sekundarstufe.

Weiterhin erfolgt die Analyse der Zusammenhänge im Hinblick auf die Frage der Domänenspezifität vs. -generalität des Arbeitsgedächtnisses. Es wird vermutet, dass die einzelnen Gruppen von Arbeitsgedächtnisaufgaben jeweils etwas höher mit Leistungen in denjenigen Schulfächern korrelieren, die am ehesten die gleiche Domäne ansprechen. So sollten beispielsweise verbale Arbeitsgedächtnisaufgaben stärker mit den Noten für Deutsch und die Fremdsprachen korrelieren als mit der Mathematiknote.

Da die Testungen im zweiten Schulhalbjahr stattfanden, handelt es sich um die Noten aus den Halbjahreszeugnissen, so dass auch die Schülerinnen und Schüler der Klassenstufe 5 bereits Noten aus der weiterführenden Schule angeben konnten. Es werden die bivariaten Korrelationen zwischen den Schulnoten in den Hauptfächern sowie dem Notenschnitt (über alle naturwissenschaftlichen und sprachlichen Fächer) und den Arbeitsgedächtnis-Scores berichtet. Zusätzlich werden über multiple Regressionen von den einzelnen Noten (bzw. Notenmittelwerten über bestimmte Fächer) auf die Arbeitsgedächtnisleistungen die standardisierten Beta-Gewichte ermittelt. Da die betrachteten Variablen in nicht unerheblichem Maße miteinander korrelieren, wird außerdem für jeden Aufgabenscore die semipartielle Korrelation

mit dem jeweiligen Kriterium berichtet. Die semipartielle Korrelation ist ein Maß für die Assoziationsstärke der Kriteriumsvariable und einer Prädiktorvariable, bei der der Einfluss der jeweils anderen Prädiktoren auspartialisiert ist.

In der Primarstufe finden sich für alle Arbeitsgedächtnis-Scores hochsignifikante bivariate Korrelationen mit den Noten in den Fächern Mathematik, Deutsch und Sachkunde sowie mit dem Notenschnitt aus diesen Fächern (Ausnahme: kein signifikanter Zusammenhang zwischen der Leistung in „Memory Updating Figural“ und der Deutschnote; siehe jeweils obere Zeile in Tabelle 7.20).

*Tabelle 7.20: Korrelationen der Arbeitsgedächtnisaufgaben mit den Schulnoten in den Hauptfächern und dem Notenschnitt sowie die standardisierten Beta-Gewichte der Arbeitsgedächtnis-Scores für die einzelnen Noten und den Notenschnitt (Primarstufe)*

| <b>Arbeitsgedächtnis<br/>Aufgabentyp</b> | <b>Note<br/>Deutsch<br/>N = 113</b> | <b>Note<br/>Mathe<br/>N = 113</b> | <b>Note<br/>Sachkunde<br/>N = 111</b> | <b>Noten-<br/>Schnitt<sup>1</sup><br/>N = 113</b> |
|--|-------------------------------------|-----------------------------------|---------------------------------------|---|
| Reading Span                             | -.54                                | -.34                              | -.46                                  | -.56  |
|  | -.38**                              | -.10                              | -.33**                                | -.34**  |
|  | -.27                                | -.07                              | -.24                                  | -.25  |
| Verbal Span                              | -.49                                | -.31                              | -.39                                  | -.49  |
|  | -.28**                              | -.00                              | -.15                                  | -.19  |
|  | -.21                                | -.00                              | -.11                                  | -.14  |
| Computation Span                         | -.39                                | -.52                              | -.42                                  | -.54  |
|  | -.12                                | -.42**                            | -.26*                                 | -.32**  |
|  | -.10                                | -.34                              | -.20                                  | -.26  |
| Backw. Digit Span                        | -.35                                | -.31                              | -.24                                  | -.37  |
|  | .08                                 | .16                               | .29*                                  | .22*  |
|  | .05                                 | .11                               | .21                                   | .15   |
| Memory Upd. Fig.                         | -.11 n.s.                           | -.41                              | -.32                                  | -.34  |
|  | .14                                 | -.22*                             | -.16                                  | -.09  |
|  | .12                                 | -.19                              | -.13                                  | -.08  |
| Spatial WM                               | -.17*                               | -.44                              | -.27                                  | -.35  |
|  | -.06                                | -.27**                            | -.13                                  | -.19*   |
|  | -.05                                | -.24                              | -.10                                  | -.16  |
| <b>R<sup>2</sup> (korr.)</b>             | .33                                 | .38                               | .30                                   | .44   |

Anmerkung. Obere Zeile: bivariate Korrelationen; mittlere Zeile: standardisierte Regressionsgewichte (Regression der einzelnen Noten bzw. Notenschnitt auf alle sechs Aufgaben); untere Zeile: semipartielle Korrelationen; <sup>1</sup>: Mittelwert der Noten in den Fächern Deutsch, Mathematik und Sachkunde; alle bivariaten Korrelationen:  $p < .01$ . Beta-Gewichte: \*\*  $p < .01$ . \*  $p < .05$ . Kursiv gesetzt: aufgrund der gleichen Domänen am höchsten erwartete Zusammenhänge.

Tendenziell entspricht das Muster der Zusammenhänge den Erwartungen: Es wurden höhere Zusammenhänge innerhalb einer Domäne als zwischen verschiedenen Domänen erwartet. So korreliert die Deutschnote am stärksten mit den verbalen Arbeitsgedächtnisaufgaben, etwas geringer mit den numerischen und nur schwach mit den figuralen Arbeitsgedächtnisaufgaben. Die Mathematiknote hingegen korreliert am stärksten mit der numerischen Arbeitsgedächtnisaufgabe „Computation Span“ und etwas schwächer mit der numerischen Aufgabe „Backward Digit Span“ und den verbalen Aufgaben. Für die Mathematiknote finden sich auch deutliche Zusammenhänge mit den figuralen Aufgaben. (Aufgrund der vielfältigen Inhalte des Fachs Sachkunde werden hier keine domänenspezifischen Zusammenhänge erwartet.)

Betrachtet man die standardisierten Beta-Gewichte aus einer multiplen Regression der einzelnen Noten und des Notenschnitts auf die Arbeitsgedächtnis-Scores (siehe jeweils mittlere Zeile in Tabelle 7.20), so fällt auf, dass sich dieses Muster bei der multiplen Betrachtungsweise noch deutlicher zeigt: Nur die verbalen Arbeitsgedächtnis-Scores liefern einen signifikanten Beitrag zur Erklärung der Varianz der Deutschnote und nur die numerische Aufgabe „Computation Span“ sowie die beiden figuralen Aufgaben einen signifikanten Beitrag zur Erklärung der Mathematiknote. Die Varianz der Sachkunde-Note wird signifikant nur durch die Aufgaben „Reading Span“ und „Computation Span“ erklärt. Die Aufgabe „Backward Digit Span“ weist hier ein positives Beta-Gewicht auf.

Zur Aufklärung der Varianz des Notenschnitts in der Primarstufe liefern vor allem die Scores der Aufgabengruppen „Reading Span“ und „Computation Span“ einen signifikanten Beitrag, in einem geringeren Ausmaß auch die Aufgabengruppe „Spatial Working Memory“. Die Aufgabengruppen sind somit dazu geeignet, einen Beitrag zur Erklärung von Schulleistungen zu liefern. Lediglich die Aufgabengruppe „Backward Digit Span“ scheint hierzu in der Grundschule wenig tauglich. Die Arbeitsgedächtnisaufgaben können gemeinsam zwischen 30 und 44% der Varianz der Noten aufklären.

In der Sekundarstufe finden sich ebenfalls für alle Arbeitsgedächtnis-Scores hochsignifikante bivariate Korrelationen mit den Schulnoten sowie mit dem Notenschnitt aus den sprachlichen Fächern, den mathematisch-naturwissenschaftlichen Fächern und dem Gesamtnotenschnitt (siehe jeweils obere Zeile von Tabelle 7.21). Erwartungsgemäß fallen die Zusammenhänge insgesamt für die Sekundarstufe etwas geringer aus als für die Primarstufe. Auch hier entspricht das Muster der Zusammenhänge jedoch tendenziell den Erwartungen bezüglich domänenspezifischer Zusammenhänge. So korreliert auch in der Sekundarstufe die Deutschnote am stärksten mit den verbalen Arbeitsgedächtnisaufgaben, etwas geringer mit den numerischen und nur schwach mit den figuralen Arbeitsgedächtnisaufgaben. Die Zusammenhänge zwischen der Fremdsprachennote und den Arbeitsgedächtnis-Scores fallen ähnlich aus. Die Mathematiknote hingegen korreliert zwar am stärksten mit den numerischen Arbeitsgedächtnisaufgaben, jedoch liegen die Korrelationen mit den anderen vier Aufgaben in vergleichbarer Höhe.

Tabelle 7.21: Korrelationen der Arbeitsgedächtnisaufgaben mit den Schulnoten in den Hauptfächern und dem Notenschnitt, die standardisierten Beta-Gewichte der Arbeitsgedächtnis-Scores für die einzelnen Noten und den Notenschnitt sowie die semipartiellen Korrelationen (Sekundarstufe)

| Arbeitsgedächtn. Aufgabentyp | Note Deutsch   | Note Fremdsprache <sup>1</sup> | Note Mathe     | Note Chem. / Physik | Note Biol.     | Notenschnitt Naturwiss. <sup>2</sup> | Ges.-notenschnitt <sup>3</sup> |
|------------------------------|----------------|--------------------------------|----------------|---------------------|----------------|--------------------------------------|--------------------------------|
|                              | <i>N</i> = 256 | <i>N</i> = 252                 | <i>N</i> = 257 | <i>N</i> = 165      | <i>N</i> = 196 | <i>N</i> = 251                       | <i>N</i> = 258                 |
| Reading                      | -.41           | -.44                           | -.42           | -.40                | -.45           | -.46                                 | -.51                           |
| Span                         | -.17*          | -.15                           | -.12           | -.30**              | -.32**         | -.37**                               | -.23**                         |
|                              | -.12           | -.10                           | -.09           | -.21                | -.23           | -.26                                 | -.16                           |
| Verbal                       | -.41           | -.48                           | -.36           | -.31                | -.32           | -.33                                 | -.47                           |
| Span                         | -.19*          | -.29**                         | -.00           | -.09                | -.02           | -.07                                 | -.17*                          |
|                              | -.13           | -.20                           | -.00           | -.06                | -.01           | -.05                                 | -.12                           |
| Comp.                        | -.36           | -.36                           | -.44           | -.19                | -.29           | -.25                                 | -.41                           |
| Span                         | -.12           | -.05                           | -.23**         | .09                 | -.01           | .06                                  | -.09                           |
|                              | -.09           | -.04                           | -.18           | .07                 | -.01           | .04                                  | -.07                           |
| Backw.                       | -.35           | -.43                           | -.43           | -.28                | -.35           | -.32                                 | -.45                           |
| Digit Sp.                    | -.05           | -.18*                          | -.14           | .02                 | -.12           | -.06                                 | -.13                           |
|                              | -.04           | -.12                           | -.10           | .01                 | -.08           | -.04                                 | -.09                           |
| Memory                       | -.21           | -.16                           | -.28           | -.27                | -.23           | -.27                                 | -.28                           |
| Upd. Fig.                    | .04            | -.16*                          | .04            | -.09                | -.01           | -.05                                 | .04                            |
|                              | .03            | -.13                           | .03            | -.07                | -.01           | -.04                                 | .04                            |
| Spatial                      | -.27           | -.30                           | -.41           | -.35                | -.29           | -.28                                 | -.37                           |
| WM                           | -.10           | -.12                           | -.25**         | -.16                | -.10           | -.06                                 | -.16                           |
|                              | -.08           | -.10                           | -.20           | -.13                | -.08           | -.05                                 | -.13                           |
| R <sup>2</sup> (korr.)       | .21            | .29                            | .29            | .17                 | .20            | .21                                  | .32                            |

Anmerkung. Obere Zeile: bivariate Korrelationen; mittlere Zeile: standardisierte Regressionsgewichte (Regression der einzelnen Noten bzw. Notenschnitte auf alle sechs Aufgaben); untere Zeile: semipartielle Korrelationen. <sup>1</sup>: Fremdsprachennote (Englisch, Französisch, Latein), bei zwei Fremdsprachen Mittelwert; <sup>2</sup>: Mittelwert der Noten Chemie, Physik, Biologie; <sup>3</sup>: Mittelwert der Noten in den naturwissenschaftlichen und sprachlichen Fächern; alle bivariaten Korrelationen:  $p < .01$ . Beta-Gewichte: \*\*  $p < .01$ . \*  $p < .05$ . Kursiv gesetzt: aufgrund der gleichen Domänen am höchsten erwartete Zusammenhänge.

Die naturwissenschaftlichen Fächer korrelieren überraschenderweise am stärksten mit der verbalen Aufgabe „Reading Span“. Dies gilt sowohl für die eher „harten“ Naturwissenschaften Chemie und Physik als auch für das „weichere“ Fach Biologie. Dies ist möglicherweise darauf zurückzuführen, dass in der Unterstufe in den Fächern Chemie und Physik sprachliche Fähigkeiten wichtiger als mathematische Fähigkeiten sein könnten. Die Zusammenhänge der Aufgabengruppen „Verbal Span“, „Backward Digit Span“, „Memory Updating Figural“ und „Spatial Working



Memory“ mit den naturwissenschaftlichen Fächern liegen alle in vergleichbarer Höhe (um  $r = .30$ ), lediglich die Zusammenhänge der Aufgabengruppe „Computation Span“ fallen etwas geringer aus.

In der multiplen Regression weist nur die „Reading Span“ ein signifikantes Beta-Gewicht für die naturwissenschaftlichen Fächer auf. Der Notenschnitt wird bedeutsam nur durch die beiden verbalen Aufgabengruppen vorhergesagt, alle anderen Aufgaben liefern keinen signifikanten Beitrag zur Varianzaufklärung. Insgesamt können die Arbeitsgedächtnisaufgaben gemeinsam zwischen 17 und 29% der Varianz der Schulnoten und 32% der Varianz des Notenschnitts in der Sekundarstufe aufklären. Die Aufgabe „Reading Span“ erweist sich insgesamt in beiden Schulstufen als der relativ beste Prädiktor für den Gesamtnotenschnitt.

Diese Ergebnisse deuten erneut auf domänenspezifische Aspekte des Arbeitsgedächtnisses hin, hier jedoch auf separate verbale und numerische Aspekte des Arbeitsgedächtnisses, wie sie in Abschnitt 3.3.3 diskutiert wurden. Zwar können schulische Leistungen diesen Befunden zufolge generell am besten durch die verbalen Aufgabentypen vorhergesagt werden. Dies ist auf die hohe Bedeutung sprachlicher Fähigkeiten in den meisten Schulfächern zurückzuführen. Die Mathematikleistung wird hingegen in beiden Schulstufen bedeutsam nur durch die Leistungen in der numerischen Aufgabe „Computation Span“ – die sehr einfache Rechenoperationen erfordert – und in der figuralen Aufgabe „Spatial Working Memory“ vorhergesagt.

## 7.5 Testtheoretische Analysen

Im Folgenden werden die sechs Gruppen von Arbeitsgedächtnisaufgaben im Hinblick darauf untersucht, inwiefern sie sich für die Zusammenstellung zu einem Test eignen. Zunächst wird ein Überblick über das Vorgehen bei der Analyse der Testdaten gegeben.

### 7.5.1 Vorgehen bei der Analyse nach der Klassischen Testtheorie und der Item Response Theorie

Die Arbeitsgedächtnisaufgaben werden entsprechend der Klassischen Testtheorie analysiert und auf ihre Passung auf probabilistische Modelle der Item Response Theorie (IRT) untersucht. Gegenstand der Untersuchung sind das 1-Parameter logistische Modell (1PL-Modell), das 2-Parameter logistische Modell (2PL-Modell) und das 3-Parameter logistische Modell (3PL-Modell; vgl. Abschnitt 5.3.2). Die Prüfung entspricht dem üblichen Vorgehen bei der Testentwicklung im Rahmen der IRT (vgl. z. B. Rost, 2004) und erfolgt in den folgenden Schritten:

### 1. *Prüfung der Dimensionalität der Arbeitsgedächtnisaufgaben*

Als erster Schritt bei der Analyse der Daten im Hinblick auf die Zusammenstellung eines Tests wird eine exploratorische Faktorenanalyse über alle 70 Arbeitsgedächtnisaufgaben gerechnet, um Aufschluss über die zugrunde liegenden Dimensionen zu erhalten. Es steht zu vermuten, dass sich auch hier – ebenso wie bei den Analysen zur Untersuchung des Konstrukts in den vorangehenden Abschnitten – eine zweifaktorielle Struktur ergibt: ein Faktor, auf den die verbalen und numerischen Aufgaben hoch laden und ein weiterer Faktor, der die Leistungen in den figuralen Aufgaben bestimmt. Auf der Grundlage dieser Analyse können die Aufgaben zu einer oder mehreren Skalen zusammengefasst werden.

### 2. *Bestimmung der Aufgabenkennwerte nach der Klassischen Testtheorie und Schätzung der Modellparameter und Prüfung des Aufgaben-Fit für das 1PL- und das 2PL-Modell*

In diesem Schritt werden zunächst die Kennwerte Schwierigkeit und Trennschärfe der Arbeitsgedächtnisaufgaben nach der Klassischen Testtheorie ermittelt. Anschließend werden die Aufgabenparameter für das 1PL-, 2PL- und 3PL-Modell über die Marginal Maximum Likelihood (MML) mit dem Programm BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 1996) geschätzt. Der Vorteil der MML gegenüber anderen Schätzmethoden besteht darin, dass sie eine Schätzung auch für Modelle mit mehreren Parametern erlaubt. Zudem wird geprüft, ob die einzelnen Testaufgaben den Erwartungen entsprechen, die das jeweilige IRT-Modell stellt. Hierzu können für das 1PL-Modell sowohl residuen-basierte als auch likelihood-basierte Fit-Maße verwendet werden (Rost, 2004). Residuen-basierte Maße gehen von den standardisierten Abweichungen zwischen den beobachteten und den nach dem Modell erwarteten Antworten aus. So kann geprüft werden, ob die Daten zu einer bestimmten Testaufgabe der Form der Item-Response-Funktion entsprechen, die durch das jeweilige Modell vorgegeben ist. Es wird dabei so vorgegangen, dass für verschiedene Intervalle auf der Fähigkeitsdimension bestimmt wird, wie viele Pbn die Aufgabe richtig gelöst haben. Dieser beobachtete Wert wird zu dem nach dem Modell erwarteten Wert in Beziehung gesetzt. Der Fit jeder Aufgabe kann somit über einen  $\chi^2$ -Test, der beobachtete gegen erwartete Werte testet, beurteilt werden. Die Ergebnisse der  $\chi^2$ -Tests sind jedoch aus mehreren Gründen mit Vorsicht zu interpretieren:  $\chi^2$ -Tests sind in starkem Maße abhängig von der Stichprobengröße, bei einem großen  $N$  wird das Ergebnis eher signifikant. Embretson und Reise (2000) raten daher von  $\chi^2$ -Tests als alleiniger Entscheidungsgrundlage ab. Rost (2004) weist darauf hin, dass Residualmaße dazu tendieren, Modellabweichungen zu überschätzen. Der Hintergrund ist, dass die Antworten der Pbn auf Testaufgaben stets ganzzahlige Werte sind, die Erwartungswerte jedoch in der Regel zwischen zwei Antwortalternativen angesiedelt sind.

Aufgrund der eingeschränkten Interpretierbarkeit der  $\chi^2$ -Tests werden häufig weitere Maße verwendet. In der vorliegenden Arbeit werden für das 1PL-Modell zusätzlich Signifikanztests anhand des likelihood-basierten Q-Index (Rost & von

Davier, 1994; Rost, 2004) berichtet. Der Q-Index ist ein Aufgaben-Fitmaß, das auf der Log-Likelihood der beobachteten Item-Pattern basiert. Der Fit einer Aufgabe wird hier anhand der bedingten Wahrscheinlichkeit des beobachteten Item-Response-Vektors beurteilt. Q variiert zwischen 0 und 1, wobei ein Wert von 0 einen perfekten Fit (Guttman-Pattern) und ein Wert von 1 einen perfekten Misfit (Anti-Guttman-Pattern) anzeigt. Ein Wert von .5 deutet auf die Unabhängigkeit von Aufgabe und Fähigkeit hin, das heißt er spiegelt ein zufälliges Antwortmuster wider. Der Signifikanztest für den Q-Index erlaubt die Bestimmung von Underfit ( $p < .05$ ) und Overfit ( $p > .95$ ) einer Aufgabe. Die Berechnung des Q-Index erfolgt mit dem Programm Winmira 2001 (von Davier, 2001).

Das 1PL- und das 2PL-Modell nehmen an, dass die Testaufgaben nicht durch Raten gelöst werden. Bei den hier verwendeten Testaufgaben ist Raten relativ unwahrscheinlich. In den verbalen Aufgabentypen „Reading Span“ und „Verbal Span“ kann zwar im Entscheidungsteil der Aufgaben geraten werden (Wahrscheinlichkeit von  $p = .50$  bzw.  $p = .25$ , richtig zu raten), die zu erinnernden Wörter jedoch können nicht geraten werden. Bei den numerischen Aufgaben können die Zahlen (1-9) zwar geraten werden, da allerdings auch die richtige Reihenfolge der Wiedergabe relevant ist, besteht nur eine äußerst geringe Wahrscheinlichkeit, eine Aufgabe durch Raten korrekt zu beantworten. Zusätzlich wurden die Pbn instruiert, eine Null einzugeben, wenn sie sich an eine Zahl nicht erinnern konnten. Auch bei den figuralen Aufgaben hilft Raten beim Finden der richtigen Lösung kaum. Beim Aufgabentyp „Spatial Working Memory“ ist es bereits sehr unwahrscheinlich, durch Raten das richtige Muster zu erhalten, zusätzlich ist auch hier die Wiedergabe der richtigen Reihenfolge der Muster relevant. Beim Aufgabentyp „Memory Updating Figural“ besteht bei jedem einzelnen Item eine Wahrscheinlichkeit von  $p = .25$ , durch bloßes Raten die richtige Lösung zu finden. Jedoch müssen auch hier mindestens zwei Items richtig beantwortet werden, um eine Aufgabe korrekt zu lösen. Selbst bei den einfachsten Aufgaben dieses Typs liegt somit die Ratewahrscheinlichkeit nur bei  $p = .25 * .25 = .06$ . Die Annahme eines Rateparameters (3PL-Modell) scheint für die hier verwendeten Skalen somit nicht nötig, dennoch wird die Passung der Daten auch auf das 3PL-Modell überprüft.

### 3. Überprüfung der Modellgültigkeit mittels verschiedener Modellgeltungstests

Im nächsten Schritt wird die Passung eines jeweiligen IRT-Modells auf den gesamten Datensatz untersucht. Hier werden zwei Vorgehensweisen verwendet: Zunächst erfolgt eine Modellprüfung über den Vergleich der relativen Passung verschiedener IRT-Modelle, die auf der Likelihood bzw. der Log-Likelihood (LogL) der Daten unter dem jeweiligen Modell basiert. Generell gilt, dass ein Modell umso besser passt, je größer die Likelihood ist, also je wahrscheinlicher die Daten unter dem betrachteten Modell sind. Da die Likelihood jedoch von der Anzahl der Parameter eines Modells abhängt, werden Maße benötigt, die die Likelihood zu der Parameteranzahl in Beziehung setzen und so den Vergleich zwischen Modellen mit unterschiedlich vielen Parametern erlauben. Dies leisten so

genannte informationstheoretische Maße. Es handelt sich dabei um Straffunktionen, die die Verwendung von Modellen mit vielen Parametern (die zwar besser auf die Daten passen, aber wenig sparsam sind) „bestrafen“. Diese Maße basieren auf der LogL multipliziert mit (-2), so dass hier kleinere Werte prinzipiell für eine bessere Erklärung der Daten durch das Modell sprechen. Zusätzliche Parameter führen bei diesen Maßen zu einer Erhöhung des Werts. Das relativ beste Modell erhält somit den niedrigsten Wert. Typische informationstheoretische Maße, die auch in dieser Arbeit berichtet werden, sind die *Indices Akaikes Information Criterion* (AIC), *Best Information Criterion* (BIC) und *Consistent AIC* (CAIC), wobei BIC und CAIC zusätzlich zur Anzahl geschätzter Parameter auch die Stichprobengröße berücksichtigen. Die zweite Vorgehensweise besteht darin, die empirischen Häufigkeiten der Antwortpattern mit den nach dem jeweiligen Modell erwarteten Patternhäufigkeiten in Beziehung zu setzen.

Die Passung des 1PL-Modells kann zusätzlich über spezielle Tests geprüft werden. Geprüft werden dazu zwei grundlegende Annahmen des 1PL-Modells, die Annahme der Personenhomogenität und die Annahme der Item- bzw. Aufgabenhomogenität. Die Personenhomogenität bezieht sich darauf, dass die Testskala bei allen Pbn dieselbe latente Fähigkeit erfasst. Diese wird mit einem graphischen Modelltest und dem Likelihoodquotiententest nach Andersen geprüft. Dieser Aspekt spielt in der vorliegenden Arbeit eine besondere Rolle: Es ist zu prüfen, ob die Aufgaben sowohl bei Pbn mit durchschnittlicher als auch mit überdurchschnittlicher Intelligenz dieselbe Fähigkeit erfassen und somit für beide Gruppen gleichermaßen geeignet sind. Die Aufgabenhomogenität erfordert, dass verschiedene Aufgabensubgruppen des Tests dieselbe latente Dimension erfassen und wird mit dem Martin-Löf-Test überprüft.

#### 4. Prüfung der Reliabilität und der externen Validität der Skalen

Aufgrund der Analysen in Schritt 2 und 3 müssen ggf. einzelne Aufgaben eliminiert werden, um eine hinreichende Passung an das am besten passende IRT-Modell zu erzielen. Die auf diese Weise reduzierten Skalen werden im Folgenden erneut auf ihre Messgüte untersucht.

### 7.5.2 Dimensionalität der Arbeitsgedächtnisaufgaben

Zur Prüfung der Dimensionalität der Arbeitsgedächtnisaufgaben wird zunächst über alle 70 Arbeitsgedächtnisaufgaben eine exploratorische Faktorenanalyse gerechnet (Extraktion: Hauptachsenanalyse). Da angenommen wird, dass ggf. vorhandene mehrere Faktoren nicht unabhängig voneinander sind, wird eine oblique Rotation vorgenommen. Es resultieren 20 Faktoren, die einen Eigenwert größer 1 aufweisen, der Scree-Test empfiehlt jedoch die Extraktion von einem oder zwei Faktoren (der erste Faktor klärt 17%, der zweite 4% und der dritte 3% der Varianz auf, siehe Abbildung 7.8). Da die Strukturanalysen auf der Basis der Summenscores (vgl. Abschnitt 7.2.2) eine Faktorstruktur von zwei inhaltspezifischen korrelierenden Faktoren nahe legen, wird anschließend die Extraktion von zwei Faktoren voreingestellt.

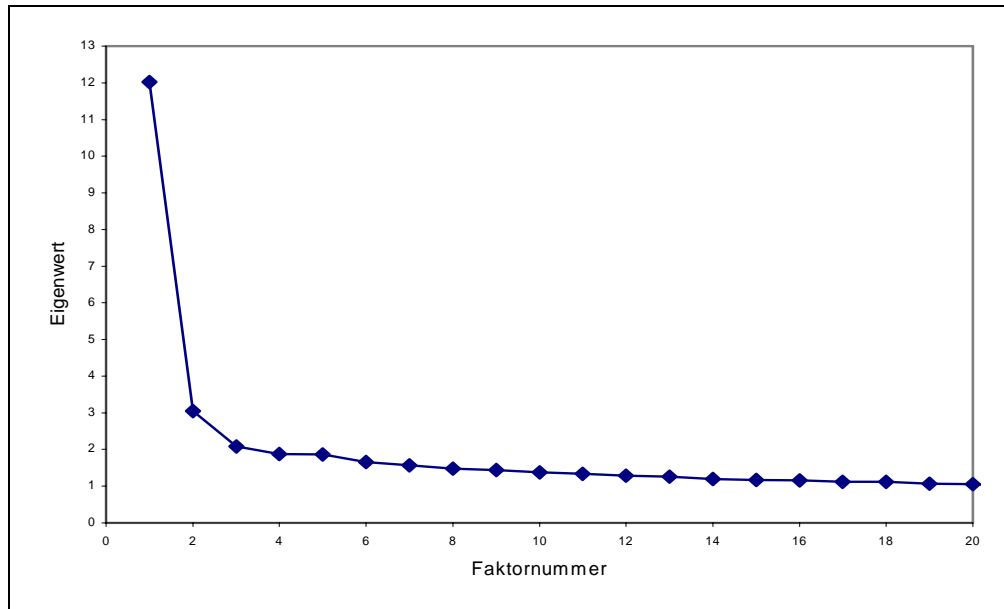


Abbildung 7.8: Screeplot: Eigenwertverlauf der 20 größten Werte der Faktorenanalyse über die 70 Aufgaben

Bei der Extraktion von zwei Faktoren ergibt sich ein klar zu interpretierendes Ladungsmuster: Alle verbalen und numerischen Aufgaben laden hoch auf dem ersten, aber niedrig auf dem zweiten Faktor; die figuralen Aufgaben laden hingegen hoch auf dem zweiten und niedrig auf dem ersten Faktor (siehe Tabelle 7.22). Die einzige Ausnahme bildet hier die Aufgabe BDS 1, die relativ niedrig auf beiden Faktoren lädt. Beide Faktoren korrelieren zu  $r = .52$  miteinander. Es können somit aus den untersuchten Testaufgaben zwei Skalen gebildet werden: eine verbal-numerische Skala, bestehend aus den Aufgabentypen „Reading Span“, „Verbal Span“, „Computation Span“ und „Backward Digit Span“, und eine figurale Skala, bestehend aus den Aufgabentypen „Memory Updating Figural“ und „Spatial Working Memory“.

Tabelle 7.22: Faktorladungen der 70 Arbeitsgedächtnisaufgaben auf zwei Faktoren  
(Hauptachsenanalyse, Rotation: oblimin; Mustermatrix)

| Arbeitsged.<br>Aufgabe | Faktor 1:<br>verbal-num. | Faktor 2:<br>figural | Aufgabe | Faktor 1:<br>verbal-num. | Faktor 2:<br>figural |
|------------------------|--------------------------|----------------------|---------|--------------------------|----------------------|
| RS 1                   | .43                      | .10                  | BDS 5   | .48                      | .10                  |
| RS 2                   | .46                      | -.03                 | BDS 6   | .52                      | .09                  |
| RS 3                   | .59                      | .06                  | BDS 7   | .50                      | .12                  |
| RS 4                   | .49                      | -.08                 | BDS 8   | .49                      | .12                  |
| RS 5                   | .43                      | .06                  | BDS 9   | .38                      | .20                  |
| RS 6                   | .65                      | -.04                 | BDS 10  | .35                      | .07                  |
| RS 7                   | .43                      | .11                  | BDS 11  | .41                      | .09                  |
| RS 8                   | .47                      | .07                  | BDS 12  | .26                      | .08                  |
| RS 9                   | .28                      | .00                  | MUF 1   | -.05                     | .43                  |
| RS 10                  | .37                      | .02                  | MUF 2   | .18                      | .37                  |
| RS 11                  | .33                      | -.05                 | MUF 3   | .15                      | .26                  |
| VS 1                   | .45                      | .00                  | MUF 4   | .12                      | .32                  |
| VS 2                   | .50                      | -.01                 | MUF 5   | -.06                     | .40                  |
| VS 3                   | .58                      | -.02                 | MUF 6   | .07                      | .16                  |
| VS 4                   | .51                      | .00                  | MUF 7   | .11                      | .22                  |
| VS 5                   | .36                      | .05                  | MUF 8   | .09                      | .34                  |
| VS 6                   | .50                      | .06                  | MUF 9   | .06                      | .25                  |
| VS 7                   | .53                      | .03                  | MUF 10  | .06                      | .36                  |
| VS 8                   | .51                      | -.09                 | MUF 11  | .15                      | .34                  |
| VS 9                   | .41                      | .01                  | MUF 12  | .08                      | .32                  |
| VS 10                  | .42                      | -.12                 | MUF 13  | .16                      | .22                  |
| CS 1                   | .42                      | -.05                 | MUF 14  | .10                      | .23                  |
| CS 2                   | .34                      | -.09                 | SWM 1   | .04                      | .31                  |
| CS 3                   | .33                      | .09                  | SWM 2   | -.10                     | .53                  |
| CS 4                   | .35                      | .02                  | SWM 3   | .05                      | .48                  |
| CS 5                   | .44                      | .00                  | SWM 4   | -.11                     | .53                  |
| CS 6                   | .36                      | .04                  | SWM 5   | .08                      | .39                  |
| CS 7                   | .51                      | .01                  | SWM 6   | .06                      | .47                  |
| CS 8                   | .54                      | -.03                 | SWM 7   | -.05                     | .47                  |
| CS 9                   | .21                      | .06                  | SWM 8   | -.06                     | .55                  |
| CS 10                  | .49                      | .00                  | SWM 9   | .02                      | .31                  |
| BDS 1                  | .21                      | .22                  | SWM 10  | -.06                     | .47                  |
| BDS 2                  | .34                      | .10                  | SWM 11  | -.01                     | .41                  |
| BDS 3                  | .50                      | -.02                 | SWM 12  | .05                      | .30                  |
| BDS 4                  | .53                      | .02                  | SWM 13  | -.03                     | .26                  |

Anmerkung. RS = Reading Span, VS = Verbal Span, CS = Computation Span, BDS = Backward Digit Span, MUF = Memory Updating Figural, SWM = Spatial Working Memory.

### 7.5.3 Bestimmung der Aufgabenkennwerte nach der Klassischen Testtheorie und Schätzung der Aufgabenparameter nach IRT-Modellen

Im Rahmen der Klassischen Testtheorie werden die Schwierigkeiten (Lösungshäufigkeiten  $p$ ) und Trennschärfen (punktbiseriale Korrelation  $r_{it}$ , jeweils bezogen auf die gesamte Skala) der einzelnen Testaufgaben berechnet. Diese Berechnung erfolgt zunächst für die Gesamtstichprobe ( $N = 374$ ). Da diese Größen jedoch vom Fähigkeitsniveau der Stichprobe abhängig sind und die Gesamtstichprobe einen großen Teil überdurchschnittlich fähiger Pbn enthält, werden die Werte zusätzlich für die beiden Teilstichproben der unausgelesenen Pbn (UA;  $N = 172$ ) und der hoch Intelligenten (HI;  $N = 170$ ) berechnet (zur Stichprobenbildung vgl. Abschnitt 6.3). Die Kennwerte für die Schwierigkeiten und Trennschärfen der Testaufgaben der beiden Skalen werden im Folgenden separat für beide Skalen dargestellt. Zur Abschätzung der inneren Konsistenz wird für beide Skalen Cronbachs Alpha berechnet. Die verbal-numerische Skala erreicht einen Wert von  $\alpha = .92$ , die figurale Skala einen Wert von  $\alpha = .82$ . Beide Skalen können daher als hinreichend reliabel betrachtet werden.

Weiterhin werden für beide Skalen für das 1PL-Modell die Schwierigkeitsparameter der Aufgaben ( $\sigma$ ), die Standardfehler der Schwierigkeitsparameter ( $SE_{\sigma}$ ) sowie zwei Fit-Maße für die Aufgaben berichtet. Es werden der  $\chi^2$ -Test und ein Signifikanztest anhand des likelihood-basierten Q-Index dargestellt, der die Bestimmung von Underfit sowie Overfit einer Aufgabe erlaubt. Für das 2PL-Modell werden ebenfalls die Schwierigkeitsparameter ( $\sigma$ ) sowie die zugehörigen Standardfehler ( $SE_{\sigma}$ ) angegeben, weiterhin die Trennschärfeparameter der Aufgaben ( $\beta$ ) mit den entsprechenden Standardfehlern ( $SE_{\beta}$ ) und der  $\chi^2$ -Test zur Abschätzung des Aufgaben-Fit. Die Parameter des 3PL-Modells wurden ebenfalls für beide Skalen geschätzt. Da der Rateparameter jedoch erwartungsgemäß für alle Aufgaben nahe Null beträgt und die Parameter  $\sigma$  und  $\beta$  bis auf relativ geringe Abweichungen denen im 2PL-Modell entsprechen, werden sie hier nicht berichtet.

#### *Verbal-numerische Skala*

Die klassisch berechneten Schwierigkeiten der 43 verbalen und numerischen Aufgaben streuen relativ breit, wie es für Tests nach der Klassischen Testtheorie gefordert wird ( $p = .06$  bis  $p = .81$ ; siehe Tabelle 7.23; vgl. z. B. Lienert & Raatz, 1994). Tendenziell sind jedoch anteilig mehr schwierige als leichte Aufgaben enthalten, da die Aufgaben auch für sehr fähige Pbn noch ausreichend schwierig sein sollen. Die verbal-numerische Skala enthält insgesamt 13 Aufgaben mit einer Schwierigkeit von  $p < .20$ , hingegen nur eine Aufgabe mit einem  $p > .80$ . Die mittlere Schwierigkeit für die verbal-numerische Skala, berechnet für die Gesamtstichprobe, beträgt  $p = .38$  ( $SD = .23$ ). Erwartungsgemäß sind die Aufgaben für die unausgelesenen Pbn schwieriger als für die hoch intelligenten Pbn. Insgesamt scheinen die Aufgaben jedoch auch für die Gruppe der hoch Intelligenten noch hinreichend schwierig zu sein, sieben Aufgaben der Skala erweisen sich auch für diese Gruppe als sehr schwierig ( $p < .20$ ). Die nach dem 1PL-Modell geschätzten Schwierigkeitsparameter der verbal-numerischen Skala liegen zwischen  $\sigma = -1.46$  und  $\sigma = 2.70$  ( $M = .65$ ,  $SD = 1.19$ ; siehe Tabelle 7.24).

Tabelle 7.23: Klassische Schwierigkeiten und Trennschärfen der Aufgaben der verbal-numerischen Skala (wird fortgesetzt)

| Verbal-<br>num.<br>Skala                | Level   | Ges.stichprobe<br>(N = 374) |                       | Unausgelesene<br>(N = 172) |                       | Hoch Intelligente<br>(N = 170) |                       |
|---|---------|-----------------------------|-----------------------|----------------------------|-----------------------|--------------------------------|-----------------------|
|   |         | <i>p</i>                    | <i>r<sub>it</sub></i> | <i>p</i>                   | <i>r<sub>it</sub></i> | <i>p</i>                       | <i>r<sub>it</sub></i> |
| <i>Aufgaben „Reading Span“ (RS)</i>     |         |                             |                       |                            |                       |                                |                       |
| RS 1                                    | RS I    | .57                         | .47                   | .42                        | .32                   | .74                            | .38                   |
| RS 2                                    |         | .72                         | .44                   | .65                        | .40                   | .85                            | .34                   |
| RS 3                                    | RS II   | .48                         | .59                   | .30                        | .51                   | .68                            | .44                   |
| RS 4                                    |         | .56                         | .43                   | .42                        | .37                   | .70                            | .35                   |
| RS 5                                    |         | .61                         | .45                   | .49                        | .43                   | .77                            | .31                   |
| RS 6                                    | RS III  | .33                         | .59                   | .15                        | .47                   | .49                            | .51                   |
| RS 7                                    |         | .14                         | .46                   | .08                        | .41                   | .22                            | .41                   |
| RS 8                                    |         | .20                         | .47                   | .12                        | .42                   | .29                            | .46                   |
| RS 9                                    | RS IV   | .06                         | .26                   | .01                        | .12                   | .10                            | .22                   |
| RS 10                                   |         | .08                         | .36                   | .04                        | .27                   | .13                            | .32                   |
| RS 11                                   |         | .06                         | .28                   | .04                        | .29                   | .09                            | .24                   |
| <i>Aufgaben „Verbal Span“ (VS)</i>      |         |                             |                       |                            |                       |                                |                       |
| VS 1                                    | VS I    | .66                         | .44                   | .55                        | .44                   | .78                            | .25                   |
| VS 2                                    | VS IIa  | .62                         | .49                   | .49                        | .54                   | .76                            | .32                   |
| VS 3                                    |         | .51                         | .55                   | .40                        | .56                   | .65                            | .44                   |
| VS 4                                    | VS IIb  | .45                         | .49                   | .36                        | .48                   | .54                            | .48                   |
| VS 5                                    |         | .50                         | .38                   | .40                        | .33                   | .67                            | .20                   |
| VS 6                                    | VS IIIa | .42                         | .50                   | .31                        | .44                   | .57                            | .41                   |
| VS 7                                    |         | .43                         | .52                   | .29                        | .51                   | .59                            | .38                   |
| VS 8                                    | VS IIIb | .15                         | .44                   | .10                        | .46                   | .25                            | .39                   |
| VS 9                                    |         | .18                         | .38                   | .12                        | .27                   | .28                            | .26                   |
| VS 10                                   | VS IV   | .13                         | .33                   | .11                        | .28                   | .18                            | .22                   |
| <i>Aufgaben „Computation Span“ (CS)</i> |         |                             |                       |                            |                       |                                |                       |
| CS 1                                    | CS I    | .76                         | .39                   | .68                        | .39                   | .84                            | .28                   |
| CS 2                                    |         | .76                         | .30                   | .74                        | .40                   | .79                            | .30                   |
| CS 3                                    | CS II   | .70                         | .38                   | .62                        | .39                   | .74                            | .42                   |
| CS 4                                    |         | .66                         | .37                   | .58                        | .38                   | .77                            | .25                   |
| CS 5                                    | CS III  | .53                         | .43                   | .42                        | .36                   | .67                            | .33                   |
| CS 6                                    |         | .45                         | .38                   | .37                        | .32                   | .58                            | .24                   |
| CS 7                                    | CS IV   | .24                         | .49                   | .18                        | .58                   | .37                            | .36                   |
| CS 8                                    |         | .26                         | .50                   | .18                        | .52                   | .34                            | .39                   |
| CS 9                                    | CS V    | .07                         | .24                   | .08                        | .25                   | .10                            | .21                   |
| CS 10                                   |         | .12                         | .46                   | .06                        | .34                   | .20                            | .46                   |



Tabelle 7.23 (Fortsetzung): Klassische Schwierigkeiten und Trennschärfen der Aufgaben der verbal-numerischen Skala

| Verbal-<br>num.<br>Skala                    | Level  | Ges.stichprobe<br>(N = 374) |                       | Unausgelesene<br>(N = 172) |                       | Hoch Intelligente<br>(N = 170) |                       |
|---|--------|-----------------------------|-----------------------|----------------------------|-----------------------|--------------------------------|-----------------------|
|   |        | <i>p</i>                    | <i>r<sub>it</sub></i> | <i>p</i>                   | <i>r<sub>it</sub></i> | <i>p</i>                       | <i>r<sub>it</sub></i> |
| <i>Aufgaben „Backward Digit Span“ (BDS)</i> |        |                             |                       |                            |                       |                                |                       |
| BDS 1                                       | BDS I  | .81                         | .33                   | .73                        | .28                   | .91                            | .18                   |
| BDS 2                                       |        | .68                         | .39                   | .61                        | .37                   | .81                            | .36                   |
| BDS 3                                       | BDS II | .45                         | .47                   | .36                        | .46                   | .55                            | .35                   |
| BDS 4                                       |        | .48                         | .52                   | .38                        | .53                   | .59                            | .47                   |
| BDS 5                                       | BDS    | .21                         | .50                   | .12                        | .42                   | .34                            | .45                   |
| BDS 6                                       | III    | .30                         | .53                   | .19                        | .39                   | .44                            | .49                   |
| BDS 7                                       |        | .26                         | .54                   | .16                        | .43                   | .41                            | .50                   |
| BDS 8                                       |        | .22                         | .52                   | .09                        | .34                   | .37                            | .46                   |
| BDS 9                                       | BDS    | .17                         | .46                   | .09                        | .31                   | .28                            | .47                   |
| BDS 10                                      | IV     | .10                         | .37                   | .06                        | .32                   | .14                            | .38                   |
| BDS 11                                      |        | .17                         | .44                   | .08                        | .19                   | .27                            | .40                   |
| BDS 12                                      |        | .06                         | .28                   | .04                        | .27                   | .09                            | .23                   |

Anmerkung. Level: RS I = 3 Sätze; RS II = 4 Sätze; RS III = 5 Sätze; RS IV = 6 Sätze; VS I = 3 Wörter, 2 Entscheidungen; VS IIa = 4 Wörter, 2 Entscheidungen; VS IIb = 4 Wörter, 3 Entscheidungen; VS IIIa = 5 Wörter, 2 Entscheidungen; VS IIIb = 5 Wörter, 3 Entscheidungen; VS IV = 6 Wörter, 2 Entscheidungen; CS I = 3 Gleichungen; CS II = 4 Gleichungen; CS III = 5 Gleichungen; CS IV = 6 Gleichungen; CS V = 7 Gleichungen; BDS I = 4 Ziffern; BDS II = 5 Ziffern; BDS III = 6 Ziffern; BDS IV = 7 Ziffern.

Insgesamt sechs Aufgaben weisen einen Schwierigkeitsparameter von  $\sigma > 2$  auf und sind daher als sehr schwierig einzuschätzen. Die klassisch berechneten Trennschärfen (punktbiseriale Korrelationen) der verbalen und numerischen Aufgaben liegen mit Werten zwischen  $r_{it} = .24$  und  $r_{it} = .59$  ( $M = .43$ ,  $SD = .09$ ) in einem angemessenen Bereich. Da die Trennschärfen nicht sehr breit streuen, spricht dieser Befund zunächst nicht gegen eine Passung des 1PL-Modells, das gleiche Trennschärfen für alle Aufgaben einer Skala annimmt. Der Trennschärfeparameter im 1PL-Modell beträgt für alle Aufgaben  $\beta = .79$  ( $SE_{\beta} = .01$ ). Nach dem 2PL-Modell liegen die Schwierigkeitsparameter zwischen  $\sigma = -1.70$  und  $\sigma = 2.79$  ( $M = .54$ ,  $SD = 1.15$ ) und die Trennschärfeparameter zwischen  $\beta = .52$  und  $\beta = 1.43$  ( $M = .87$ ,  $SD = .22$ ).

Die Konstruktion der Aufgaben zielte darauf ab, dass die einzelnen Aufgaben mit zunehmendem Level innerhalb einer Aufgabengruppe schwieriger werden. Die Schwierigkeiten der Aufgaben innerhalb eines Levels sollten jedoch etwa gleich ausfallen, da sie dieselbe Struktur aufweisen und sich nur in den Iteminhalten unterscheiden. Insgesamt spiegeln sowohl die klassisch berechneten Schwierigkeiten als auch die geschätzten Schwierigkeitsparameter die Tendenz einer ansteigenden Schwierigkeit innerhalb der einzelnen Aufgabengruppen wider. In der Regel erwei-

sen sich die Aufgaben eines höheren Levels auch empirisch tatsächlich als schwieriger als die Aufgaben eines niedrigeren Levels. So sind beispielsweise die Aufgaben des Levels RS III deutlich schwieriger als die Aufgaben des Levels RS I. Die einzelnen Aufgaben innerhalb eines Levels sind in der Regel von einem vergleichbaren Schwierigkeitsniveau. Einzelne Aufgaben weichen jedoch von diesem Trend ab. Auffällig ist beispielsweise die Aufgabe RS 7, die deutlich schwieriger ist als die beiden strukturell äquivalenten Aufgaben RS 6 und RS 8, alle drei Aufgaben erfordern die Bearbeitung von fünf Sätzen.

Tabelle 7.24 (wird fortgesetzt): Aufgabenparameter und Fit der verbalen und numerischen Aufgaben für das 1PL- und 2PL-Modell

| Verbal-<br>num.<br>Skala | 1 PL-Modell |               |            |                     | 2 PL-Modell |         |               |              |            |
|--------------------------|-------------|---------------|------------|---------------------|-------------|---------|---------------|--------------|------------|
|                          | $\sigma$    | $SE_{\sigma}$ | $\chi^2 p$ | $p$<br>( $X > Zq$ ) | $\sigma$    | $\beta$ | $SE_{\sigma}$ | $SE_{\beta}$ | $\chi^2 p$ |
| RS 1                     | -0.28       | 0.09          | .97        | .27                 | -0.28       | 0.72    | 0.10          | 0.09         | .65        |
| RS 2                     | -0.93       | 0.10          | .74        | .45                 | -0.94       | 0.76    | 0.13          | 0.10         | .89        |
| RS 3                     | 0.11        | 0.10          | <u>.00</u> | <u>.97</u>          | 0.12        | 1.11    | 0.07          | 0.13         | <u>.02</u> |
| RS 4                     | -0.22       | 0.09          | .36        | .11                 | -0.23       | 0.65    | 0.11          | 0.08         | .84        |
| RS 5                     | -0.47       | 0.09          | .06        | .33                 | -0.47       | 0.75    | 0.11          | 0.08         | .19        |
| RS 6                     | 0.73        | 0.10          | .24        | <u>.99</u>          | 0.63        | 1.16    | 0.08          | 0.15         | .14        |
| RS 7                     | 1.79        | 0.13          | <u>.03</u> | <u>.98</u>          | 1.42        | 1.25    | 0.11          | 0.20         | .60        |
| RS 8                     | 1.39        | 0.11          | .06        | .88                 | 1.21        | 1.04    | 0.11          | 0.14         | .79        |
| RS 9                     | 2.70        | 0.18          | .86        | .59                 | 2.55        | 0.85    | 0.41          | 0.21         | .91        |
| RS 10                    | 2.30        | 0.16          | .54        | .89                 | 1.92        | 1.08    | 0.21          | 0.22         | .85        |
| RS 11                    | 2.62        | 0.18          | .56        | .63                 | 2.34        | 0.93    | 0.32          | 0.21         | .72        |
| VS 1                     | -0.68       | 0.10          | .71        | .28                 | -0.71       | 0.70    | 0.12          | 0.09         | .81        |
| VS 2                     | -0.49       | 0.10          | .28        | .53                 | -0.47       | 0.80    | 0.10          | 0.09         | .29        |
| VS 3                     | -0.03       | 0.10          | .36        | .83                 | -0.01       | 0.95    | 0.08          | 0.11         | .56        |
| VS 4                     | 0.20        | 0.10          | .27        | .43                 | 0.23        | 0.78    | 0.09          | 0.10         | .66        |
| VS 5                     | 0.03        | 0.09          | .15        | <u>.01</u>          | 0.05        | 0.57    | 0.12          | 0.08         | .16        |
| VS 6                     | 0.31        | 0.09          | .35        | .58                 | 0.31        | 0.85    | 0.09          | 0.11         | .71        |
| VS 7                     | 0.30        | 0.09          | .26        | .69                 | 0.30        | 0.86    | 0.09          | 0.10         | .91        |
| VS 8                     | 1.73        | 0.12          | .21        | .90                 | 1.50        | 1.02    | 0.14          | 0.17         | .57        |
| VS 9                     | 1.51        | 0.11          | .75        | .28                 | 1.50        | 0.79    | 0.17          | 0.12         | .93        |
| VS 10                    | 1.89        | 0.13          | .44        | .33                 | 1.86        | 0.80    | 0.24          | 0.14         | .90        |

Anmerkung. RS = Reading Span, VS = Verbal Span, CS = Computation Span, BDS = Backward Digit Span;  $\chi^2 p$ : Signifikanztest auf Basis des  $\chi^2$ -Test;  $p$  ( $X > Zq$ ): Signifikanztest auf Basis des Q-Index; eine signifikante Abweichung wird bei Werten  $<.05$  (Underfit) und Werten  $>.95$  (Overfit) angenommen. Signifikante Werte sind unterstrichen. Datenbasis: 371 bearbeitete Aufgabensets.

Tabelle 7.24 (Fortsetzung): Aufgabenparameter und Fit der verbalen und numerischen Aufgaben für das 1PL- und 2PL-Modell

| Verbal-<br>num.<br>Skala | 1 PL-Modell |               |            |                     | 2 PL-Modell |         |               |              |            |
|--------------------------|-------------|---------------|------------|---------------------|-------------|---------|---------------|--------------|------------|
|                          | $\sigma$    | $SE_{\sigma}$ | $\chi^2 p$ | $p$<br>( $X > Zq$ ) | $\sigma$    | $\beta$ | $SE_{\sigma}$ | $SE_{\beta}$ | $\chi^2 p$ |
| CS 1                     | -1.17       | 0.10          | .43        | .35                 | -1.25       | 0.69    | 0.17          | 0.10         | .91        |
| CS 2                     | -1.17       | 0.10          | .07        | <u>.04</u>          | -1.52       | 0.52    | 0.23          | 0.08         | .80        |
| CS 3                     | -0.84       | 0.09          | .13        | .14                 | -1.01       | 0.58    | 0.17          | 0.08         | .73        |
| CS 4                     | -0.69       | 0.09          | .08        | .05                 | -0.86       | 0.55    | 0.16          | 0.07         | .58        |
| CS 5                     | -0.12       | 0.09          | .30        | .10                 | -0.12       | 0.63    | 0.11          | 0.08         | .69        |
| CS 6                     | 0.22        | 0.09          | .05        | <u>.01</u>          | 0.28        | 0.54    | 0.12          | 0.08         | .41        |
| CS 7                     | 1.15        | 0.11          | <u>.01</u> | <u>.87</u>          | 1.04        | 0.98    | 0.10          | 0.14         | <u>.03</u> |
| CS 8                     | 1.08        | 0.11          | .61        | .88                 | 0.97        | 0.99    | 0.10          | 0.13         | .96        |
| CS 9                     | 2.48        | 0.16          | .98        | .16                 | 2.79        | 0.65    | 0.45          | 0.13         | .97        |
| CS 10                    | 1.98        | 0.14          | <u>.00</u> | <u>.99</u>          | 1.49        | 1.43    | 0.10          | 0.24         | .43        |
| BDS 1                    | -1.46       | 0.10          | .84        | .15                 | -1.70       | 0.62    | 0.21          | 0.09         | .70        |
| BDS 2                    | -0.78       | 0.10          | .16        | .12                 | -0.88       | 0.63    | 0.14          | 0.08         | .27        |
| BDS 3                    | 0.21        | 0.09          | .48        | .31                 | 0.23        | 0.77    | 0.09          | 0.10         | .43        |
| BDS 4                    | 0.06        | 0.09          | .19        | .62                 | 0.08        | 0.84    | 0.09          | 0.09         | .86        |
| BDS 5                    | 1.34        | 0.12          | .35        | <u>.96</u>          | 1.14        | 1.12    | 0.10          | 0.15         | .95        |
| BDS 6                    | 0.87        | 0.10          | .82        | .93                 | 0.78        | 1.01    | 0.09          | 0.13         | .97        |
| BDS 7                    | 1.05        | 0.11          | <u>.01</u> | <u>.98</u>          | 0.90        | 1.12    | 0.09          | 0.14         | .60        |
| BDS 8                    | 1.27        | 0.11          | .05        | <u>.98</u>          | 1.07        | 1.14    | 0.09          | 0.14         | .69        |
| BDS 9                    | 1.58        | 0.12          | .47        | .92                 | 1.36        | 1.05    | 0.13          | 0.15         | .67        |
| BDS 10                   | 2.13        | 0.15          | .37        | .84                 | 1.78        | 1.08    | 0.17          | 0.18         | .87        |
| BDS 11                   | 1.56        | 0.12          | <u>.01</u> | .81                 | 1.43        | 0.92    | 0.14          | 0.13         | .87        |
| BDS 12                   | 2.66        | 0.18          | .81        | .73                 | 2.37        | 0.93    | 0.33          | 0.20         | .93        |

Anmerkung. RS = Reading Span, VS = Verbal Span, CS = Computation Span, BDS = Backward Digit Span;  $\chi^2 p$ : Signifikanztest auf Basis des  $\chi^2$ -Test;  $p$  ( $X > Zq$ ): Signifikanztest auf Basis des Q-Index; eine signifikante Abweichung wird bei Werten  $<.05$  (Underfit) und Werten  $>.95$  (Overfit) angenommen. Signifikante Werte sind unterstrichen. Datenbasis: 371 bearbeitete Aufgabensets.

Möglicherweise spielen hier Reihenfolgeeffekte im Sinne von Übungs- und Ermüdungseffekten eine Rolle. Die erste Aufgabe eines neuen Levels erweist sich oft als etwas schwieriger als die weiteren Aufgaben desselben Levels. Die Aufgabe RS 1 beispielsweise fiel den Pbn deutlich schwerer als die äquivalente Aufgabe RS 2, dies kann vermutlich auf Startschwierigkeiten zurückgeführt werden (es ist die erste Testaufgabe insgesamt). Bei den sehr schwierigen Aufgaben (z. B. Level BDS IV) kommt möglicherweise ein Ermüdungseffekt zum Tragen: Die letzte BDS-Aufgabe dieses Levels wird von deutlich weniger Pbn richtig gelöst als die erste Aufgabe desselben Levels. Es wird jedoch vermutet, dass diese Effekte alle Pbn in ähnlichem

Maße betreffen, so dass sie keine Verletzung der lokalen stochastischen Unabhängigkeit darstellen sollten (s. u.).

Betrachtet man die Passung der Aufgaben auf das 1PL-Modell in Tabelle 7.24, so zeigt sich, dass 31 der 43 Aufgaben weder nach dem  $\chi^2$ -Test noch nach dem Q-Index eine signifikante Abweichung von den Vorgaben des 1PL-Modells aufweisen. Zwölf Aufgaben weichen jedoch nach einem der beiden oder beiden Maßen signifikant vom 1PL-Modell ab. Nach dem  $\chi^2$ -Test betrifft dies die Aufgaben RS 3 und CS 10 (Alpha-Niveau von 1%) und die Aufgaben RS 7, CS 7, BDS 7 und BDS 11 (Alpha-Niveau von 5%). Der Q-Index zeigt einen Underfit auf 1%-Niveau für die Aufgaben VS 5 und CS 6. Einen Overfit zeigen die Aufgaben RS 3, RS 7, BDS 5, BDS 7 und BDS 8 (5%-Niveau) sowie die Aufgaben RS 6 und CS 10 (1%-Niveau).

Für diese Aufgaben ist zu prüfen, ob die Abweichung vom 1PL-Modell inhaltlich zu erklären ist. Hierzu werden die Aufgaben daraufhin inspiziert, ob einzelne Items der Aufgaben auffällig sind (z. B. im Hinblick auf die Lösungshäufigkeiten der einzelnen Items) oder ob die Position der Aufgabe in der Darbietungsreihenfolge eine mögliche Erklärung für die Modellabweichung geben kann. Letztlich sind Interpretationen der Ergebnisse im Sinne von Reihenfolgeeffekten hier jedoch spekulativ. Da alle Pbn die Aufgaben in derselben Reihenfolge bearbeitet haben, kann hier nicht entschieden werden, ob die Ergebnisse auf Eigenschaften der Aufgabe an sich oder auf Positionseffekte zurückgehen. Es kann jedoch vermutet werden, dass sich Unterschiede in den Kennwerten bei den Aufgaben eines Levels vor allem über solche Effekte erklären lassen, da die Variation der Inhalte bei dieser Art von Aufgaben nicht so sehr ins Gewicht fallen dürfte.

Bei der Aufgabe RS 3 haben Pbn mit einer mittleren Fähigkeitsausprägung eine deutlich zu hohe Lösungswahrscheinlichkeit als nach der Funktion des 1PL-Modells zu erwarten wäre. Die beobachtete Lösungswahrscheinlichkeit liegt für diese Pbn außerhalb des 95%-Konfidenzintervalls um die Item Response Funktion. Zusätzlich fällt auf, dass RS 3 für das Level relativ schwierig ist. Über den Inhalt der einzelnen Items lässt sich das jedoch nicht erklären. Die Aufgabe RS 3 ist die erste Aufgabe im gesamten Test, bei der die Anzahl der Items erhöht wird. Zunächst werden zwei Aufgaben mit je drei Items bearbeitet, bei RS 3 müssen dann vier Items bearbeitet werden. Obwohl dieses Prinzip der kontinuierlich steigenden Anzahl von Items bereits in den drei Übungsaufgaben eingeführt wurde, ist es möglich, dass einige Pbn davon überrascht wurden, dass ein Item mehr als in der Aufgabe zuvor dargeboten wurde. Während bei den Items 1 bis 3 jeweils zwischen 88 und 91% der Pbn die richtige Entscheidung trafen, waren es bei Item 4 nur noch 82%. Die Aufgabe RS 3 hat (gemeinsam mit RS 6, die nach dem Q-Index einen Overfit aufweist) die höchste Trennschärfe der gesamten Skala, was die Verletzung des 1PL-Modells erklären könnte. Auch die vom 1PL-Modell abweichende Aufgabe RS 7 erweist sich – im Vergleich zu den anderen Aufgaben des Levels – als relativ schwierig. Bei RS 7 haben Pbn mit einer eher höheren Fähigkeitsausprägung eine zu hohe Lösungswahrscheinlichkeit. Ebenso wie bei Aufgabe RS 3 lassen sich auch hier keine inhaltlichen Besonderheiten der einzelnen Items ausmachen, die die Abweichung erklären könnten. Für die Abweichung von Aufgabe VS 5, die nach dem Q-Index einen Underfit

aufweist und die strukturell zur Aufgabe VS 4 äquivalent ist, findet sich ebenfalls keine inhaltlich begründete Erklärung.

Bei der Aufgabe CS 2 (Underfit nach Q-Index) zeigt die ICC für das 1PL-Modell, dass die Aufgabe im oberen Fähigkeitsbereich nur wenig trennscharf ist. Im höheren Fähigkeitsbereich liegen die Lösungswahrscheinlichkeiten unter den Erwartungen nach der Item Response Funktion. Auch bei den Aufgaben CS 4 und CS 6 bleibt die Lösungswahrscheinlichkeit der Pbn mit relativ hoher Fähigkeit unter den Erwartungen. Im schwierigsten Level scheint sich diese Beziehung jedoch umzukehren: Hier löst ein größerer Anteil der hoch intelligenten Pbn die zweite Aufgabe (CS 10), die auch deutlich trennschärfer ist als die erste Aufgabe (CS 9). Bei der Aufgabe CS 10 ist die Lösungswahrscheinlichkeit für relativ fähige Pbn höher als erwartet. Für die Abweichung der Aufgabe CS 7 vom 1PL-Modell lässt sich hingegen keine vergleichbare oder inhaltlich begründete Erklärung finden.

Bei den BDS-Aufgaben finden sich Modellabweichungen bei den mittelschweren Aufgaben (Level BDS III) und bei der sehr schwierigen Aufgabe BDS 11. Diese Aufgaben haben relativ hohe Trennschärfen. Bei der Betrachtung der jeweiligen ICC zeigt sich, dass sie gerade im höheren Fähigkeitsbereich sehr trennscharf sind.

Diejenigen Aufgaben, die vom 1PL-Modell abweichen und eine hohe Trennschärfe aufweisen, sind möglicherweise mit dem 2PL-Modell, das keine gleichen Trennschärfen voraussetzt, besser zu beschreiben. Tatsächlich zeigt sich, dass bei der Parameterschätzung nach dem 2PL-Modell nur noch zwei Aufgaben der Skala signifikant abweichen (RS 3 und CS 7). Die Abweichung der Aufgabe RS 3 vom 1PL-Modell kann somit nicht nur über eine höhere Trennschärfe erklärt werden.

Insgesamt kann somit das 2PL-Modell einige der verbalen und numerischen Aufgaben besser anpassen als das 1PL-Modell. Erwartungsgemäß gelingt die Anpassung durch das 3PL-Modell nicht besser, ebenso wie bereits beim 1PL-Modell werden auch hier die Aufgaben RS 6, RS 7, CS 7 und BDS 11 auffällig.

### *Figurale Skala*

Auch für die figurale Skala werden zunächst die Schwierigkeiten der Aufgaben betrachtet (siehe Tabelle 7.25). Es findet sich ebenso wie bei der verbal-numerischen Skala eine breite Streuung der Aufgabenschwierigkeiten ( $p = .05$  bis  $p = .86$ ;  $M = .38$ ,  $SD = .22$ ). Von den 27 Aufgaben weisen sieben Aufgaben eine Schwierigkeit von  $p < .20$  auf, jedoch nur eine Aufgabe eine Schwierigkeit von  $p > .80$ . Sechs Aufgaben wurden auch in der Gruppe der hoch Intelligenten nur von weniger als 20% richtig gelöst. Die Schwierigkeitsparameter, geschätzt nach dem 1PL-Modell, liegen für die figuralen Aufgaben zwischen  $\sigma = -2.11$  und  $\sigma = 3.31$  ( $M = .72$ ,  $SD = 1.35$ , siehe Tabelle 7.26). Fünf figurale Aufgaben haben einen Schwierigkeitsparameter  $> 2$ . Die klassisch berechneten Trennschärfen der figuralen Aufgaben streuen von  $r_{it} = .18$  bis  $r_{it} = .45$  ( $M = .35$ ,  $SD = .07$ ).

Tabelle 7.25: Klassische Schwierigkeiten und Trennschärfen der figuralen Skala

| Figurale Skala                                  | Level    | Ges.stichprobe<br>(N = 374) |                       | Unausgelesene<br>(N = 172) |                       | Hoch Intelligente<br>(N = 170) |                       |
|---|----------|-----------------------------|-----------------------|----------------------------|-----------------------|--------------------------------|-----------------------|
|   |          | <i>p</i>                    | <i>r<sub>it</sub></i> | <i>p</i>                   | <i>r<sub>it</sub></i> | <i>p</i>                       | <i>r<sub>it</sub></i> |
| <i>Aufgaben „Memory Updating Figural“ (MUF)</i> |          |                             |                       |                            |                       |                                |                       |
| MUF 1   | MUF Ia   | .64                         | .38                   | .58                        | .34                   | .74                            | .33                   |
| MUF 2   |          | .47                         | .44                   | .34                        | .44                   | .61                            | .35                   |
| MUF 3   | MUF Ib   | .77                         | .33                   | .74                        | .27                   | .83                            | .30                   |
| MUF 4   |          | .66                         | .37                   | .57                        | .38                   | .75                            | .20                   |
| MUF 5   | MUF IIa  | .38                         | .35                   | .34                        | .44                   | .49                            | .18                   |
| MUF 6   |          | .11                         | .18                   | .08                        | .09                   | .14                            | .20                   |
| MUF 7   | MUF IIb  | .35                         | .27                   | .30                        | .33                   | .45                            | .21                   |
| MUF 8   | MUF IIc  | .47                         | .38                   | .37                        | .26                   | .62                            | .32                   |
| MUF 9   |          | .42                         | .28                   | .37                        | .32                   | .47                            | .24                   |
| MUF 10  |          | .34                         | .37                   | .25                        | .33                   | .44                            | .35                   |
| MUF 11  |          | .40                         | .41                   | .30                        | .43                   | .51                            | .32                   |
| MUF 12  |          | .26                         | .35                   | .20                        | .30                   | .32                            | .33                   |
| MUF 13  | MUF IIIa | .12                         | .28                   | .04                        | .04                   | .18                            | .25                   |
| MUF 14  | MUF IIIb | .12                         | .25                   | .06                        | .17                   | .15                            | .23                   |
| <i>Aufgaben „Spatial Working Memory“ (SWM)</i>  |          |                             |                       |                            |                       |                                |                       |
| SWM 1   | SWM I    | .86                         | .29                   | .80                        | .28                   | .94                            | .19                   |
| SWM 2   |          | .52                         | .41                   | .40                        | .30                   | .63                            | .33                   |
| SWM 3   |          | .47                         | .45                   | .34                        | .42                   | .63                            | .38                   |
| SWM 4   | SWM II   | .58                         | .42                   | .48                        | .38                   | .69                            | .46                   |
| SWM 5   |          | .60                         | .38                   | .48                        | .20                   | .76                            | .35                   |
| SWM 6   |          | .30                         | .43                   | .22                        | .35                   | .42                            | .43                   |
| SWM 7   |          | .56                         | .38                   | .46                        | .30                   | .64                            | .24                   |
| SWM 8   | SWM III  | .37                         | .43                   | .26                        | .34                   | .45                            | .34                   |
| SWM 9   |          | .16                         | .27                   | .06                        | .19                   | .22                            | .27                   |
| SWM 10  |          | .24                         | .39                   | .16                        | .30                   | .34                            | .32                   |
| SWM 11  |          | .10                         | .34                   | .05                        | .11                   | .14                            | .31                   |
| SWM 12  | SWM IV   | .05                         | .29                   | .02                        | .26                   | .07                            | .27                   |
| SWM 13  |          | .06                         | .20                   | .03                        | .17                   | .07                            | .17                   |

Anmerkung. Level: MUF Ia = 2 Zellen, 3 Operationen; MUF Ib = 2 Zellen, 4 Operationen; MUF IIa = 3 Zellen, 2 Operationen; MUF IIb = 3 Zellen, 3 Operationen, MUF IIc = 3 Zellen, 4 Operationen; MUF IIIa = 4 Zellen, 2 Operationen; MUF IIIb = 4 Zellen, 4 Operationen. SWM I = 1 Muster; SWM II = 2 Muster; SWM III = 3 Muster; SWM IV = 4 Muster.

Tabelle 7.26: Aufgabenparameter und Fit der figuralen Aufgaben für das 1PL- und 2PL-Modell

| Figurale Skala | 1 PL-Modell |               |            |                    | 2 PL-Modell |         |               |              |            |
|----------------|-------------|---------------|------------|--------------------|-------------|---------|---------------|--------------|------------|
|                | $\sigma$    | $SE_{\sigma}$ | $\chi^2 p$ | $\frac{p}{(X>Zq)}$ | $\sigma$    | $\beta$ | $SE_{\sigma}$ | $SE_{\beta}$ | $\chi^2 p$ |
| MUF 1          | -0.66       | 0.12          | .67        | .41                | -0.67       | 0.61    | 0.14          | 0.08         | .47        |
| MUF 2          | 0.14        | 0.12          | .27        | .77                | 0.14        | 0.71    | 0.10          | 0.10         | .55        |
| MUF 3          | -1.41       | 0.13          | .74        | .36                | -1.45       | 0.59    | 0.21          | 0.09         | .84        |
| MUF 4          | -0.77       | 0.12          | .64        | .39                | -0.78       | 0.60    | 0.15          | 0.09         | .74        |
| MUF 5          | 0.57        | 0.11          | .26        | .24                | 0.62        | 0.54    | 0.14          | 0.08         | .09        |
| MUF 6          | 2.36        | 0.16          | .57        | .06                | 2.93        | 0.46    | 0.54          | 0.09         | .92        |
| MUF 7          | 0.72        | 0.11          | .18        | <u>.02</u>         | 0.92        | 0.44    | 0.19          | 0.07         | .70        |
| MUF 8          | 0.17        | 0.11          | .24        | .44                | 0.18        | 0.58    | 0.12          | 0.09         | .17        |
| MUF 9          | 0.40        | 0.11          | .22        | <u>.03</u>         | 0.52        | 0.42    | 0.16          | 0.07         | .13        |
| MUF 10         | 0.78        | 0.12          | .31        | .46                | 0.78        | 0.61    | 0.14          | 0.09         | <u>.02</u> |
| MUF 11         | 0.53        | 0.12          | .30        | .66                | 0.52        | 0.64    | 0.12          | 0.09         | .92        |
| MUF 12         | 1.25        | 0.13          | .94        | .51                | 1.26        | 0.61    | 0.18          | 0.10         | .85        |
| MUF 13         | 2.30        | 0.17          | .91        | .72                | 2.18        | 0.66    | 0.33          | 0.12         | .99        |
| MUF 14         | 2.27        | 0.16          | .95        | .41                | 2.31        | 0.60    | 0.35          | 0.11         | .97        |
| SWM 1          | -2.11       | 0.16          | .74        | .39                | -2.02       | 0.66    | 0.26          | 0.10         | .56        |
| SWM 2          | -0.07       | 0.11          | .59        | .58                | -0.06       | 0.69    | 0.10          | 0.10         | .63        |
| SWM 3          | 0.14        | 0.12          | .26        | .80                | 0.13        | 0.76    | 0.10          | 0.10         | .42        |
| SWM 4          | -0.36       | 0.12          | .09        | .66                | -0.32       | 0.70    | 0.11          | 0.10         | .06        |
| SWM 5          | -0.50       | 0.11          | .52        | .38                | -0.49       | 0.63    | 0.12          | 0.09         | .36        |
| SWM 6          | 1.02        | 0.13          | <u>.03</u> | .92                | 0.86        | 0.82    | 0.12          | 0.11         | .25        |
| SWM 7          | -0.25       | 0.11          | .78        | .39                | -0.24       | 0.64    | 0.11          | 0.09         | .93        |
| SWM 8          | 0.67        | 0.12          | .05        | .84                | 0.58        | 0.78    | 0.10          | 0.12         | .12        |
| SWM 9          | 1.95        | 0.15          | .99        | .36                | 1.99        | 0.60    | 0.30          | 0.11         | .99        |
| SWM 10         | 1.35        | 0.13          | .33        | .86                | 1.16        | 0.78    | 0.15          | 0.12         | .95        |
| SWM 11         | 2.51        | 0.19          | .13        | <u>.98</u>         | 1.84        | 1.02    | 0.20          | 0.17         | .38        |
| SWM 12         | 3.31        | 0.25          | .13        | <u>.98</u>         | 2.22        | 1.19    | 0.27          | 0.28         | .57        |
| SWM 13         | 3.14        | 0.22          | .63        | .59                | 2.75        | 0.74    | 0.45          | 0.17         | .66        |

Anmerkung. MUF = Memory Updating Figural, SWM = Spatial Working Memory;  $\chi^2 p$ : Signifikanztest auf Basis des  $\chi^2$ -Test;  $p (X>Zq)$ : Signifikanztest auf Basis des Q-Index; eine signifikante Abweichung wird bei Werten  $<.05$  (Underfit) und Werten  $>.95$  (Overfit) angenommen. Signifikante Werte sind unterstrichen. Datenbasis: 372 bearbeitete Aufgabensets.

Der Trennschärfeparameter nach dem 1PL-Modell beträgt für alle figuralen Aufgaben  $\beta = .61$  ( $SE_{\beta} = .02$ ). Nach dem 2PL-Modell streuen die Schwierigkeitsparameter zwischen  $\sigma = -2.02$  und  $\sigma = 2.93$  ( $M = .66$ ,  $SD = 1.25$ ) und die Trennschärfeparameter zwischen  $\beta = .42$  und  $\beta = 1.19$  ( $M = .67$ ,  $SD = .16$ ).

Auch bei den figuralen Aufgaben wurde aufgrund der strukturellen Äquivalenz der Aufgaben eines Levels erwartet, dass die empirischen Schwierigkeiten innerhalb eines Levels ungefähr gleich ausfallen sollten. Die Aufgaben in höheren Levels sollten zudem schwieriger sein als in niedrigeren Levels. Für die MUF-Aufgaben zeigen sich empirisch drei grobe Kategorien von Schwierigkeiten, diese entsprechen den Levels I, II und III. Die weitere Unterteilung der Levels in Ia und Ib usw. spiegelt sich in den Daten nicht wider. Somit scheint in diesem Aufgabentyp vor allem die Anzahl parallel zu bearbeitender Zellen relevant zu sein, die Anzahl auszuführender Operationen hingegen scheint keinen Einfluss auf die Schwierigkeit zu haben.

Tendenziell sind die Aufgaben innerhalb desselben Levels ähnlich schwierig. Hinsichtlich der Schwierigkeit fallen zwei MUF-Aufgaben jedoch besonders auf: Die Aufgabe MUF 3 ist deutlich einfacher als die anderen Aufgaben in Level I. MUF 6 (Level II) hingegen ist die schwierigste Aufgabe der ganzen Skala, somit auch schwieriger als die Aufgaben in Level III, nur 11% aller Pbn konnten sie richtig lösen. Beide Auffälligkeiten können vermutlich über einen ähnlichen Effekt erklärt werden.

MUF 3 ist die einzige Aufgabe des Levels, bei der in beiden aktiven Zellen je zwei Operationsanforderungen erfolgen. Die Positionen der Punkte in beiden Zellen müssen aufgrund der neuen Operationsanforderungen von den Pbn kognitiv aktiviert werden, das heißt sie müssen gewissermaßen wieder „aufgefrischt“ werden. In den restlichen Aufgaben des Levels muss stets nur die linke Zelle kognitiv erneut aktiviert werden, die Information aus der rechten Zelle könnte daher im Kurzzeitspeicher der Pbn eher zerfallen. Beispielsweise konnten sich bei MUF 2 nur 50% der Pbn die richtige Position des Punktes in der rechten Zelle merken, jedoch gaben 86% die richtige Position für die linke Zelle an. In der Aufgabe MUF 3 finden sich diese Unterschiede nicht, 86% machten eine korrekte Angabe für die linke Zelle und 82% für die rechte Zelle.

In der Aufgabe MUF 6 werden drei Punkte in drei aktiven Zellen dargeboten und auch die Ergebnisse in den drei Zellen abgefragt. Jedoch erfolgt jeweils nur in den Zellen 1 und 3 eine Operation, so dass die Position des Punktes in Zelle 2 nicht mehr kognitiv aktiviert wird. MUF 6 ist die einzige Aufgabe, bei der in der mittleren Zelle keine Operationsanforderung erfolgt. Während die Schwierigkeit der linken Zelle mit  $p = .81$  in etwa der Lösungshäufigkeit für einzelne Items von Aufgaben desselben Levels entspricht, gab es deutlich weniger richtige Eingaben in der mittleren und der rechten Zelle ( $p = .36$  und  $p = .27$ ). Vermutlich waren die Pbn auch von der anderen Reihenfolge der Operationsanforderungen irritiert. In den vorangegangenen Aufgaben erfolgte die Darbietung der Pfeile stets von links nach rechts, ohne dass dabei eine Zelle ausgelassen wurde.

Bei den SWM-Aufgaben fällt auf, dass sich die Schwierigkeiten der Aufgaben im ersten Level deutlich voneinander unterscheiden. Die Aufgabe SWM 1 wurde von fast doppelt so vielen Pbn gelöst wie die Aufgabe SWM 3. Dieser große Unterschied ist sicherlich auf die Struktur der jeweiligen Muster zurückzuführen. Das zu rotierende Muster in SWM 1 ist weniger komplex und kann leichter mental rotiert werden



als die Muster in SWM 2 und 3. Eine zweite Auffälligkeit ist, dass die SWM-Aufgaben auf Level IV außerordentlich schwierig sind – selbst in der Gruppe der hoch Intelligenten konnten sie nur 7% der Pbn richtig lösen. Jedoch erwiesen sich die Aufgaben auf Level III in einem Vortest für sehr fähige Pbn als noch zu leicht. Auch in dieser Untersuchung wurde beispielsweise die Aufgabe SWM 8 noch von 45% der hoch Intelligenten Pbn richtig gelöst, so dass weitere Aufgaben mit stärkeren Anforderungen nötig waren.

Insgesamt vier figurale Aufgaben weichen nach dem Q-Index signifikant vom 1PL-Modell ab: Die Aufgaben MUF 7 und MUF 9 zeigen einen Underfit (Alpha-Niveau von 5%) und die Aufgaben SWM 11 und SWM 12 einen Overfit (Alpha-Niveau von 5%). Die Aufgabe SWM 6 weicht nach dem  $\chi^2$ -Test signifikant vom 1PL-Modell ab.

Eine Betrachtung der ICC der Aufgaben MUF 7 und MUF 9 zeigt, dass die Aufgaben lediglich im unteren Fähigkeitsbereich trennscharf sind, im mittleren und oberen Fähigkeitsbereich jedoch nicht. Die Lösungswahrscheinlichkeiten der fähigeren Pbn fallen geringer aus als erwartet. Die Aufgaben MUF 7 und MUF 9 enthalten beide drei parallele Zellen, bei beiden erfolgen zunächst Operationen in jeder der Zellen, bei MUF 9 zusätzlich eine zweite Operation in der letzten Zelle. Es lässt sich keine inhaltliche Begründung finden, warum diese beiden Aufgaben vom 1PL-Modell abweichen. Die Aufgaben SWM 11 und SWM 12 sind beide sehr schwierig, nur Pbn mit einem Fähigkeitsparameter von mindestens etwa  $\sigma = 0$  haben überhaupt eine Chance, diese Aufgaben richtig zu lösen. Im höheren Fähigkeitsbereich zwischen etwa 1 und 2 ist insbesondere die Aufgabe SWM 11 jedoch recht trennscharf: Ein Pb mit einem Fähigkeitsparameter von 1.5 hat eine deutlich höhere Wahrscheinlichkeit, die Aufgabe zu lösen, als ein Pb mit einer Fähigkeitsausprägung von 1. Bei der Aufgabe SWM 6, der einzigen Aufgabe, die nach dem  $\chi^2$ -Test signifikant vom 1PL-Modell abweicht, passt das 1PL-Modell die Daten im niedrigen und mittleren Fähigkeitsbereich gut an. Sehr fähige Pbn haben hier jedoch eine deutlich zu hohe Lösungswahrscheinlichkeit, als nach dem 1PL-Modell zu erwarten wäre. Die drei vom 1PL-Modell abweichenden SWM-Aufgaben zeichnen sich somit insgesamt durch eine an sich – für eine Skala, die im oberen Fähigkeitsbereich differenzieren soll – wünschenswerte Trennschärfe im höheren Fähigkeitsbereich aus. Diese verletzt jedoch die Vorgaben des 1PL-Modells. Entsprechend können diese drei Aufgaben, ebenso wie die beiden schlecht passenden MUF-Aufgaben, durch das 2PL-Modell gut angepasst werden. Jedoch führt auch das 2PL-Modell nicht zu einer guten Anpassung aller figuralen Aufgaben. Bei einer Schätzung der Parameter nach dem 2PL-Modell ergibt sich für die Aufgabe MUF 10, die durch das 1PL-Modell gut angepasst werden konnte, ein schlechter Fit.

### *Fazit*

Insgesamt kann festgehalten werden, dass sich die meisten Aufgaben beider Skalen durch das 1PL-Modell gut anpassen lassen. Bei einigen Aufgaben können inhaltliche Anhaltspunkte der Aufgaben oder ihre Position in der Darbietungsreihenfolge als mögliche Ursache für die Modellabweichung herangezogen werden, bei manchen Aufgaben gelingt dies jedoch nicht. Prinzipiell sollten die Aufgaben mit signifikan-

ten Modellabweichungen eliminiert werden, um eine Rasch-Modell konforme Skala zu erhalten. Da insgesamt 70 Aufgaben für zwei Testskalen vorhanden sind, wäre auch nach der Eliminierung der nicht modellkonformen Aufgaben noch eine ausreichende Anzahl an Testaufgaben vorhanden. Dennoch werden die Aufgaben mit einem schlechten Fit zunächst nicht aus den Skalen herausgenommen. Einige Aufgaben weichen von den Vorgaben des 1PL-Modells deshalb ab, da sie an verschiedenen Punkten auf der Fähigkeitsdimension unterschiedlich trennscharf sind. Wenn eine besonders hohe Trennschärfe im oberen Fähigkeitsbereich gegeben ist, wäre es an sich wünschenswert, für einen Test zur Messung im hohen Fähigkeitsbereich entsprechende Aufgaben einzusetzen. Daher wird zunächst geprüft, ob die vollständigen Skalen verschiedenen Modellgeltungstests standhalten.

#### 7.5.4 Modellgeltungstests

Eine Prüfung der Modellgeltung erfolgt zunächst über einen Vergleich der Passung der Daten auf das 1PL-, 2PL- und 3PL-Modell. Tabelle 7.27 zeigt die Anzahl der unabhängigen Modellparameter, die Log-Likelihood multipliziert mit (-2) und die informationstheoretischen Maße AIC, BIC und CAIC für die beiden Skalen.

*Tabelle 7.27: Modellvergleiche für die zwei Arbeitsgedächtnis-Skalen anhand der Log-Likelihood und informationstheoretischer Maße*

| Skala/<br>Modell        | Parameter<br>$n_p$ | -2 Log L | AIC             | BIC             | CAIC            |
|-------------------------|--------------------|----------|-----------------|-----------------|-----------------|
| <i>verbal-numerisch</i> |                    |          |                 |                 |                 |
| 1 PL                    | 85                 | 14504.56 | 14674.56        | <i>15006.06</i> | <i>15091.06</i> |
| 2 PL                    | 128                | 14367.86 | <i>14623.86</i> | 15123.06        | 15251.06        |
| 3 PL                    | 171                | 14503.04 | 14845.04        | 15511.94        | 15682.94        |
| <i>figural</i>          |                    |          |                 |                 |                 |
| 1 PL                    | 53                 | 10199.84 | <i>10305.84</i> | <i>10512.54</i> | <i>10565.54</i> |
| 2 PL                    | 80                 | 10161.12 | 10321.12        | 10633.12        | 10713.12        |
| 3 PL                    | 107                | 10324.17 | 10538.17        | 10955.47        | 11062.47        |

Anmerkung: Der jeweils kleinste AIC-, BIC- und CAIC-Wert pro Skala ist kursiv gesetzt.

Die informationstheoretischen Maße sprechen überwiegend für eine beste Passung des 1PL-Modells. Der AIC-Index fällt in der verbal-numerischen Skala zwar für das 2PL-Modell am geringsten aus, bei einer Berücksichtigung des Stichprobenumfangs (siehe BIC und CAIC) schneidet jedoch das 1PL-Modell am besten ab. Bei der figuralen Skala deuten alle drei Maße auf die relativ beste Passung des 1PL-Modells hin.

Zusätzlich wird die Modellpassung des geschätzten 1PL-Modells überprüft, indem die empirischen Häufigkeiten der Antwortpattern mit den nach dem 1PL-Modell erwarteten Patternhäufigkeiten verglichen werden. Da in den Daten keine ausreichende Anzahl von Antwortpattern enthalten ist, bzw. für zu viele Antwortpattern

keine empirische Häufigkeit vorliegt, kann nicht davon ausgegangen werden, dass die entsprechenden Prüfgrößen ( $\chi^2$ -Test nach Pearson sowie nach Cressie Read) tatsächlich  $\chi^2$ -verteilt sind. Es wird daher ein Bootstrap-Verfahren durchgeführt, um eine Häufigkeitsverteilung von  $\chi^2$ -Werten zu erhalten. Hierzu werden mit dem Programm Winmira 2001 aus den Parameterschätzungen auf Basis der Testdaten 40 artifizielle Datensätze resimuliert. Es können dann die geschätzten  $\chi^2$ -Werte aus den empirischen Daten zu den  $\chi^2$ -Werten aus den simulierten Datensätzen in Beziehung gesetzt werden.

Für die verbal-numerische Skala ergibt der  $\chi^2$ -Test nach Pearson einen empirischen p-Wert von .050 und der  $\chi^2$ -Test nach Cressie Read einen empirischen p-Wert von .025. Für die Entscheidung über eine Modellakzeptanz wird ein Schwellenwert von  $p = .025$  verwendet (von Davier, 1997; Langeheine, Pannekoek & van de Pol, 1995). Die  $\chi^2$ -Werte aus den Daten weichen somit weder nach dem  $\chi^2$ -Test nach Cressie Read noch nach dem  $\chi^2$ -Test nach Pearson signifikant von der Verteilung der Likelihoods aus den simulierten Daten ab. Die geschätzte Likelihood der figuralen Skala weicht ebenfalls nach keinem der beiden Tests signifikant von der Verteilung der  $\chi^2$ -Werte ab ( $\chi^2$ -Test nach Cressie Read:  $p = .175$ ;  $\chi^2$ -Test nach Pearson:  $p = .300$ ).

Weiteren Aufschluss über die Passung der Modelle geben weiterhin Testinformationfunktionen, die das Ausmaß an Information beschreiben, die die gesamte Skala für verschiedene Punkte der Fähigkeitsdimension liefert. Diese Information verhält sich komplementär zur Größe des Standardfehlers: Je geringer der Standardfehler an einem Punkt auf der Fähigkeitsdimension, desto größer das Ausmaß an Information. Die Informationsfunktion wird beeinflusst sowohl durch die Qualität als auch durch die Quantität von Aufgaben für den jeweiligen Punkt auf der Fähigkeitsdimension. Die Abbildungen 7.9 bis 7.12 zeigen die Testinformationfunktionen für die beiden Skalen, jeweils für das 1PL- und das 2PL-Modell. Abgetragen ist auf der linken vertikalen Achse die Testinformation und auf der rechten vertikalen Achse der Standardfehler, die horizontale Achse stellt die Fähigkeit der Pbn dar.

Die Funktionen nach beiden Modellen zeigen für die verbal-numerische Skala das größte Ausmaß an Information und den geringsten Standardfehler im höheren Fähigkeitsbereich. Die Modellierung nach dem 1PL-Modell erbringt über einen breiten Bereich der Fähigkeitsdimension relativ genaue Messungen mit einem kleinen Standardfehler. Das 2PL-Modell hingegen erbringt die beste Messung im Bereich der Fähigkeit um 1. Im 2PL-Modell ist der Bereich einer genauen Messung jedoch recht schmal: unterhalb einer Fähigkeit von 0 und oberhalb einer Fähigkeit von 2 steigt der Standardfehler rapide.

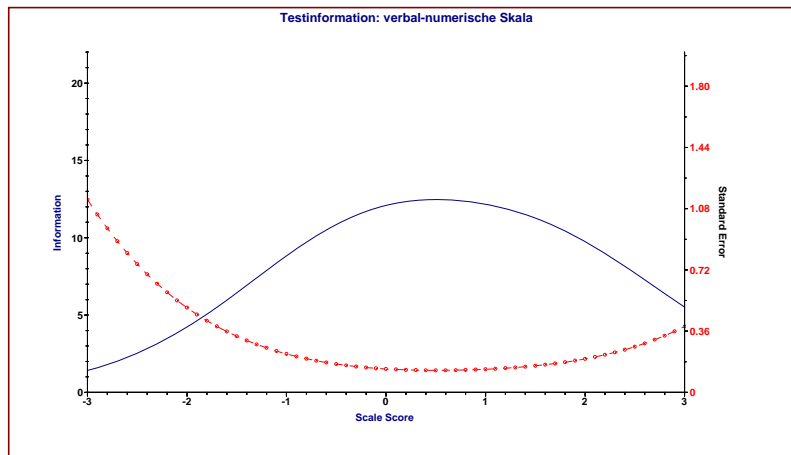


Abbildung 7.9: Testinformationsfunktion für die verbal-numerische Skala nach dem 1PL-Modell

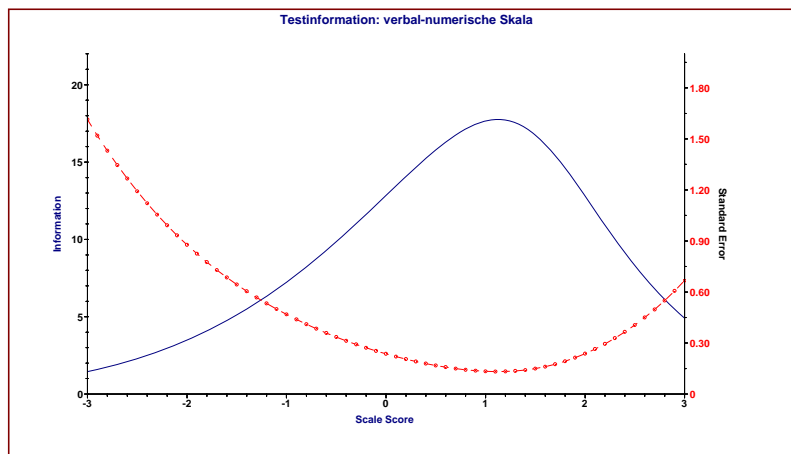


Abbildung 7.10: Testinformationsfunktion für die verbal-numerische Skala nach dem 2PL-Modell

Für die figurale Skala zeigt sich ebenfalls, dass sie die meisten Informationen bei Messungen im höheren Fähigkeitsbereich erbringt. Das 1PL- und das 2PL-Modell zeigen einen ähnlichen Verlauf des Standardfehlers und der Testinformation: Der Fehler fällt im Fähigkeitsbereich zwischen 0 und 2 am geringsten aus. Das 2PL-Modell erbringt für diesen Fähigkeitsbereich jedoch insgesamt etwas mehr Informationen als das 1PL-Modell.

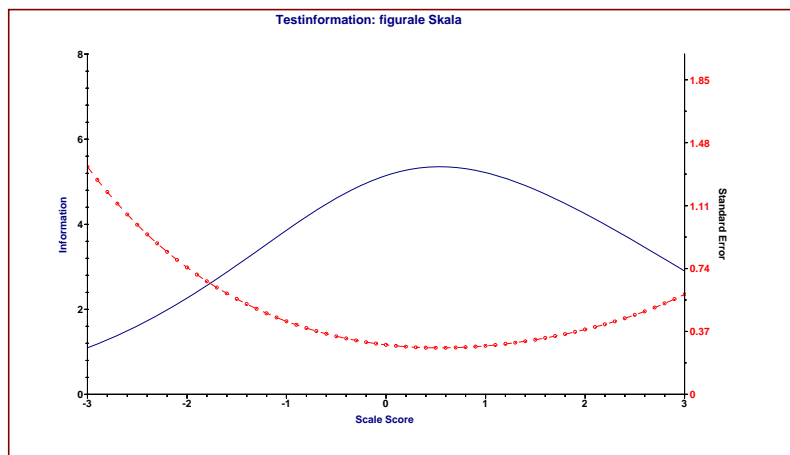


Abbildung 7.11: Testinformationsfunktion für die figurale Skala nach dem 1PL-Modell

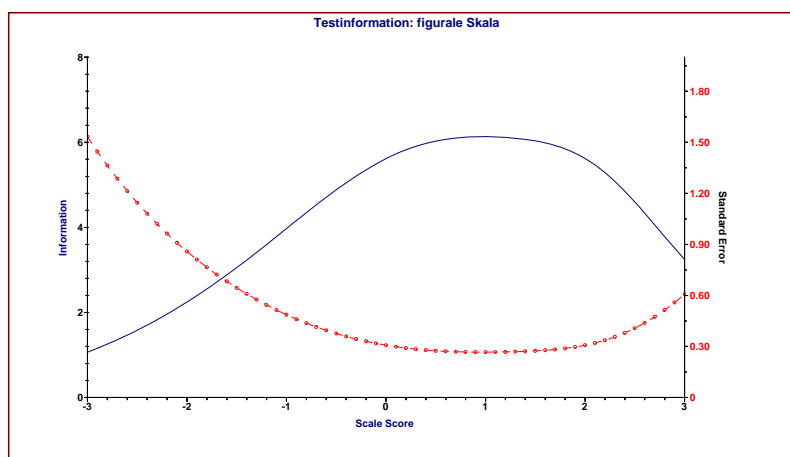


Abbildung 7.12: Testinformationsfunktion für die figurale Skala nach dem 2PL-Modell

Zusammenfassend kann für die verbal-numerische Skala festgehalten werden, dass sowohl der Modellvergleich über die informationstheoretischen Maße als auch der Vergleich der empirischen und der erwarteten Patternhäufigkeiten im Bootstrap und die Auswertung der Testinformationsfunktion für eine relativ gute Anpassung der Daten durch das 1PL-Modell sprechen. Im Modellvergleich schneidet das 1PL-Modell für die verbal-numerische Skala am besten ab, wenn der Stichprobenumfang mit berücksichtigt wird (siehe BIC und CAIC). Nach dem 2PL-Modell erbringt die Skala zwar im Fähigkeitsbereich um 1 deutlich mehr Information als das 1PL-Modell, jedoch ist nach dem 2PL-Modell der Standardfehler im höheren Fähigkeitsbereich deutlich größer. Da die Skala jedoch gerade auch im oberen Fähigkeitsbereich reliable Messungen ermöglichen soll, scheint hier das 1PL-Modell zu den besseren Schätzungen zu führen.

Für die figurale Skala präferieren die Ergebnisse des Modellvergleichs über die informationstheoretischen Maße eindeutig das 1PL-Modell. Die  $\chi^2$ -Tests im Bootstrap deuten ebenfalls darauf hin, dass die Daten hinreichend gut durch das 1PL-Modell zu beschreiben sind. Nach dem Verlauf der Testinformationsfunktion und des Standardfehlers wäre hier jedoch eher das 2PL-Modell zu bevorzugen, da das 2PL-Modell über einen größeren Fähigkeitsbereich mehr an Informationen erbringt.

### 7.5.5 Spezifische Modelltests für das 1PL-Modell

Für das 1PL-Modell existieren – im Gegensatz zum 2PL- und 3PL-Modell – zusätzlich verschiedene explizite Tests, die die Modellgeltung prüfen. Untersucht werden dabei zwei zentrale Annahmen des 1PL-Modells, die Personen- und die Aufgabenhomogenität. Die Ergebnisse dieser Prüfungen werden im Folgenden dargestellt. Zunächst erfolgen jedoch einige Überlegungen zu einer weiteren zentralen Annahme des Rasch-Modells, der lokalen stochastischen Unabhängigkeit.

#### *Lokale stochastische Unabhängigkeit*

Die lokale stochastische Unabhängigkeit erfordert, dass die Antworten der einzelnen Pbn unabhängig sind, wenn ihre Fähigkeit statistisch kontrolliert wird (Rost, 2004). Das bedeutet, dass die Lösungswahrscheinlichkeit eines Probanden für ein Item nur von der Fähigkeit des Pb auf der betreffenden latenten Dimension und von der Schwierigkeit des Items abhängt. Durch das Antwortverhalten des Pb auf vorhergehende oder nachfolgende Items wird die Lösungswahrscheinlichkeit hingegen nicht beeinflusst. Eine besondere Gefährdung der lokalen stochastischen Unabhängigkeit stellen Lerneffekte während der Aufgabenbearbeitung im Sinne von Einsichts- oder Verstärkungslernen (nicht im Sinne reiner Übung) dar. Die Übungsaufgaben der vorliegenden Testbatterie erlauben jedoch ein relativ ausführliches Vertrautmachen mit dem jeweiligen Aufgabentyp. Verständnisprobleme werden in der Übungsphase von den Testleitern entdeckt und können behoben werden, da der Pb erst mit den Testaufgaben beginnen kann, wenn er alle Übungsaufgaben richtig bearbeitet hat. Es kann so davon ausgegangen werden, dass alle Pbn – trotz möglicher unterschiedlicher Ausgangsbedingungen – in etwa die gleichen Startbedingungen haben. Einsichtslernen sollte somit in den Testaufgaben nicht mehr stattfinden.

#### *Prüfung der Personenhomogenität*

Das 1PL-Modell geht mit der Annahme der Personenhomogenität davon aus, dass alle Pbn die Aufgaben der Skala aufgrund derselben Fähigkeit bearbeiten, also bei allen Pbn dieselbe Personenvariable erfasst wird (Rost, 2004). Die Rangreihe der Schwierigkeitsparameter sollte somit in allen Teilstichproben konstant sein. Diese Annahme wird für die vorliegenden Testskalen überprüft, indem die Gesamtstichprobe nach verschiedenen Kriterien in jeweils zwei Teile geteilt wird und die Schwierigkeitsparameter für beide Subgruppen separat geschätzt werden. Anschließend wird mit dem Programm LPCM-Win 1.0 (Fischer & Ponocny-Seliger, 1998)

über graphische Modelltests und Likelihoodquotiententests nach Andersen (1973) überprüft, ob sich die Schwierigkeitsparameter in beiden Gruppen bedeutsam unterscheiden. Die Aufgabenparameter werden in LPCM-Win 1.0 mit (-1) multipliziert, so dass schwierigere Aufgaben geringere Werte aufweisen als einfachere Aufgaben (von etwa -4 = sehr schwierig bis +4 = sehr einfach). Als Splitkriterien werden der Median des Gesamtscores der jeweiligen Skala, das Geschlecht und das Alter (Median: 11.4 Jahre) verwendet. Von besonderem Interesse ist hier zudem ein weiteres Kriterium: Unterscheiden sich die Schwierigkeitsschätzungen der Aufgaben für durchschnittlich und hoch intelligente Probanden? Zur Prüfung dieser Frage wird die Teilgruppenkonstanz für die Teilstichproben der „hoch Intelligenten“ (HI) und der „nicht hoch Intelligenten“ (NHI) untersucht. Wenn die Annahme der Personenhomogenität erfüllt ist, dürfen die Itemparameter in jeweils verschiedenen Subgruppen nicht bedeutsam voneinander abweichen.

Im graphischen Modelltest werden die separat geschätzten Schwierigkeitsparameter für zwei Teilgruppen in einem Streudiagramm gegeneinander abgetragen. Eine gute Übereinstimmung ist gegeben, wenn die Parameter möglichst nah an einer 45-Grad-Linie im Diagramm liegen. Zusätzlich vermittelt die graphische Darstellung der Parameterschätzungen einen Eindruck darüber, wie sich die Aufgaben auf der Fähigkeitsdimension verteilen.

Für die verbal-numerische Skala zeigt sich eine gute Übereinstimmung der Parameter für die Teilgruppen Mädchen vs. Jungen und die Teilgruppen jüngere vs. ältere Pbn. Bei den Kriterien „Score“ und „Intelligenz“ (HI vs. NHI) weichen jedoch einzelne Aufgaben deutlich von der Diagonalen ab (siehe Abbildungen 7.13 bis 7.16). Da für den graphischen Modelltest keine signifikanzstatistische Kennwerte vorliegen, gibt es hier kein explizites Kriterium dafür, wann eine Aufgabe zu stark abweicht. Alle graphischen Modelltests ergeben, dass die verbal-numerischen Aufgaben breit über die Fähigkeitsdimension verteilt sind, wobei die Parameter für einen relativ großen Anteil der Aufgaben sehr niedrig geschätzt werden. Das zeigt erneut, dass ein relativ großer Anteil sehr schwieriger Aufgaben in der Testskala enthalten ist.

Tendenziell haben die leichteren Aufgaben etwas höhere Schwierigkeitsparameter bei den Jungen, die schwierigeren Aufgaben etwas höhere Schwierigkeitsparameter bei den Mädchen. Das bedeutet, dass die einfacheren Aufgaben für die Jungen, die schwierigeren Aufgaben hingegen für die Mädchen etwas leichter sind. Bei der Teilung nach dem Alter hat die Aufgabe RS 7 eine höhere, die Aufgaben VS 7 und BDS 1 haben hingegen eine niedrigere Ausprägung des Schwierigkeitsparameters in der Gruppe der älteren Pbn als in der Gruppe der jüngeren Pbn. Beim Kriterium „Score“ fällt eine deutliche Abweichung von der Diagonale auf für die Aufgaben VS 5 und VS 8; CS 2, CS 6, CS 7 und CS 10 sowie für BDS 5, BDS 7, BDS 8 und BDS 11. Die insgesamt leichteren Aufgaben (CS 2, CS 6, VS 5) fallen hier den Pbn mit einem geringeren Gesamtscore etwas leichter, die schwierigeren Aufgaben hingegen den Pbn mit einem höheren Gesamtscore (CS 7, CS 10, VS 8 sowie die schwierigeren BDS-Aufgaben). Diese Abweichungen entsprechen überwiegend den

Abweichungen, die bereits mit den beiden Itemstatistiken  $\chi^2$ -Test und Q-Index (vgl. Tabelle 7.24) festgestellt wurden.

Die Teilung der Gruppe nach „hoch Intelligenten“ und „nicht hoch Intelligenten“ ergibt deutliche Abweichungen nur für die Aufgaben der CS-Aufgabengruppe: Die Parameter der relativ einfachen Aufgaben CS 2 und CS 3 werden in der Gruppe der Pbn mit einem geringeren IQ höher (das heißt: leichter) geschätzt, die schwierige Aufgabe CS 10 hingegen wird in der HI-Gruppe höher geschätzt.

Insgesamt kann festgestellt werden, dass vor allem die numerischen Aufgaben zu abweichenden Parameterschätzungen in verschiedenen, nach den Fähigkeiten aufgeteilten Teilgruppen der Stichprobe führen. Insbesondere die einfacheren Aufgaben der CS-Aufgabengruppe (bei denen sich die Pbn nur drei oder vier Ziffern in Folge merken müssen) erhalten tendenziell höhere Schätzwerte in der Gruppe mit der jeweils geringeren Fähigkeit. Für weniger fähige Pbn scheinen daher die insgesamt leichteren CS-Aufgaben im Vergleich mit den anderen Aufgaben der Skala eher einfach. Auf der anderen Seite zeigt sich, dass die schwierigeren Aufgaben der BDS-Aufgabengruppe (bei denen sich die Pbn sechs oder mehr Ziffern in Folge merken müssen) für die fähigeren Pbn in Relation zu den anderen Aufgaben der Skala relativ einfach sind. Möglicherweise kann dieser Befund zu den BDS-Aufgaben damit erklärt werden, dass die Verarbeitungsanforderung in diesem Aufgabentyp für sehr fähige Personen geringer ist als in anderen Aufgabentypen (vgl. Engle, Tuholski et al., 1999; Schofield & Ashman, 1986).

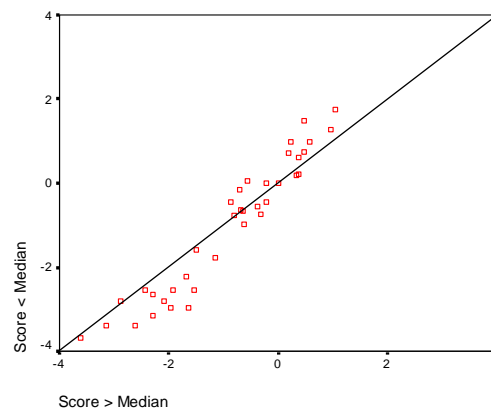


Abbildung 7.13: Graphischer Modelltest für die verbal-numerische Skala, Split-Kriterium: Score



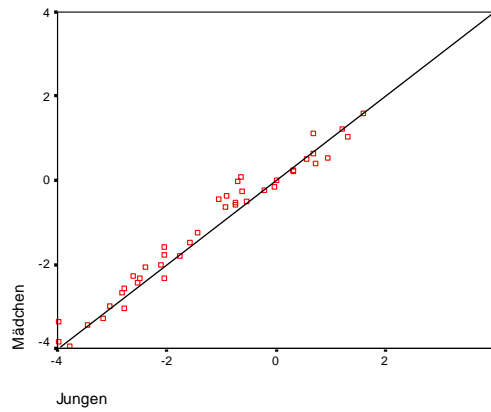


Abbildung 7.14: Graphischer Modelltest für die verbal-numerische Skala, Split-Kriterium: Geschlecht

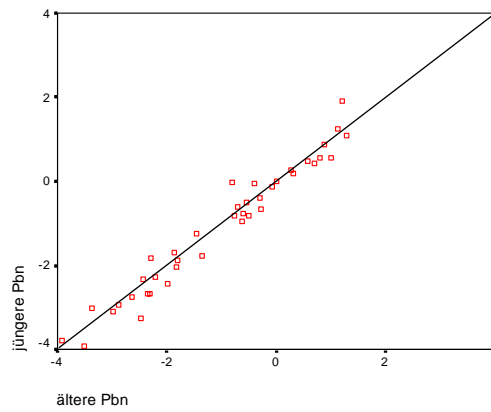


Abbildung 7.15: Graphischer Modelltest für die verbal-numerische Skala, Split-Kriterium: Alter

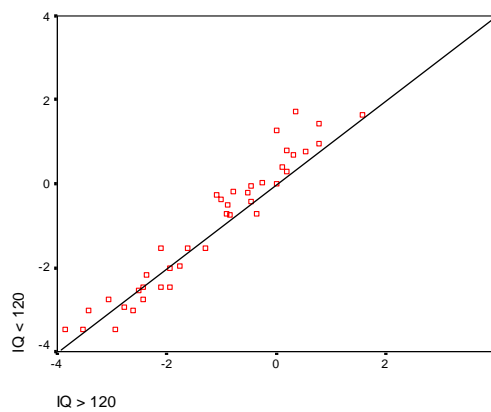


Abbildung 7.16: Graphischer Modelltest für die verbal-numerische Skala, Split-Kriterium: Intelligenz (HI vs. NHI)

Die Aufgaben der figuralen Skala verteilen sich ebenfalls über einen breiten Bereich der Fähigkeitsdimension und die Schwierigkeitsparameter liegen für die verschiedenen Split-Kriterien in den Modelltests überwiegend relativ nah entlang der Geraden (siehe Abbildungen 7.17 bis 7.21). Bei der Unterteilung nach dem Geschlecht fällt auf, dass die figuralen Aufgaben für die Mädchen tendenziell als etwas leichter geschätzt werden als für die Jungen. Insbesondere die anspruchsvollen Aufgaben MUF 13 und MUF 14 sowie SWM 9 und SWM 12 weichen deutlich in dieser Richtung von der Diagonalen ab. Bei der Teilung nach den Kriterien „Score“ und „Intelligenz“ wird die mit Abstand leichteste Aufgabe der gesamten Skala, Aufgabe SWM 1, bei der nur ein Muster rotiert und behalten werden muss, für die jeweils fähigere Teilgruppe als relativ leichter geschätzt als für die jeweils weniger fähige Gruppe.

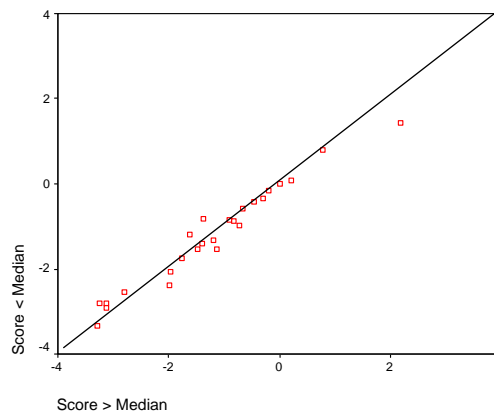


Abbildung 7.17: Graphischer Modelltest für die figurale Skala, Split-Kriterium: Score

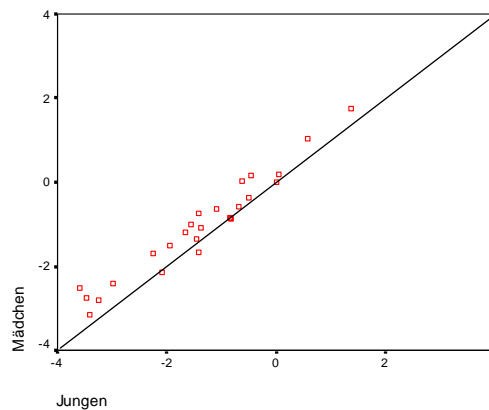


Abbildung 7.18: Graphischer Modelltest für die figurale Skala, Split-Kriterium: Geschlecht

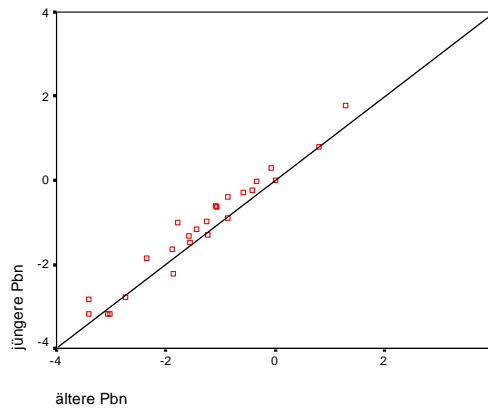


Abbildung 7.19: Graphischer Modelltest für die figurale Skala, Split-Kriterium: Alter

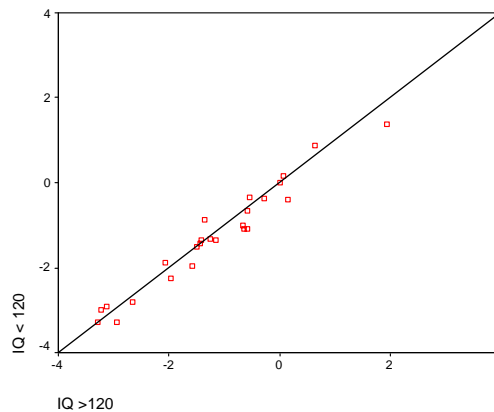


Abbildung 7.20: Graphischer Modelltest für die figurale Skala, Split-Kriterium: Intelligenz (HI vs. NHI)

Eine signifikanzstatistische Prüfung der Aufgabenhomogenität für die Skalen erfolgt mit dem Andersen-Test. Das Prinzip dieser Modelltestung besteht in der Prüfung, ob die Daten besser durch bedingte Likelihoodfunktionen aus Parameterschätzungen auf der Basis des gesamten Datensatzes oder auf der Basis von Teilgruppen erklärt werden können (Rost, 2004). Es wird dazu ein Quotient aus der bedingten Likelihood des restriktiveren Modells (das 1PL-Modell gilt für die gesamte Stichprobe) im Zähler und der bedingten Likelihood des weniger restriktiven Modells (das 1PL-Modell gilt innerhalb einzelner Teilgruppen, diese können unterschiedliche Aufgabenparameter aufweisen) im Nenner gebildet (cLR). Je heterogener die Gruppen sind, desto größer wird die Likelihood im Nenner im Verhältnis zur Likelihood im Zähler und umso eher wird die Prüfgröße  $-2\log(\text{cLR})$  signifikant. Die Ergebnisse des Andersen-Tests sind Tabelle 7.28 zu entnehmen.

Tabelle 7.28: Likelihoodquotiententests nach Andersen mit verschiedenen Teilungskriterien für die beiden Skalen

| Teilungskrit.                  | <i>N</i> | Gr. 1     | <i>n</i> <sub>1</sub> | Gr. 2     | <i>n</i> <sub>2</sub> | $\chi^2$ | <i>df</i> | <i>p</i> |
|--------------------------------|----------|-----------|-----------------------|-----------|-----------------------|----------|-----------|----------|
| <b>Verbal-numerische Skala</b> |          |           |                       |           |                       |          |           |          |
| Score (Median)                 | 371      | hoch      | 177                   | niedrig   | 194                   | 124.67   | 42        | .00      |
| Geschlecht                     | 370      | Mädchen   | 168                   | Jungen    | 202                   | 42.58    | 42        | .45      |
| Alter (Median)                 | 371      | < 11.4 J. | 185                   | > 11.4 J. | 186                   | 51.56    | 42        | .15      |
| Intelligenz<br>(CFT-IQ 120)    | 371      | HI        | 202                   | NHI       | 169                   | 84.25    | 42        | .00      |
| <b>Figurale Skala</b>          |          |           |                       |           |                       |          |           |          |
| Score (Median)                 | 372      | hoch      | 180                   | niedrig   | 192                   | 22.89    | 26        | .64      |
| Geschlecht                     | 371      | Mädchen   | 168                   | Jungen    | 203                   | 33.82    | 26        | .14      |
| Alter (Median)                 | 372      | < 11.4 J. | 184                   | > 11.4 J. | 188                   | 26.16    | 26        | .45      |
| Intelligenz<br>(CFT-IQ 120)    | 372      | HI        | 202                   | NHI       | 170                   | 29.08    | 26        | .31      |

Anmerkungen: HI = hoch Intelligente, Pbn der Gesamtstichprobe mit einem CFT-IQ  $\geq$  120; NHI = nicht hoch Intelligente, Pbn der Gesamtstichprobe mit einem CFT-IQ < 120.

Für die verbal-numerische Skala ergibt sich, dass sich die Parameterschätzungen für Mädchen und Jungen sowie für jüngere und ältere Pbn der Stichprobe nicht signifikant unterscheiden. Bei Gruppenbildungen nach den eher fähigkeitsbezogenen Kriterien „Score“ und „Intelligenz“ kann hingegen keine hinreichende Teilgruppenkonstanz festgestellt werden. Bei der figuralen Skala führt keines der Split-Kriterien zu einer signifikanten Abweichung, auch der im graphischen Modelltest beobachtete Unterschied zwischen Mädchen und Jungen lässt sich mit dem Andersen-Test nicht bestätigen.

Die Personenhomogenität kann somit für die figurale Skala uneingeschränkt angenommen werden. Für die verbal-numerische Skala wird im Weiteren untersucht, ob sich durch die Eliminierung abweichender Aufgaben eine Personenhomogenität für die verbleibenden Aufgaben der Skala herstellen lässt.

Die stärksten Abweichungen von verbal-numerischen Aufgaben finden sich im graphischen Modelltest für das Kriterium „Score“. Daher wird zunächst versucht, durch die sukzessive Eliminierung von Aufgaben eine Teilgruppenkonstanz für Pbn mit hohem und mit niedrigem Score zu erreichen. Zunächst werden aufgrund der Beobachtung, dass insbesondere die einfachen CS-Aufgaben und die schwierigen BDS-Aufgaben zu einer Verletzung der Personenhomogenität führen, nur die im Modelltest abweichenden numerischen Aufgaben eliminiert (CS 2, CS 6, CS 7, CS 10, BDS 5, BDS 7, BDS 8 und BDS 11). Die verbleibenden Aufgaben führen jedoch weiterhin zu einem signifikanten Befund im Andersen-Test für das Kriterium „Score“ ( $\chi^2 = 78.41$ ,  $df = 34$ ,  $p = .00$ ).

Es zeigt sich, dass es erforderlich ist, alle nach dem 1PL-Modell schlecht passenden Aufgaben herauszunehmen, um eine hinreichend gute Teilgruppenkonstanz zu erreichen. Im Folgenden werden daher zusätzlich die schlecht passenden verbalen Aufgaben herausgenommen. Insgesamt werden die Aufgaben RS 3, RS 6, RS 7, VS 5, VS 8 (Abweichung im graphischen Modelltest, jedoch guter Fit nach der  $\chi^2$ -Itemstatistik), CS 2, CS 6, CS 7, CS 10, BDS 5, BDS 7, BDS 8 und BDS 11 eliminiert. Es resultiert im Andersen-Test ein  $\chi^2 = 48.57$  ( $df = 29$ ), der Andersen-Test wird daher auf einem Alpha-Niveau von 1% nicht signifikant ( $p = .01$ ). Für die verbleibenden Aufgaben der Skala besteht nun auch eine hinreichende Teilgruppenkonstanz für „hoch Intelligente“ und „nicht hoch Intelligente“ ( $\chi^2 = 35.54$ ,  $df = 29$ ,  $p = .19$ ).

Es verbleiben 30 verbal-numerische Aufgaben, für die eine Teilgruppenkonstanz entsprechend der verschiedenen Splitkriterien festgestellt werden kann. Diese 30 Aufgaben verteilen sich über alle vier Aufgabengruppen, und es bleiben Aufgaben aus jedem (a priori) angenommenen Schwierigkeitslevel erhalten. Um eine befriedigende Personenhomogenität herzustellen, mussten jedoch auch einige Aufgaben eliminiert werden, die sich im hohen Fähigkeitsbereich als besonders trennscharf herausgestellt hatten (z. B. die Aufgaben CS 10 und BDS 11).

#### *Prüfung der Aufgabenhomogenität*

Im Folgenden wird die Annahme des 1PL-Modells geprüft, dass verschiedene Aufgabensubgruppen einer Skala dieselbe latente Dimension erfassen. Hierzu wird der Modelltest nach Martin-Löf (1973) im Programm LPCM-Win 1.0 (Fischer & Ponocny-Seliger, 1998) verwendet, der untersucht, inwieweit die Summenscores der beiden Testteile übereinstimmen. Der zugehörige Signifikanztest ist ein modifizierter Likelihoodquotiententest, der auf den bedingten Likelihoods der beiden Testteile beruht (Rost, 2004). Wenn der Test signifikant wird, sind die Aufgaben nicht homogen. Für jede Skala werden die Teilungskriterien „Aufgabentyp“, „Aufgabennummer“ (gerade vs. ungerade Aufgaben) und „Schwierigkeit“ verwendet. Da in der verbal-numerischen Skala vier verschiedene Aufgabentypen enthalten sind, erfolgt hier einmal eine Teilung nach dem Inhalt (verbale Aufgabengruppen vs. numerische Aufgabengruppen) und einmal nach der Struktur der Aufgaben („Reading Span“ und „Computation Span“ vs. „Verbal Span“ und „Backward Digit Span“). Die Teilung nach der Struktur bezieht sich auf die strukturelle Äquivalenz der Reading Span und der Computation Span. Das Kriterium „Schwierigkeit“ teilt die Skala in Aufgaben mit einem Schwierigkeitsparameter unterhalb und oberhalb des Skalenmittelwertes von  $\sigma_{1PL} = .65$  (verbal-numerische Skala) bzw.  $\sigma_{1PL} = .72$  (figurale Skala).

Für die verbal-numerische Skala erfolgt die Prüfung zum einen mit sämtlichen 43 Aufgaben und zum anderen mit den 30 Aufgaben, die sich entsprechend des Andersen-Tests als hinreichend personenhomogen erwiesen haben. Die Ergebnisse der Modelltests nach Martin-Löf finden sich in Tabelle 7.29.

Die Ergebnisse zeigen, dass die Aufgabenhomogenität sowohl für die gesamte verbal-numerische Skala als auch für die reduzierte Skala mit 30 Aufgaben uneinge-

schränkt angenommen werden kann. Keines der Teilungskriterien deutet auf eine Verletzung der Aufgabenhomogenität hin.

Tabelle 7.29: Modelltests nach Martin-Löf zur Prüfung der Aufgabenhomogenität

| Teilungskriterium   | Gr. 1                | $k_1$ | Gr. 2                   | $k_2$ | $\chi^2$ | df  | p    |
|---|----------------------|-------|-------------------------|-------|----------|-----|------|
| <i>Verbal-numerische Skala (gesamt)</i>                       |                      |       |                         |       |          |     |      |
| Aufgabentyp: Inhalt   | verbal               | 21    | numerisch               | 22    | 384.84   | 461 | 1.00 |
| Aufgabentyp: Struktur   | RS/CS                | 21    | VS/BDS                  | 22    | 339.80   | 461 | 1.00 |
| Aufgabennummer  | gerade               | 21    | ungerade                | 22    | 199.27   | 461 | 1.00 |
| Schwierigkeit   | $\sigma_{IPL} < .65$ | 22    | $\sigma_{IPL} \geq .65$ | 21    | 252.33   | 461 | 1.00 |
| <i>Verbal-numerische Skala (30 personenhomogene Aufgaben)</i> |                      |       |                         |       |          |     |      |
| Aufgabentyp: Inhalt   | verbal               | 16    | numerisch               | 14    | 226.20   | 223 | .43  |
| Aufgabentyp: Struktur   | RS/CS                | 14    | VS/BDS                  | 16    | 182.43   | 223 | .98  |
| Aufgabennummer  | gerade               | 14    | ungerade                | 16    | 100.56   | 223 | 1.00 |
| Schwierigkeit   | $\sigma_{IPL} < .65$ | 18    | $\sigma_{IPL} \geq .65$ | 12    | 95.93    | 215 | 1.00 |
| <i>Figurale Skala (gesamt)</i>                                |                      |       |                         |       |          |     |      |
| Aufgabentyp   | MUF                  | 14    | SWM                     | 13    | 274.11   | 181 | .00  |
| Aufgabentyp<br>(reduzierte Skala)                             | MUF                  | 12    | SWM                     | 10    | 196.00   | 119 | .00  |
| Aufgabennummer  | gerade               | 13    | ungerade                | 14    | 146.94   | 181 | .97  |
| Schwierigkeit   | $\sigma_{IPL} < .72$ | 15    | $\sigma_{IPL} \geq .72$ | 12    | 116.82   | 179 | 1.00 |

Anmerkung. Reduzierte Skala: Figurale Skala ohne die Aufgaben MUF 7, MUF 9, SWM 6, SWM 11 und SWM 12.

Die Kriterien „Aufgabennummer“ und „Schwierigkeit“ erbringen für die figurale Skala ebenfalls keinen Hinweis auf eine mangelnde Aufgabenhomogenität. Jedoch zeigt der  $\chi^2$ -Test eine signifikante Abweichung für das Kriterium „Aufgabentyp“, die Aufgaben der Gruppen „Memory Updating Figural“ und „Spatial Working Memory“ sind somit nicht homogen. Eine wiederholte Analyse ohne die figuralen Aufgaben, die signifikant vom IPL-Modell abweichen, zeigt keine verbesserte Übereinstimmung. Es ist somit fraglich, ob – trotz der Befunde in den vorherigen Modellgeltungstests, die insgesamt für eine gute Passung des IPL-Modells auf die Skala sprechen – eine Zusammenfassung der beiden figuralen Aufgabengruppen zu einer gemeinsamen Skala tatsächlich sinnvoll ist. Möglicherweise ist die Verletzung der Aufgabenhomogenität durch unterschiedliche Anforderungen in den Verarbeitungskomponenten in beiden Aufgabentypen zu erklären: Während bei „Memory Updating Figural“ die Position von Punkten innerhalb einer Zelle mental verschoben werden muss, also die Vorstellung von Bewegungen einzelner Punkte verlangt, erfordert die „Spatial Working Memory“ eine mentale Rotation von zum Teil relativ komplexen Strukturen. Auch die Wiedergabe der gespeicherten Informationen durch den Pb ist in den beiden Aufgabentypen unterschiedlich anspruchsvoll und fehleranfällig: Bei der „Spatial Working Memory“ müssen die gemerkten Muster aufge-

zeichnet werden, während bei der „Memory Updating Figural“ ein Tastendruck für die Beantwortung eines Items genügt.

### 7.5.6 Reliabilität und externe Validität der Rasch-Modell konformen Skalen

Im Folgenden werden die klassischen Gütekriterien Reliabilität und externe Validität für die reduzierten Testskalen berichtet. Aufgrund der Ergebnisse der IRT-Analysen wurden nicht-modellkonforme Aufgaben eliminiert, so dass an dieser Stelle erneut die Messgüte der nun reduzierten Skalen zu prüfen ist.

Die Analysen nach den IRT-Modellen ergaben, dass für 30 Aufgaben der verbal-numerischen Skala von einer hinreichenden Konformität für das Rasch-Modell ausgegangen werden kann. Für diese 30 modellkonformen Aufgaben der verbal-numerischen Skala stellt daher der Summenscore über alle Aufgaben eine erschöpfende Statistik der Fähigkeit der Pbn dar. Bei der figuralen Skala zeigten sich für fünf Aufgaben signifikante Abweichungen von den Vorgaben des 1PL-Modells (vgl. Tabelle 7.26).

In den verschiedenen Modelltests erreichte die gesamte figurale Skala zwar überwiegend eine gute Passung auf das 1PL-Modell, jedoch führte der Test zur Prüfung der Aufgabenhomogenität für das Kriterium „Aufgabentyp“ zu einem signifikanten Befund. Somit ist die Modellkonformität der figuralen Skala nicht gesichert. Eine Aufteilung der Skala in zwei Subskalen, von denen die eine die Vorstellung einfacher Bewegungsfolgen erfordert („Memory Updating Figural“) und die andere mentale Rotationen verlangt („Spatial Working Memory“), wäre eine mögliche Lösung dieses Problems. Diese Subskalen würden dann jedoch recht kurz und würden sich nicht mehr zu einer hinreichend differenzierten Messung von Fähigkeiten eignen. Die bisherige figurale Skala konnte sich jedoch in den anderen Modellgeltungstests bewähren, auch fanden sich sehr hohe Korrelationen zwischen dem figuralen Faktor und der fluiden Intelligenz im Gesamtmodell kognitiver Faktoren (Abschnitt 7.3.3). Somit weist die Skala verschiedene günstige Eigenschaften auf, obwohl die Aufgaben nicht homogen im Sinne des 1PL-Modells sind. Daher wird auch hier im Folgenden ein Skalenscore über die 22 figuralen Aufgaben berechnet, die nach der  $\chi^2$ -Statistik und dem Q-Index nicht signifikant vom 1PL-Modell abweichen. Die Skalenscores der verbal-numerischen Skala und der figuralen Skala korrelieren zu  $r = .53$  ( $p < .01$ ).

Die Verkürzung der verbal-numerischen Skala auf 30 Aufgaben führt nicht zu einer drastischen Verringerung der Reliabilität: Die innere Konsistenz sinkt lediglich von  $\alpha = .92$  (43 Aufgaben) auf  $\alpha = .88$  (30 Aufgaben), so dass auch für die reduzierte Skala von einer ausreichenden Reliabilität ausgegangen werden kann. Die mittlere, nach der Klassischen Testtheorie berechnete Schwierigkeit der Skala verändert sich durch die Herausnahme der 13 Aufgaben nur minimal, die mittlere Lösungshäufigkeit steigt von  $p = .38$  ( $SD = .23$ ) auf  $p = .41$  ( $SD = .25$ ).

Bei der figuralen Skala führt die Reduktion von 27 auf 22 Aufgaben ebenfalls nur zu einer geringfügigen Verminderung der inneren Konsistenz. Die Skala mit 22 Aufgaben weist mit einem Wert von  $\alpha = .80$  eine noch ausreichende Reliabilität auf

gaben weist mit einem Wert von  $\alpha = .80$  eine noch ausreichende Reliabilität auf (bei 27 Aufgaben:  $\alpha = .82$ ). Die mittlere Lösungshäufigkeit der Aufgaben der Skala verändert sich durch die Elimination der fünf Aufgaben nur minimal und beträgt nun  $p = .39$  ( $SD = .22$ ; bei 27 Aufgaben:  $p = .38$ ,  $SD = .22$ ).

In Abschnitt 7.3.1 wurden bereits die Korrelationen der Scores der sechs Gruppen von Arbeitsgedächtnisaufgaben mit den Intelligenztests und in Abschnitt 7.4 mit den Schulnoten dargestellt. Um jedoch die Übereinstimmungsvalidität der beiden nun gebildeten Testskalen abschätzen zu können, werden im Folgenden diese Zusammenhänge noch einmal auf der Basis der Skalenscores berichtet (siehe Tabelle 7.30). Ergänzt werden weiterhin die Korrelationen der Skalen mit den Merkfähigkeitsaufgaben.

*Tabelle 7.30: Partialkorrelationen der Scores der verbal-numerischen und der figuralen Skala mit den Intelligenztests und Merkfähigkeitsaufgaben (Rohwerte, Alter kontrolliert) und bivariate Korrelationen der Skalen mit den Schulnoten in den Kernfächern*

| Tests und Schulnoten              | Skalenscore verbal-numerisch | Skalenscore figural |
|-----------------------------------|------------------------------|---------------------|
| CFT 20                            | .56                          | .61                 |
| CFT-WS                            | .59                          | .44                 |
| CFT-ZF                            | .57                          | .57                 |
| Wörter merken (WM)                | .43                          | .24                 |
| Zweistellige Zahlen (ZZ)          | .28                          | .19                 |
| Wege erinnern (WE)                | .09 n.s.                     | .26                 |
| <i>Primarstufe</i>                |                              |                     |
| Deutsch ( $N = 113$ )             | -.56                         | -.17                |
| Mathematik ( $N = 113$ )          | -.45                         | -.50                |
| Sachkunde ( $N = 111$ )           | -.46                         | -.35                |
| Gesamtnotenschnitt ( $N = 113$ )  | -.61                         | -.41                |
| <i>Sekundarstufe</i>              |                              |                     |
| Deutsch ( $N = 256$ )             | -.44                         | -.27                |
| Mathematik ( $N = 257$ )          | -.47                         | -.39                |
| Naturwissenschaften ( $N = 251$ ) | -.40                         | -.32                |
| Gesamtnotenschnitt ( $N = 258$ )  | -.53                         | -.37                |

Anmerkung. CFT-WS = Wortschatztest des CFT 20, CFT-ZF = Zahlenfolgentest des CFT 20, Naturwissenschaften = Mittelwert der Noten für Chemie, Physik, Biologie; alle Korrelationen:  $p < .01$ ; n.s.:  $p > .05$ .

Die Korrelationen der Arbeitsgedächtnisskalen mit den drei Intelligenztests liegen in mittlerer Höhe, sind jedoch durchaus vergleichbar mit den Korrelationen dieser Intelligenztests mit anderen konstrukt<sup>n</sup>ahen Intelligenztests (vgl. Abschnitt 5.3.2). Die Zusammenhänge des CFT 20 sowie seiner Zusatzmodule mit eher konstrukt<sup>f</sup>ernen Intelligenztests liegen im Mittel tendenziell unter den hier gefundenen Werten. Dies



kann – auch angesichts der Unterschiede in der Darbietungsweise (computergestützt vs. Paper-Pencil-Format) und in den Aufgabeninhalten – als guter Beleg für die Übereinstimmungsvalidität der Arbeitsgedächtnisskalen interpretiert werden.

Die Merkfähigkeitsaufgaben aus dem BIS-HB korrelieren ebenfalls positiv mit den Testskalen, der höchste Zusammenhang findet sich zwischen der verbalen Merkaufgabe „Wörter merken“ und der verbal-numerischen Skala. Die figurale Aufgabe „Wege erinnern“ korreliert zwar signifikant mit der figuralen Testskala, nicht jedoch mit der verbal-numerischen Skala. Dieser Befund kann als weiterer Beleg für die teilweise Domänenspezifität des Arbeitsgedächtnisses angesehen werden.

Erneut zeigt sich, dass die Zusammenhänge zwischen Arbeitsgedächtnis und Schulleistungen in der Primarstufe etwas höher ausfallen als in der Sekundarstufe. Vergleichbare Befunde sind aus der Untersuchung der Zusammenhänge von Intelligenztestleistungen und Schulnoten bekannt (Jensen, 1980, 1998). Die Höhe der gefundenen Zusammenhänge ist vergleichbar mit den Korrelationen der standardisierten Arbeitsgedächtnistests, über die in Abschnitt 4.3 berichtet wurde. Die Skala „Zentrale Exekutive“ korreliert in der WMTB-C von Pickering und Gathercole (2001) zwischen  $r = .49$  und  $r = .61$  mit sprachlichen und mathematischen Leistungen von sieben- und achtjährigen Kindern. Für den SCT von de Jong und Das-Smaal (1990, 1995), der aus nur einem Aufgabentyp besteht, finden sich bei neunjährigen Kindern Korrelationen zwischen  $r = .29$  und  $r = .35$  mit Schulleistungstests.

Vergleicht man die hier gefundenen Korrelationen mit der Höhe der Korrelationen aus Studien zum Zusammenhang von Intelligenztests und Schulnoten, so wird deutlich, dass die Arbeitsgedächtnisskalen in etwa ebenso stark mit den Schulleistungen zusammenhängen wie unterschiedliche klassische Intelligenztests. Aus verschiedenen Metaanalysen geht hervor, dass der mittlere Zusammenhang zwischen Intelligenz und Schulleistung zwischen etwa  $r = .34$  und  $r = .51$  liegt (Fraser et al., 1987; Amelang & Bartussek, 1997; Jensen, 1998). Mit dem BIS-HB (Jäger et al., 2005, vgl. Abschnitt 2.5) wurden für Schülerinnen und Schüler zwischen 12 und 16 Jahren Korrelationen der Skala „Verarbeitungskapazität“ mit den Noten in den Kernfächern zwischen  $r = -.40$  und  $r = -.44$  ( $N = 1297$ ) ermittelt, mit dem Gesamtnotenschnitt korrelierte die Skala zu  $r = -.50$  ( $N = 1320$ ). Die über den BIS-HB ermittelte allgemeine Intelligenz korrelierte zu  $r = -.54$  mit dem Gesamtnotenschnitt.

Insgesamt weisen die beiden Arbeitsgedächtnisskalen somit nach den Kriterien der Klassischen Testtheorie im Hinblick auf die Reliabilität und die externe Validität eine hinreichende Messqualität auf. Die innere Konsistenz der Skalen ist ausreichend hoch, und die Zusammenhänge der Skalen mit den Außenkriterien „Intelligenztests“ liegen in einem Bereich, der für klassische Intelligenztests üblich ist. Auch die gefundenen Korrelationen zwischen den Skalen und den Schulnoten liegen – insbesondere für die verbal-numerische Skala – in einem Bereich, der auch für Korrelationen zwischen Intelligenztests und Noten typisch ist.

Zur Beurteilung der internen Validität der Skalen kann auf die konfirmatorischen Faktorenanalysen in Abschnitt 7.2 zurückgegriffen werden: Es zeigte sich, dass ein Zwei-Faktoren-Modell mit den Faktoren „verbal-numerisches Arbeitsgedächtnis“

und „figurales Arbeitsgedächtnis“ nicht nur für die Gesamtstichprobe bestätigt werden kann, sondern dass das Modell auch für Kinder verschiedenen Alters und verschiedener intellektueller Begabung sowie für Jungen und Mädchen gleichermaßen gilt.

### 7.5.7 Zusammenfassung der testtheoretischen Analysen und Fazit

Die Befunde aus den testtheoretischen Analysen werden im Folgenden noch einmal kurz zusammengefasst: Auf der Grundlage einer exploratorischen Faktorenanalyse wurden die 70 Arbeitsgedächtnisaufgaben zu zwei Skalen, einer verbal-numerischen (43 Aufgaben) und einer figuralen (27 Aufgaben) Skala zusammengestellt. Die Ermittlung der Aufgabenkennwerte nach der Klassischen Testtheorie für beide Skalen zeigt angemessene Trennschärfen für alle Aufgaben und eine breite Streuung der Schwierigkeiten, wobei in jeder Skala auch ein relativ großer Anteil sehr schwieriger Aufgaben enthalten ist. Auch für die Teilstichprobe der „hoch Intelligenten“ erweisen sich in jeder Skala noch mehrere Aufgaben als sehr schwierig, so dass Deckeneffekte nicht zu befürchten sind. Zusätzlich wurden die Aufgabenparameter für das 1PL- und das 2PL-Modell geschätzt und der Aufgaben-Fit anhand des  $\chi^2$ -Test und des Q-Index bestimmt. Insgesamt zwölf verbale oder numerische Aufgaben und fünf figurale Aufgaben zeigten nach einem der beiden Fit-Maße eine signifikante Abweichung vom 1PL-Modell.

Verschiedene Modellgeltungstests (Vergleich der Modellpassung des 1PL-, 2PL- und 3PL-Modells über informationstheoretische Maße, der Vergleich der nach dem 1PL-Modell erwarteten mit den empirisch gefundenen Antwortpattern im Bootstrap sowie die Inspektion der Testinformationsfunktionen für das 1PL- und das 2PL-Modell) sprechen für beide Skalen überwiegend für eine gute Passung des 1PL-Modells. Nach den Ergebnissen des Andersen-Tests erfüllt die figurale Skala die Voraussetzung der Personenhomogenität uneingeschränkt. Für die verbal-numerische Skala konnte die Personenhomogenität nur für einen reduzierten Aufgabensatz von 30 Aufgaben gezeigt werden. Die Voraussetzung der Aufgabenhomogenität kann sowohl für die gesamte (43 Aufgaben) als auch für die reduzierte (30 Aufgaben) verbal-numerische Skala uneingeschränkt angenommen werden. Bei der figuralen Skala erweisen sich jedoch die beiden Aufgabengruppen „Memory Updating Figural“ und „Spatial Working Memory“ als nicht homogen. Eine Eliminierung von fünf nicht Rasch-Modell konformen Aufgaben erbrachte hier keine Verbesserung der Übereinstimmung der Schätzungen. In Folgeuntersuchungen könnte daher geprüft werden, wie sich eine homogene figurale Skala gestalten ließe (z. B. durch die Beschränkung auf einen der beiden Aufgabentypen, jedoch mit mehr einzelnen Aufgaben dieses Aufgabentyps).

Für die auf 30 Aufgaben reduzierte verbal-numerische Skala kann somit eine hinreichende Passung auf das 1PL-Modell angenommen werden. Es kann daher die Summe aller richtig gelösten Aufgaben als Fähigkeitsschätzung für die einzelnen Pbn verwendet werden. Bei der figuralen Skala hingegen ist die Voraussetzung der Aufgabenhomogenität durch die Verwendung zweier unterschiedlicher Aufgabentypen

pen verletzt. Die restlichen Modelltests bescheinigten der Skala jedoch keine bedeutende Abweichung vom 1PL-Modell. Daher wurde im Folgenden auch für die figurale Skala der Summenscores für jeden Pb berechnet.

Anschließend wurden für die beiden reduzierten, aber (weitgehend) Rasch-Modell konformen Skalen die Reliabilität und externe Validität nach der Klassischen Testtheorie bestimmt. Es zeigte sich, dass die Skalenverkürzung nur zu einer geringfügigen Verminderung der inneren Konsistenz führte. Die Korrelationen der Skalen mit den Schulnoten und den Intelligenztests liegen in mittlerer Höhe und damit in einem Bereich, der für die Zusammenhänge von klassischen Intelligenztests mit anderen Tests und Schulleistungen üblich ist.

Es konnte zudem auch für alle sechs einzelnen Aufgabentypen gezeigt werden, dass sie sich für den Einsatz bei Kindern und bei unterschiedlichen Fähigkeitsniveaus eignen. Insbesondere für den Aufgabentyp „Backward Digit Span“ war aufgrund verschiedener Studien fraglich, ob er sich für den Einsatz bei intellektuell sehr fähigen Pbn bewähren würde (vgl. Abschnitt 4.1.2). In der vorliegenden Arbeit war die „Backward Digit Span“ jedoch durchaus dazu in der Lage, auch bei hoch intelligenten Pbn die Leistungsunterschiede in Intelligenztests in gewissem Maße zu erklären. Die IRT-Analysen ergaben zudem, dass die schwierigeren BDS-Aufgaben gerade im höheren Fähigkeitsbereich besonders trennscharf sind. Lediglich für die Erklärung von schulischen Leistungen scheinen BDS-Aufgaben nicht besonders gut geeignet. Der Aufgabentyp „Reading Span“ erwies sich insgesamt als der Aufgabentyp, der am höchsten mit den verschiedenen Intelligenzmaßen korrelierte und die Schulleistungen am besten erklären konnte.

Die Analysen zeigen somit, dass sich die hier verwendeten Arbeitsgedächtnisaufgaben prinzipiell für die Konstruktion eines Tests zur Messung kognitiver Fähigkeiten – insbesondere im höheren Fähigkeitsbereich – eignen. Trotz der simplen Struktur der Aufgabentypen erweisen sich alle sechs Aufgabengruppen auch für die Stichprobe der hoch Intelligenten als hinreichend schwierig. In der Aufgabengruppe „Reading Span“ gelang es keinem Pb, alle elf Aufgaben richtig zu lösen. Auch in der Aufgabengruppe „Spatial Working Memory“ fand kein Pb für alle 13 Aufgaben die richtige Lösung. Für einen Großteil der Aufgaben konnte gezeigt werden, dass sie Rasch-Modell konform sind. Die Eliminierung von nicht-modellkonformen Aufgaben führte nicht zu einer zu starken Verkürzung der Skalen, es konnten genügend schwierige Aufgaben beibehalten werden.

Im Rahmen dieser Arbeit konnte die Retest-Reliabilität nicht bestimmt werden. Als Voraussetzung für einen Einsatz der Skalen in der psychologischen Diagnostik wäre als nächster Schritt somit die Stabilität der Messwerte zu prüfen. Weitere Belege für die externe Validität wären wünschenswert, insbesondere sollte die prognostische Validität der Skalen untersucht werden. Anschließend wäre eine Normierung der Skalen an durchschnittlich und überdurchschnittlich intelligenten Kindern erforderlich.

## 8. Zusammenfassung und Diskussion

Diese Arbeit verfolgte zwei Ziele: Das erste Ziel bestand darin, die Grundlagen für ein computergestütztes Testverfahren für die Diagnostik von hoher und sehr hoher intellektueller Begabung bei Kindern im Alter zwischen 8 und 13 Jahren zu entwickeln. Spezielle Testverfahren, die sich auch für die Messung hoher und sehr hoher intellektueller Fähigkeiten bei Kindern bis zum Alter von 12-13 Jahren eignen, fehlen bisher. Als Aufgabenmaterial für diesen Test wurden klassische Aufgaben zur Erfassung der Arbeitsgedächtniskapazität erprobt. Die Grundlage dieses Ansatzes bildet die zunehmend durch empirische Erkenntnisse gestützte Ansicht, dass das Arbeitsgedächtnis einen zentralen Stellenwert in der menschlichen Informationsverarbeitung einnimmt und die Basis jeglicher höherer kognitiver Leistungen darstellt. Der Vorteil in der Verwendung des Konzepts des Arbeitsgedächtnisses für die Diagnostik von intellektuellen Fähigkeiten gegenüber anderen Konzepten der Intelligenzforschung wie Verarbeitungskapazität oder Reasoning besteht vor allem darin, dass das Arbeitsgedächtniskonzept nicht vorwiegend auf den Ergebnissen von Faktorenanalysen beruht, sondern dass es auf einer theoretischen Grundlage und auf dem Hintergrund umfangreicher experimenteller Forschung entwickelt wurde (Kyllonen, 1996).

Dennoch sind bisher viele Aspekte des Konstrukts „Arbeitsgedächtnis“ noch nicht hinreichend geklärt. So liegen bisher sehr widersprüchliche Befunde zur Faktorstruktur des Arbeitsgedächtnisses vor. Verschiedene Studien belegen einen generellen Faktor, der den Leistungen in unterschiedlichsten Arbeitsgedächtnisaufgaben mit verbalen, numerischen und figuralen Inhalten zugrunde liegt (z. B. Ackerman et al., 2002). Andere Studien stützen eher die Hypothese, dass sich im Arbeitsgedächtnis verschiedene domänenspezifische Komponenten differenzieren lassen (z. B. Miyake & Shah, 1996).

Auch die Einordnung des Arbeitsgedächtnisses in das Netzwerk der verschiedenen kognitiven Konstrukte ist in der differentiellen Arbeitsgedächtnisforschung bisher umstritten. Insbesondere die Enge des Zusammenhangs von Arbeitsgedächtnis und fluider Intelligenz ist Gegenstand vieler Untersuchungen. Die gefundenen Zusammenhänge reichen hier von bereits hohen Korrelationen im Bereich um  $r = .70$  (z. B. de Jong & Das-Smaal, 1995; Ackerman et al., 2002) bis zu fast perfekten Übereinstimmungen (Kyllonen, 1996). Es wurde diskutiert, dass sich die unterschiedlichen Befunde zum Teil über Unterschiede in den untersuchten Stichproben einerseits und im verwendeten Aufgabenmaterial andererseits erklären lassen.

Zwar wurden entsprechende Analysen auch bereits mit Kindern durchgeführt, jedoch wurde in der Regel entweder nur eine Altersgruppe oder kleine Gruppen verschiedener Jahrgänge untersucht (z. B. de Jong & Das-Smaal, 1995; Fry & Hale, 2000). Daher liegen bisher kaum aussagekräftige Befunde zur Faktorstruktur im Kindesalter

und insbesondere nicht zu möglichen Veränderungen der Struktur im Verlauf der kognitiven Entwicklung vor. Hoch begabte Kinder wurden bisher mit Arbeitsgedächtnisaufgaben nicht untersucht. Aufgrund der hohen Zusammenhänge zwischen Intelligenz- und Arbeitsgedächtnismaßen war zu erwarten, dass hoch begabte Kinder deutlich bessere Leistungen in Arbeitsgedächtnisaufgaben zeigen als durchschnittlich begabte Kinder. Es stellte sich hier jedoch die Frage, ob die klassischen und üblicherweise verwendeten Arbeitsgedächtnisaufgaben auch bei sehr fähigen Pbn dazu in der Lage sind, das Konstrukt „Arbeitsgedächtnis“ angemessen abzubilden und höhere kognitive Leistungen vorherzusagen, wie sie sich in klassischen Intelligenztests oder den schulischen Leistungen zeigen.

Das zweite Ziel dieser Arbeit bestand daher darin, die Struktur des Arbeitgedächtnisses und seine Zusammenhänge mit anderen Konstrukten der Intelligenzforschung wie fluide und kristalline Intelligenz, Speed und Kurzzeitgedächtnis bei Kindern verschiedener intellektueller Fähigkeiten näher zu explorieren. Insgesamt wurden somit zwei Fragenkomplexe untersucht: Zum einen sollten die Faktorstruktur des Arbeitsgedächtnisses für Kinder verschiedener Altersgruppen und unterschiedlicher intellektueller Begabung sowie die Zusammenhänge des Arbeitsgedächtnisses mit fluider und kristalliner Intelligenz, Speed und Merkfähigkeit bei Kindern untersucht werden. Zum anderen sollte anhand einer Batterie verschiedener klassischer Arbeitsgedächtnisaufgaben erprobt werden, ob sich diese für die Erfassung überdurchschnittlich hoher intellektueller Fähigkeiten von Kindern eignen.

Da bereits vielfältige Arbeitsgedächtnisaufgaben in der Literatur beschrieben sind, war es nicht erforderlich, völlig neuartige Aufgabentypen zu entwickeln. Stattdessen konnte auf die Befunde von Oberauer et al. (2000) zurückgegriffen werden, die eine umfassende Batterie verschiedener Arbeitsgedächtnisaufgaben untersucht und nach ihren Anforderungen klassifiziert haben. Aus dieser Batterie wurden insgesamt sechs verschiedene Aufgabentypen ausgewählt, die die bisher am besten belegte Arbeitsgedächtnisfunktion der „simultanen Speicherung und Verarbeitung“ erfassen und die über eine gute psychometrische Qualität verfügen. Es handelt sich um zwei Aufgabentypen mit verbalem Material („Reading Span“ und „Verbal Span“), zwei Aufgabentypen mit numerischem Material („Computation Span“ und „Backward Digit Span“) sowie zwei Aufgabentypen mit figuralem Material („Memory Updating Figural“ und „Spatial Working Memory“). Alle diese Aufgabentypen können computergestützt dargeboten werden und eignen sich für Gruppentestungen. Die Verwendung von Aufgaben mit verschiedenem Inhaltmaterial begründet sich zum einen auf dem stabilen Befund aus der Intelligenzforschung, nach dem sich kognitive Fähigkeiten in die drei Bereiche „verbale Fähigkeiten“, „numerische Fähigkeiten“ und „räumliche Fähigkeiten“ gliedern lassen. Zum anderen sollte mit der Testbatterie die Möglichkeit bestehen, die Frage der Domänengeneralität vs. -spezifität des Arbeitsgedächtnisses, die in der Forschung kontrovers diskutiert wird, empirisch prüfen zu können. In vier Vorstudien waren die Aufgabeninhalte zuvor an verschiedenen Stichproben, die sich zum einen aus hoch begabten Kindern und zum anderen aus nicht nach Begabung ausgelesenen Kindern verschiedener Altersgruppen zusammensetzten, entwickelt und erprobt worden.

Die so entwickelte Testbatterie, bestehend aus insgesamt 70 Aufgaben, wurde schließlich bei 374 Kindern zwischen 8;6 und 13;11 Jahren eingesetzt (der größte Teil der Pbn war jedoch zwischen neun und zwölf Jahren alt). Zusätzlich wurden mit den Pbn verschiedene kognitive Aufgaben und Testverfahren durchgeführt, um die fluide und kristalline Intelligenz, die Informationsverarbeitungsgeschwindigkeit und die Merkfähigkeit zu erfassen. Von den Pbn besuchten 202 Kinder spezielle Einrichtungen für hoch Begabte, waren Mitglied in einem Hochbegabtenverein oder waren von ihrer Schule aufgrund ihrer hohen Leistungen für die Teilnahme an einer Schülerakademie vorgeschlagen worden („Hochbegabtenstichprobe“). Die restlichen Pbn waren Schülerinnen und Schüler der vierten Klasse einer regulären Grundschule oder der fünften oder sechsten Klasse einer Orientierungsstufe (Teilstichprobe „Unausgelesene“,  $N = 172$ ). Es wurde angestrebt, dass die Gruppe der Unausgelesenen – insbesondere im Hinblick auf die kognitiven Fähigkeiten – möglichst repräsentativ für Kinder dieser Altersgruppe ist.

In der Hochbegabtenstichprobe zeigten die Ergebnisse der Intelligenzdiagnostik, dass 37% dieser Pbn lediglich über einen IQ von  $< 120$  (CFT 20) verfügen. Es erschien daher nicht sinnvoll, die Unterschiede zwischen allen Pbn aus Hochbegabtengruppen und solchen Pbn aus regulären Schulen zu analysieren, da dies tatsächlich vorhandene Unterschiede in fähigkeitsbezogenen Maßen wie den Arbeitsgedächtnismaßen unterschätzen würde. Stattdessen wurde aus der Gesamtstichprobe nach dem Befund im CFT 20 in einer „Hochintelligentenstichprobe“ alle die Pbn zusammengefasst, die einen IQ von mindestens 120 aufweisen und somit zu den etwa 10% Intelligentesten der Population gehören ( $N = 170$ ). Alle Pbn mit einem IQ unter 120 wurden in der Substichprobe der „Nicht hoch Intelligenten“ ( $N = 204$ ) zusammengefasst.

Die Analyse der Daten gliederte sich entsprechend den beiden Zielen dieser Arbeit in zwei Teile. Zunächst wurden auf der Basis der sechs Aufgabengruppen Untersuchungen zum Konstrukt des Arbeitsgedächtnisses vorgenommen. Zur Berechnung der Scores für jede Aufgabengruppe wurde – abweichend von dem in der Literatur vorherrschenden Scoring auf der Ebene der Sets oder einem nach Gedächtnisbelastung gewichteten Scoring – ein einheitsgewichtetes Scoring auf der Ebene der Aufgaben verwendet. Die Analyse der Korrelationen der Aufgaben untereinander sowie die Ergebnisse von konfirmatorischen Faktorenanalysen führten zu dem Schluss, dass sich das Arbeitsgedächtnis am besten über ein obliques Zwei-Faktorenmodell beschreiben lässt. Es ließen sich ein verbal-numerischer und ein figuraler Faktor belegen, die miteinander zu  $r = .73$  korrelieren. Ein domänengenerelles Ein-Faktor-Modell ergab einen signifikant schlechteren Fit. Ein Drei-Faktoren-Modell, das zusätzlich zwischen verbalen und numerischen Fähigkeiten differenziert, wurde aufgrund der sehr hohen Korrelation zwischen dem verbalen und dem numerischen Faktor ( $r = .95$ ) verworfen, da es gegenüber dem Zwei-Faktoren-Modell keinen bedeutsamen Informationsgewinn erbringt.

Die Zwei-Faktoren-Struktur erwies sich im Folgenden als invariant für die Gruppe der jüngeren und der älteren Pbn der Stichprobe sowie für Jungen und Mädchen. Der Nachweis der Invarianz der Faktorstruktur für die beiden Intelligenzgruppen gelang

hingegen nicht ausreichend. Lediglich für die Gruppe der hoch Intelligenten ließ sich ein guter Modell-Fit feststellen, nicht jedoch für weniger intelligenten Pbn.

Die hoch intelligenten Pbn schnitten in allen Arbeitsgedächtnisaufgaben im Mittel signifikant besser ab als die restlichen Pbn (Effektstärken zwischen  $d = 0.71$  und  $d = 1.02$ ), ebenso die älteren Pbn im Vergleich mit den jüngeren Pbn (Effektstärken zwischen  $d = 0.27$  und  $d = 0.53$ ). Mädchen und Jungen unterschieden sich in ihrer mittleren Leistung generell nicht, lediglich in der unausgelesenen Stichprobe schnitten die Mädchen in der Aufgabengruppe „Verbal Span“ etwas besser ab als die Jungen (Effektstärke von  $d = 0.43$ ).

Vor der Untersuchung der Zusammenhänge des Arbeitsgedächtnisses mit den anderen kognitiven Konstrukten in einem gemeinsamen Modell korrelierender Faktoren war es zunächst erforderlich, die jeweilige Binnenstruktur der Konstrukte „Intelligenz“ und „Speed“ zu explorieren. Die verwendeten Intelligenzmaße ergaben erwartungsgemäß ein obliquus Zwei-Faktor-Modell, in dem der eine Faktor als fluide Intelligenz (gebildet durch die vier Subtests des CFT 20), der andere Faktor als kristalline Intelligenz (gebildet durch den Wortschatz- und den Zahlenfolgentest des CFT 20) interpretiert werden kann.

Es wurden verschiedene Speed-Maße eingesetzt, von denen angenommen wurde, dass sie sich in ihren Anforderungen an eine kognitive Verarbeitung der Informationen unterscheiden. Von den eingesetzten Speed-Tests, die auch in der Intelligenzdiagnostik Verwendung finden (ZVT sowie die Untertests „Zahlen-Symbol-Test“ und „Symbolsuche“ aus dem HAWIK III), wurde angenommen, dass sie eher komplexe Anforderungen an die Verarbeitung stellen. Weiterhin wurde eine *Elementary Cognitive Task* eingesetzt, es handelte sich dabei um eine Paper-Pencil-Version von Aufgaben nach dem Hick-Paradigma. Von den Hick-Aufgaben wurde angenommen, dass sie im Vergleich zu den anderen Speed-Tests eher geringere Anforderungen an die kognitive Verarbeitung stellen. Außerdem wurde angenommen, dass sich auch die einzelnen Hick-Aufgaben hinsichtlich ihrer Anforderungen an eine kognitive Verarbeitung der Informationen unterscheiden. Insgesamt konnte auch hier ein obliquus Zwei-Faktoren-Modell die Speed-Daten am besten erklären. Die Speed-Aufgaben aus der Intelligenzdiagnostik laden dabei auf dem einen, die Hick-Aufgaben auf dem anderen Faktor. Ein gemeinsames Ein-Faktor-Modell erbrachte einen signifikant schlechteren Fit.

Weiterhin wurde die Frage untersucht, ob die Arbeitsgedächtnisaufgaben stärker mit solchen Speed-Aufgaben korrelieren, die eine komplexere Informationsverarbeitung erfordern. Da für die einzelnen Hick-Aufgaben die Menge an zu verarbeitender Information pro Item genau spezifiziert ist, wurde diese Fragestellung durch Korrelationen zwischen den Arbeitsgedächtnis-Scores und den einzelnen Hick-Aufgaben untersucht. Es zeigte sich, dass die Aufgabe Hick-0, in der keine Wahlreaktion erfolgt, sondern nur die Zeichengeschwindigkeit des Pb erfasst wird, nicht oder nur sehr gering mit den Arbeitsgedächtnisaufgaben korrelierte. Für die komplexeren Hick-Aufgaben, die die Verarbeitung von ein bis drei Bit erfordern, konnten hingegen signifikante Korrelationen in mittlerer Höhe festgestellt werden.

Schließlich wurden die Korrelationen des Arbeitsgedächtnisses mit den anderen kognitiven Konstrukten in einem Gesamtmodell latenter Faktoren geschätzt. Aufgrund des Forschungsstands wurde erwartet, dass die Zusammenhänge der Arbeitsgedächtnisfaktoren mit der fluiden Intelligenz am höchsten ausfallen und mindestens eine Höhe von ca.  $r = .70$  erreichen. Die Zusammenhänge mit der kristallinen Intelligenz sollten etwas geringer ausfallen. Ferner wurde vermutet, dass der Faktor „Speed-Arbeitsgeschwindigkeit“ (bestimmt durch die Maße „Zahlen-Symbol-Test“, „Symbolsuche“ und ZVT) höher mit dem Arbeitsgedächtnis korreliert als der Faktor „Speed-Hick“ (bestimmt durch die vier Hick-Aufgaben).

Es ergaben sich positive Zusammenhänge zwischen allen Konstrukten. Die Korrelationen zwischen den Arbeitsgedächtnisfaktoren und der fluiden Intelligenz fielen mit  $r = .79$  und  $r = .92$  erwartungsgemäß hoch aus. Jedoch fanden sich ebenfalls hohe Zusammenhänge mit der kristallinen Intelligenz ( $r = .85$  und  $r = .86$ ). Es zeigte sich, dass der figurale Arbeitsgedächtnisfaktor stärker mit der fluiden Intelligenz korrespondiert und der verbal-numerische Arbeitsgedächtnisfaktor stärker mit der kristallinen Intelligenz. Da die kristalline Intelligenz über einen verbalen und einen numerischen Test (Zusatzmodule des CFT 20) und die fluide Intelligenz über einen figuralen Test gemessen wurde, kann dieser Befund auch im Sinne einer Domänenspezifität des Arbeitsgedächtnisses interpretiert werden: Das verbal-numerische Arbeitsgedächtnis korreliert stärker mit der verbal-numerischen Intelligenz und das figurale Arbeitsgedächtnis stärker mit der figuralen Intelligenz. Weiterhin zeigte sich im Gesamtmodell, dass die beiden Speed-Faktoren in unterschiedlichem Maße mit den Arbeitsgedächtnisfaktoren korrelieren: Die Zusammenhänge des Arbeitsgedächtnisses mit „Speed-Arbeitsgeschwindigkeit“ fielen höher aus als mit „Speed-Hick“.

Zusätzlich wurde die Fragestellung untersucht, ob alle sechs verwendeten Arbeitsgedächtnisaufgaben sowohl für unausgelesene als auch für hoch intelligente Pbn einen positiven Zusammenhang mit Intelligenzmaßen aufweisen. Diese Analysen wurden durch Befunde motiviert, dass insbesondere Rückwärtsspannen-Aufgaben wie die „Backward Digit Span“ bei sehr fähigen Pbn nicht als Indikatoren des Arbeitsgedächtnisses dienen, bzw. keine Vorhersagen von allgemeiner Intelligenz erlauben, da die Verarbeitungsanforderung dieser Aufgaben zu gering ist (Schofield & Ashman, 1986; Engle, Tuholski et al., 1999). Es zeigte sich, dass sowohl in der Gesamtstichprobe als auch in der Gruppe der hoch Intelligenten alle Arbeitsgedächtnisaufgaben signifikant mit dem CFT 20 und seinen Zusatzmodulen zur Messung der kristallinen Intelligenz korrelieren. Lediglich die figurale Aufgabe „Spatial Working Memory“ korrelierte in der Gruppe der hoch Intelligenten nicht signifikant mit dem Wortschatztest.

Im Weiteren wurde die Vorhersagegüte der Arbeitsgedächtnisaufgaben für die schulischen Leistungen in den Kernfächern untersucht. Sowohl in der Primarstufe als auch in der Sekundarstufe fanden sich bedeutsame Korrelationen in mittlerer Höhe zwischen den Arbeitsgedächtnis-Scores und den Zeugnisnoten. Tendenziell fielen die Zusammenhänge höher aus, wenn das jeweilige Schulfach und die Arbeitsgedächtnisaufgabe Leistungen in denselben Domänen erfordern (z. B. verbale Arbeitsge-



dächtnisaufgaben und sprachliche Schulfächer wie Deutsch und Fremdsprachen). Dies kann als ein weiterer Beleg für domänenspezifische Aspekte im Arbeitsgedächtnis interpretiert werden. Über multiple Regressionen der Noten auf die Arbeitsgedächtnis-Scores ließen sich in der Grundschule zwischen 30 und 44% und in der Orientierungsstufe zwischen 17 und 32% der Varianz der Noten aufklären. Insgesamt erwies sich hier die „Reading Span“ als der relativ beste Prädiktor für die Schulleistungen.

Der zweite Teil der statistischen Analysen bezog sich auf die Überprüfung der psychometrischen Qualität der eingesetzten Arbeitsgedächtnisaufgaben und die Entwicklung von Testskalen. Die einzelnen Testaufgaben wurden auf der Basis der Ergebnisse einer exploratorischen Faktorenanalyse zu zwei Skalen – einer verbal-numerischen und einer figuralen Skala – zusammengestellt. Die Kennwerte der Aufgaben wurden zunächst nach den Maßgaben der klassischen Testtheorie bestimmt. Die Schwierigkeiten der Aufgaben streuen breit, da jedoch insbesondere eine Messung im oberen Fähigkeitsbereich angestrebt wurde, enthalten beide Skalen auch einen größeren Teil sehr schwieriger Aufgaben. Die innere Konsistenz der verbal-numerischen Skala betrug  $\alpha = .92$ , die innere Konsistenz der figuralen Skala  $\alpha = .82$ . Anschließend wurden die Aufgabenparameter nach dem 1PL-, 2PL- und 3PL-Modell der IRT geschätzt und die Passung der Aufgaben auf das jeweilige Modell untersucht. Von den insgesamt 43 verbalen und numerischen Aufgaben zeigten 31 und von den 27 figuralen Aufgaben 22 Aufgaben eine hinreichend gute Anpassung an das 1PL-Modell.

Verschiedene Modellgeltungstests sprechen überwiegend für eine gute Passung des 1PL-Modells auf die Daten. Die Testinformationsfunktionen der beiden Skalen zeigten, dass die Skalen jeweils das größte Ausmaß an Informationen und den geringsten Standardfehler im oberen Bereich der Fähigkeitsdimension (etwa zwischen  $\sigma = 0$  und  $\sigma = 2$ ) erbringen. Während die figurale Skala eine uneingeschränkte Personenhomogenität aufweist, ist diese für die verbal-numerische Skala bei einer Bildung von Teilgruppen nach dem Score oder der Intelligenz nicht gegeben. Nach der Eliminierung der signifikant vom 1PL-Modell und im graphischen Modelltest abweichenden Aufgaben konnte für die verbleibenden 30 Aufgaben eine hinreichende Personenhomogenität festgestellt werden.

Die Aufgabenhomogenität konnte für die verbal-numerische Skala über die Vergleiche verschiedener Aufgabensubgruppen bestätigt werden. Die Aufgaben der figuralen Skala hingegen erwiesen sich als nicht homogen. Eine Verletzung der Aufgabenhomogenität stellte hier die Zusammenfassung der beiden Aufgabentypen „Spatial Working Memory“ und „Memory Updating Figural“ zu einer Skala dar. Auch durch die Eliminierung der fünf nicht modellkonformen Aufgaben konnte hier keine Aufgabenhomogenität erreicht werden. Offenbar erfassen die beiden figuralen Aufgabentypen etwas unterschiedliche Fähigkeiten. Die Unterschiede in den Anforderungen beider figuraler Aufgaben korrespondieren mit dem neueren Verständnis des klassischen Baddeley-Modells, dass der visuell-räumliche Skizzenblock aus zwei Komponenten besteht: einem „inner eye“, zuständig für die Erkennung statischer

Muster, und einem „inner scribe“, zuständig für die räumliche Speicherung von Bewegungsfolgen (Logie & Pearson, 1997). Da jedoch nicht genügend Aufgaben vorhanden waren, um ausreichend lange Skalen aus den jeweils einzelnen Aufgabentypen zu erstellen, und die figurale Skala außerdem verschiedene günstige Eigenschaften aufwies, wurde hier dennoch ein Skalenscore über beide Aufgabentypen hinweg berechnet (aus den 22 figuralen Aufgaben, die weder nach dem  $\chi^2$ -Test noch nach dem Q-Index bedeutsam vom IPL-Modell abwichen).

Die auf die modellkonformen Aufgaben reduzierten Skalen à 30 bzw. 22 Aufgaben sind ausreichend reliabel (verbal-numerische Skala:  $\alpha = .88$ ; figurale Skala:  $\alpha = .80$ ) und extern valide (Korrelationen der Skalen mit Intelligenztests zwischen  $r = .56$  und  $r = .61$ ; Korrelationen mit dem Notenschnitt zwischen  $r = -.37$  und  $-.61$ ). Die interne Validität der Skalen kann durch die gute Passung des Zwei-Faktoren-Modells des Arbeitsgedächtnisses als gesichert angesehen werden.

Insgesamt kann der Schluss gezogen werden, dass sich die verwendeten Arbeitsgedächtnisaufgaben zur Diagnostik bei Kindern zwischen 8 und 13 Jahren prinzipiell eignen. Beide Skalen weisen eine ausreichende psychometrische Güte auf und sind (die figurale Skala mit Einschränkungen) Rasch-Modell konform. Nach der Überprüfung weiterer Kennwerte (insbesondere der Retest-Reliabilität) und einer Normierung könnten die Skalen für die Individualdiagnostik verwendet werden. Abschließend werden nun noch einmal drei übergreifende zentrale Fragestellungen dieser Arbeit zusammenfassend diskutiert.

#### *Domänengeneralität vs. Domänenspezifität des Arbeitsgedächtnisses?*

Die Ergebnisse der vorliegenden Arbeit zeigen, dass für Kinder der untersuchten Altersgruppe ein domänengenerelles Modell des Arbeitsgedächtnisses nicht angenommen werden kann. Die Frage nach der Domänengeneralität vs. Domänenspezifität wurde in der vorliegenden Arbeit auf verschiedenen Ebenen untersucht: auf der Ebene der bivariaten Korrelationen der einzelnen Aufgabengruppen, auf der Ebene latenter Faktoren sowie über die Zusammenhänge der Aufgaben bzw. Skalen mit verschiedenen anderen kognitiven Maßen derselben und unterschiedlicher Domänen.

Die konfirmatorischen Faktorenanalysen sprechen gegen ein domänengenerelles Ein-Faktor-Modell, und die Korrelationen mit verschiedenen anderen Maßen zeigen, dass die kognitiven Leistungen innerhalb derselben Domäne tendenziell stärker miteinander korrelieren als über Domänen hinweg. Insgesamt lässt sich aus diesen Befunden schließen, dass für Kinder zwischen 8 und 13 Jahren eine teilweise Domänenspezifität des Arbeitsgedächtnisses anzunehmen ist. Verbal-numerische Arbeitsgedächtnisleistungen unterscheiden sich relativ deutlich von figuralen Arbeitsgedächtnisleistungen. In geringerem Maße zeigte sich auch eine Differenzierung zwischen verbalen und numerischen Leistungen (siehe die Analyse der Zusammenhänge mit den Schulnoten), diese fällt jedoch relativ gering aus, wie die hohe Korrelation des verbalen und des numerischen Faktors im Drei-Faktoren-Modell zeigt.

Die hier gefundene Domänenspezifität lässt sich jedoch nicht auf die in Abschnitt 3.3.4 diskutierten methodischen Ursachen zurückführen: In der vorliegenden Studie wurde eine ausgewogene Aufgabenauswahl verwendet, und es wurde eine Stichprobe mit einer breiten Fähigkeitsstreuung untersucht. Den Argumenten verschiedener Autoren (z. B. Kane et al., 2004) zufolge sollten diese Voraussetzungen eher zu einer Bestätigung der Generalitätshypothese führen. Dass diese hier – im Gegensatz zu anderen Studien (z. B. Ackerman et al., 2002) – nicht bestätigt werden kann, könnte daher auf das Alter der Pbn zurückzuführen sein. Möglicherweise lassen sich in diesem Alter verschiedene Komponenten des Arbeitsgedächtnisses noch stärker differenzieren als im Erwachsenenalter. Auch Kane et al. (2004) fanden eine gute Passung für ein entsprechendes Zwei-Faktoren-Modell für Erwachsene, präferierten jedoch aufgrund der hohen Faktorkorrelation von  $r = .93$  bzw.  $r = .84$  das ebenfalls gut passende domänengenerelle Modell. Die Unterschiede zwischen Kindern und Erwachsenen sind daher möglicherweise geringer als es die Festlegung auf eine bestimmte Anzahl Faktoren erkennen lässt. Zur genaueren Untersuchung dieser Annahme wäre jedoch die Testung von mehreren Altersgruppen mit demselben Aufgabenmaterial erforderlich.

### *Arbeitsgedächtnis und Intelligenz*

Es konnte für eine fähigkeitsheterogene Gruppe von Kindern zwischen 8 und 13 Jahren gezeigt werden, dass die Arbeitsgedächtniskapazität hoch mit der fluiden und der kristallinen Intelligenz (gemessen über klassische Intelligenztests) korreliert. Die gefundenen Zusammenhänge liegen in einer Größenordnung, die der von methodisch vergleichbaren Studien mit erwachsenen Pbn entspricht. Der geringere Zusammenhang zwischen Intelligenz und Arbeitsgedächtnis in der methodisch vergleichbaren Studie von de Jong und Das-Smaal (1995) mit Kindern ist daher vermutlich nicht auf Entwicklungsaspekte, sondern auf die dort verwendete heterogenere Aufgabenzusammenstellung zurückzuführen. Die Einschätzung verschiedener Autoren, die Arbeitsgedächtniskapazität entspreche weitgehend der allgemeinen oder fluiden Intelligenz, kann somit auf der Basis dieser Ergebnisse auch für Kinder getroffen werden.

Eine domänenspezifische Betrachtungsweise des Arbeitsgedächtnisses erwies sich auch hier als sinnvoll, um näheren Aufschluss über die Beziehung zwischen Arbeitsgedächtnis und Intelligenz zu erhalten: Der figurale Arbeitsgedächtnisfaktor korrelierte höher mit der über figurale Aufgaben gemessenen Intelligenz, der verbal-numerische Faktor jedoch höher mit der über verbale und numerische Aufgaben erfassten Intelligenz.

Die Korrelationen der beiden Testskalen mit den Intelligenztests liegen zudem in einer Höhe, wie sie aufgrund von bisherigen Validitätsstudien (z. B. Weiß, 1998) für zwei Tests, die das gleiche oder ein ähnliches Intelligenzkonstrukt erfassen, erwartet werden können. Die Grundidee dieser Arbeit, die intellektuelle Begabung von Kindern über Arbeitsgedächtnisaufgaben zu erfassen, erscheint somit auf der Basis der empirischen Ergebnisse insgesamt gerechtfertigt.

*Arbeitsgedächtnis und intellektuelle Hochbegabung*

In dieser Arbeit wurden erstmals intellektuell hoch begabte Kinder mit Arbeitsgedächtnisaufgaben untersucht. Bisher lagen keine Erkenntnisse dazu vor, inwiefern sich hoch begabte Kinder in ihrer Arbeitsgedächtniskapazität von durchschnittlich begabten Kindern unterscheiden. Die Ergebnisse der vorliegenden Arbeit zeigen, dass für hoch intelligente Kinder ein domänenspezifisches Zwei-Faktoren-Modell des Arbeitsgedächtnisses angenommen werden kann.

Erwartungsgemäß zeigten die hoch intelligenten Pbn in dieser Untersuchung im Mittel deutlich bessere Leistungen in den Arbeitsgedächtnisaufgaben als nicht hoch intelligente Pbn. Die gefunden Mittelwertsunterschiede für die einzelnen Aufgabengruppen sind mit Effektstärken im Bereich von  $d = .71$  bis  $d = 1.02$  überwiegend als groß zu bezeichnen.

Schließlich zeigte sich, dass sich die klassischen Aufgabentypen auch für die Messung der Arbeitsgedächtniskapazität im hohen Fähigkeitsbereich eignen. Die hier verwendeten Aufgaben waren ausreichend schwierig, um auch in einer Gruppe sehr intelligenter Kinder noch zu einer Streuung der Leistungen zu führen. In der Literatur diskutierte Befürchtungen, dass sich insbesondere Rückwärtsspannen-Aufgaben nicht für die Untersuchung zentral-exekutiver Prozesse bei sehr fähigen Pbn eignen könnten, werden somit nicht bestätigt.

## Literatur

- Ackerman, P. L., Beier, M. E. & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131(4), 567-589.
- Amelang, M. & Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung* (4. Aufl.). Stuttgart: Kohlhammer.
- Andersen, E. B. (1973). A goodness of fit test for the Rasch Model. *Psychometrika*, 38, 123-140.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Arbuckle, J. L. (1999). *AMOS 4.0* [Computer software]. Chicago: Smallwaters.
- Atkinson, R. C. & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, 225(2), 82-90.
- Babcock, R. L. (1994). Analysis of adult age differences on the Raven's Advanced Progressive Matrices test. *Psychology and Aging*, 9, 303-314.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. (1995). Working memory or working attention? In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness & control* (pp. 152-170). Oxford: Clarendon.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-422.
- Baddeley, A. D., Emslie, H., Kolodny, J. & Duncan, J. (1998). Random generation and the executive control of memory. *Quarterly Journal of Experimental Psychology*, 51A, 819-852.
- Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47-90). New York: Academic Press.
- Baddeley, A. D. & Logie, R. H. (1999). Working memory. The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 28-59). Cambridge: University Press.
- Bardon, G. (1999). *Relations entre modes de traitement préférentiels en mémoire de travail et dans l'apprentissage de la lecture* [Zusammenhänge zwischen den bevorzugten Verarbeitungsmodi im Arbeitsgedächtnis und beim Lesenlernen]. Unveröffentlichte Dissertation. Paris: Université René Descartes. Series V.
- Baving, L. (2002). *Intelligenzdiagnostik*. Vortrag auf dem XXVII. Kongress der Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie und Psychotherapie, 3.-6. April 2002, Berlin.
- Bayliss, D. M., Jarrold, C., Gunn, D. M. & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, 132(1), 71-92.

- Bickley, P. G., Keith, T. Z. & Wolfe, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across life span. *Intelligence, 20*, 309-328.
- Binet, A. & Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *Année Psychologique, 11*, 191-244.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Reading/Mass: Addison-Wesley.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Heidelberg: Springer.
- Brainerd, C. J. & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology, 17*, 210-247.
- Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung* (S. 13-42). Lengerich: Pabst.
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 445-455). Newbury Park, CA: Sage.
- Bull, R. & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, task switching and working memory. *Developmental Neuropsychology, 19*, 273-293.
- Butler-Por, N. (1995). Gifted children: Who is at risk for underachievement and why? In M. W. Katzko & F. J. Mönks (Eds.), *Nurturing talent: Individual needs and social ability* (pp. 252-261). Assen: Van Gorcum.
- Byrne, B. (2001). *Structural Equation Modeling with AMOS. Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Carlson, J. S. & Wiedl, K. H. (1990). The dynamic assessment of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 38-52). Hillsdale, NJ: Erlbaum.
- Carpenter, P. A., Just, M. A. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 3, 404-431.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, MA: Cambridge University Press.
- Case, R. & Globerson, T. (1974). Field independence and central computing space. *Child Development, 45*, 772-778.
- Case, R., Kurland, D. M. & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.
- Cattell, R. B. (1973). *Measuring intelligence with the Culture Fair Tests*. Champaign, IL: Institute for Personality and Ability Testing.
- Chiappe, P., Hasher, L. & Siegel, L. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28*(1), 8-17.

- Colom, R., Flores-Mendoza, C. & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences*, 34, 33-39.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, 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.
- Conway, A. R. A. & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, 4(6), 577-590.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford: University Press.
- Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Daneman, M. & Merikle, Ph. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422-433.
- Daneman, M. & Tardif, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 491-508). Hove: Erlbaum.
- Das, J. P., Kirby, J. R. & Jarman, R. F. (1979). *Simultaneous and successive cognitive processes*. New York: Academic.
- de Jong, P. F. & Das-Smaal, E. A. (1990). The Star Counting Test: An attention test for children. *Personality and Individual Differences*, 11(6), 597-604.
- 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.
- de Jonge, P. & de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences*, 21(6), 1007-1020.
- Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N. & Wilson, L. (1999). Pattern Span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, 37, 1189-1199.
- Demetriou, A., Christou, C., Spanoudis, G. & Platsidou, M. (2002). *The development of mental processing: Efficiency, working memory, and thinking*. Oxford: Blackwell Publishing.
- Detterman, D. K. & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, 13, 349-359.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93(1), 179-197.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Engle, R. W., Cantor, J. & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 972-992.

- Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 102-134). Cambridge: University Press.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E. & 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.
- Feger, B. & Prado, T. M. (1998). *Hochbegabung: Die normalste Sache der Welt*. Darmstadt: Primus Verlag.
- Feldhusen, J. F. & Jarwan, F. A. (2000) Identification of gifted and talented youth for educational programs. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 233- 252). Oxford: Pergamon.
- Fischer, G. H. & Ponocny-Seliger, E. (1998). *LPCM-Win 1.0* [computer software]. Groningen, NL: ProGamma.
- Fleischmann, U. M. (1989). *Gedächtnis und Alter*. Bern: Huber.
- Flynn, J. R. (1987). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.
- Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 147-252.
- Fry, A. & Hale, S. (2000). Relationship among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, 54, 1-34.
- Gathercole, S. E. (1995). The assessment of phonological memory skills in preschool children. *British Journal of Educational Psychology*, 65, 155-164.
- Gathercole, S. E. & Pickering, S. J. (2000a). Assessment of Working Memory in six- and seven-year-old children. *Journal of Educational Psychology*, 92(2), 377-390.
- Gathercole, S. E. & Pickering, S. J. (2000b). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, 70, 177-194.
- Gathercole, S. E., Pickering, S. J., Knight, C. & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18, 1-16.
- Gilhooly, K. J., Logie, R. H., Wetherick, N. E. & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory and Cognition*, 21, 115-124.
- Golombok, S. & Rust, J. (1992). *Manual of the Wechsler Intelligence Scale for Children – Third Edition UK*. Kent: The Psychological Corporation.
- Greenfield, P. M. (1997). You can't take it with you. Why ability assessments don't cross cultures. *American Psychologist*, 52(10), 1115-1124.



- Gross, M. U. M. (2000). Issues in the cognitive development of exceptionally and profoundly gifted individuals. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 179-192). Oxford: Pergamon.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Hacker, W., Handrick, S., Heimann, I., Oehm, D., Richter, F., Sachse, P. & Schneider M. (1999). Individuelle Unterschiede im Arbeitsgedächtnis für numerisches Material: Rechenspanne. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20(4), 290-308.
- Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Hanley, J. R., Young, A. W. & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *Quarterly Journal of Experimental Psychology*, 43A, 101-125.
- Hanses, P. & Rost, D. H. (1998). Das „Drama“ der hochbegabten Underachiever – „Gewöhnliche“ oder „außergewöhnliche“ Underachiever? *Zeitschrift für Pädagogische Psychologie*, 12(1), 53-71.
- Hany, E. A. (1987). Psychometrische Probleme bei der Identifikation Hochbegabter. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8(3), 173-191.
- Hasher, L. & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 193-225). San Diego, CA: Academic Press.
- Heinbokel, A. (1996). *Hochbegabte. Erkennen, Probleme, Lösungswege*. Baden Baden: Nomos.
- Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 93-101.
- Heller, K. A. (1992). *Hochbegabung im Kindes- und Jugendalter*. Göttingen: Hogrefe.
- Heller, K. A., Gaedike, A. K. & Weinläder, H. (1985). *Kognitiver Fähigkeitstest für 4. bis 13. Klassen (KFT 4-13+)*. Weinheim: Beltz.
- Heller, K. A. & Geisler, H. J. (1983). *Kognitiver Fähigkeitstest für 1. bis 3. Klassen (KFT 1-3)*. Göttingen: Hogrefe.
- Heller, K. A., Kratzmeier, H. & Lengfelder, A. (1998). *Matrizen Test Manual, Band 2. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices*. Weinheim: Beltz.
- Heller, K. A., Perleth, Ch. & Hany, E. A. (1994). Hochbegabung – ein lange Zeit vernachlässigtes Forschungsthema. *Einsichten – Forschung an der Ludwig-Maximilians-Universität München*, 1, 18-22.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4, 11-26.
- Hines, M. (1990). Gonadal hormones and human cognitive development. In J. Balthazart (Ed.), *Brain and behaviour in vertebrates 1: Sexual differentiation, neuroanatomical aspects, neurotransmitters and neuropeptides* (pp. 51-63). Basel: Karger.

- Hitch, G. J., Towse, J. N. & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, 130(2), 184-198.
- Holling, H. & Kanning, U. P. (1999). *Hochbegabung. Forschungsergebnisse und Fördermöglichkeiten*. Göttingen: Hogrefe.
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik*. Göttingen: Hogrefe.
- Holling, H., Vock, M. & Preckel, F. (2001). Schulische Begabtenförderung in den Ländern der Bundesrepublik Deutschland. Befunde und Empfehlungen. In Forum Bildung (Hrsg.), *Finden und Fördern von Begabungen*. Fachtagung des Forum Bildung in Berlin, 2001 (S. 45-75). Bonn: Herausgeberin.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236-248.
- Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139-153.
- Hyde, J. S. & Linn, M. C. (1988). Gender differences in verbal abilities: A meta-analysis. *Psychological Bulletin*, 104, 53-69.
- Jäger, A. O. (1982). Mehrdimensionale Klassifikation von Intelligenztestleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 145-226.
- Jäger, A. O., Holling, H., Preckel, F., Schulze, R., Vock, M., Süß, H.-M. & Beauducel, A. (2005). *Berliner Intelligenzstrukturtest für Jugendliche: Begabungs- und Hochbegabungsdiagnostik (BIS-HB)*. Göttingen: Hogrefe.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test, Form 4 (BIS-4)*. Göttingen: Hogrefe.
- Jarvis, H. L. & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, 20(3), 123-140.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Jensen, A. R. (1998). *The g-factor. The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R. & Figueroa, R. A. (1975). Forward and backward digit span interaction with race and IQ: Predictions from Jensen's theory. *Journal of Educational Psychology*, 67, 882-893.
- Johnsen, S. K. & Corn, A. L. (2001). *Screening Assessment for Gifted Elementary and Middle School Students - Second Edition (SAGES-2)*. Austin, Texas: pro-ed.
- Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology*, 87(1), 93-102.
- Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122-149.

- Kail, R. & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, 86, 199-225.
- Kane, H. D. (2000). A secular decline in Spearman's g: Evidence from the WAIS, WAIS-R and WAIS-III. *Personality and Individual Differences*, 29, 561-566.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A. & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130(2), 169-183.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W. & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189-217.
- Karnes, F. A. & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. *Journal of Educational Psychology*, 72(2), 197-199.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F. & Salthouse, T. A. (1999). Models of working memory. Eight questions and some general issues. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 412-441). Cambridge: University Press.
- Klein, K. & Fiss, W. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers*, 31(3), 429-432.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)*. Göttingen: Beltz.
- Kyllonen, P. C. (1985). *Dimensions of information processing speed* (No. AFHRL-TP-84-56). Brooks Air Force Base, TX: Air Force Systems Command.
- Kyllonen, P. C. (1993). Aptitude testing based on information processing: A test of the four-sources model. *Journal of General Psychology*, 120, 375-405.
- Kyllonen, P. C. (1994). Information processing. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 581-588). New York: Macmillan.
- Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In I. Dennis & P. Tapsfield (Eds.), *Human abilities. Their nature and measurement* (pp. 49-75). Mahwah, NJ: Lawrence Erlbaum.
- Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- Langeheine, R., Pannekoek, J. & van de Pol, F. (1995). Bootstrapping Goodness-of-Fit measures in categorical data analysis. *Sociological Methods & Research*, 24(4), 492-516.
- Leather, C. V. & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology*, 58, 88-111.
- Lehto, J. (1995). Working memory and school achievement in the ninth form. *Educational Psychology*, 15(3), 271-281.

- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479-1498.
- Linn, M. C. & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67-101). Baltimore: Johns Hopkins University Press.
- Loehlin, J. C. (1998). *Latent variable models. An introduction to factor, path, and structural analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176-196). Cambridge: Cambridge University Press.
- Logie, R. H. & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial Working Memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *9*(3), 241-257.
- Lustig, C., May, C. & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199-207.
- Mackintosh, N. J. & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence*, *31*, 519-531.
- Marland, S. P. (1972). *Education of the gifted and talented: Report to the Congress of the United States by the U.S. Commissioner of Education*. Washington: U.S. Government Printing Office.
- Marshalek, B., Lohman, D. F. & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107-127.
- Martin-Löf, P. (1973). *Statistica modeller. Anteckningar från seminarier läsåret 1969-70 utarbetade av Rolf Sundberg*. (2. Aufl.). [Statistische Modelle. Notizen aus Seminaren der Semester 1969-70, ausgearbeitet von Rolf Sundberg]. Institutet för Försäkringsmatematik och Matematisk Statistik, Universit t Stockholm.
- May, C. P., Hasher, L. & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, *27*, 759-767.
- McClelland, S. (1982). *A verbal/performance analysis of the Stanford-Binet intelligence scale and the development of high IQ preschoolers*. Unpublished dissertation, University of Washington.
- Melchers, P. & Preuß, U. (1994). *Kaufman-Assessment Battery for Children (K-ABC)*. Göttingen: Hogrefe.
- Miller, G. A., Galanter, E. & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt.
- Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General*, *130*(2), 163-168.
- Miyake, A. & Shah, P. (1999). Models of working memory. An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 1-27). Cambridge: University Press.

- Mönks, F. J. (1990). Hochbegabtenförderung als Aufgabe der Pädagogischen Psychologie. *Psychologie in Erziehung und Unterricht*, 37, 243-250.
- Norman, D. A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In Davidson, R. J., Schwartz, G. E. & Shapiro, D. (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1-18). New York: Plenum Press.
- Oberauer, K. (1993). Die Koordination kognitiver Operationen - eine Studie über die Beziehung zwischen Intelligenz und „working memory“. *Zeitschrift für Psychologie*, 201, 57-84.
- Oberauer, K. (2005). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring Intelligence* (pp. 393-407). Thousand Oaks, CA: Sage.
- Oberauer, K. & Süß, H.-M. (1996). *WMC 2000* [computer software]. Universität Mannheim.
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O. & Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences*, 29, 1017-1045.
- Oberauer, K., Süß, H.-M., Wilhelm, O. & Wittmann, W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31, 167-193.
- Oswald, W. D. (1998). Entwicklung der Intelligenz. In E. Roth (Hrsg.), *Intelligenz. Grundlagen und neuere Forschung* (S. 79-100). Stuttgart: Kohlhammer.
- Oswald, W. D. & Roth, E. (1987). *Zahlen-Verbindungs-Test (ZVT)*. Göttingen: Hogrefe.
- Park, D. C., Lautenschlager, G. & Hedden, T. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology & Aging*, 17(2), 299-320.
- Pellegrino, J. W. & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence*, 3, 187-214.
- Perleth, Ch., Schatz, T. & Mönks, F. J. (2000). Early identification of high ability. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 297- 316). Oxford: Pergamon.
- Perfetti, C. A. & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior*, 14, 33-42.
- Pickering, S. J. & Gathercole, S. E. (2001). *Working Memory Test Battery for Children* (WMTB-C). London: Psychological Corporation.
- Pickering, S. J., Gathercole, S. E., Hall, M. & Lloyd, S. (2001). Development of memory for pattern and path: Further evidence for the fractionation of visuo-spatial short-term memory. *Quarterly Journal of Experimental Psychology*, 54A, 397-420.

- Prado, T. M. & Schiebel, W. (1996). *Entwicklung und Erprobung eines Modells zur Förderung besonders begabter Schülerinnen und Schüler durch Fördermaßnahmen zur Verkürzung der individuellen Schulzeit. Schlußbericht*. Behörde für Schule, Jugend, Berufsbildung, Amt für Schule Hamburg.
- Preckel, F. (2003). *Diagnostik intellektueller Hochbegabung. Testentwicklung zur Erfassung der fluiden Intelligenz*. Göttingen: Hogrefe.
- Proctor, T. B., Black, K. N. & Feldhusen, J. F. (1986). Early admission of selected children to elementary school. A review of the research literature. *Journal of Educational Research*, 80, 70-76.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Danish Institute for Educational Research.
- Raven, J. C. (1962a). *Advanced Progressive Matrices (APM)*. London: Lewis.
- Raven, J. C. (1962b). *Coloured Progressive Matrices (CPM)*. London: Lewis.
- Raven, J. C. (1962c). *Standard Progressive Matrices (SPM)*. London: Lewis.
- Renzulli, J. S. (1979). *What makes giftedness: A reexamination of the definition of the gifted and talented*. Ventura Ca.
- Reuter-Lorenz, P. A., Jonides, J., Smith, E. E., Hartley, A., Miller, A., Marshuetz, C. & Koeppel, R. A. (2000). Age differences in frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of Cognitive Neuroscience*, 12, 174-187.
- Richardson, J. T. E. (1996). Evolving concepts in working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus & R. T. Zacks (Eds.), *Working memory and human cognition* (pp. 3-30). Oxford: Oxford University Press.
- Roberts, R. D., Pallier, G., & Stankov, L. (1999). *The Processing Speed-Accuracy Test (PAST) Battery*. Unpublished Battery. Sydney: University of Sydney.
- Roberts, R. D., Stankov, L. & Pallier, G. (1999). *Processing Speed-Accuracy Test (PAST) Battery: Underlying theory, description, and validation*. Special Symposium on Intelligence Assessment. The Fifth European Conference on Psychological Assessment, Patras, Greece, August, 25-29, 1999.
- Robinson, N. M. & Janos, P. M. (1987). The contribution of intelligence tests to the understanding of special children. In J. D. Day & J. B. Borkowski (Eds.), *Intelligence and exceptionality: New directions for the theory, assessment, and instructional practices* (pp. 21-56). Norwood, NJ: Ablex Publishing.
- Rost, D. H. (2000). *Hochbegabte und hochleistende Jugendliche*. Münster: Waxmann.
- Rost, D. H. & Hanes, P. (1993). Zur Brauchbarkeit des Zahlen-Verbindungs-Tests (ZVT) bei Kindern der 3. Jahrgangsstufe – psychometrische Überprüfung und Neustandardisierung. *Diagnostica*, 39, 80-95.
- Rost, J. (2000). Haben ordinale Rasch-Modelle variierende Trennschärfen? Eine Antwort auf die Wiener Repliken. *Psychologische Rundschau*, 51(1), 36-37.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Huber.
- Rost, J. & von Davier, M. (1994). A conditional item fit index for Rasch models. *Applied Psychological Measurement*, 18(2), 171-182.

- Ryser, G. R. & Johnsen, S. K. (1998). *Test of Mathematical Abilities for Gifted Students* (TOMAGS). Austin, Texas: pro-ed.
- Salthouse, T. A., Babcock, R. L. & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging*, 6(1), 118-127.
- Sapp, G. L., Chissom, B. & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: A replication and comparison. *Psychological Reports*, 57, 947-951.
- Schofield, N. J. & Ashman, A. F. (1986). The relationship between digit span and cognitive processes across ability groups. *Intelligence*, 10, 59-73.
- Schweizer, K. (1995). *Kognitive Korrelate der Intelligenz*. Göttingen: Hogrefe.
- Schweizer, K. & Moosbrugger, H. (2000). Das Arbeitsgedächtnis als Schnittstelle zwischen quantitativen und qualitativen Leistungen. In K. Schweizer (Hrsg.), *Intelligenz und Kognition. Die kognitiv-biologische Perspektive der Intelligenz* (S. 105-124). Landau: Verlag Empirische Pädagogik.
- 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(1), 4-27.
- Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development*, 17(1), 109-124.
- Spearman, C. (1904). *General intelligence, objectively determined and measured*. New York: Macmillan.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- Speck, N. C. (1979). Sex differences in factors of intelligence on the WISC-R in gifted children. *Dissertation Abstracts International*, 40, 199-200.
- Stapf, A. (1999). Psychologische Diagnostik bei hochbegabten Kindern. In T. Fitzner, W. Stark, H. P. Kagelmacher & T. Müller (Hrsg.), *Erkennen, anerkennen und fördern von Hochbegabten* (S. 16-34). Stuttgart: Klett.
- Stuart-Hamilton, I. (1996). *Dictionary of cognitive psychology*. London: Jessica Kingsley.
- Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung*. Lengerich: Pabst.
- Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O. & Schulze, R. (2002). Working memory capacity explains reasoning ability - and a little bit more. *Intelligence*, 30, 261-288.
- Swanson, H. L. (1992). Generality and modifiability of Working Memory among skilled and less skilled readers. *Journal of Educational Psychology*, 84(4), 473-488.
- Swanson, H. L. (1996). *Swanson Cognitive Processing Test (S-CPT)*. Austin, Texas: pro-ed.
- Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental Psychology*, 35(4), 986-1000.

- Tarver, S. G. & Hallahan, D. P. (1974). Attention deficits in children with learning disabilities: A review. *Journal of Learning Disabilities*, 17, 560-572.
- Taylor, C. W. (1978). How many types of giftedness can your program tolerate? *Journal of Creative Behavior*, 12, 39-51.
- Tewes, U. (1983). *Hamburg-Wechsler-Intelligenztest für Kinder – Revision (HAWIK-R)*. Bern: Huber.
- Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision (HAWIE-R)*. Bern: Huber.
- Tewes, U., Rossmann, P. & Schallberger, U. (2000). *Hamburg-Wechsler-Intelligenztest für Kinder III (HAWIK III)*. Göttingen: Hogrefe.
- Thorndike, H. L. & Hagen, E. (1987). *Cognitive Abilities Test*. Boston: Houghton-Mifflin.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University Press.
- Tirre, W. C. & Pena, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology*, 84(4), 462-472.
- Torgesen, J. K. & Houck, D. G. (1980). Processing deficiencies of learning-disabled children who perform poorly on the digit span test. *Journal of Educational Psychology*, 72(2), 141-160.
- Towse, J. N., Hitch, G. J. & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39, 195-217.
- Towse, J. N., Hitch, G. J. & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, 28(3), 341-348.
- Tuholski, S. W., Engle, R. W. & Baylis, G. C. (2001). Individual differences in working memory capacity and enumeration. *Memory & Cognition*, 29(3), 484-492.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory* (pp. 381-450). New York: Academic Press.
- Tulving, E. (1986). How many memory systems are there? *American Psychologist*, 40, 385-398.
- Turner, M. L. & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.
- Van Melis-Wright, M. & Stone, W. (1986). Materials review: A comparison of the K-ABC global scales and the Stanford-Binet with young gifted children. *Topics in Early Childhood Special Education*, 6(1), 88-91.
- Verguts, T. & De Boeck, P. (2002). On the correlation between working memory capacity and performance on intelligence tests. *Learning and Individual Differences*, 13, 37-55.
- Vernon, P. E. (1987). The demise of the Stanford-Binet scale. *Canadian Psychology/Psychologie Canadienne*, 28(3), 251-258.
- von Davier, M. (1997). *Methoden zur Prüfung probabilistischer Testmodelle*. Kiel: Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel.
- von Davier, M. (2001). *Winmira 2001* [computer software]. Groningen, NL: ProGamma.



- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249-294.
- Waters, G. S. & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49 A, 51-74.
- Waters, G. S. & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers*, 35(4), 550-564.
- Weiß, R. H. (1998). *Grundintelligenztest Skala 2 (CFT 20)*. Göttingen: Hogrefe.
- Whipple, G. M. (1921). *Manual of mental and physical tests. Part 1: Simpler processes; Part 2: Complex processes*. Baltimore: Warwick & York.
- Wieczerkowski, W. & Wagner, H. (1985). Diagnostik von Hochbegabung. In R. S. Jäger, R. Horn & K. Ingenkamp (Hrsg.), *Tests und Trends 4 – Jahrbuch der Pädagogischen Diagnostik* (S. 109-134). Basel: Beltz.
- Willson, V. L., Reynolds, C. R. & Gilbert, J. A. (1982). *Factor structure of the WISC-R for gifted children: A failure to replicate*. Paper presented at the Southwest Education Research Association.
- Zimowski, M. F., Muraki, E., Mislevy, R. J. & Bock, R. D. (1996). *BILOG-MG: Multiple group IRT analysis and test maintenance for binary items* [computer software]. Chicago, IL: Scientific Software International.

# Anhang

- A Fragebogen (Primarstufe, Sekundarstufe)
- B Faktorstruktur des Arbeitsgedächtnisses im Gruppenvergleich (Abbildungen)

# Fragebogen Primarstufe

Ein paar Fragen über Dich ...

CODE: \_\_\_\_\_

Alter: \_\_\_\_\_ Jahre

Geburtsdatum: \_\_\_\_\_

Klasse: \_\_\_\_\_

Ich bin...  weiblich  männlich

Schule: \_\_\_\_\_

Meine Muttersprache ist:

Deutsch oder

eine andere Sprache, nämlich: \_\_\_\_\_

Ich spreche Deutsch seit \_\_\_\_\_ Jahren

Ich bin...  Rechtshänder/in

Linkshänder/in

Beruf meines Vaters: \_\_\_\_\_

Beruf meiner Mutter: \_\_\_\_\_

Spielst Du manchmal Computerspiele?

Ja

Nein

Hast Du schon mal eine E-Mail geschrieben?

Ja

Nein

Hast Du schon mal etwas anderes auf einem  
Computer geschrieben (eine Geschichte,  
einen Brief oder etwas ganz anderes)?

Ja

Nein

Warst Du schon einmal mit Deiner Klasse im  
Computerraum?

Ja

Nein

Im letzten Zeugnis hatte ich diese **Schulnoten**:

In Deutsch hatte ich eine \_\_\_\_\_.

In Mathe hatte ich eine \_\_\_\_\_.

In Sachkunde hatte ich eine \_\_\_\_\_.

**-- Bitte erst nach Aufforderung umblättern! --**

# Fragebogen Sekundarstufe

Ein paar Fragen über Dich ...

CODE: \_\_\_\_\_

Alter: \_\_\_\_\_ Jahre      Geburtsdatum: \_\_\_\_\_      Klasse: \_\_\_\_\_

Ich bin...     weiblich     männlich      Schule: \_\_\_\_\_

Meine Muttersprache ist:

Deutsch    oder     eine andere Sprache, nämlich: \_\_\_\_\_

Ich spreche Deutsch seit \_\_\_\_\_ Jahren

Ich bin...     Rechtshänder/in     Linkshänder/in

Beruf meines Vaters: \_\_\_\_\_

Beruf meiner Mutter: \_\_\_\_\_

Spielst Du manchmal Computerspiele?       Ja       Nein

Hast Du schon mal eine E-Mail geschrieben?       Ja       Nein

Hast Du schon mal etwas anderes auf einem  
Computer geschrieben (eine Geschichte,  
einen Brief oder etwas ganz anderes)?       Ja       Nein

Warst Du schon einmal mit Deiner Klasse im  
Computerraum?       Ja       Nein

Bitte gib nun einige Deiner **Schulnoten aus Deinem letzten Schulzeugnis** an.

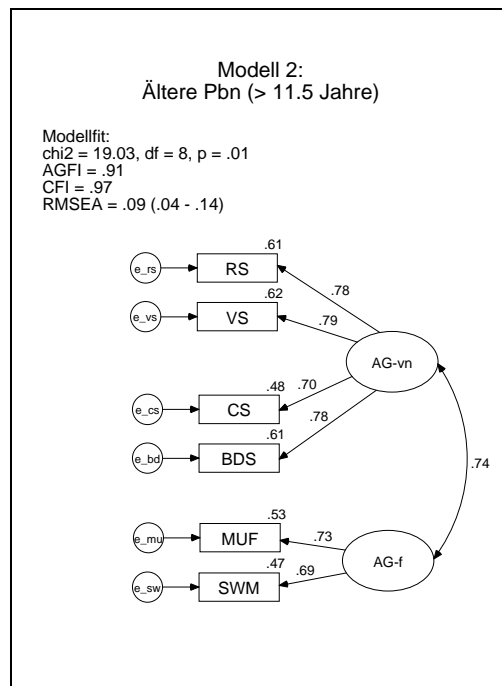
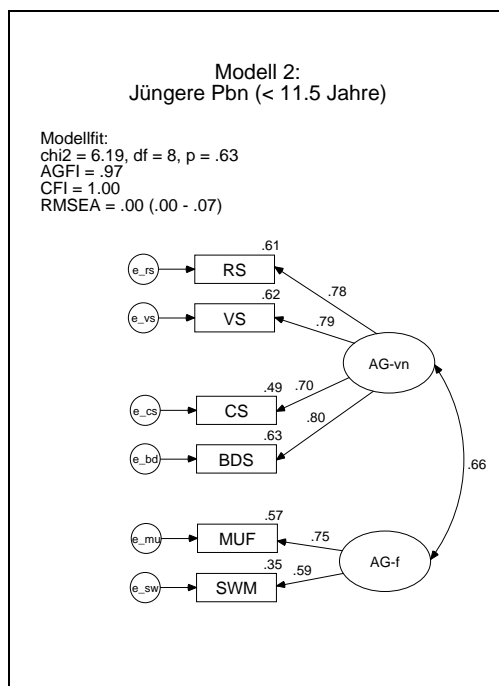
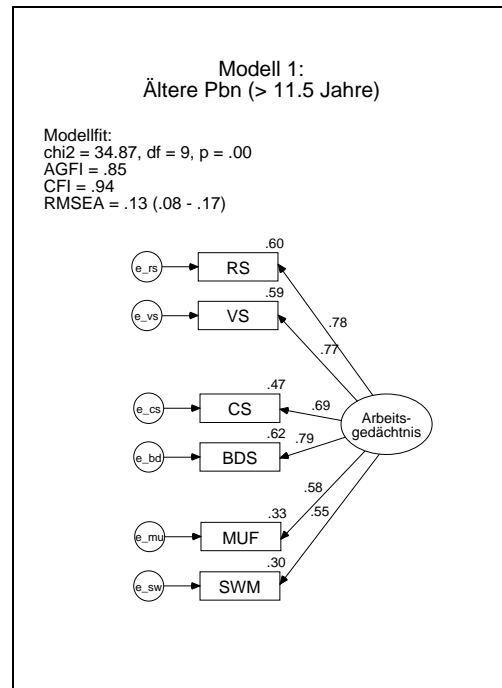
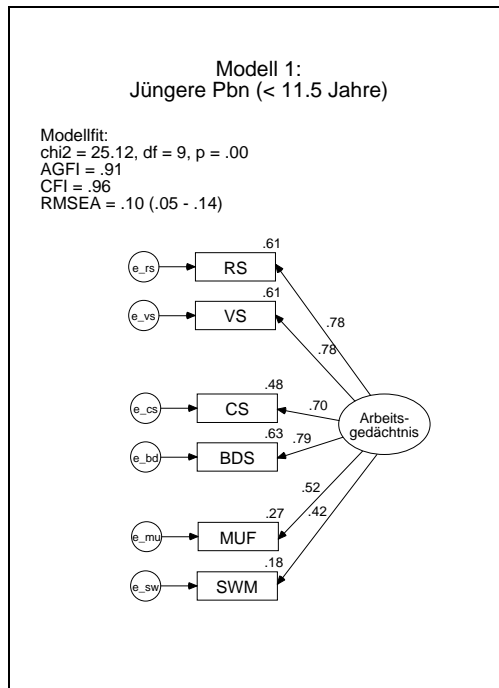
|                                |  |              |  |
|--------------------------------|--|--------------|--|
| Deutsch                        |  | Physik       |  |
| Englisch                       |  | Chemie       |  |
| Französisch                    |  | Biologie     |  |
| Latein                         |  | Informatik   |  |
| Geschichte                     |  | Musik        |  |
| Gemeinschafts- od. Sozialkunde |  | Kunst        |  |
| Religion                       |  | Sport        |  |
| Erdkunde                       |  | Sonstiges 1: |  |
| Mathematik                     |  | Sonstiges 2: |  |

-- Bitte erst nach Aufforderung umblättern! --

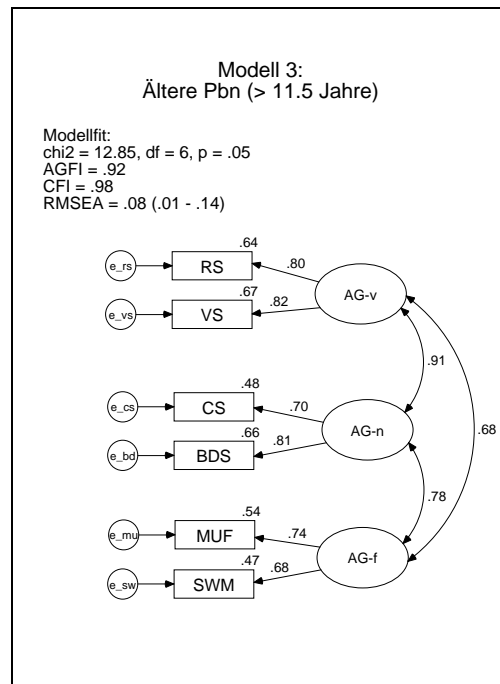
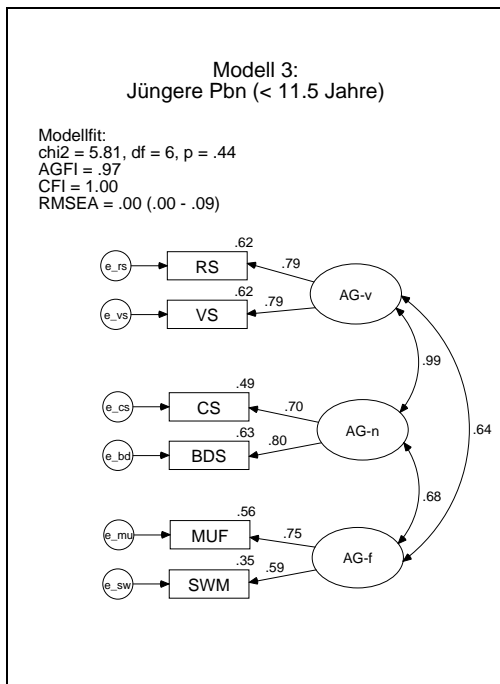
# Anhang B

## Faktorstruktur des Arbeitsgedächtnisses im Gruppenvergleich

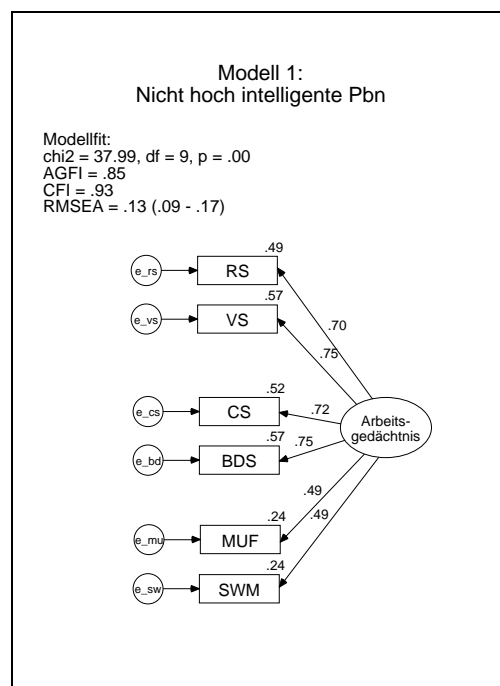
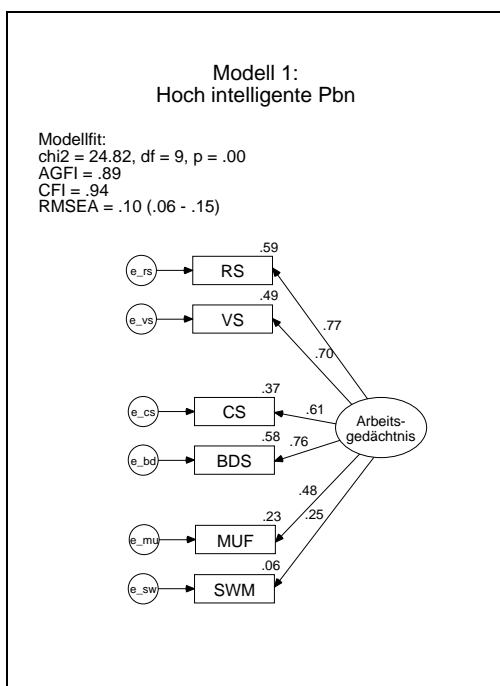
1. Prüfung der drei hypothetischen Arbeitsgedächtnismodelle für die jüngeren (< 11.5 Jahre, N = 195) und die älteren Pbn (> 11.5 Jahre, N = 178)



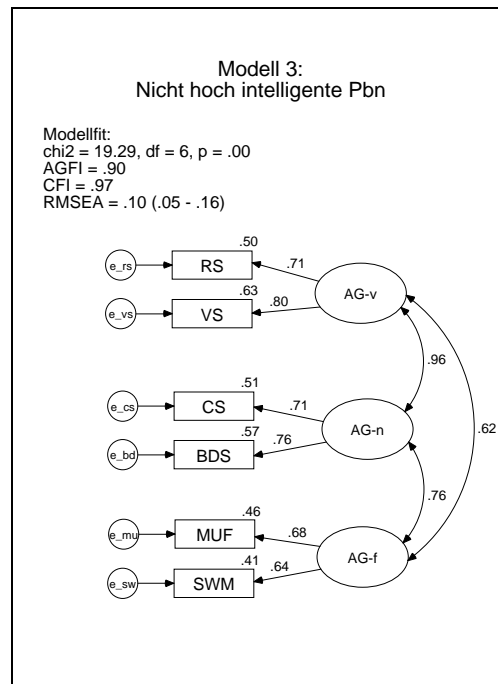
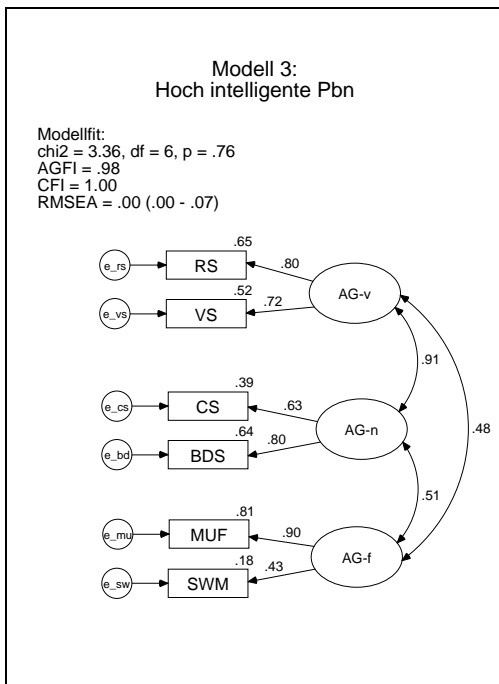
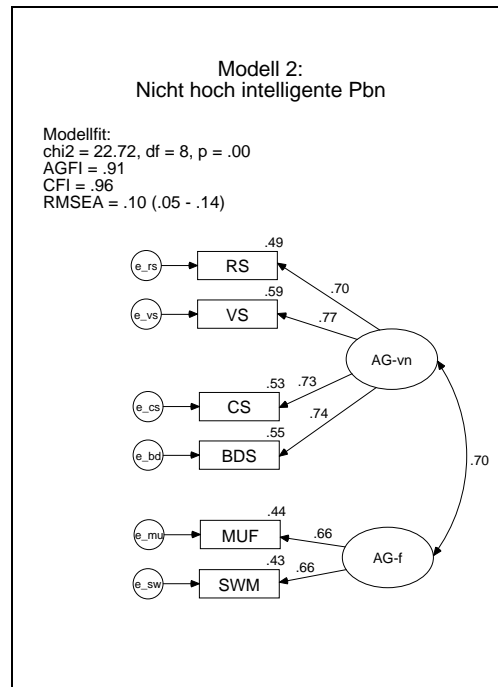
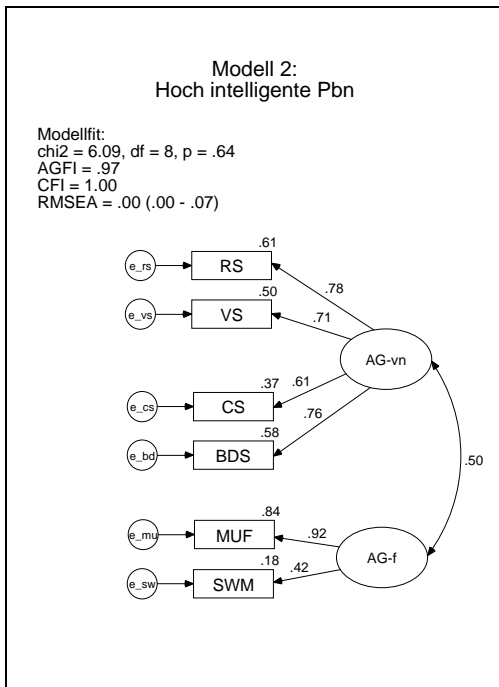
## Anhang B



2. Prüfung der drei hypothetischen Arbeitsgedächtnismodelle für die hoch intelligenten Pbn (HI-Gruppe,  $N = 170$ ) und die nicht hoch intelligenten Pbn (NHI-Gruppe,  $N = 204$ )

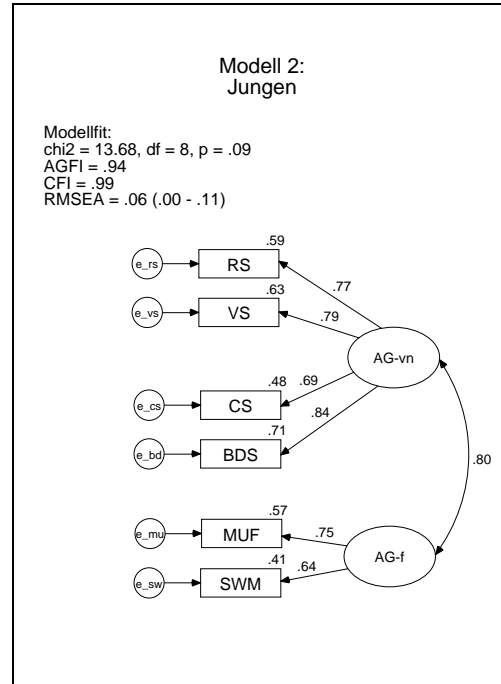
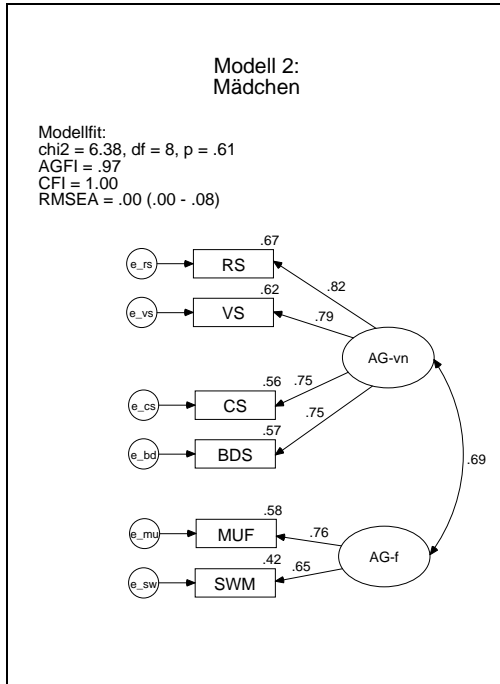


## Anhang B



## Anhang B

### 3. Prüfung von Modell 2 des Arbeitsgedächtnisses für Mädchen ( $N = 168$ ) und Jungen ( $N = 205$ )





## Danksagung

Mein Dank gebührt an erster Stelle Herrn Prof. Dr. Holling, der diese Arbeit betreut und mich in allen Phasen der Arbeit stets großzügig unterstützt, ermutigt und immer gut beraten hat.

Weiter gilt mein Dank allen Kolleginnen und Kollegen des Psychologischen Instituts IV der Universität Münster, die mir mit Rat und Tat beiseite standen. Vor allem danken möchte ich Frau Dr. Franzis Preckel und Herrn Dr. Ralf Schulze für ihre Ideen, Anregungen und Hilfestellungen.

Bedanken möchte ich mich weiterhin bei allen Schülerinnen und Schülern, die dazu bereit waren, mir einen Blick in ihr Arbeitsgedächtnis zu gewähren, und bei ihren Lehrerinnen und Lehrern, die den organisatorischen Rahmen dazu geschaffen haben.

Ich danke allen Forschungspraktikantinnen und -praktikanten, die mich bei den Untersuchungsdurchführungen und den Testauswertungen unterstützt haben, für ihren unermüdlichen Einsatz.

Danke an meine Familie und an Anneke für all die Unterstützung bis hierher.

