Diplomarbeit

Ist der Mental Arithmetic Test (MAT) Rasch homogen und valide? Eine Überprüfung 20 weiterer Items

Verfasst von

Gregor Voigt

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2014

Studienkennzahl:	298
Studienrichtung:	Psychologie
Betreuer:	Ao. Univ.-Prof. Dr. Georg Gittler

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Das Kopfrechnen spielt eine größere Rolle im Leben als man denkt. In Zeiten von Hochglanzmagazinen, vermeintlich perfekt geformten Körpern und Photoshop beschäftigen sich die Menschen zunehmend mit ihrem Äußeren und ihrem Gewicht. Das Zählen von Kalorien wird dabei durch die Kopfrechenfähigkeit bewerkstelligt. Aber auch im Beruf ist die Kopfrechenfähigkeit ständig im Einsatz. Hier muss zum Beispiel eine Restaurantbesitzerin die Preise für einzelne Gerichte berechnen. Führt man eine Suche nach offenen Stellen durch, so findet man schnell viele Angebote, die Kopfrechnen als Anforderung an den Bewerber stellen. Gibt man z. B. das Stichwort „Kopfrechnen“ in der Jobbörse der deutschen Arbeitsagentur ein, so bekommt man 41 Inserate angeboten (www.jobboerse.arbeitsagentur.de , 12.05.2014). Gibt man das gleiche Stichwort auf einer Internetseite zu Kleinanzeigen im Bereich Jobs und Karriere ein, so erhält man 31 Jobangebote, die Kopfrechnen als Anforderung listen (www.markt.de , 12.05.2014). Die angebotenen Arbeitsstellen kommen dabei aus einem breiten Feld. Nicht nur KassiererInnen, KellnerInnen und VerkäuferInnen aus verschiedensten Bereichen werden gesucht, sondern ebenfalls SchlosserInnen, KommissioniererInnen und Hotelfachkräfte. Auch eher außergewöhnliche Jobs, wie Poker-Dealer und ProduktionshelferInnen in der Chemie- und Pharmatechnik sind gelistet.

Das Anliegen dieser Arbeit ist die testtheoretische Analyse eines, sich in der Phase der Testkonstruktion befindlichen, diagnostischen Instruments, dem Mental Arithmetic Test[1] (MAT, deutsch: Kopfrechentest, Gabriel, 2004, Gittler, 2008 & Franz 2009). Dabei ist das Ziel einen Rasch homogenen Test zur Erfassung der Kopfrechenfähigkeit zu konstruieren. Dafür werden sowohl Verfahren der probabilistischen als auch der klassischen Testtheorie mit einbezogen. Im genaueren werden 20 Items des MAT auf Rasch Homogenität überprüft. Dieselben Items werden mithilfe eines Untertests der Sprachkompetenzanalyse (SKA), den Analogien, auf diskriminante Validität überprüft. Dieser verbale Test erscheint für diesen Zweck geeignet, da er wenig Anforderungen an die Versuchspersonen stellt, die gleichzeitig auch vom MAT gefordert werden. Die diskriminante Validität wird hier anstelle der konvergenten Validität gewählt, um keinen zweiten Kopfrechentest zu applizieren. Da der MAT sehr anstrengend ist, soll den Versuchspersonen kein zweiter Kopfrechentest zugemutet werden.

Zusätzlich werden erste Hinweise auf Validität der Items gesammelt, indem Gruppen, die sich in der Kopfrechenfähigkeit unterscheiden sollten, verglichen werden.

Es wird die Frage nach Geschlechtsunterschieden im Kopfrechnen untersucht, und exploriert, ob

[1] Die generativen Regeln für die Konstruktion der MAT-Items und einige der MAT-Items selbst wurden mir dankenswerterweise von meinem Betreuer zur Verfügung gestellt.

jüngere oder ältere Personen besser im Kopfrechnen sind.

Die Arbeit bietet zunächst theoretischen Hintergrund zum Kopfrechnen und zu Kopfrechentests. Sie erklärt im Methodenteil den MAT im Detail. Dann folgt sie dem üblichen Aufbau einer wissenschaftlichen Arbeit mit den Ergebnissen und schließt mit der Diskussion.

2 Theoretischer Hintergrund

2.1 Aktueller Forschungsstand zur Kopfrechenfähigkeit

Im 20. Jahrhundert wurde viel Forschung zum Thema Geschlechtsunterschiede im Bereich Mathematik und Kopfrechen publiziert. Feingold (1993) betrachtete Literatur aus den Jahren 1941 bis 1988 und stellte fest, dass Geschlechtsunterschiede in arithmetischen Fähigkeiten über die Zeit geringer wurden. Zusätzlich spielte das Alter eine große Rolle. Während es bei Kindern keine Geschlechtsunterschiede im Kopfrechnen gab, zeigte sich bei Erwachsenen ein Vorteil für männliche Teilnehmer. Die Meta-Analyse von Hyde, Fennema und Lamon (1990) kommt zu einem ähnlichen Ergebnis. 100 Studien aus den Jahren 1967 bis 1987 wurden analysiert und zeigten ebenfalls ein Absinken des Geschlechtsunterschieds über die Zeit. Es zeigte sich ein ähnliches Muster, was das Alter betrifft. In der Altersgruppe 5 bis 14 war der Geschlechtsunterschied kaum vorhanden. Die Mädchen waren hier ein wenig besser. In den Altersgruppen 15 und älter wurde der Geschlechtsunterschied konstant größer. Hierbei schnitten immer die männlichen Teilnehmer besser ab. Neben einem Vorteil für männliche Schüler im Kopfrechnen berichten Krinzinger et al. (2012) von keinem Unterschied zwischen den Kulturen. Der männliche Vorteil scheint zumindest in Österreich, Deutschland, Belgien und Frankreich robust zu sein.

Andere Literatur beschäftigt sich mit den Gründen für das Zustandekommen eines Geschlechtsunterschieds. Spencer, Steele und Quinn (1999) fanden z. B. heraus, dass allein die Instruktion ein Effekt auf die mathematische Leistung von Männern und Frauen haben kann. Hier schnitten Frauen schlechter ab, sobald die Instruktion Geschlechtsunterschiede in der mathematischen Leistung postulierte. Aber auch Umgebungsvariablen, wie Schulkultur (Ayalon & Livneh, 2013, Kenney-Benson, Patrick, Pomerantz & Ryan, 2006) und Schulbildung der Eltern (Penner & Paret, 2008) spielen eine Rolle. Ayalon und Livneh (2013) berichten von einem Einfluss der Schulstandardisierung, hier gemessen durch das Vorhandensein von einheitlichen, nationalen Prüfungen. Hierbei zeigte sich, dass männliche und weibliche Schüler ähnlicher in mathematischen Leistungen abschnitten, wenn auch die Schulstandardisierung höher war. Penner und Paret (2008) zeigen einen interessanten Effekt. Zunächst zeigen sie, dass Geschlechtsunterschiede bereits im Kindergarten nachweisbar sind. Weiter zeigen sie, dass Mädchen besser sind, aber nur in der mathematisch leistungsschwächeren Teilstichprobe. In der leistungsstärkeren Gruppe sind die Jungen besser.

Weiter wurden Persönlichkeitsfaktoren untersucht. Interessanterweise berichten Devine, Fawcett, Szücs und Dowker (2012) von keinem Geschlechtsunterschied bei Schülern von 12 bis 15 Jahren.

Und dies obwohl weibliche Jugendliche sowohl höhere Mathematik-Ängstlichkeit, als auch höhere Testängstlichkeit zeigten. Auch Rubinsten, Bialik und Solar (2012) berichten bei Studierenden von keinem Geschlechtsunterschied im Kopfrechnen.

Andere Intelligenzfaktoren als Ursache für Geschlechtsunterschiede im Kopfrechnen wurden ebenfalls untersucht. Zum Beispiel fanden Sluis et al. (2006) heraus, dass der männliche Vorteil in der für Kopfrechnen zentralen Arbeitsgedächtnisleistung nicht durch einen generellen Intelligenzfaktor (g) erklärt werden kann. Zu einem anderen Ergebnis kommen Jackson und Rushton (2006) mit der Methode der Faktorenanalyse. Hier lag der generelle Intelligenzfaktor g der mathematischen Leistung zugrunde.

Schlussendlich zeigen Lynn und Irwing (2008) in ihrer Metaanalyse, ähnlich, wie das Literatur-Review von Feingold (1993), dass Geschlechtsunterschiede zumindest im Kopfrechnen generell männliche Teilnehmer bevorzugen und dass dieser Unterschied mit steigendem Alter größer wird. Studien, die explizit das Alter als Einflussvariable untersuchten, fanden heraus, dass ältere Menschen generell langsamer im Kopfrechnen sind, aber gleichzeitig nicht mehr Fehler dabei machen (Allen, Ashcraft & Weber, 1992, Salthouse & Kersten, 1993). Zusätzlich fanden Allen, Ashcraft und Weber (1992) heraus, dass beim Einschätzen von simplen arithmetischen Gleichungen auf Richtigkeit (z. B. 7 x 4 = 28), als richtig eingeschätzte Gleichungen weniger Zeit benötigten. Auch Gleichungen mit größeren Ergebnissen als 20 benötigten mehr Zeit, wobei das Alter dabei keine Rolle spielte. Charness und Campbell (1988) zeigen ein gegenteiliges Ergebnis für das Multiplizieren einfacher Gleichungen. Hier machten die älteren Menschen weniger Fehler, und waren nicht langsamer, als die Jüngeren. In einer Übungsphase, in der das Quadrieren von Zahlen gelernt wurde, waren Ältere allerdings wieder langsamer in der Kopfrechengeschwindigkeit. Ältere Menschen waren ebenfalls bei Thevenot, Castel, Danjon, Fanget und Fayol (2012) langsamer als Jüngere. Dieser Unterschied nahm mit zunehmender Aufgabenschwierigkeit zu.

Andere neuere Literatur beschäftigt sich auch zunehmend mit neuronalen Korrelaten des Kopfrechnens (Fehr, Weber, Willmes & Herrmann, 2010, Feng, Wang, Chen & Liu, 2012, Pivik, Tennal, Chapman & Gu, 2012, Price, Mazzocco & Ansari, 2013, Van Impe, Coxon, Goble, Wenderoth & Swinnen, 2011). Zum Beispiel fanden Van Impe et al. (2011) heraus, dass ältere Menschen während einer Kopfrechenaufgabe in keiner Gehirnregion eine höhere Aktivierung aufweisen, obwohl diese in der Aufgabe schlechter abschnitten. Ebenfalls fanden sie heraus, dass Ältere und Jüngere in einer Dual-Task-Aufgabe die gleichen Gehirnregionen aktivierten und ebenfalls gleich gut abschnitten. Die Untersuchung von Fehr et al. (2010) fand heraus, dass ein Kopfrechenkünstler in einigen Gehirnregionen höhere Aktivierung zeigt. Er aktivierte zwar andere Regionen, aber auch die gleichen, wie normale Personen. Daraus schließen die Autoren, dass solche Savants nicht grundsätzlich anders als normale Kopfrechner sind, sondern, dass sie einfach mehr

Übung haben.
Welche anderen Einflüsse auf die Kopfrechenleistung werden von der aktuellen Literatur berichtet? Bürolärm sowohl mit, als auch ohne Hintergrundgespräche verschlechtern die Kopfrechenleistung (Perham, Hodgetts & Banbury, 2013). Depressive Patienten schneiden im Kopfrechnen schlechter ab, als gesunde Menschen (Feng, Wang, Chen & Liu, 2012). Bei Schülerinnen sagt deren Raumvorstellungsfähigkeit nicht nur deren Kopfrechengenauigkeit voraus, sondern ebenfalls deren Strategieeinsatz beim Kopfrechnen (Laski et al., 2013). Sogar Handbewegungen und die Bewegung der Kopfrechenaufgabe selbst haben Einfluss darauf, wie gut ein Mensch kopfrechnet (Wiemers, Bekkering & Lindemann, in Druck). Man rechnet besser, wenn man bei Additionen den Arm nach oben oder rechts bewegt und bei Subtraktionen, wenn man den Arm nach unten oder links bewegt. Hinweise darauf, dass das Fingerzählsystem beim Kopfrechnen auch in Erwachsenen noch verankert ist, fanden Klein, Moeller, Willmes, Nuerk und Domahs (2011). StudentInnen brauchten länger eine simple Additionsaufgabe zu lösen, wenn diese eine Berechnung enthielt, bei der ein „5er-Übertrag" enthalten war.
Kinder lösten Kopfrechenaufgaben häufiger korrekt, wenn sie ein Frühstück aßen, im Vergleich zu Kindern, die das nicht taten (Pivik et al., 2012). Eine interessante Methode, die Kopfrechenfähigkeit zu steigern zeigen Hauser, Rotzer, Grabner, Mérillat und Jäncke (2013). In ihrer Studie benutzten die Autoren die transkranielle Gleichstromstimulation (Nitsche et al., 2008) und verbesserten die Reaktionszeiten bei StudentInnen, während diese Subtraktionen bearbeiteten. Dies war der Fall, wenn die linke Hemisphäre stimuliert wurde. Eine Verbesserung der Genauigkeit konnte nicht erzielt werden. Wie sich eine gleichzeitig zu bearbeitende Aufgabe auf die Kopfrechenleistung auswirkt wurde ebenfalls untersucht (Imbo & LeFevre, 2009, Imbo & LeFevre, 2010, Imbo & Vandierendonck, 2010, Ketelsen & Welsh, 2010). Der Konsens der Studien ist, dass die Reaktionszeiten beim Kopfrechnen durch das Ausführen einer zweiten Aufgabe verlangsamt sind. Die Fehlerhäufigkeit ist weniger eindeutig interpretierbar. Bei Imbo und LeFevre (2010) sind die Fehlerhäufigkeiten nicht verändert, bei Imbo und Vandierendonck (2010) sind diese erhöht. Hier wird ebenfalls berichtet, dass die Fehlerhäufigkeiten mit steigender Problemgröße und niedrigerem Fähigkeitsniveau höher sind.
Nimmt man alle Befunde zusammen, so kann gesagt werden, dass sich eine Vielzahl von Einflüssen positiv, wie negativ auf das Kopfrechnen auswirken, und dass männliche Personen besser im Kopfrechnen abschneiden.

2.2 Kopfrechenfähigkeit als Teil der Intelligenz

Die ersten Untersuchungen zu intellektuellen Unterschieden zwischen Personen fanden bereits lange vor den ersten wissenschaftlichen Untersuchungen des 19. Und 20. Jahrhunderts statt (Eberwein, 1991). Auch vor der Veröffentlichung des ersten „echten“ Intelligenztests von Binet und Simon im Jahre 1905, fanden Testungen der Intelligenz statt. Der französische Arzt Edouard Séguin benutzte Mitte des 18. Jahrhunderts sogenannte „form boards“, um Kinder mit intellektueller Behinderung zu trainieren (Boake, 2002). Und ebenfalls Sir Francis Galton beschäftigte sich bereits im Jahre 1879 intensiv mit dem Messen von psychologischen Konstrukten, i. e. der Psychometrik (Galton, 1879).

Der Binet-Simon-Test für Kinder kann als der erste Intelligenztest angesehen werden. Die zwei im Test enthaltenen numerischen Tests „Répétiton de trois chiffres“ (Binet & Simon, 1904) und „unverzügliche Wiederholung von Ziffern“ (Funke, 2006) zeigen hier, dass numerische Fähigkeiten seit Anbeginn der Intelligenzmessung eine Rolle gespielt haben.

Später, im Jahre 1927, veröffentlichte Spearman, basierend auf Korrelationsstudien, sein Generalfaktormodell oder Zweifaktorenmodell (Stemmler, Hagemann, Amelang & Bartussek, 2011) der Intelligenz. Dieses sagte aus, dass intelligentes Verhalten auf einen einzelnen Faktor, genannt „g“, zurückführbar ist, sowie, dass jeder Test noch zusätzlich zum generellen Faktor etwas für sich Spezifisches, genannt „s“, misst. Die Kopfrechenfähigkeit wird in diesem Modell durch einen spezifischen Faktor „s“ des generellen Intelligenzfaktors „g“ repräsentiert.

Schnell wurde jedoch klar, dass der g-Faktor nicht ausreicht, um intelligentes Verhalten erschöpfend zu erklären. Vernon stellte auch deswegen sein hierarchisches Modell der Intelligenz auf. Dabei sind intelligente Faktoren in verschiedene hierarchische Gruppen aufgeteilt. Ganz oben in der Hierarchie steht dabei Spearmans Generalfaktor „g“. Inhaltliche Dimensionen waren hierbei verbal-numerisch-schulische und praktisch-mechanisch-räumlich-physikalische Faktoren. Schließlich gab es einen Faktor, der sich „arithmetic“ nannte, und der auf Einzeltestebene diese zu messen versuchte (Rost, 2009). Hier ist ebenfalls die Dimension Kopfrechenfähigkeit einzuordnen.

Einen anderen Weg ging im Jahr 1938 der amerikanische Psychologe Louis L. Thurstone. Er ging nicht von einem Generalfaktor der Intelligenz aus, sondern von mehreren Primärfaktoren der Intelligenz, den berühmten „primary mental abilities“. Hierbei postulierte Thurstone, dass die Faktoren, die intelligentes Verhalten bedingen, relativ unabhängig voneinander sind (Schmitt & Altstötter-Gleich, 2010). Ein Generalfaktor ist also nicht mehr verantwortlich für alles intelligente Verhalten, sondern unabhängige Faktoren. Zunächst waren dies neun Faktoren, die anschließend auf sieben reduziert wurden. Diese sind laut Schmitt und Altstötter-Gleich (2010) verbal comprehension, word fluency, number, space, memory, perceptual speed und reasoning. Der Faktor

„number“ kann wieder mit „rechnerische Fähigkeiten“ übersetzt werden und zeigt, dass auch in diesem sehr prominenten Modell die Kopfrechenfähigkeit einen Teil der Intelligenz darstellt. Genauer kann man dies erkennen, wenn man sich den Intelligenz-Struktur-Test 2000-R (I-S-T 2000-R) von Liepmann, Beauducel, Brocke und Amthauer (2007) detaillierter ansieht. Dieser wurde nämlich in seiner Ursprungsform von Amthauer konstruiert, wobei er sich dabei auf die originalen Arbeiten von Thurstone (Stemmler et al., 2011) bezog. In dem Test gibt es eine Bandbreite von numerischen Aufgaben, aber eben auch einfache bis schwierige Kopfrechenaufgaben, die logischerweise die Kopfrechenfähigkeit messen sollen.
Raymond B. Cattell stützte, ebenso wie Thurstone, sein Modell auf faktorenanalytische Auswertungen. Zunächst wurden 1966 ebenfalls Primärfaktoren extrahiert. Primärfaktor bedeutet hier lediglich, dass es sich um die Faktoren der ersten, bei dem betreffenden Datensatz, durchgeführten Faktorenanalyse handelt. In diesem Modell sind 23 Primärfaktoren aufgelistet (Stemmler et al. 2011), wobei auch einer davon mit rechnerischen Fähigkeiten übersetzt werden kann, nämlich der Faktor „Umgang mit Zahlen“. Dieser Faktor enthält in diesem Modell die Dimension Kopfrechnen. Darauf aufbauend, wurde mit den jetzt vorhandenen Primärfaktoren eine neuerliche Faktorenanalyse berechnet, um Faktoren höherer Ordnung zu extrahieren. Aus diesen Berechnungen kamen die zwei berühmten Sekundärfaktoren „fluide Intelligenz“ und „kristallisierte Intelligenz“ hervor. Mit fluider Intelligenz ist dabei die primär genetisch determinierte Fähigkeit gemeint, sich mit neuen Situationen auseinanderzusetzen und sich diesen anzupassen. Ebenfalls ist sie sensitiver in Bezug auf Hirnschädigungen und das Alter. Je älter man wird und je mehr Hirnschädigungen man erlitten hat, desto schlechter wird die fluide Intelligenz ausgeprägt sein (Horn, 1968). Anders ist dies bei der kristallisierten Intelligenz. Sie ist mehr das Ausmaß an deklarativem Wissen, welches man sich im gesamten Leben angeeignet hat, und sinkt nicht mit dem Alter ab, sondern bleibt bis ins hohe Lebensalter stabil und kann sogar noch steigen. Die Kopfrechenfähigkeit ist hier bei der kristallisierten Intelligenz einzuordnen. Charness und Campbell (1988) zeigen nämlich, dass ältere Menschen gleich schnell und sogar genauer im Kopfrechnen sind als Jüngere.
Guilford postulierte 1967 ein komplexes Modell mit drei Kategorien, den Operationen, Produkten und Inhalten (Eberwein, 1991). Bei den Operationen handelt es sich dabei um jene Tätigkeit, die eine Person gerade im Moment des intelligenten Verhaltens durchführt. Bezogen auf die Kopfrechenfähigkeit wäre das zum Beispiel das Gedächtnis, mit dem man direkt auf einfache, auswendig gelernte Ergebnisse von einfachen Rechnungen, wie z. B. 5 x 3 zugreifen kann. Produkte sind die Ergebnisse des intelligenten Verhaltens, also das Endresultat einer intelligenten Tätigkeit. Das Resultat einer Kopfrechnung, im Sinne des MAT, ist immer eine Einheit. Und schließlich sind Inhalte die Arten von Informationen, die durch intelligentes Verhalten verarbeitet werden sollen

(Stemmler et al., 2011). Im Fall des Kopfrechnens sind dies immer Symbole, da Zahlen immer ein Symbol für eine dahinter stehende Menge sind. Und die mathematischen Operatoren sind ebenfalls ein Symbol dafür, wie Zahlen miteinander zu verrechnen sind. Aus diesen postulierten drei Kategorien lassen sich 5 x 6 x 4 = 120 Intelligenzfähigkeiten ableiten, die theoretisch unabhängig voneinander sind. 100 davon gelten als empirisch bestätigt (Eberwein, 1991).

Das Berliner Intelligenzstrukturmodell von Jäger aus dem Jahre 1984 zeichnet sich durch einige Verbesserungen zu vorhergehenden Intelligenzmodellen aus. Außerdem vereinigt es Elemente „von Spearman, Thurstone und Guilford unter Berücksichtigung allgemeiner Erkenntnisse der Intelligenzforschung und sehr gezielt durchgeführter gesonderter Erhebungen bzw. Analysen" (Stemmler et al., 2011, S. 158). Wieder durch faktorenanalytische Auswertungen gelangte Jäger zu vier Operationsfacetten und drei Inhaltsfacetten (Rost, 2009). Jede Operationsfacette repräsentiert dabei ein Bündel von Fähigkeiten, welches eine Person bei der Bearbeitung von Aufgaben einsetzen kann. Für die Operationsfacette „Bearbeitungsgeschwindigkeit" sind diese Fähigkeiten das Arbeitstempo, wie leicht man etwas auffasst (Auffassungsleichtigkeit) und die Konzentrationfähigkeit (Rost, 2009, S.67). Die anderen drei Operationsfacetten heißen Gedächtnis, Einfallsreichtum und Verarbeitungskapazität. Inhaltsfacetten repräsentieren ebenfalls ein Bündel von Fähigkeiten. Diese Fähigkeiten sind hier jedoch an den Inhalt des Aufgabenmaterials gebunden. So umfasst die Inhaltsfacette „zahlengebundenes Denken" alle Fähigkeiten, die bei der Bearbeitung von Material, das Zahlen enthält, eingesetzt werden (Rost, 2009, S.67). Dazu zählt ebenfalls die Kopfrechenfähigkeit. Die übrigen zwei Inhaltsfacetten heißen sprachgebundenes Denken und anschauungsgebundenes Denken. Man kann sehen, dass in diesem vorbildlich konstruierten Modell die Kopfrechenfähigkeit in einer von nur sieben Facetten intelligenten Verhaltens enthalten ist, nämlich der Inhaltsfacette zahlengebundenes Denken. Bezogen auf die Operationsfacetten kann die Kopfrechenfähigkeit in jeder einzelnen verortet werden.

Zum Abschluss sei noch gesagt, dass es sich bei den Intelligenzleistungen von Jäger nicht um Primärfaktoren, wie im Modell von Guilford handelt, sondern um eine Zusammensetzung bzw. ein Zusammenwirken von zwei Intelligenzleistungen, die zu einer Leistung „verschmelzen". Weiter sei „die Kombination von gleichermaßen hypothesengeleitetem wie ausgesprochen methodenkritischem Vorgehen [als] absolut beispielhaft, und zwar auch im internationalen Vergleich" (Stemmler et al., 2011, S.159) noch einmal unterstrichen.

Das letzte hier vorgestellte Intelligenzmodell ist auch gleichzeitig das aktuellste von den hier beschriebenen. Gardner (2002) gibt einen Überblick über sein Modell der multiplen Intelligenzen aus dem Jahre 1999. Das Neue am Modell ist die Ausweitung einer einzelnen, der klassischen Intelligenz, auf andere, selten in Betracht gezogene Bereiche. Insgesamt werden sieben Intelligenzen genannt und dazu jeweils einige Berufe, in denen die Intelligenz von Vorteil wäre. Die

sieben Intelligenzen heißen laut Gardner (2002) sprachliche, logisch-mathematische, musikalische, körperlich-kinästhetische, räumliche, interpersonelle und intrapersonelle Intelligenz.
Die Kopfrechenfähigkeit ist hier bei der logisch-mathematischen Intelligenz einzuordnen.
Gardner bemüht sich zusätzlich um eine Integration von anderen Wissenschaften in die Intelligenzforschung. Zu nennen sind hier die Zellbiologie und die Genetik, die untersuchen sollen, welches Zusammenwirken welcher Gene, für welche Intelligenzfunktionen verantwortlich sind (Gardner, 2002, S. 254). Weiter führt er einige Überlegungen zum Nachweis der Autonomie der einzelnen Intelligenzen an (Gardner, 2002, S. 125f.).
An den aufgeführten und kurz erklärten Intelligenzmodellen ist zu erkennen, dass die Fähigkeit zum Kopfrechnen schon seit Anbeginn der Intelligenzmessung, aber auch in methodisch hochgelobten und auch sehr aktuellen Intelligenzmodellen einen, in verschiedenen Modellen größeren oder auch kleineren Teil der Überlegungen ausmachten. Dadurch, dass die Kopfrechenfähigkeit in jedem einzelnen Modell verortet ist, lässt sich schlussfolgern, dass diese in der psychologischen Intelligenzmessung von großer Bedeutung ist.

2.3 Tests zur Messung der Kopfrechenfähigkeit

Die Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern – revidierte Fassung (ZAREKI-R) soll eine Dyskalkuliediagnose bei Kindern ermöglichen, um so eine frühzeitige Förderung im Bereich Zahlenverarbeitung und Rechnen zu indizieren (Aster, Weinhold-Zulauf & Horn, 2006).
Der Untertest „Kopfrechnen“ misst die Kopfrechenfähigkeit mittels mündlich vorgegebenen Additions-, Subtraktions- und Multiplikationsaufgaben. Der Untertest „Textaufgaben“ misst die Kopfrechenfähigkeit mittels vorgelesenen Additions- und Subtraktionsaufgaben, die in Text eingekleidet sind.
Der Test zur Diagnose von Dyskalkulie (TeDDy-PC) ist ein Computertest, der mathematische Kompetenzen von Kindern im Grundschulalter diagnostizieren soll. Insbesondere soll er Dyskalkulie und Hochbegabung diagnostizieren helfen (Schroeders, & Schneider, 2008).
Der Untertest „Kopfrechnen“ misst die Kopfrechenfähigkeit über akustisch dargebotene Aufgaben, die im Kopf gelöst werden müssen. Der Untertest „Sachaufgaben“ misst die Kopfrechenfähigkeit durch akustisch vorgegebene Textaufgaben, die im Kopf gelöst werden müssen.
Der Test zur Erfassung numerisch-rechnerischer Fertigkeiten vom Kindergarten bis zur 3. Klasse (TEDI-MATH) ist ein Test zur Erkennung mathematischer Stärken und Schwächen zu einem frühen Zeitpunkt. Er erlaubt eine differenzierte Dyskalkuliediagnostik, Verlaufsdiagnostik, Interventionsplanung und Interventionsevaluation (Kaufmann et al., 2009).

Die Untertests „Addition“, „Unvollständige Addition“, „Subtraktion“, „Unvollständige Subtraktion“ und „Multiplikation“ messen die Kopfrechenfähigkeit entweder durch Rechenaufgaben in der entsprechenden Grundrechenart im normalen Format, oder durch ein ungewohntes Format. Bei diesem wird nicht das Ergebnis gesucht, sondern ein fehlender Operand aus der Gleichung gilt als richtige Antwort.

Das Rechenfertigkeiten- und Zahlenverarbeitungs-Diagnostikum für die 2. bis 6. Klasse (RZD 2-6) soll primär Rechenstörungen, unter Hinzunahme von weiteren Testverfahren, diagnostizieren. Zusätzlich liefert es Hinweise auf basale Informationsverarbeitungsstörungen (Jacobs, & Petermann, 2005).

Der Untertest „Kopfrechnen“ misst die Kopfrechenfähigkeit mittels schriftlich vorgelegter Aufgaben zu den vier Grundrechenarten. Der Untertest „Textaufgaben“ misst die Kopfrechenfähigkeit durch die visuelle Präsentation von Textaufgaben. Diese werden dem Kind laut vorgelesen und es soll die Aufgabe im Kopf lösen.

Der Test Zahlenrechnen 4+ (ZR 4+) ist ein Test, der es dem Lehrer erleichtern soll, sowohl eine Einzeldiagnose, als auch eine Kollektivdiagnose der gesamten Klasse über die Rechenlücken und Fehlerarten der Kinder zu stellen. Der Anwendungsbereich reicht dabei von dem Ende der vierten bis zum Anfang der fünften Klasse (Krueger, Hylla, & Bargmann, 1970).

Mit 30 Aufgaben misst der Test die Kopfrechenfähigkeit in allen vier Grundrechenarten. Weiter misst ein Teil dieser Aufgaben die Kopfrechenfähigkeit, indem in der Gleichung das Ergebnis steht, jedoch ein Operand fehlt.

Der Allgemeine Schulleistungstest für die 4. Schulstufe – Österreichische Version (AST 4/A) soll die Leistung in den Fächern Mathematik, Deutsch und Sachkunde in der zweiten Hälfte des vierten Schuljahrs erfassen. Dies soll den Lehrer bei der weiteren Schullaufbahnberatung unterstützen (Fippinger, Steidl, & Eher, 1972).

Der Untertest „Kopfrechnen“ misst die Kopfrechenfähigkeit durch die Vorgabe von Aufgaben zu den vier Grundrechenarten. Ebenfalls enthält der Test Aufgaben zum Bruchrechnen und zum Ergänzungsrechnen.

Der Hamburg-Wechsler-Intelligenztest für Kinder – IV stellt ein Einzeltestverfahren für die „Beurteilung der kognitiven Fähigkeiten von Kindern und Jugendlichen im Alter von 6;0 bis16;11 Jahren dar“ (Petermann & Petermann 2007, S.19).

Der Untertest „Rechnerisches Denken“ soll die Kopfrechenfähigkeit der Kinder mittels der vier Rechenoperationen testen. Zusätzlich werden hier die Aufgaben mündlich vorgegeben.

Das adaptive Intelligenz Diagnostikum 2 ist eine „Testbatterie für Kinder und Jugendliche [im Alter von 6;0 bis 15;11 Jahren] zur Erfassung komplexer und basaler Kognitionen“ (Kubinger & Wurst, 2000, S.8).

Der Untertest „Angewandtes Rechnen“ misst Kopfrechenfähigkeit, indem der Testleiter die Aufgabe, die in Text eingebettet ist, vorliest.
„Der Kognitive Fähigkeitstest (KFT) ist ein differentieller [Gruppen- und Einzel-] Intelligenztest zur Ermittlung der kognitiven Ausstattung von Schülern der 4. bis 13. Klassen“ (Heller, Gaedike & Weinläder, 1985, S.6). Er ist besonders zum Einsatz bei der Bildungs- und Schullaufbahnberatung geeignet.
Der Untertest „Gleichungenbilden“ misst die Kopfrechenfähigkeit, indem Gleichungen durch das Einsetzen der korrekten Rechenoperation richtig gelöst werden müssen.
Weitere Tests, die das Kopfrechnen messen, sind die Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern – Kindergartenversion (ZAREKI-K, Aster, Bzufka & Horn, 2009), der Deutsche Mathematiktest für erste Klassen (DEMAT 1+, Krajewski, Kuespert & Schneider, 2002) und der Test für schulrelevante Fähigkeiten 6-7 (TSF 6-7, Schallberger, & Trier, 1978).

Der Zahlenverarbeitungs- und Rechentest (ZRT) ermöglicht es, ein differenziertes Leistungsprofil von hirngeschädigten, erwachsenen Personen zu erheben. Weiter erlaubt er eine effektive Therapieplanung und die Einschätzung der Schwere der Schädigung (Kalbe, Brand, & Kessler, 2002).
Der Untertest „Kopfrechnen“ misst die Kopfrechenfähigkeit durch mündlich vorgegebene Rechenaufgaben in allen Grundrechenarten. Die Zahlen sind dabei maximal dreistellig und die Aufgabe kann maximal einmal wiederholt angehört werden.
Die Intelligenz-Struktur-Batterie (INSBAT) ist eine Testbatterie zur Erfassung des Intelligenzniveaus und der Intelligenzstruktur. Ihre Hauptanwendungsgebiete sind die Personalpsychologie, die Flugpsychologie und die pädagogische Psychologie (Arendasy et al., 2009).
Der Untertest „Arithmetische Kompetenz“ misst Kopfrechnen mittels der vier Rechenoperationen unter Zeitdruck. Im Untertest „Arithmetische Schätzfähigkeit“ soll nicht mehr exakt im Kopf gerechnet werden, sondern es soll das Ergebnis ohne Zeitdruck geschätzt werden. Der Untertest „Arithmetische Flexibilität“ fordert dagegen das Kopfrechnen dadurch, dass die Eingabe der richtigen Rechenoperation gefordert wird. Alle drei Untertests werden adaptiv vorgegeben.
Der Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) ist ein Test zur Messung der Intelligenz (Amthauer, Brocke, Liepmann & Beauducel, 2001). Er wird unter anderem in der differentiell-psychologischen Forschung eingesetzt. Er ist nach der klassischen Testtheorie konstruiert worden.
Die Aufgabengruppe „Rechenaufgaben“ misst Kopfrechnen mittels der vier Rechenoperationen unter Zeitdruck. Die Aufgabengruppe „Rechenzeichen“ misst die Kopfrechenfähigkeit durch die

Anforderung, die korrekten Rechenzeichen in eine Gleichung einzusetzen.
Der Wilde-Intelligenz-Test 2 „dient der differenzierten Erfassung klar unterscheidbarer und theoretisch fundierter kognitiver Fähigkeiten von Jugendlichen und Erwachsenen“ (Kersting, Althoff & Jäger, 2008, S.11). Sein Anwendungsbereich reicht von der allgemeinen Intelligenzdiagnostik über psychologische Forschung bis zur Eignungsdiagnostik. Beide Untertests „Grundrechnen“ und „Eingekleidete Rechenaufgaben“ sollen das Kopfrechnen mittels der vier Rechenoperationen messen. Bei letzterem Test sind die Aufgaben zusätzlich noch in Text eingekleidet.
„Der [Wechsler-Intelligenztest für Erwachsene] WIE ist ein Individualtest zur Untersuchung der kognitiven Fähigkeiten von Erwachsenen von 16 bis 89 Jahren“ (Aster, Neubauer & Horn 2006, S.8). Er soll unterschiedliche Aspekte der Intelligenz erfassen, darunter ebenfalls die Kopfrechenfähigkeit.
Der Untertest „Rechnerisches Denken“ misst die Kopfrechenfähigkeit durch mündliche Vorgabe von Rechenaufgaben. Diese enthalten die vier Rechenoperationen und sollen im Kopf gelöst werden.

Die meisten der Tests sind für Kinder und Jugendliche konstruiert, wurden nach der klassischen Testtheorie konstruiert und beinhalten eine Rechenoperation pro Aufgabe. Der MAT hingegen ist für die erwachsene Normalbevölkerung und nach Item-Response-Theorie konstruiert worden. Weiter enthält eine Aufgabe des MAT drei verschiedene Rechenoperationen und fünf Operanden in der Angabe. Somit erfüllt der MAT einen Bedarf, der von den derzeitig auf dem Markt befindlichen Kopfrechentests nicht erfüllt wird.

2.4 Benötigte Fähigkeiten zur Lösung des MAT

In diesem Abschnitt werden psychologische Konstrukte erklärt, die verantwortlich für das Lösen des Mental Arithmetic Tests sind. Zu den wichtigsten und hauptverantwortlichsten Konstrukten bzw. Fähigkeiten werden prominente Modelle präsentiert und im Detail erläutert. Der Abschnitt beginnt mit einer exemplarischen Lösung eines MAT-Beispiels, jedoch wird diesmal der Fokus des Lösungsweges auf die psychologischen Fähigkeiten gelegt. Nach der Identifizierung aller beteiligten Fähigkeiten werden diese im Detail beleuchtet und näher erklärt. Hierbei wird versucht, den Fokus wiederum auf Fähigkeiten zu legen, die für das Kopfrechnen notwendig sind, und nicht auf solche Fähigkeiten, die nur der Vollständigkeit halber erwähnt werden (z. B. Lesefähigkeit). Es wird versucht den gesamten Prozess des MAT-Item-Lösens von Anfang bis Ende sukzessive durchzuarbeiten. Dies geschieht anhand eines fiktiven Beispiels, in dem eine KlientIn vorschriftsgemäß zu einem Termin zur Testung ihrer Kopfrechenfähigkeit in einer psychologisch-diagnostischen Einrichtung eingeladen wird (zu ethischen Richtlinien der psychologischen Diagnostik in Österreich siehe Beiglböck, 2003). Auf die Erwähnung solcher, für das Anliegen dieser Arbeit unwichtig erscheinenden Fähigkeiten, wie die Fähigkeit zur Fortbewegung und Grundintelligenz, wird verzichtet. Die als wichtig erachteten Fähigkeiten sind immer kursiv geschrieben.

Man stelle sich nun eine KlientIn vor, die durch die Tür zum Testleiter hereintritt. Sie wird ordnungsgemäß begrüßt und instruiert, sich an den Computer zu setzen, wo alles Weitere erklärt wird. Sie beginnt die Instruktion des MAT zu lesen. Zum Verständnis der Instruktion sind selbstverständlich eine ausreichende Lesefähigkeit Voraussetzung, sowie andere Fähigkeiten, die ebenfalls zur Lösung eines MAT-Items nötig sind (z. B. das Verständnis der arabischen Ziffer als Symbol für eine dahinter stehende Menge). Man gehe nun pragmatisch davon aus, dass die KlientIn die Instruktion korrekt verstanden hat und mit dem ersten Item beginnt.

Als Allererstes muss sie verstehen, was die mathematischen Operatoren + und – bedeuten. Sie muss gelernt und begriffen haben, dass das Symbol „+“, wenn es zwischen zwei Zahlen steht, die folgende oder ähnliche Bedeutung bekommt: „Zähle die zwei Zahlen zusammen“. Gleich verhält es sich mit dem „-“. Dieses Symbol sollte bei ihr eine Handlungsanweisung in so oder so ähnlicher Form wachrufen: „Ziehe die Zahl, die rechts steht, von der Zahl, die links steht, ab“. Diese Fähigkeit wird hier als *Verständnis von Strichrechnungen* bezeichnet.

Zum Verstehen der Strichrechnungen hinzu kommt natürlich noch, dass die KlientIn die Bedeutungen dieser Symbole erinnern muss. Sie muss also die Assoziationen zu den Operatoren aus dem Gedächtnis abrufen können. Diese Fähigkeit wird hier als *semantisches Gedächtnis* bezeichnet. Semantisch deshalb, weil im semantischen Gedächtnis die Bedeutungen von Symbolen und Zeichen gespeichert werden. Bei den folgenden Erklärungen zu den benötigten Fähigkeiten, ist das semantische Gedächtnis immer beteiligt.

Jetzt muss die KlientIn natürlich bei jeder Zahl und Ziffer wissen, welche exakte Menge sich hinter der betreffenden Zahl verbirgt. Sie muss für jede arabische Zahl die genau richtige Menge repräsentieren. Bei den meisten erwachsenen Menschen geschieht das hochgradig automatisiert und subjektiv aufwandslos, jedoch zeigen Mejias, Grégoire und Noël (2012), dass auch erwachsene Menschen noch Schwierigkeiten mit dieser Aufgabe haben können. Diese Fähigkeit sei hier als *exakte Zahl-Menge-Relation* bezeichnet.

Nun folgt die eigentliche Verrechnung der beiden Zahlen. Es müssen Strategien vorhanden sein, um die Zahlen entsprechend den mathematischen Operationen zu verrechnen. Für die Strichrechnungen sei hier je kurz ein Beispiel angeführt. Bei einfachen Additionen wie 20 + 13 wird oftmals die Strategie des direkten Abrufs aus dem Gedächtnis verwendet. Hierbei wurde das Ergebnis der

Rechnung bereits viele Male im Langzeitgedächtnis gespeichert und ist deswegen relativ aufwandslos abrufbar. Bei einfachen Subtraktionen 29 – 7 wird oftmals einfach die Strategie des Herunterzählens verwendet (Thompson & Smith, 1999). Es wird von der 29 in Einerschritten heruntergezählt. Weitere Strategien, die auch häufiger von Erwachsenen benutzt werden, werden weiter unten im Detail besprochen. Diese Fähigkeit soll hier *Strategieeinsatz bei Strichrechnungen* heißen.

Weiter muss sich die KlientIn natürlich auch wieder an diese Strategien erinnern können. Ganz analog zum semantischen Gedächtnis muss sie also die Strategien aus dem Gedächtnis abrufen. Hierfür wird aber ein anderer Teil des Gedächtnisses beansprucht, und zwar das *prozedurale Gedächtnis*. Dieses speichert nicht die Bedeutung von Symbolen, sondern es speichert Prozeduren, was in diesem Fall einfach nur den beschriebenen Strategien entspricht.

Nun hat die KlientIn also das Ergebnis aus der ersten Rechnung errechnet. Dieses muss sie sich nun merken. Hierfür wird das Arbeitsgedächtnis in Anspruch genommen. Sie muss sich die Zahl merken, und zwar während sie die zweite Strichrechnung durchführt. Zur Veranschaulichung dessen, was in diesen Momenten des Merkens und Rechnens passiert, wird hier auf das später noch im Detail zu erklärende Arbeitsgedächtnismodell von Baddeley und Hitch (Baddeley, 2012) vorgegriffen. Dieses enthält nämlich einen sogenannten phonologischen Speicher und eine sogenannte zentrale Exekutive. Ersterer ist hier verantwortlich für die Zwischenspeicherung des ersten Zwischenergebnisses. Dieser phonologische Speicher tut nichts anderes, als das Ergebnis subvokal zu wiederholen. Zur Veranschaulichung wird hier ein konkretes Beispiel verwendet. Jeder Mensch hat eine innere Stimme, die er zum Beispiel beim Lesen eines Buches oder beim Nachdenken benutzt (Ridgway, 2009). Und diese innere Stimme wird beim Rechnen ebenfalls benutzt. Die KlientIn hat im Beispiel nun das Ergebnis der ersten Rechnung im Kopf, z. B. die Zahl 34. Nun wiederholt die innere Stimme (die übersetzt in psychologische Begriffe dem phonologischen Speicher entspricht) das Ergebnis ganz simpel immer wieder: „34, 34, 34, 34, 34, ...“. Währenddessen berechnet die zentrale Exekutive mittels der oben beschriebenen Fähigkeit *Strategieeinsatz bei Strichrechnungen* das zweite Zwischenergebnis. Die hier eingesetzten Fähigkeiten werden als *Arbeitsgedächtnis* betitelt.

Im nächsten Schritt müssen beide Zwischenergebnisse sowohl mittels phonologischer Schleife im Arbeitsgedächtnis aufrechterhalten werden, als auch miteinander verglichen werden. Dafür sind Kenntnisse der exakten Mengen der beiden arabischen Zahlen notwendig (*exakte Zahl-Menge-Relation*). Nun müssen nur noch die Mengen der arabischen Ziffern nach Anzahl der Einheiten (die arabische Ziffer 8 repräsentiert acht Einheiten) verglichen werden. Die hier neu dazu kommende Fähigkeit wird in Anlehnung an Landerl und Kaufmann (2008) als *Fähigkeit zum Zahlengrößenvergleich* bezeichnet.

Je nachdem, wie der Zahlengrößenvergleich ausfiel, muss die KlientIn nun wieder eine Additions- oder Subtraktionsaufgabe lösen. Welche Fähigkeiten daran beteiligt sind, ist bereits weiter oben erklärt worden. Die KlientIn berechnet also das dritte Zwischenergebnis. Dieses muss nun, in einem letzten Schritt, entweder multipliziert oder dividiert werden. Hierbei ist es wieder von Nöten, zu verstehen, wofür die Symbole der mathematischen Operatoren für Multiplikation (x) und Division (/) stehen. Im ersten Fall sollte eine Assoziation der Form „Addiere die Zahl, die links steht mit sich selbst, und zwar so oft, wie sie die Zahl, die rechts steht, angibt, oder umgekehrt“ entstehen. Im letzteren Fall sollte eine, der folgenden Aussage mindestens ähnliche, Aussage wachgerufen werden: „Überlege, wie oft die rechte Zahl in die linke Zahl hineinpasst“. Diese Fähigkeit soll hier *Verständnis von Punktrechnungen* heißen.
Schließlich ist die letzte neue Fähigkeit, die bei der Berechnung eines einzelnen MAT-Items zum Einsatz kommt der *Strategieeinsatz bei Punktrechnungen*. Hier soll vorerst wieder je ein Beispiel pro Operator genügen. Bei Multiplikationen kann die Strategie Count-by eingesetzt werden. Bei einer einfachen Rechnung wie z. B. 8 x 4 wird einfach die Sequenz der Achterfolge hochgezählt (also 8, 16, 24, 32) um auf das Ergebnis zu kommen (Sherin & Fuson, 2005). Bei Divisionen kann die Strategie der Umformung der Aufgabe in eine Multiplikationsaufgabe angewandt werden. Hierbei wird z. B. bei der Aufgabe 288 / 12 zunächst 12 x 10 = 120 berechnet, dann 120 x 2 = 240 um auf ein Zwischenergebnis von 20 zu kommen. Anschließend wird nun noch der Rest zu 288, in diesem Fall 48, geteilt durch 12 berechnet. Hier bekommt man die Antwort 4 und kommt somit auf das richtige Ergebnis 24. Genau genommen werden in dieser Berechnung ebenfalls die Operationen Addition und Subtraktion durchgeführt. Der Vollständigkeit halber sei noch angemerkt, dass hier natürlich wieder das *Arbeitsgedächtnis* sowie die *exakte Zahl-Menge-Relation* mit im Einsatz sind. Schließlich muss nun noch das von der KlientIn berechnete Endergebnis mittels phonologischer Schleife kurz im Arbeitsgedächtnis behalten werden, um es anschließend in das betreffende Eingabefeld am Computer einzutippen.
Somit sind bei der Lösung eines MAT-Items neun Fähigkeiten im Einsatz. Das sind in „chronologischer“ Reihenfolge *das Verständnis von Strichrechnungen, das semantische Gedächtnis,* eine *exakte Zahl-Menge-Relation,* der *Strategieeinsatz bei Strichrechnungen,* das *prozedurale Gedächtnis, das Arbeitsgedächtnis,* die *Fähigkeit zum Zahlengrößenvergleich,* das *Verständnis von Punktrechnungen* und schließlich der *Strategieeinsatz bei Punktrechnungen*.

Im folgenden Abschnitt werden, wo nötig, nun die einzelnen Fähigkeiten im Detail näher beleuchtet. Zum Strategieeinsatz bei Punkt- und Strichrechnungen werden diesmal verschiedene Strategien anhand von Beispielen erklärt. Hingegen wird auf die Erklärung der Fähigkeit zum Zahlengrößenvergleich, dem Verständnis von Strich- und Punktrechnungen sowie der exakten Zahl-

Menge-Relation verzichtet. Das geschieht aufgrund der bereits ausreichenden Erklärung, sowie mangelnder Komplexität dieser Fähigkeiten. Zunächst wird jedoch auf die wichtigste Fähigkeit im Kopfrechenprozess eingegangen, auf das Arbeitsgedächtnis.

2.4.1 Das Arbeitsgedächtnis

Das Arbeitsgedächtnis wird durch das Arbeitsgedächtnismodell von Baddeley und Hitch aus dem Jahr 1974 beschrieben. Es besteht aus drei Komponenten (siehe Abbildung 1): Der zentralen Exekutive, der phonologischen Schleife und dem visuell-räumlichen Notizblock.

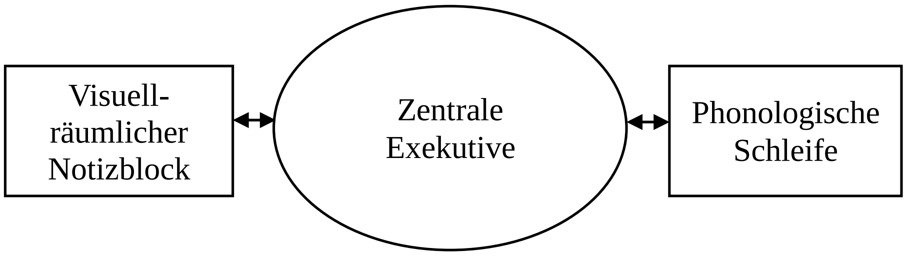


Abbildung 1: Arbeitsgedächtnismodell von Baddeley und Hitch (1974)

Die beiden letztgenannten Komponenten dienen als Zwischenspeicher für Informationen, die im Moment einer zu erfüllenden Aufgabe benötigt werden. Die phonologische Schleife übernimmt dabei die Zwischenspeicherung von verbalem, also sprachlichem Material. Der visuell-räumliche Notizblock speichert visuelle, also bildliche, aber auch räumliche Informationen. Die zentrale Exekutive übernimmt dabei die Aufgaben der Kontrolle und Regulation der beiden Zwischenspeicher. Sie koordiniert die beiden Systeme, ist verantwortlich für die Aufmerksamkeitssteuerung und verbindet das Arbeitsgedächtnis mit dem Langzeitgedächtnis (Baddeley & Logie, 1999).
Baddeley, Eysenck und Anderson (2009) bringen ein anschauliches Beispiel, um diese drei Subsysteme zu illustrieren. Man stelle sich vor, man ist nicht bei sich zu Hause und bekommt die Aufgabe gestellt, zu berichten, wie viele Fenster man in seinem Haus oder Apartment hat. Zunächst stellt man sich zum Beispiel sein eigenes Wohnzimmer vor. Hierfür wird der visuell-räumliche Notizblock eingesetzt. Dieser produziert vor dem geistigen Auge, ein mehr oder weniger präzises und detailreiches Bild vom eigenen Wohnzimmer. Er bekommt den Auftrag dazu von der zentralen Exekutive. Nun wird das Bild vor dem geistigen Auge aufrechterhalten und es werden die einzelnen Fenster abgezählt. Welche Strategie dabei gewählt wird, bestimmt wieder die zentrale Exekutive. Man zählt also zum Beispiel von rechts nach links. Vor dem geistigen Auge dreht man sich also von

rechts nach links und zählt. Hierbei kommt die phonologische Schleife zum Einsatz. Die innere Stimme zählt von eins hoch, bis alle Fenster gezählt sind. Die Überwachungsfunktion der zentralen Exekutive ist dabei ständig aktiv, sie vergleicht ständig das aktuelle Ergebnis (z. B. vier gezählte Fenster) mit dem finalen Ziel, dem Zählen aller Fenster im Wohnzimmer. Ist dieses erreicht und von der zentralen Exekutive bestätigt, so wird die Aufgabe im nächsten Raum fortgesetzt. Dabei muss die phonologische Schleife ständig das Endergebnis des Wohnzimmers mittels innerer Stimme wiederholen, damit das Ergebnis nicht verloren geht. Nun werden vom visuell-räumlichen Notizblock räumliche Informationen des gesamten Hauses repräsentiert. Keller bis Dachgeschoss werden in räumliche Beziehungen gebracht und vor dem geistigen Auge vorgestellt. Hier entscheidet wieder die zentrale Exekutive, welche Strategie angewandt wird. „Zuerst im Keller nachzählen oder weiter im ersten Stock?“. Es wird sich wieder das entsprechende Zimmer vorgestellt und der Prozess wiederholt sich so lange, bis alle Fenster gezählt sind. Die zentrale Exekutive überprüft ein letztes Mal die Zielerreichung und schließt die Aufgabe ab. Zu diesem Beispiel sei noch angemerkt, dass die meisten beschriebenen Prozesse höchst automatisch und schnell ablaufen.
Es werden nun die Einzelkomponenten des Modells näher beleuchtet und erklärt. Zusätzlich wird kurz auf eine Erweiterung des Modells eingegangen.

2.4.1.1 Die Zentrale Exekutive (nach Baddeley, Eysenck & Anderson, 2009)

Die zentrale Exekutive (kurz: zE) hat, wie schon gesagt, die Rolle eines Aufpassers, der bei einer Aufgabe alles kontrolliert und aufpasst, dass nichts schief läuft. Dabei hat die zE zwei typische Möglichkeiten, wie Verhalten kontrolliert werden kann. Der eine Modus ist der automatische Modus, der andere ist der der bewussteren, aufmerksameren Kontrolle des gerade ausgeübten Verhaltens. Letztgenannter Modus beansprucht mehr kognitive Ressourcen. Der zweite Modus ist also sehr viel anstrengender als der erste. Bei automatisierten Aufgaben, wie es bei geübten Autofahrern der Fall ist, läuft die Verhaltenskontrolle meist automatisch ab. Hier weiß die Autofahrerin bereits in den meisten Situationen, was zu tun ist und was aus ihrer Sicht das richtige und beste Verhalten ist (und damit aus Sicht des Arbeitsgedächtnisses). Zum Beispiel weiß sie, dass sie bei Rot anhalten muss, um keinen Unfall zu verursachen. Diese Entscheidung fordert kaum kognitive Ressourcen, weil sie automatisiert ist. Kommt es jedoch zu einer ungewohnten, neuen Situation, wie z. B. aufsteigendem Rauch aus der eigenen Motorhaube, so kommt das „supervisory attentional system“ (SAS) zum Einsatz. Übersetzt heißt dieses in etwa „überwachendes Aufmerksamkeitssystem“. In Anbetracht des Rauches, der aus der Motorhaube kommt, hat nun die Autofahrerin mehrere Optionen. Sie kann weiterfahren und größeren Schaden riskieren oder sie

kann anhalten und nachschauen, ob sie das Problem beseitigen kann. Das SAS entscheidet hier, welche Option die wahrscheinlich bessere ist. Auch ist das System verantwortlich für die Suche nach weiteren Handlungsoptionen, wie z. B. anzuhalten und einen Pannendienst anzurufen.
An diesem Beispiel kann man ebenfalls die überwachende Aufgabe der zE erkennen. Ständig wird das aktuelle Verhalten darauf überprüft, ob es angemessen und richtig ist. Fällt die zE aus, so kann es zu unangemessenem Verhalten kommen. Baddeley, Eysenck und Anderson (2009) schildern hier ein Beispiel eines Frontallappenpatienten, der mit seiner Frau redet und gleichzeitig bestreitet verheiratet zu sein. Nachdem sie ihm deren gemeinsame Hochzeitsbilder zeigt, beharrt der Patient aber weiter darauf, nicht verheiratet zu sein.
Und schließlich hat die zE zwei weitere Aufgaben, und zwar erstens die bewusste Aufmerksamkeitszuwendung auf einen bestimmten Stimulus, hier genannt selektive Aufmerksamkeitssteuerung, und zweitens die Aufteilung der Aufmerksamkeit auf mehrere Stimuli, hier genannt geteilte Aufmerksamkeitssteuerung. Bei der selektiven Aufmerksamkeitssteuerung muss aus einem komplexen Umfeld, wie z. B. der Umgebung aus der Sicht eines Autofahrers, immer ein bestimmter Bereich ausgesucht werden, auf den die Aufmerksamkeit gelegt werden soll. Bei richtig funktionierender zE wird das verhaltensrelevanteste Merkmal des Umfelds ausgesucht. In einem Beispiel zum Autofahren wäre das die Ampel, um in Erfahrung zu bringen, ob angehalten werden sollte oder weitergefahren werden kann.
Die geteilte Aufmerksamkeitssteuerung funktioniert ähnlich, nur dass aus dem komplexen Umfeld zwei oder auch mehrere Stimuli ausgesucht werden, die mit Aufmerksamkeit belegt werden. Dies geschieht immer wieder im Beispiel des Telefonierens mit einem Handy während des Autofahrens. Die Aufmerksamkeit ist auf das Autofahren gelegt, aber gleichzeitig wird dem, was der Gesprächspartner sagt, aufmerksam gefolgt.

2.4.1.2 Die phonologische Schleife (nach Baddeley, 2012)

Die phonologische Schleife ist verantwortlich für die kurzfristige Zwischenspeicherung verbaler Information. Damit sind sprachliche Informationen gemeint, also gesprochene und gehörte Wörter und Sätze. Die Speicherung geschieht dabei durch vokale oder subvokale Wiederholung der zu speichernden Information. Das Erstere bedeutet, dass man die Information laut ausspricht und fortlaufend wiederholt, um sich den Inhalt zu merken, des Letztere bedeutet das Gleiche, mit dem Unterschied, dass der Inhalt nicht laut, sondern stumm, mittels innerer Stimme wiederholt wird.
Es wurde, und wird noch immer, die laut Baddeley (2012) bis heute ungeklärte Frage behandelt, auf welchem Mechanismus das Vergessen in Zusammenhang mit der phonologischen Schleife basiert. Zwei Mechanismen stehen hier zur Auswahl, und zwar das Zerfallen von Gedächtnisspuren und die

Interferenz. Bei Erstgenanntem bilden sich kurzfristig Gedächtnisspuren des potenziell zu speichernden Inhalts. Diese werden für extrem kurze Zeit in einem passiven phonologischen Speicher gespeichert, jedoch verwischen oder Zerfallen diese Spuren dann sehr schnell wieder. Durch die subvokale oder vokale Wiederholung können diese Spuren wieder aufgefrischt werden. Bei dem Mechanismus der Interferenz werden die kurzfristig gespeicherten Gedächtnisspuren durch das aktive Überschreiben durch andere Spuren gelöscht. Das heißt, dass direkt nachdem erste Spuren durch einen ersten Inhalt kreiert wurden, ein neuer Inhalt in die phonologische Schleife kommt und die Spuren des ersten Inhalts verwischt. Dabei kann das tatsächlich durch das Überschreiben der ersten Inhalte geschehen, oder aber es geschieht durch das Verdrängen oder Verschieben (engl.: displacement) der Inhalte.
Eine weitere Eigenschaft der phonologischen Schleife ist dessen limitierte Kapazität. Es kann nur eine gewisse Menge an Informationen zwischenspeichern. Ein Beispiel hierfür ist der Wortlängeneffekt. Er besagt, dass eine Person sich längere Wörter weniger gut merken kann als kurze Wörter. Dies wird damit erklärt, dass bei längeren Wörtern das Wiederholen der Wörter mehr Zeit benötigt und deswegen auch der Zerfall von Gedächtnisspuren länger vor sich geht. Allerdings könnten auch Interferenzen für diesen Effekt verantwortlich sein. Kurzum kann gesagt werden, dass man sich so viele Wörter merken kann, wie man in zwei Sekunden vorlesen kann.
Weiter kann die phonologische Schleife sich die Sequenz mehrerer Informationen merken. Das bedeutet, sie erinnert, an welcher Position welche Information ist. Die Informationen werden also in Relation zueinander gesetzt.
Schließlich sei noch erwähnt, dass die phonologische Schleife, also das vokale oder subvokale Wiederholen von Informationen, ein sehr schneller Prozess ist, der minimale Aufmerksamkeit erfordert. Es scheint also ein sehr effizienter Weg zu sein, um sich kurzzeitig verbale Informationen zu merken.

2.4.1.3 Der visuell-räumliche Notizblock (nach Baddeley, 1996)

Da der visuell-räumliche Notizblock bei der Lösung von Kopfrechenaufgaben nur eine untergeordnete Rolle zu spielen scheint (Logie, Gilhooly & Wynn, 1994), wird dieser nicht, wie die anderen Komponenten des Arbeitsgedächtnismodells, im Detail erläutert.
Der visuell-räumliche Notizblock ist verantwortlich für die Zwischenspeicherung von visuellen, das sind bildliche, Informationen. Ebenfalls sorgt er dafür, dass räumliche Informationen zwischengespeichert werden. Hierbei handelt es sich um Informationen über die räumlichen Anordnungen oder Relationen von Objekten. Ein anschauliches Beispiel ist hier die Information, dass Rom südlich, Hamburg nördlich, Moskau östlich und Porto westlich von Wien liegen. Es wird

vermutet, dass es innerhalb des Notizblocks zwei Systeme gibt, die für die temporäre Speicherung von jeweils einer der beiden Kategorien von Informationen verantwortlich ist. Diese Systeme interagieren dabei sehr stark miteinander. Hinweise für diese Theorie liefern vor allem neurologische Patienten, bei denen die Fähigkeit zur Zwischenspeicherung räumlicher Information gestört ist, aber nicht die Speicherung visueller Information und umgekehrt.
Weiter scheint der visuell-räumliche Notizblock weit weniger eingesetzt zu werden und auch weniger automatisch abzulaufen, als die phonologische Schleife. Schließlich ist auch dessen Ausführung wesentlich anstrengender und vereinnahmt mehr kognitive Ressourcen, als es die phonologische Schleife zu tun scheint. Letztendlich weiß man, auch aus diesen Gründen, noch relativ wenig über diesen Teil des Arbeitsgedächtnisses, speziell über die Frage, wie die visuelle und räumliche Information wiederholt wird, um so die Zwischenspeicherung zu gewährleisten.

2.4.1.4 Erweiterung des Modells: Der episodische Puffer (nach Baddeley, 2000)

Der episodische Puffer wurde neu in das Modell eingefügt, um einige unerklärte Phänomene zu erklären (siehe Abbildung 2). Dieser speichert ebenfalls Informationen für eine kurze Zeit, integriert aber Informationen aus mehreren Quellen. Es sind also im episodischen Puffer visuelle, räumliche und verbale Informationen integriert verfügbar.

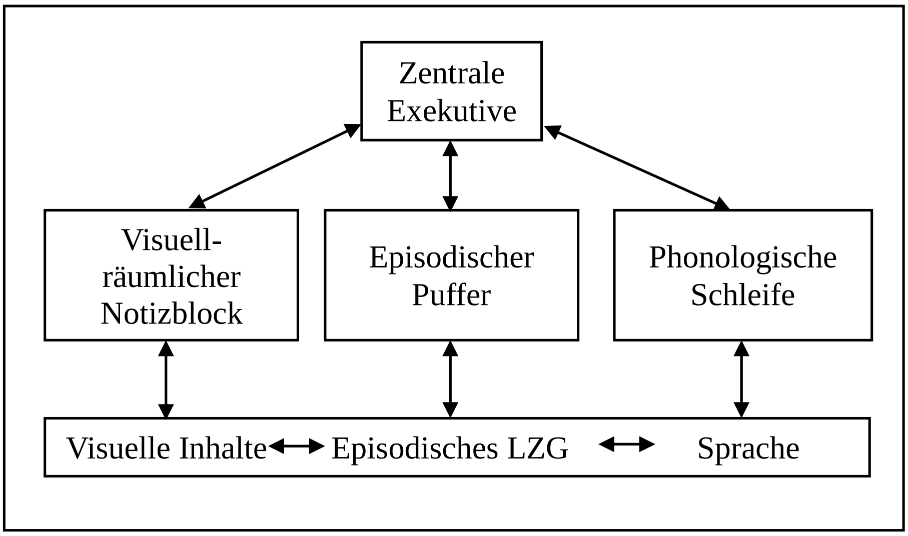


Abbildung 2: Erweitertes Arbeitsgedächtnismodell von Baddeley (2000)

Es werden dabei insofern Episoden kreiert, als die verschiedenen Informationen quer durch Raum und potenziell ebenfalls quer durch Zeit integriert werden. Dabei ist auch dieses System von limitierter Kapazität, da der Aufwand des Produzierens einer solchen Episode, enorm ist. Der episodische Puffer wird dabei von der zentralen Exekutive kontrolliert. Diese kann die

Informationen, die vom Puffer bereitgestellt werden, abrufen und modifizieren. Dies gelingt durch bewusste Aufmerksamkeit, die auf die Episode im Puffer gelegt wird. Es kann sich der episodische Puffer also als „Interface“ zwischen phonologischer Schleife, visuell-räumlichem Notizblock und Langzeitgedächtnis vorgestellt werden, auf den die zentrale Exekutive, durch Belegung dessen, mit bewusster Aufmerksamkeit, zugreifen kann. Letztere übernimmt ebenfalls das Integrieren der Information aus verschiedenen Quellen. Der Puffer dient lediglich zur Aufrechterhaltung.

2.4.2 Das semantische Gedächtnis (nach Baddeley, Eysenck & Anderson, 2009)

Das semantische Gedächtnis ist einer von zwei Teilen des deklarativen Gedächtnisses und dieses ist wiederum Teil des gesamten Gedächtnisses. Der zweite Teil des deklarativen Gedächtnisses ist das episodische Gedächtnis, auf das hier nicht näher eingegangen wird. Welche Art von Information wird in diesem Teil des Gedächtnisses gespeichert? Es ist zum größten Teil das Wissen um die Bedeutung von Symbolen, wie z. B. Buchstaben und Zahlen, und generelles Allgemeinwissen. Dazu gehören geographische Begebenheiten, Wissen über politische Figuren und Ämter, Wissen über alle Arten von Kunst (z. B. Film, Musik), Wissen über die Eigenschaften von Tieren und vieles mehr. Es würde schlicht ewig dauern, alle Inhalte aufzulisten. Diese Inhalte sind alle bewusst abrufbar und bewusst aussprechbar. Gleichzeitig ist ihnen gemein, dass das auch die generelle Erscheinungsweise von semantischem Wissen ist, nämlich als ausgesprochener Satz. Wichtig dabei ist, dass jeder Mensch ein großes, überlappendes Wissen besitzt. Zum Beispiel weiß fast jeder Mensch, dass New York City in den Vereinigten Staaten von Amerika liegt. Oder, dass eine Tochter immer eine Mutter hat. Gleichzeitig hat aber auch jede Person ihr eigenes Wissen, welches so nicht jeder andere Mensch besitzt. Hierbei geht es meistens um das Wissen, welches man sich im eigenen Beruf und durch Hobbys angeeignet hat. Ein Filmregisseur wird sicherlich wissen, wer die Oscargewinner der Kategorie „Bester Regisseur“ der letzten zehn Jahre sind. Viele andere Menschen werden das nicht wissen. Ebenfalls wird ein Mensch, der Gitarre spielen als Hobby ausübt, wissen, wie man einen F-Dur-Akkord spielt, ohne dass eine große Menge an Menschen außerhalb der Stichprobe von Gitarrenspielern das ebenfalls weiß.
Der Speicher des semantischen Gedächtnis scheint dabei riesig zu sein, denn professionelle Schachspieler, zum Beispiel, haben zwischen 10.000 und 100.000 Stücke an Information, nur über Muster, die sich auf dem Schachbrett zeigen können. Hinzu kommen dann noch zwischen 20.000 und 100.000 Wörter nur der eigenen Muttersprache, die jeder Mensch kennt. Und mit diesen Informationen ist nur ein kleiner Teil der menschlichen Realität abgedeckt. Es kann also davon ausgegangen werden, dass der Speicher noch sehr viel größer ist.

2.4.3 Das prozedurale Gedächtnis (nach Anderson, 2007)

Das prozedurale Gedächtnis ist Teil des nichtdeklarativen Gedächtnisses und dieses ist wieder Teil des gesamten Gedächtnisses. Nichtdeklarativ spricht dabei bereits eine zentrale Eigenschaft des prozeduralen Gedächtnisses an. Die normale Erscheinungsweise von prozeduralem Wissen ist ganz im Gegensatz zu semantischem Wissen implizit, also nicht explizit artikuliert. Eine weitere Eigenschaft ist, dass es sich bei prozeduralen Wissensinhalten um Wissen um Vorgehensweisen und typische Abläufe handelt. Sie werden oft gelernt, ohne sich dessen bewusst zu sein, was genau gelernt wird oder ohne sich die einzelnen Schritte einer Vorgehensweise genau zu vergegenwärtigen.

Anderson (2007) gibt ein anschauliches Beispiel. Fahrrad fahren haben wohl die meisten Menschen gelernt. Hierbei wurde wahrscheinlich aber nicht zuerst gelernt, sich den Sitz richtig einzustellen und sich dann auf den Sitz zu setzen, während man den Lenker festhält. Auch wurde wahrscheinlich nicht gelernt, dass man dann abwechselnd mit dem linken und rechten Fuß in die Pedale treten muss. Nein, die meisten werden sich einfach auf das Fahrrad gesetzt haben und sind mit Stützrädern oder Hilfe der Mutter losgefahren, bis sie es allein konnten.

Anderson (2007) gibt ein weiteres Beispiel für ein eher kognitives Wissen um eine Vorgehensweise. TeilnehmerInnen sollten in einem Experiment die produzierte Zuckermenge (Z) einer Fabrik in einem Bereich von 8000 bis 10000 Tonnen halten. Alles, was sie dafür angeben mussten, war die Menge an eingesetzten Arbeitern (A). Die produzierte Menge an Zucker wurde dabei nach der immer gleichen Formel ($Z = 2 \times A - Z_1$) berechnet, wobei Z_1 die produzierte Menge aus der vorherigen Periode war. Die TeilnehmerInnen konnten nach 60 Durchgängen das Ziel recht gut erreichen. Jedoch konnten sie nicht sagen, nach welcher Vorgehensweise sie die Anzahl der Arbeiter aussuchten. Sie sagten etwa, dass sie eben ein Gefühl dafür entwickelt hätten. Die Formel, also die Vorgehensweise, nach der die Produktionsmenge bestimmt wurde, konnte keiner benennen. Es wurde hier von den TeilnehmerInnen also implizites, prozedurales Wissen über den Zusammenhang zwischen Arbeitskräften und produzierter Zuckermenge erworben, ohne das gleichzeitig auch explizites Wissen darüber erlernt wurde.

Weitere Beispiele für prozedurales Wissen sind das Wissen um den typischen Ablauf in einem Restaurant, das Wissen, wie man schwimmt und das Wissen, wie man große Zahlen multipliziert.

2.4.4 Strategieeinsatz bei Strichrechnungen

Für die Addition und Subtraktion werden hier je drei Beispiele angeführt, wie Menschen Kopfrechnen können. Dabei ist zu erwähnen, dass für viele kognitive Aufgaben eine Kombination von Strategien eingesetzt wird und diese daraufhin ausgewählt werden, wie effizient sie sind (Geary, Frensch & Wiley, 1993).

Bei Additionen werden laut Geary und Wiley (1991) folgende Strategien von jüngeren und älteren Erwachsenen benutzt. Dabei benutzen ältere Erwachsene nur die beiden letztgenannten Strategien.

1. Retrieval (direkter Abruf): Das Ergebnis der einfachen Addition 8 + 5 wird direkt aus dem Langzeitgedächtnis abgerufen.
2. Verbal Counting (Hochzählen): Die Aufgabe 6 + 9 wird durch ein Ziffer-für-Ziffer Hochzählen gelöst. Dabei wird die höhere Ziffer 9 als Ausgangswert benutzt (9, 10, 11 usw.).
3. Decomposition (Zerlegung): Die Aufgabe 6 + 9 wird in Teilrechnungen aufgeteilt:

    6 – 1 = 5, dann 9 + 1 = 10 und letztlich 5 + 10 = 15.

Bei den Subtraktionsstrategien geben Geary, Frensch und Wiley (1993) einen Überblick über die gewöhnlich verwendeten Strategien, die auch von Erwachsenen benutzt werden.

1. Decomposition (Zerlegung): Für die Lösung der Aufgabe 32 – 3 wird der Subtrahend 3 in zwei kleinere Ziffern zerlegt und dann nacheinander vom Minuend 32 abgezogen:

    3 = 2 + 1, dann 32 – 2 = 30 und letztlich 30 – 1 = 29.

2. Rule (Regel): Für die Lösung der Aufgabe 32 – 9 wird der Subtrahend 9 auf 10 erhöht. Anschließend wird 10 vom Minuend 32 abgezogen. Im Anschluss wird nun die Differenz zwischen 10 und dem Subtrahend 9 zum Zwischenergebnis addiert:

    9 + 1 = 10, dann 32 – 10 = 22, dann 10 – 9 = 1 und letztlich 22 + 1 = 23.

3. Columnar Retrieval (Spaltenweiser direkter Abruf): Für die Lösung der Aufgabe 32 - 9 werden aus dem Langzeitgedächtnis arithmetische Fakten abgerufen. Dies geschieht Spalte für Spalte:

    30 – 10 = 20, dann wird sich das Ergebnis 20 zwischen gemerkt. Im Anschluss 10 + 2 = 12, dann 12 – 9 = 3 und letztlich 20 + 3 = 23

2.4.5 Strategieeinsatz bei Punktrechnungen

Strategien bei Multiplikationen werden bei Heirdsfield, Cooper, Mulligan und Irons (1999) aufgelistet. Einschränkend muss hier gesagt werden, dass es sich dabei um Strategien handelt, die von Kindern in den Schuljahren vier bis sechs angewendet wurden. Allerdings zeigt Gabriel (2004, S. 39), dass auch noch Erwachsene die vermeintlich simpelste Strategie des Hochzählens benutzen. Deswegen kann davon ausgegangen werden, dass Erwachsene ebenfalls die komplizierteren

Strategien anwenden.
1. Counting (Hochzählen): Die Aufgabe 5 x 8 wird durch Hochzählen in 5er Schritten gelöst (5, 10, 15 usw.).
2. RL seperated (Rechts-Links getrennt): Für die Lösung der Aufgabe 5 x 19 wird zuerst die rechte Ziffer mit 5 multipliziert. Dann wird die linke Ziffer mit 5 multipliziert und anschließend werden die Ergebnisse aus beiden Teilrechnungen addiert:

    5 x 9 = 45, dann 5 x 10 = 50 und letztlich 45 + 50 = 95

3. Wholistic (Ganzheitlich): Bei der Lösung der Aufgabe 25 x 19 werden die Zahlen als Ganzheit betrachtet:

    4 x 25 = 100, dann 4 x 4 = 16, dann 4 x 100 = 400 und schließlich werden noch 3 x 25 = 75 addiert (400 + 75 = 475).

Campbell und Xue (2001) teilten die Rechenstrategien von StudentInnen a priori in vier Kategorien ein, wobei eine davon lediglich als Restkategorie verwendet wurde. Die Kategorien können für alle vier Grundrechenarten angewandt werden, werden hier aber exemplarisch für Strategien bei Divisionen erklärt.
1. Transform (Transformieren): Die Aufgabe 72 / 9 wird gelöst, weil die Aufgabe in die Aufgabe 8 x 9 = 72 transformiert werden kann.
2. Count (Hochzählen): Die Aufgabe 72 / 9 wird gelöst, indem in 9er Schritten hochgezählt wird (9, 18, 27 usw.) und dabei die Anzahl der einzelnen Hochzählschritte mitgezählt wird.
3. Remember (Direkter Abruf): Das Ergebnis der Aufgabe 72 / 9 wird einfach direkt aus dem Langzeitgedächtnis abgerufen.

Generell kann gesagt werden, dass die meisten Strategien auf dem direkten Abruf und auf der Zerlegung einer Aufgabe in Teilaufgaben basieren. Die dritte, häufig eingesetzte Strategie ist auch bei Erwachsenen das Hochzählen.

2.5 Nutzen und Anwendung des Mental Arithmetic Test

Ein Nutzen des MAT ist, dass dieser in der allgemeinen Intelligenzdiagnostik eingesetzt werden kann. Bei der allgemeinen Intelligenzdiagnostik geht es darum, möglichst viele Facetten der Intelligenz mithilfe von Verfahren zu erfassen, die mittels psychologischer Gütekriterien konstruiert und abgesichert wurden. Kopfrechnen ist, wie bereits erwähnt, ein Teil der Intelligenz. Und somit ist der MAT überall dort einsetzbar, wo die allgemeine Intelligenz diagnostiziert werden soll.
Ein weiterer Nutzen ist, dass der MAT im Prozess einer Dyskalkuliediagnostik eingesetzt werden

kann. Hierbei muss sowohl die allgemeine Intelligenz festgestellt werden, als auch die Rechenfähigkeit.

Der MAT kann auch helfen, Fragestellungen der Berufs-, Studien- und allgemeinen Eignungsdiagnostik zu beantworten. Es gibt bestimmte Berufe, in denen die Kopfrechenfähigkeit eine spezielle Anforderung für jede BewerberIn darstellt. Für jeden Beruf gibt es ein Anforderungsprofil, in dem Anforderungen an die BewerberInnen aufgelistet sind. Ist in einem Anforderungsprofil auch die Anforderung rechnerische Fähigkeiten enthalten, so kann der MAT hier wiederum eingesetzt werden, um die Kopfrechenfähigkeit zu überprüfen.
Höft, Stelling und Maschke (2010) bringen ein Beispiel zur Berufseignungsdiagnostik. Beim Auswahlprozess eines Kopiloten gibt es mehrere Selektionsstufen, die eine jede BewerberIn überstehen muss. In der zweiten werden neben anderen Fähigkeiten auch rechnerische Fähigkeiten verlangt. Der MAT könnte an dieser Stelle des Selektionsprozesses eingesetzt werden. Dazu sei noch erwähnt, dass es neben dem Berufsfeld des Kopiloten noch viele, weitere Berufsbilder gibt, die eine gute Kopfrechenfähigkeit als Anforderung stellen.
Analog verhält es sich mit der Studieneignungsdiagnostik**.** Ist es in einem Studiengang wichtig, gute rechnerische Fähigkeiten zu besitzen, so kann man mithilfe des MAT überprüfen, ob dies bei einer interessierten Person der Fall ist oder nicht.
Schließlich sei der potentielle Anwendungsbereich des MAT innerhalb der Eignungsdiagnostik noch auf alle Bereiche, in denen eine Auswahl aus einer bestimmten Menge an Personen vollzogen wird, erweitert. Dies ist der Fall, wenn z.B. Zugangsbeschränkungen zu privaten Hochschulen oder Ausbildungsplätzen existieren.

Genutzt werden kann der MAT auch, um die eigene Kopfrechenfähigkeit zu verbessern. Übungseffekte können schon durch das bloße, mehrmalige Wiederholen des gleichen Tests erwartet werden (Kubinger, 2009b). Sowohl in der Allgemeinbevölkerung, als auch im klinischen Bereich, kann der MAT so Anwendung finden.

2.6 Das Rasch-Modell

Georg William Rasch schlug im Jahre 1960 das bekannte Rasch-Modell zur Leistungs- und Intelligenzmessung vor (Rasch, 1960). Sein dichotom, logistisches Testmodell geht davon aus, dass je höher die Fähigkeit einer Person (Personenparameter) und je niedriger die Schwierigkeit eines Items (Itemparameter) ist, desto höher ist die Wahrscheinlichkeit, dass eine Person ein Item löst

(Lösungswahrscheinlichkeit) und umgekehrt. Wenn die Fähigkeit einer Person gleich ausgeprägt ist, wie die Schwierigkeit eines Items, so ist in diesem Fall die Lösungswahrscheinlichkeit 50 %.
Eine Veranschaulichung dieses Zusammenhangs zwischen Itemparameter, Personenparameter und Lösungswahrscheinlichkeit, geben die sogenannten Item-Characteristic-Curves (ICCs) oder Itemcharakteristikkurven (siehe Abbildung 3).

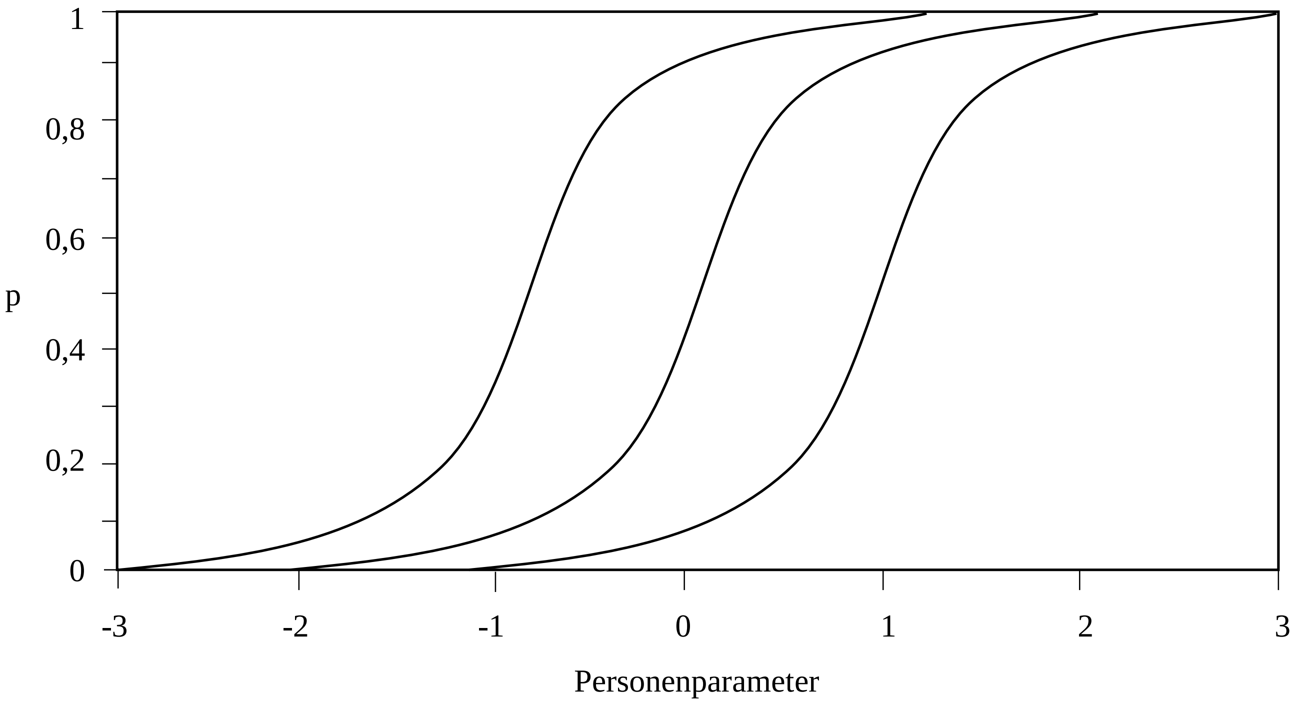


Abbildung 3: Parallele Itemcharakteristikkurven von drei verschieden schwierigen Items

Die X-Achse repräsentiert die möglichen Ausprägungen des Personenparameters und die Y-Achse repräsentiert die Lösungswahrscheinlichkeit. Die drei Kurven repräsentieren drei verschieden schwierige Items. Dabei weisen die Items, die weiter rechts abgebildet sind, eine höhere Schwierigkeit auf, als jene, die weiter links stehen. Oder anders ausgedrückt: Bei den weiter rechts abgebildeten Itemcharakteristikkurven braucht eine Person eine höhere Fähigkeit, um die gleiche Lösungswahrscheinlichkeit zu erreichen, wie bei einem Item, das weiter links abgebildet ist.
Eine ausführliche und mathematische Erklärung des Rasch-Modells bieten Fischer und Molenaar (1995).

2.6.1 Modelltests

2.6.1.1 Likelihood-Ratio-Test

Der Likelihood-Ratio-Test (LRT) wurde 1973 von Andersen vorgeschlagen (Andersen, 1973) und wird deswegen auch Andersen‘s LRT genannt. Dieser schätzt zunächst die Itemparameter für die gesamte Stichprobe und berechnet dessen Likelihood. Likelihood bedeutet hier so etwas, wie die begründete Wahrscheinlichkeit der Daten. Im Anschluss wird diese Berechnung noch einmal durchgeführt und zwar getrennt für zwei Teilstichproben, die nach einem bestimmten Teilungskriterium getrennt wurden. Das häufigste Kriterium ist dabei der Median des Testrohwerts (Bühner, 2011, S. 531). Hierbei wird die Stichprobe in fähigere (die Personen haben einen hohen Rohwert) und weniger fähige Personen (die Personen haben einen niedrigen Rohwert) unterteilt. Andere Teilungskriterien sind das Geschlecht, das Alter (hier wird die Stichprobe in zwei Altersgruppen geteilt) und die höchste abgeschlossene Schulbildung. Man kann jetzt sagen, dass zwei Modelle vorhanden sind. Einmal das Modell der gesamten Stichprobe und das Modell der geteilten Stichprobe, jeweils mit verschiedenen Likelihoods und verschiedenen Itemparametern. Im Modell der gesamten Stichprobe haben alle Items der Gesamtstichprobe die gleiche Schwierigkeit. Im Modell der geteilten Stichprobe haben die gleichen Items zwei verschiedene Schwierigkeiten, eine für die fähigeren und eine für die weniger fähigen Personen. Das zweite Modell ist also etwas komplexer, es erklärt mehr und beschreibt die Itemschwierigkeiten präziser. Der LRT überprüft nun, wie stark die beiden Modelle voneinander abweichen und ob sie dies signifikant tun. Das geschieht mittels folgender Formel:

$$\chi 2 = -2 \cdot ln\left(\frac{L_0}{cL_1 \cdot cL_2}\right) \tag{1}$$

Hier wird der natürliche Logarithmus (*ln*) gezogen. Und zwar von der Likelihood des Modells für die gesamten Daten (L_0) geteilt durch das Produkt der Likelihoods der Modelle für die geteilten Daten (cL_1 und cL_2). Schließlich wird alles noch mit -2 multipliziert.

Weichen nun die Modelle signifikant voneinander ab, so bedeutet dies, dass die Itemschwierigkeiten der zwei Teilstichproben stark voneinander abweichen. Man kann hier die Analogie zur Messung einer Fähigkeit ziehen. Wenn in den Teilstichproben eine Fähigkeit gemessen wird, so sollten die Itemschwierigkeiten annähernd gleich sein. Ist dies der Fall, so unterscheiden sich die Modelle auch nicht signifikant voneinander, der LRT wird ebenfalls nicht signifikant.

2.6.1.2 Wald-Test

Der Wald-Test basiert ebenso auf der Idee, dass sich die Itemschwierigkeiten der beiden gebildeten Teilstichproben nicht unterscheiden sollten (Strobl, 2012). Im Gegensatz zum LRT werden hier die Itemschwierigkeiten der einzelnen Teilstichproben direkt und einzeln miteinander verglichen. Weichen die Schwierigkeiten signifikant voneinander ab, so spricht das dafür, dass das betreffende Item unterschiedliche Dimensionen für die Teilstichproben misst bzw. das Item nicht fair ist. Der Wald-Test kann eingesetzt werden, nachdem der LRT signifikant wurde, um herauszufinden, welche Itemschwierigkeiten besonders stark voneinander abweichen.

2.6.1.3 Martin-Löf-Test

Der Martin-Löf-Test (MLT) teilt, im Gegensatz zum LRT, die Stichprobe nicht nach Personen auf, sondern nach Items (Bühner, 2011). Hierbei ist es wichtig, dass dies nach inhaltlichen Kriterien geschieht. Hier kann z. B. die Itemschwierigkeit als Teilungskriterium dienen. Dadurch entstehen zwei Testteile, einer mit den leichten und einer mit den schwierigen Items. Der MLT prüft nun, ob die Testteile homogen sind bzw. ob sie die gleiche Dimension messen (Bühner, 2011).

2.6.1.4 Vorteile und Anwendungen des Rasch-Modells

Das Rasch-Modell als Hilfsmittel zur Konstruktion von psychologischen Tests bietet einige Vorteile gegenüber Tests, die nach klassischer Testtheorie konstruiert wurden. Einer wurde bereits weiter oben erwähnt. Gilt das Rasch-Modell für einen Test, so werden spezifisch objektive Vergleiche möglich. Das erlaubt Personenparameter und Itemparameter zu vergleichen und das jeweils unabhängig von den eingesetzten Items und der getesteten Stichprobe.
Weiter ist empirisch nachgewiesen, dass der Test eindimensional misst. Das heißt, dass lediglich eine Fähigkeit bzw. ein Bündel von Fähigkeiten für die Beantwortung des Tests verantwortlich ist. Die meisten Tests behaupten zwar, dass sie nur eine Fähigkeit messen, jedoch ist dies meist nicht empirisch überprüft worden. Ein fiktives Beispiel soll den Vorteil der Eindimensionalität veranschaulichen. Die Stichproben werden dabei durch das Teilungskriterium Nationalität getrennt. Ein Item eines Intelligenztests fragt nach der Entfernung zwischen Bregenz und Wien. Den Personen mit einer österreichischen Nationalität wird die Lösung der Aufgabe vermutlich leichter fallen als jenen Personen mit einer spanischen Nationalität. Man kann hier sagen, dass die Frage in der österreichischen Teilstichprobe das Allgemeinwissen abfragt, jedoch in der spanischen Stichprobe misst diese Frage vermutlich etwas Anderes. Hier wird wohl eher geographisches Detailwissen gemessen. So wäre dieses Item nicht fair gegenüber den Stichproben, weil es eben

nicht die gleiche Fähigkeit misst. Eindimensionalität durch Rasch Homogenität weist nach, dass ein solches Problem bei keinem der Items im Test vorliegt.
Auch kommt mit Geltung des Rasch-Modells die lokale stochastische Unabhängigkeit einher. Diese bedeutet, dass das Lösen eines Items nicht davon abhängig ist, wie und ob das direkt vorangegangene Item, oder auch andere Items, gelöst wurde. Es ist wieder nur die Fähigkeit, die für das Lösen eines Items verantwortlich ist. Die lokale, stochastische Unabhängigkeit wäre zum Beispiel dann verletzt, wenn man zum Lösen des zweiten Items das erste Item richtig gelöst haben müsste. Dort hängt die Lösung des zweiten Items direkt von der Lösung des ersten ab.
Weiter werden erschöpfende Statistiken möglich. Das bedeutet, dass der Rohwert eine ausreichende und erschöpfende Statistik bzw. Basis für die Schätzung des Personenparameters darstellt. Statt aufwändig jedes einzelne gelöste Item in Kombination mit anderen gelösten Items als Basis einer Schätzung der Fähigkeit heranzuziehen, genügt ein Wert: die Summe der gelösten Items (Rohwert). Dabei ist es unerheblich, welche Items dabei gelöst wurden. Die Anzahl allein reicht schon aus.
Ein weiterer Vorteil ist hier, dass die Fähigkeit einer Person, durch die Transformation des Rohwerts in einen Personenparameter, ebenfalls eine Transformation seines Skalenniveaus erfährt.
Rohwerte haben Ordinalskalenniveau, sie sagen nur aus, dass jemand besser als jemand anderes ist. Die genaue Differenz des „Besser-Seins“ liegt dabei im Dunklen. Anders ist das bei den Personenparametern. Diese haben Intervallskalenniveau, welches genau abbilden kann, wie groß der Fähigkeitsunterschied zwischen zwei Personen ist. Dieser wird also besser quantifizierbar und mittels genauer Werte präziser abgebildet.
Mittels der Itemkonstruktion mithilfe des Rasch-Modells wird eine besondere Anwendung möglich, das computerisierte, adaptive Testen. Hierfür ist zunächst eine aufwändige Generierung eines großen Itempools notwendig. Alle Items müssen dabei positiv auf Rasch Homogenität geprüft worden sein und es müssen Items möglichst aller Schwierigkeitsstufen vorhanden sein. Aus dieser Voraussetzung folgt, dass stichprobenunabhängige Itemparameter für jedes Item vorhanden sind.
Nun bearbeitet eine Person das erste Item am PC und löst es oder nicht. Basierend darauf, ob die Person das Item löst oder nicht, berechnet der Computer automatisch den Fähigkeitsparameter der Person. Dann wählt er ein Item mit einer Schwierigkeit aus, welche der Fähigkeit der Person entspricht. Die Person beantwortet das Item, der Computer berechnet einen neuen Fähigkeitsparameter und wählt das nächste passendste Item aus. Mit jedem vorgegebenen Item erhöht sich die Präzision des geschätzten Fähigkeitsparameters, bis dieser sich nur noch unwesentlich verändert. Ist das der Fall so wird der Test automatisch beendet.
Durch eine solche Vorgehensweise kommen weitere Vorteile zur Geltung. Es werden keine zu leichten oder zu schwierigen Items vorgegeben. Dadurch kann sich die Frustration durch Über- oder Unterforderung verringern. Die Testdauer kann dadurch ebenfalls verkürzt werden.

2.7 Diskriminante Validität und Gruppenunterschiede als Validitätshinweise

Ein Test ist diskriminant valide, wenn er eine schwache Korrelation mit einem konstruktfernen Test aufweist. Konstruktfern bedeutet, dass der Test, der zur Validierung herangezogen wird, etwas komplett anderes messen sollte, als der zu validierende Test.

Validitätshinweise können für einen Test gesammelt werden, indem sich StudienteilnehmerInnen in einer Fähigkeit selbst einschätzen. Anschließend wird die Stichprobe künstlich geteilt, um so einen Gruppenunterschied berechnen zu können. Die Teilstichproben werden nach höherer und niedrigerer Fähigkeit geteilt. Zeigt der berechnete Gruppenunterschied Signifikanz und die Gruppenmittelwerte, der sich als fähiger eingeschätzten Gruppe, sind höher, so kann das als Hinweis auf Validität interpretiert werden.

3 Methode

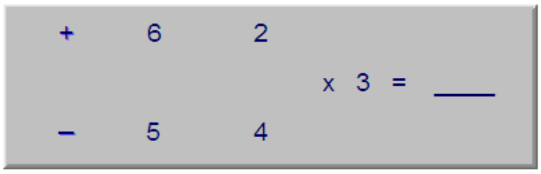
3.1 Instrumente

Alle Instrumente wurden am Computer, in der nachfolgend beschriebenen Reihenfolge, vorgegeben. Zunächst wurden demographische Daten, wie z. B. Alter und Geschlecht, abgefragt. Im Anschluss daran, wurden die TeilnehmerInnen gebeten, Selbsteinschätzungen bezüglich ihrer Matheaffinität abzugeben. Dies geschah immer auf einer sechsstufigen Skala, in der 1 für eine hohe Ausprägung und 6 für eine niedrige Ausprägung stand (sehr hoch = 1 2 3 4 5 6 = sehr gering). Es wurde zum Beispiel gefragt, wie gut sie ihre eigene Fähigkeit im Umgang mit Zahlen und Mengen oder ihr Wissen um mathematische Vorgehensweisen einschätzen.
Im Anschluss wurde der Mental Arithmetic Test vorgegeben.
Der Mental Arithmetic Test (kurz: MAT) entsprang der Initialidee bzw. dem Bedürfnis, eine Intelligenztestbatterie zu konstruieren, bei dem jeder einzelne Untertest dem Rasch-Modell entspricht. Dabei wird neben der Rasch Homogenität, und all deren Vorteilen (siehe Abschnitt 2.8.1.4), auf zwei mögliche Anwendungen dieser abgezielt. Der Test soll erstens adaptiv vorgegeben werden, das heißt, auf das individuelle Leistungsniveau der KlientInnen angepasst. Und zweitens soll bei manchen Tests die automatische Itemgenerierung zum Einsatz kommen. Dabei können durch einen Computer ständig neue Items nachkonstruiert werden, ohne jeglichen Aufwand für den Testanwender. Für beide Anwendungen ist die Rasch Homogenität jeweils Voraussetzung. Dem Urheber der Initialidee, ao. Univ. Prof. Dr. Georg Gittler, war es ein Bedürfnis das Rasch-Modell zur Testkonstruktion zu verwenden, um so eine Batterie zu erstellen, die alle Vorteile und Anwendungen des Rasch-Modells mit sich bringt.
Der MAT ist ein, sich in der Konstruktionsphase befindlicher, Test, der die Intention hat, die Fähigkeit „Kopfrechnen" zu messen. Weiter ist er ein Leistungs- und Powertest mit Work-Limit-Instruktion. Außerdem ist er ein Gruppen- und Computertest, der sich eines freien Antwortformats bedient (siehe Abschnitt 2.7).
Die Testlänge des Mental Arithmetic Tests beträgt in den beiden hier eingesetzten Versionen 14 Aufgaben pro Version. Hinzu kommen noch drei Übungsaufgaben, die aber nicht in die Auswertung mit einbezogen werden. Basierend auf ersten Probetestungen wurde die Bearbeitungszeit für eine Version mit 28 Minuten eingeschätzt. Das entspricht zwei Minuten pro Aufgabe. Dabei soll das nur einen Durchschnittswert darstellen. Items zu Beginn der Testung sollten schneller, und Items am Ende der Testung sollten langsamer gelöst werden. Inklusive Instruktion und Verständnisfragen sollte die Gesamtbearbeitungszeit für den MAT so auf geschätzte 33 Minuten kommen.

In Anbetracht dessen, dass die letzte beschriebene Instruktion bei Gabriel (2004) stark von der in dieser Arbeit verwendeten abweicht, wird nun die hier verwendete Version der Instruktion präsentiert. Die gesamte Instruktion inklusive aller Screenshots basiert auf Gittler, 2009.

Die folgenden Aufgaben erfassen Kopfrechenfähigkeit. Die Aufgaben enthalten die vier Grundrechenarten: Addieren (+), Subtrahieren (-), Multiplizieren (x) und Dividieren (/).

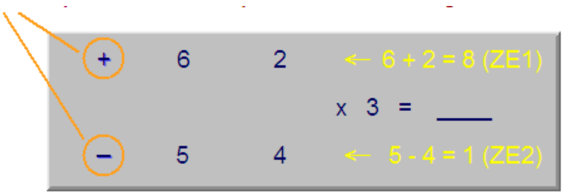
Alle Zwischen und Endergebnisse sind ganzzahlig (ohne Nachkommastellen).



Jede Aufgabe hat vier Rechenschritte, die nun für das Beispiel vorgeführt werden.

Berechnen Sie die Zwischenergebnisse (ZE1 und ZE2) aus der ersten und letzten Zeile und merken Sie sich diese.

Die mathematischen Operatoren finden Sie jeweils am Zeilenanfang.



Danach wenden Sie eine der folgenden Regeln an:

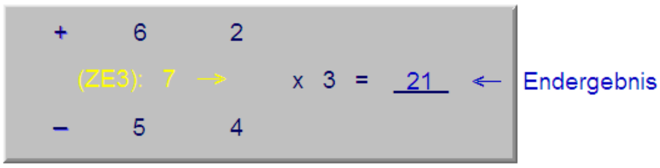
- Wenn das ZE1 größer ist als das ZE2, dann wird ZE2 von ZE1 abgezogen.
- Wenn das ZE1 kleiner ist als das ZE2, dann werden ZE1 und ZE2 zusammengezählt.

Mit anderen Worten:

Subtrahiert wird nur dann, wenn ZE1 minus ZE2 eine positive Zahl ergibt.

Im Beispiel ist ZE1 größer als ZE2. Es wird daher subtrahiert: 8 – 1 = 7 (eine positive Zahl). Das dritte Zwischenergebnis (ZE3) lautet also 7 und muss gemerkt werden, da damit weitergerechnet wird.

Um das Endergebnis zu erhalten, muss für das Zwischenergebnis 3 (ZE3 = 7) noch die in der Mitte stehende, mathematische Operation ausgeführt werden. Im Beispiel wird also 7 x 3 = 21 als Endergebnis berechnet. Das Endergebnis muss wiederum gemerkt werden, bevor es im Test als Lösung eingegeben werden kann.



Noch ein Hinweis bevor Sie selbstständig Übungsbeispiele bearbeiten:
Nebenrechnungen und Zwischenergebnisse müssen im Kopf ausgeführt und gemerkt werden, sie dürfen nicht notiert werden!

Der Einfachheit halber wird im folgenden Abschnitt derselbe Lösungsweg noch einmal in Form eines erfundenen „Protokolls“ des lauten Denkens präsentiert. Dafür wird dasselbe Beispiel, wie in der Instruktion herangezogen:
Ich addiere 6 mit 2 und merke mir das Zwischenergebnis 8. Dann subtrahiere ich 4 von 5 und merke mir das Zwischenergebnis 1. Jetzt vergleiche ich 8 mit 1 und stelle fest, dass 8 größer ist als 1. Das bedeutet, ich muss 1 von 8 abziehen. Ich ziehe 1 von 8 ab und merke mir das Ergebnis 7. 7 multipliziere ich jetzt mit 3 und schreibe das Ergebnis 21 in die dafür vorgesehene Spalte.

Alle Items des MAT wurden nach dem, in der Instruktion gezeigten Muster konstruiert. Jedes Item wurde nach einem bestimmten Itemdesign konstruiert, welches von Gabriel (2004) aus der Literatur abgeleitet wurde. Dieses theoriebasierte Itemdesign wurde anhand von Daten von Gittler (2008) erweitert und konkretisiert, so dass daraus ein Kriterienkatalog resultierte. Ziel war es dabei, generative Regeln für die Itemkonstruktion zu entwickeln, die erwarten lassen, dass damit neue, Rasch homogene Items konstruiert werden können. Das ursprüngliche Itemdesign enthielt folgende neun Kriterien, nach denen jedes Item konstruiert werden sollte.

1. Bei der oberen wie auch bei der unteren, horizontalen Präsentation wurde darauf geachtet, dass die größere vor der kleineren Zahl steht.
2. Der Zahlenraum 1 bis 999 sollte eingehalten werden.
3. Es sollten keine gleichen Zahlen in der Angabe vorkommen.
4. Die mittlere, horizontale Präsentation sollte entweder eine Divisions- oder eine Multiplikationsaufgabe sein.
5. Der Multiplikator bzw. der Divisor sollte zwischen 2 und 99 liegen.

6. Es sollten keine gleichen Ziffern in der Einerstelle in Zwischenergebnissen und Angabe enthalten sein.
7. Bei keinen Zahlen sollte in der Einerstelle eine Null stehen, mit Ausnahme vom Endergebnis.
8. Das Ergebnis ist immer eine ganze Zahl.
9. Die Zwischenergebnisse sollten nicht in der Angabe enthalten sein.

Die Kriterien werden im folgenden Abschnitt anhand des Beispiels in Abbildung 4 veranschaulicht.

+	6	2	8				
			7	x	3	=	___________
-	5	4	1				

Abbildung 4: Ein Übungsbeispiel des MAT inklusive Zwischenergebnisse

Kriterium Eins ist im Beispiel erfüllt, da 6 größer als 2 und 5 größer als 4 ist. Das zweite Kriterium ist auch erfüllt, da alle enthaltenen Zahlen, größer als 0 und kleiner als 1000 sind. Ebenfalls ist Kriterium Drei erfüllt, da alle Zahlen der Aufgabe ungleich sind. Auch die Kriterien Vier und Fünf sind erfüllt, da im letzten Rechenschritt multipliziert werden muss und weil 3 größer als 1 und kleiner als 100 ist. Das sechste Kriterium ist erfüllt, da acht verschiedene Ziffern aus der Gesamtmenge (1, 2, 3, 4, 5, 6, 7, 8, 9) auf der Einerstelle stehen. In keiner der acht Zahlen im Beispielitem steht eine Null auf der Einerstelle. Somit ist Kriterium Sieben erfüllt. Das achte Kriterium ist erfüllt, weil 21, das Ergebnis des Beispiels, eine ganze Zahl ist, in Abgrenzung zu einer gebrochenen Zahl, wie z. B. 21,5. Das neunte Kriterium ist im echten Test erfüllt, da alle Zwischenergebnisse nicht sichtbar sind.

Im Anschluss an die Fertigstellung der, nach diesem Itemdesign und dem erweiterten Itemdesign von Gittler (2008) konstruierten, Items wurden diese auf Rasch Homogenität überprüft. Wie das geschieht, kann in Abschnitt 2.8.1 nachgelesen werden.
Für die vorliegende Arbeit wurden nicht extra neue Items konstruiert, sondern es lagen bereits fertige Items vor, die mir von meinem Betreuer übergeben wurden. Zur Verfügung standen 20 Items, die noch nicht auf Rasch Homogenität überprüft worden waren, und vier Items, sogenannte Linkitems, mit denen das bereits geschehen war. Es sollten zwei MAT-Versionen erstellt werden, also mussten je zehn unüberprüfte Items pro Version ausgewählt werden. Die vier Linkitems mussten in beiden Versionen vorkommen. Jede Version sollte also 14 Items lang sein. Linkitems

erfüllen hier den Zweck, die Schwierigkeiten der Items aus beiden Versionen miteinander verrechnen zu können. Ohne die Linkitems gäbe es keinen gemeinsamen Referenzpunkt für die Itemschwierigkeiten der beiden Versionen. Die Itemschwierigkeiten beider Versionen werden berechnet und sollen später miteinander vergleichbar sein. Das soll sowohl mit den Items der beiden hier verwendeten Versionen der Fall sein, als auch mit bereits überprüften Items aus vorherigen Untersuchungen. Die Linkitems gewährleisten also, dass die Itemschwierigkeiten sämtlicher untersuchter und noch zu untersuchender Items miteinander verrechenbar sind, sie „verlinken" sie miteinander. Weiter erlaubten es die Linkitems überhaupt, dass der Test in der Länge von 14 Items vorgegeben werden konnte. Ohne die Linkitems wäre der Test wesentlich länger gewesen, da die 20 Items alle in einer Version vorgegeben hätten müssen, um gemeinsame Verrechnung zu gewährleisten. So wäre der Test, aufgrund der Länge, für die TeilnehmerInnen unzumutbar gewesen. Weiter wären Erschöpfungseffekte gegen Ende der Testung wahrscheinlicher gewesen.
Die Linkitems wurden in beiden Versionen auf die gleichen Positionen gelegt, und zwar auf die Positionen 1, 5, 9 und 14.
Wie wurden jetzt also die verbliebenen Items auf die Versionen verteilt? Es wurden zunächst Itempaare mit intuitiv gleicher Schwierigkeit gebildet. Aus diesen Paaren sollte dann je ein Item auf die gleiche Position in jeweils einer Version gelegt werden. Bei den Überlegungen zu diesen und folgenden Itemschwierigkeiten wurde auf die Größe der Zahlen und Zwischenergebnisse, insbesondere der, die als Multiplikator oder Divisor benutzt wurden, geachtet. Divisionen wurden grundsätzlich als schwieriger eingeschätzt. Auch wurde darauf geachtet, ob Überträge bei den Additionen und Subtraktionen vorhanden waren und wenn ja, wie viele. Je mehr Überträge eine Aufgabe hatte, desto schwieriger wurde sie eingeschätzt. Ebenfalls eine Rolle spielte dabei die Nähe zur Zahl 10 oder einem Vielfachen davon. Es wurde argumentiert, dass, wieder speziell bei Punktrechnungen, aber auch bei den Strichrechnungen, eine Aufgabe umso leichter sei, je näher ein Multiplikator oder Divisor bei einem Vielfachen von zehn liegt. Im Anschluss an die Paarbildung wurden nun zunächst die intuitiv einfachsten Items und dann die schwierigeren Items platziert. Es lagen pro Version gleich viele Multiplikations-, wie Divisionsaufgaben vor, deswegen wurden diese in abwechselnder Weise platziert. Es wurde gemutmaßt, dass Personen vielleicht eine Präferenz für entweder Multiplikations- oder Divisionsaufgaben haben. Die abwechselnde Platzierung sollte so mögliche Frustrationen mit einem der beiden Aufgabentypen so gering wie möglich halten.

Schließlich wurden die Analogien aus der Sprachkompetenzanalyse vorgegeben. Hier wurde eine von meinem Betreuer zur Verfügung gestellte Forschungsversion eingesetzt. Diese weist die Form aus Abbildung 5 auf und wird wie folgt gelöst:

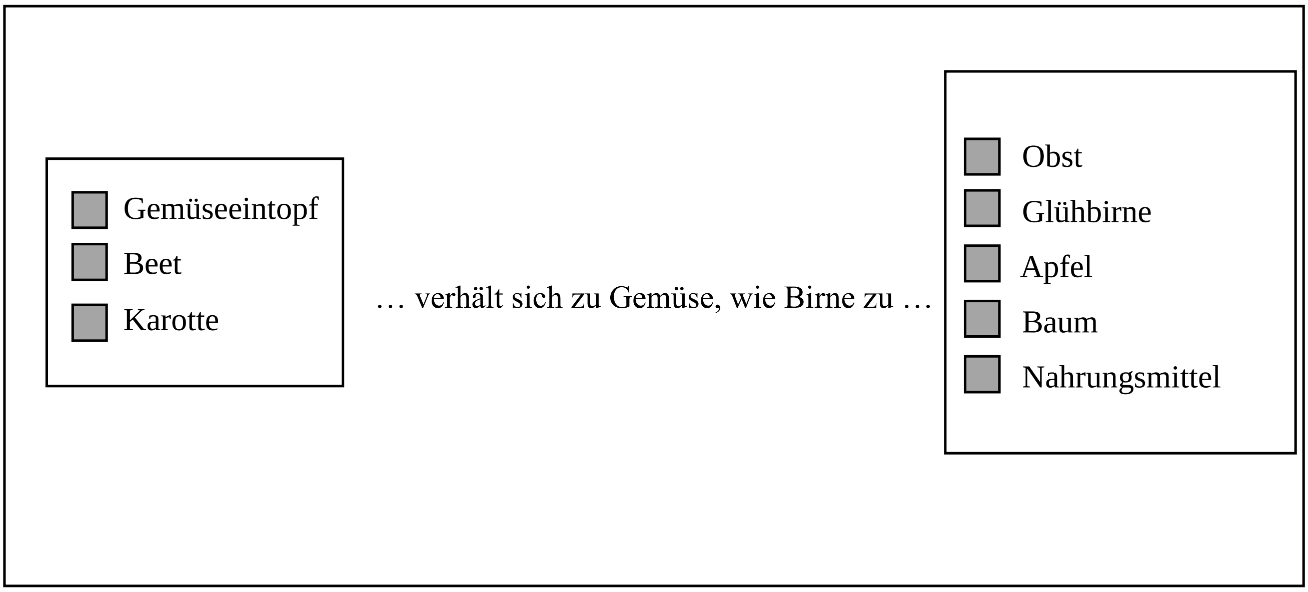


Abbildung 5: Ein Beispiel aus der Instruktion des verbalen Tests Analogien

Es muss bei jeder Aufgabe ein Satz korrekt, d. h. inhaltlich möglichst gut passend, vervollständigt werden. Es fehlen das erste und das letzte Wort. Jeder Satz enthält dabei zwei Beziehungen zwischen jeweils zwei Wörtern. Um eine Aufgabe korrekt zu lösen, muss im linken Auswahlkasten die richtige Antwort ausgewählt werden und ebenfalls im rechten Auswahlkasten. Dabei muss sich die Antwort im linken Kasten genauso zum linken Wort im Satzteil in der Mitte verhalten, wie das rechte Wort im Satzteil zu der ausgewählten Antwort im rechten Kasten. Im Beispiel verhält sich also die Karotte zu Gemüse wie eine Birne zu Obst. Die Beziehung ist dabei diejenige, dass die Karotte ein Teil der Kategorie Gemüse und die Birne ein Teil der Kategorie Obst ist. Die Birne könnte hier ebenfalls in die Kategorie der Nahrungsmittel aus dem rechten Kasten eingeordnet werden. Diese Antwort wäre trotzdem falsch, weil immer das ähnlichste Verhältnis ausgewählt werden soll. Es sind also verschiedene Antworten denkbar, jedoch wird immer nur eine als richtig gewertet.

3.2 Rekrutierung

Zunächst wurde festgelegt, dass die Stichprobe zum größten Teil aus StudentInnen, möglichst verschiedener Studienrichtungen, zusammengesetzt werden soll. Das Alter der StudienteilnehmerInnen wurde aufgrund der erhöhten Schwierigkeit gegenüber Gabriel (2004) auf 18 bis 60 begrenzt. Weiter war Deutsch als Muttersprache Grundvoraussetzung für die Teilnahme an der Studie. Dies wurde festgelegt, um sicherzustellen, dass alle TeilnehmerInnen die Instruktion gut verstehen. Ansonsten wurden keine weiteren, einschränkenden Ausschlusskriterien formuliert.

Die Stichprobenakquise teilte sich in zwei Abschnitte.
Der erste Teil der Rekrutierung begann Mitte Oktober und geschah durch Aussendungen von E-Mails mit allen relevanten Informationen und durch das Besuchen von Vorlesungen. Die Teilnahme an einer Verlosung von 40 € wurde als Incentive angeboten.
Aufgrund mangelnden Erfolgs wurde die Rekrutierungsstrategie gegen Mitte Dezember modifiziert. Diese bestand daraus, dass ich mich mit zwei Laptops in Aufenthaltsräume der nachfolgend gelisteten Institute setzte und StudentInnen um Mithilfe bat: Mathematik, Informatik, Physik, Chemie, Publizistik, Kommunikationswissenschaften, Wirtschaftswissenschaften (sowohl an der Universität Wien, als auch an der Wirtschaftsuniversität Wien), Psychologie, Kultur- und Sozialanthropologie und Lehramt von allen oben genannten Studienrichtungen, sofern es dieses gibt.

3.3 Design

Da es sich in dieser Arbeit um eine Testkonstruktionsstudie handelt und nicht um eine Überprüfung irgendeiner Art von Behandlung, wurde kein experimentelles Design mit Versuchs- und Kontrollgruppe, sowie Randomisierung und Prä- und Posttest gewählt. Stattdessen wurde, den zwei eingesetzten Versionen geschuldet, ein unvollständiges Datendesign (siehe Tabelle 1) gewählt. Das bedeutet, dass nicht jeder teilnehmenden Person jede, in der Testung eingesetzte Frage oder Aufgabe vorgegeben wurde. Tabelle 1 zeigt, an welcher Testposition jedes Item des MAT in der entsprechenden Version eingesetzt wurde. Weiter zeigt sie, welchen Personen welches Item vorgegeben wurde. Für Personen, die Version 1 (V1) bearbeitet haben, liegen Daten für Items 1 bis 14 vor, nicht jedoch für Items 15 bis 24 (Ein „+“ steht in Tabelle 1 für ein vorgegebenes Item, ein „-“ steht für ein nicht vorgegebenes Item). Für Personen, die Version 2 (V2) bearbeitet haben, liegen Daten für die Linkitems (Items 1 bis 4) und für die Items 15 bis 24 vor, nicht jedoch für die Items 5 bis 14. In der Datenmatrix scheinen also leere Datenfelder auf. Dies ist der Fall für die Daten zum MAT. Sonst wurden alle Fragen und Aufgaben allen TeilnehmerInnen vorgegeben. Das unvollständige Datendesign mit den vier Linkitems erlaubte es die Modellprüfungen für das Rasch-Modell für alle 24 Items gemeinsam durchzuführen. Dasselbe Design wurde für die restlichen vier Hypothesen verwendet.

Tabelle 1: Das unvollständige Datendesign

	Linkitems				Items (V1)										Items (V2)									
Item-nummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Test-position	1	5	9	14	2	3	4	6	7	8	10	11	12	13	2	3	4	6	7	8	10	11	12	13
Antwort (V1)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Antwort (V2)	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+

3.4 Statistische Analyse

Für alle Berechnungen zur Rasch Homogenität wird das, von meinem Betreuer zur Verfügung gestellte, Programm LpcM-Win 1.0 verwendet (Fischer & Ponocny-Seliger, 1998). Es wird die Maximum-Likelihood-Methode zur Schätzung der Itemparameter verwendet (für eine detaillierte Erklärung dieser Methode siehe Molenaar, 1995). Als Modelltest wird der Andersen Likelihood-Quotienten-Test verwendet (Andersen, 1973). Sollte dieser signifikant werden, so wird ein Wald-Test berechnet, um herauszufinden, welche Items besonders auffällig sind. Ein Martin-Löf-Test wird für die Überprüfung der Homogenität von Testteilen berechnet.

Für die zweite Hypothese werden die beiden Variablen „Rohscore des Mental Arithmetic Test“ und „Rohscore der Analogien“ mittels Kolmogorov-Smirnov-Test auf Normalverteilung überprüft. Sollte diese für beide gelten, so wird eine Pearson-Korrelation berechnet. Sollte die Normalverteilungsannahme nicht für beide Variablen gelten, so wird eine Spearman-Korrelation berechnet.

Für die dritte Hypothese werden alle 13 Fragen zur Matheaffinität analog zu Gabriel (2004) künstlich dichotomisiert, um diese als Gruppierungsvariablen benutzen zu können. Dabei werden alle Variablen in der Mitte der Skala, zwischen 3 und 4 getrennt, wobei die Ausprägungen 1-3 den Wert 1 zugewiesen bekommen und die Ausprägungen 4-6 den Wert 0. Der Wert 1 steht dabei für eine hohe Ausprägung und 0 für eine niedrige Ausprägung in der betreffenden Variable.

Anschließend wird der Rohscore des MAT auf Normalverteilung und Varianzenhomogenität geprüft. Die Gruppengrößen und Zusammensetzungen variieren dabei jeweils nach Gruppierungsvariable. Es werden hier der Kolmogorov-Smirnov-Test zur Überprüfung der Normalverteilungsannahme und der Levene-Test für die Überprüfung der Varianzenhomogenität eingesetzt. Sollten beide Voraussetzungen erfüllt sein, so wird ein t-Test für unabhängige Stichproben berechnet. Sollte mindestens eine der Voraussetzungen verletzt sein, so wird ein U-Test für unabhängige Stichproben berechnet.

Für die vierte Hypothese wird die Stichprobe nach dem Geschlecht aufgeteilt und anschließend werden beide Teilstichproben mittels Kolmogorov-Smirnov-Test auf Normalverteilung überprüft. Sollten beide Teilstichproben normalverteilt sein, wird ein t-Test berechnet, ist dies nicht der Fall wird ein U-Test berechnet.

Für die fünfte Hypothese wird die Stichprobe nach dem Alter aufgeteilt und anschließend werden beide Teilstichproben mittels Kolmogorov-Smirnov-Test auf Normalverteilung überprüft. Die Stichprobe wird dabei in Personen von 17 bis 21 Jahren und Personen von 22 bis 61 Jahren aufgeteilt. Sollten beide Teilstichproben normalverteilt sein, wird ein t-Test berechnet, ist dies nicht der Fall wird ein U-Test berechnet.

3.5 Hypothesen

Die Hypothesen der hier vorliegenden Arbeit heißen:

1. Die MAT-Items sind Rasch homogen.
2. Die Korrelation zwischen MAT-Leistungen und Analogien-Leistungen ist nicht größer als 0,35.
3. Die Personen, die sich selbst als matheaffiner einschätzen, schneiden im MAT signifikant besser ab, als die Personen, die sich matheferner einschätzen.
4. Männer schneiden im MAT signifikant besser ab, als Frauen.
5. Ältere und jüngere Personen unterscheiden sich nicht in ihrer MAT-Leistung.

Mit der ersten Hypothese wird die Rasch Homogenität der 20 neuen MAT-Items überprüft.

Die Ergebnisse der zweiten Hypothese dienen der Überprüfung der diskriminanten Validität der MAT-Items in Bezug auf den verbalen Test Analogien. Der Wert wurde gewählt, da Korrelationen die kleiner als 0,35 sind, bei Validitätsuntersuchungen von Leistungstests als nicht bedeutsam und als Hinweis auf diskriminante Validität angesehen werden können (Hagmann-von Arx, Petermann & Grob, 2013).

Die Ergebnisse der dritten Hypothese werden als erste Validitätshinweise der 20 MAT-Items interpretiert.

Die vierte Hypothese wurde aufgrund der vorhandenen Literatur so formuliert (Feingold, 1993,

Hyde et al., 1990, Krinzinger et al., 2012, Lynn & Irwing 2008). Sollte sie sich bestätigen kann dies als weiterer Hinweis dafür gelten, dass der MAT Kopfrechnen misst.
Die fünfte Hypothese wurde ebenfalls aufgrund der Literatur so formuliert (Allen et al., 1992, Salthouse & Kersten, 1993). Da es zum Alter allerdings auch gegenteilige Befunde gibt (Charness & Campbell, 1988) soll hier der Altersunterschied nochmal überprüft werden.

4 Ergebnisse

4.1 Datenaufbereitung

Es konnte nicht davon ausgegangen werden, dass sämtliche StudienteilnehmerInnen den Test instruktionsgemäß bearbeitet hatten. Das zeigte die Inspektion der Itembearbeitungszeiten. Einige dieser waren schlichtweg so gering, dass davon ausgegangen werden musste, dass sich einige TeilnehmerInnen durchgeklickt hatten. Dies wurde durch einige, seltene Berichte von TeilnehmerInnen im Voraus bereits bestätigt. Diese hatten gesagt, dass sie sich am Ende nur noch durchgeklickt hätten. Es wurden deswegen einige der 176 Personen von allen weiteren Analysen ausgeschlossen. Dies geschah mithilfe des Kriteriums der Itembearbeitungszeit. Es wurde ein willkürlich festgelegter Wert von 200 Einheiten der Systemzeit pro Item (ca. 11 Sekunden) festgelegt, ab dem gesagt wurde, dass hier nicht mehr von Durchklicken gesprochen werden kann. Beim Durchklicken wurde von den TeilnehmerInnen nicht mehr versucht den MAT ernsthaft zu bearbeiten, sondern sie klickten sich mittels Weiter-Button durch. Deswegen wurden sämtliche TeilnehmerInnen, die eine Bearbeitungszeit von elf Sekunden und geringer bei mindestens einem der Items aufwiesen, ausgeschlossen. Daraus resultierte eine Stichprobe von 160 Personen. Diese wurde für die zweite bis fünfte Hypothese verwendet. Für die Haupthypothese wurde die Stichprobe um weitere zehn Personen reduziert. Hierbei wurden alle Personen ausgeschlossen, die entweder alle Items oder kein einziges Item des MAT gelöst hatten. Das Rasch-Modell kann keine Personenparameter für solche Personen schätzen lassen, da die Information über die Fähigkeit nicht ausreicht. Sämtliche Analysen zur Rasch Homogenität wurden mit diesem verbliebenen Datensatz von 150 Personen gerechnet.

4.2 Stichprobenbeschreibung

Die Stichprobe bestand aus 110 (68,8%) männlichen und 50 (31,2%) weiblichen Personen. Der Altersdurchschnitt lag bei 23,29 Jahren bei einem Minimum von 17 und einem Maximum von 61 Jahren. Die Standardabweichung lag bei 6,73. Die höchste abgeschlossene Bildung war wie erwartet sehr homogen, mit 135 (84,4%) Personen, die entweder ihre Matura oder ihr Abitur abgeschlossen hatten. Die demographischen Daten der Stichprobe werden in Tabelle 2 illustriert. Abbildung 6 zeigt die Häufigkeitsverteilung des Alters.

Tabelle 2: Demographische Stichprobenbeschreibung

	n	%	$\overline{x}$	σ	Min	Max
Geschlecht	160	100				
Männlich	110	68,8				
Weiblich	50	31,2				
Alter	160	100	23,29	6,73	17	61
17 – 21	84	52,5	19,86	1,01	17	21
22 – 61	76	47,5	27,09	8,19	22	61
Bildung	160	100				
1	1	0,6				
2	5	3,1				
3	135	84,4				
4	6	3,8				
5	13	8,1				

1 = Hauptschule, AHS-Unterstufe; 2 = BMS (Fachschule), Lehre; 3 = Matura (AHS, BHS), Abitur; 4 = Fachhochschule, Akademie; 5 = Universität

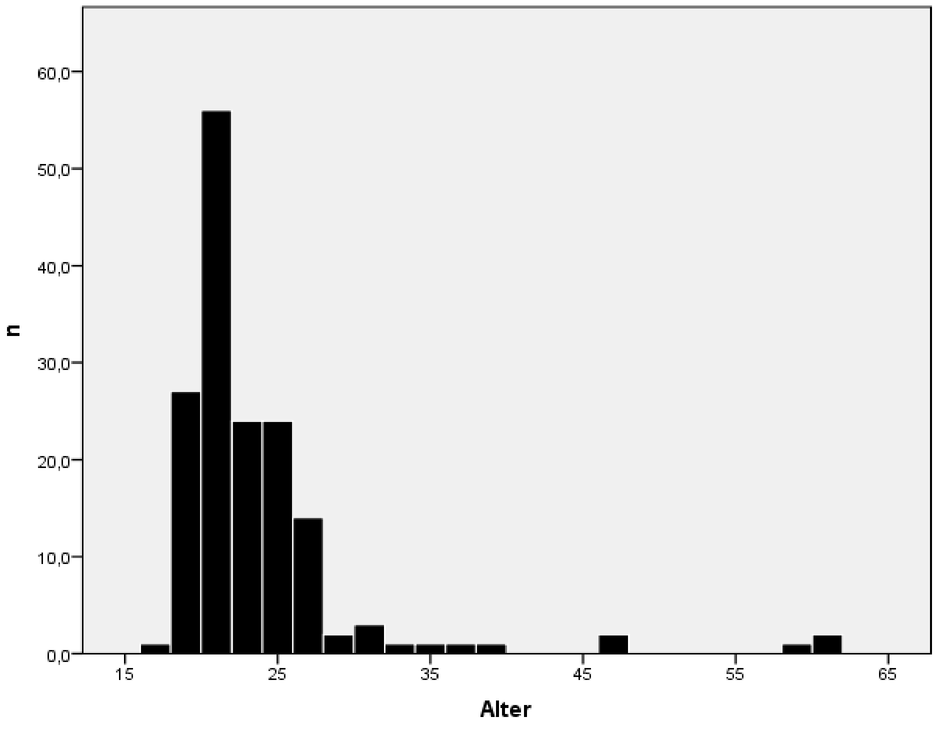


Abbildung 6: Häufigkeitsverteilung des Alters

4.3 Erste Hypothese: Rasch Homogenität

Für alle Modellprüfungen wurde ein Signifikanzniveau von 0,01 gewählt. Zur Überprüfung der Rasch Homogenität wurden die fünf Teilungskriterien Mittelwert des Rohscores, Geschlecht, Alter, Gesamtbearbeitungszeit und Version verwendet, wobei letzteres Teilungskriterium nur bei den Linkitems eingesetzt wurde.

Aufgrund des unvollständigen Datendesigns (siehe Abschnitt 3.3) war es möglich die Modellprüfungen für alle eingesetzten 24 Items aus den beiden Versionen gleichzeitig zu berechnen. Die Linkitems, die von allen TeilnehmerInnen beantwortet wurden, gewährleisteten eine gemeinsame Verrechnung aller Items.

Für das erste Teilungskriterium Rohscore wurde die Teilung durch das Programm LpcM vorgenommen. Die Itemparameter für Item 17 konnten nicht geschätzt werden, da nicht genug Information in den Subgruppen vorlag. Dieses wurde für die Analyse mit diesem Teilungskriterium ausgeschlossen. Daher ist k = 23 und df = k – 1 = 22. Der anschließende Likelihood-Ratio-Test (LRT) zeigte keine Signifikanz ($\chi^2 = 34{,}49$, $p = 0{,}04$, $df = 22$).

Bei dem Teilungskriterium Geschlecht wurde die Stichprobe nach dem, von den StudienteilnehmerInnen angegebenem Geschlecht aufgeteilt. Es konnten ebenfalls für ein Item keine Parameter geschätzt werden. Dieses war Item 23 und wurde von weiteren Analysen dieses Kriteriums ausgeschlossen. Der LRT zeigte wiederum keine Signifikanz ($\chi^2 = 10{,}94$, $p = 0{,}98$, $df = 22$), so dass mit dem nächsten Kriterium weiter gerechnet wurde.

Die Stichprobe wurde nun in jüngere (17 – 21 Jahre) und ältere Personen (22 – 61 Jahre) aufgeteilt. Alle Itemparameter konnten geschätzt werden und der LRT zeigte ebenfalls keine Signifikanz ($\chi^2 = 30{,}96$, $p = 0{,}12$, $df = 23$).

Schließlich wurde die Stichprobe nun in Personen, die eine niedrige Gesamtbearbeitungszeit und Personen, die hohe Gesamtbearbeitungszeit aufwiesen unterteilt. Dabei reicht die Bearbeitungszeit der schnellen Gruppe für alle MAT-Items von 12:36 bis 30:00 Minuten, die der langsamen Gruppe reicht von 30:01 bis 93:54 Minuten. Der LRT zeigte keine Signifikanz ($\chi^2 = 32{,}08$, $p = 0{,}10$, $df = 23$). Alle Itemparameter konnten berechnet werden.

In allen bisherigen Berechnungen waren Personen in jeweils beiden Subgruppen inkludiert, die entweder die erste Version oder die zweite Version bearbeitet hatten. Dies ist für das Teilungskriterium Version nicht mehr der Fall. Hier wären in der einen Substichprobe nur Personen, die die erste Version bearbeitet hatten, und in der anderen nur Personen, die die zweite Version bearbeitet hatten. Deswegen wurden für dieses Teilungskriterium nur die vier Linkitems verwendet, da die in beiden Versionen vorkamen. Der LRT wies auch hier keine Signifikanz auf ($\chi^2 = 5{,}44$, $p = 0{,}14$, $df = 3$). Tabelle 3 fasst die Ergebnisse zusammen.

Tabelle 3: Ergebnisse der Überprüfung der Rasch Homogenität für alle fünf Teilungskriterien

Teilungskriterium	k	n	LRT(χ^2)	p
Median	23	150	34,49	0,04
-- Lowscore		80		
-- Highscore		70		
Geschlecht	23	150	10,94	0,98
-- Weiblich		49		
-- Männlich		101		
Alter	24	150	30,96	0,12
-- jung		79		
-- alt		71		
Bearbeitungszeit	24	150	32,08	0,10
-- Schnell		75		
-- Langsam		75		
Version	4	150	5,44	0,14
-- 1		75		
-- 2		75		

Hieraus lässt sich schließen, dass sowohl die Linkitems, als auch die neuen Items Rasch homogen sind, sie also nachweislich eine Dimension messen. Es folgt daraus, dass alle in Abschnitt 2.8.1.4 beschriebenen Vorteile und Anwendungsmöglichkeiten auf diese Items zutreffen. Kleine Ausnahmen sind die Items 17 und 23, für die die Schätzung der Itemparameter in jeweils einem Teilungskriterium nicht gelang. Allerdings waren sie in drei von vier Teilungskriterien Rasch homogen.

Abschließend wurde ein Martin-Löf-Test (MLT) zur Überprüfung der Homogenität von Testteilen eingesetzt. Da dieser komplette Daten benötigt, wurde dieser wieder nur für die vier Linkitems berechnet. Die Linkitems wurden in zwei Gruppen unterteilt. Das Teilungskriterium war hier die Schwierigkeit der Items. Die erste Gruppe bestand aus den leichteren Items 1 und 3, die zweite Gruppe aus den schwierigeren Items 2 und 4. Der MLT zeigte keine Signifikanz (χ^2 = 0,68, p = 0,88, df = 3). Daraus folgt, dass die beiden Testteile homogen sind, sie also dieselbe Dimension messen.

In Abbildung 7 und Abbildung 8 sind die Personenparameter ξv in Abhängigkeit vom jeweils zugehörigen Rohwert abgebildet. Dies geschah mit den gesamten Daten, jeweils getrennt nach der ersten und zweiten Version. Für die Schätzung der Personenparameter wurde die Methode Weighted Likelihood Estimation (WLE, Warm, 1989) verwendet. Die Methode erlaubt es auch die Personenparameter für solche Personen zu schätzen, die entweder keins oder alle Items gelöst haben. Deswegen sind diese ebenfalls mit abgebildet. Man sieht, dass bei einer Zunahme von einem Rohwert die Zunahme im Personenparameter nicht immer gleichartig ist. In den Extrembereichen nimmt der Personenparameter pro Rohwert stärker zu, als im mittleren Bereich.

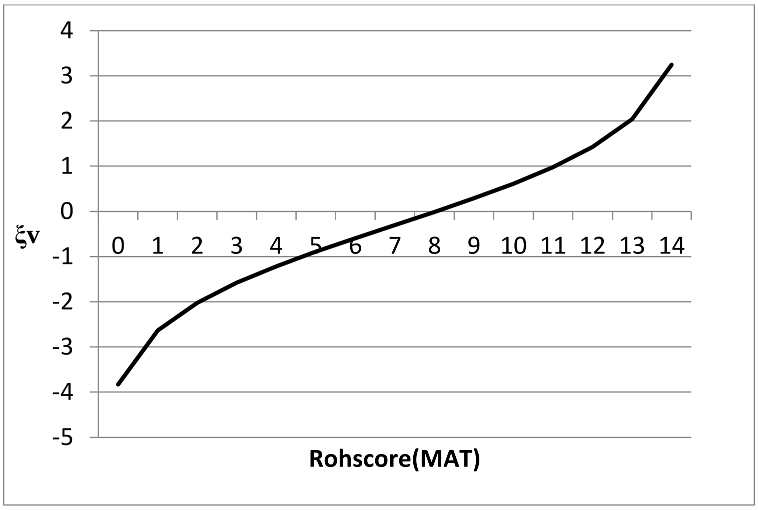


Abbildung 7: Personenparameter der Personen, die Version 1 bearbeitet haben

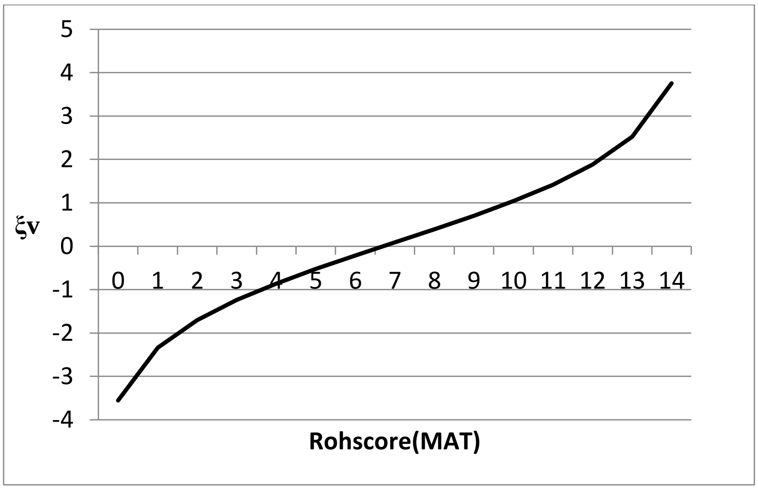


Abbildung 8: Personenparameter der Personen, die Version 2 bearbeitet haben

4.4 Zweite Hypothese: Diskriminante Validität

Beide Kolmogorov-Smirnov-Tests zeigten keine Normalverteilung der beiden Rohscores (p_{MAT} < 0,01, p_{ANA} < 0,001). Die Rohscoreverteilungen des MAT und des Tests Analogien sind in Abbildung 9 und Abbildung 10 abgebildet.

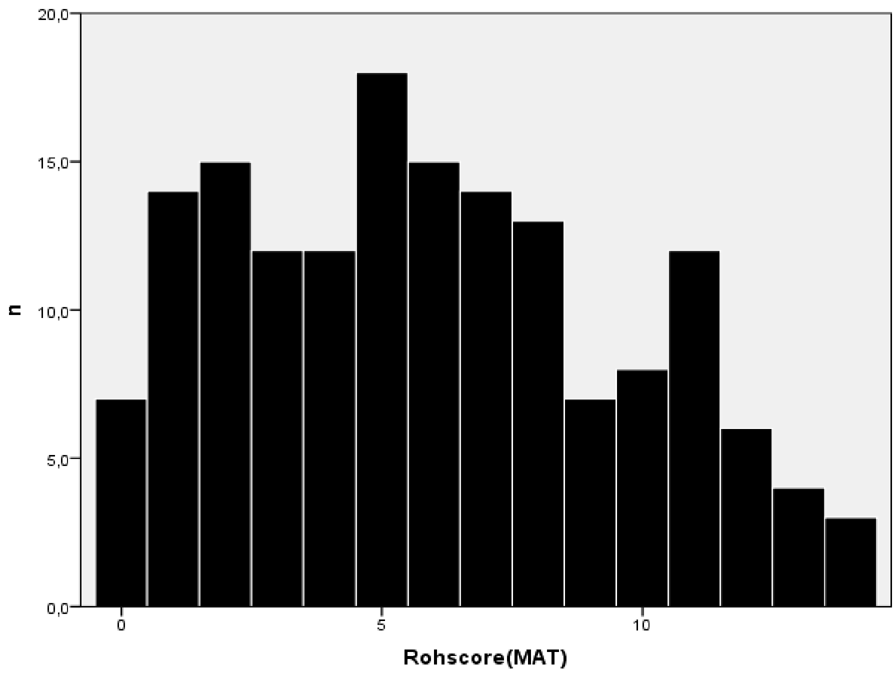


Abbildung 9: Rohscoreverteilung des MAT

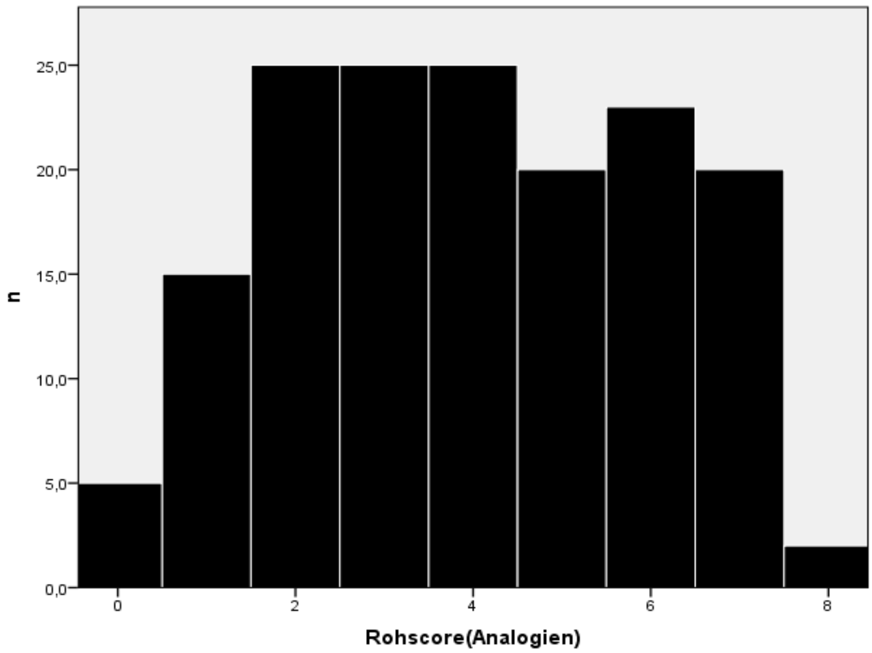


Abbildung 10: Rohscoreverteilung des Tests Analogien

Aufgrund der Verletzung der Normalverteilungsannahme wurde eine Spearman-Korrelation berechnet. Diese zeigte eine signifikante, positive Korrelation der beiden Variablen (r = 0,34, p < 0,001). Je besser also eine Person im MAT abschnitt, desto besser schnitt sie auch in den Analogien ab. Das Streudiagramm mit Regressionsgerade in Abbildung 11 veranschaulicht den

Zusammenhang. Der Zusammenhang zeigt, dass der MAT gegenüber dem Test Analogien diskriminant valide ist (Hagmann-von Arx et al., 2013).

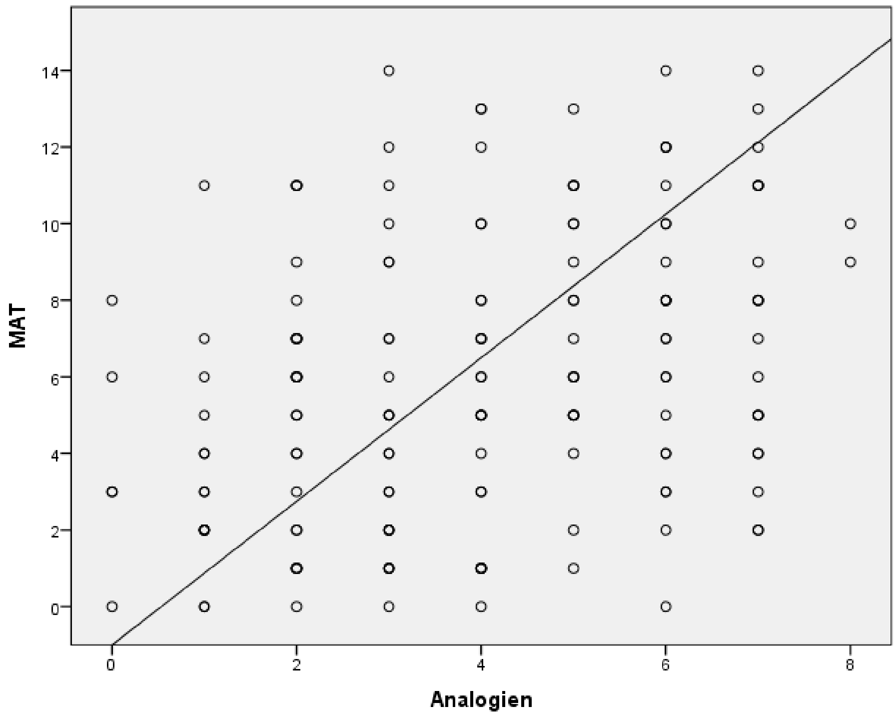


Abbildung 11: Streudiagramm mit Regressionsgerade für die Korrelation zwischen den Leistungen im MAT und den Leistungen im Test Analogien

4.5 Dritte Hypothese: Gruppenunterschiede als erste Validitätshinweise

Bei 8 von 13 der Variablen zur Matheaffinität wurde die Skala „sehr hoch = 1 2 3 4 5 6 = sehr gering" als Antwortformat gewählt. Bei allen Variablen, mit Ausnahme der Variable „Spielen des Logikrätsels Sudoku", kam immer die gleiche sechsstufige Skala zum Einsatz. Lediglich die verbalen Anker der Extremwerte lauteten anders. Diese sind bei den Ergebnissen zur jeweiligen Variable präsentiert. Bei der Variable „Spielen des Logikrätsels Sudoku" kamen die Antwortalternativen „Ja" und „Nein" zum Einsatz.

Bei sämtlichen der 13 Gruppenvergleiche, mit Ausnahme der Variable „Spielen des Logikrätsels Sudoku", wurde die sechsstufige Skala zwischen den Werten 3 und 4 geteilt. Den Ausprägungen 1, 2 und 3 wurde einheitlich der Wert 1 zugeteilt und den Ausprägungen 4, 5 und 6 wurde einheitlich der Wert 0 zugeteilt. Die Gruppe 1 berichtet somit von einer eher hohen Ausprägung auf der entsprechenden Eigenschaft und die Gruppe 0 berichtet von einer eher niedrigen Ausprägung. Die Variable „Spielen des Logikrätsels Sudoku" wurde nach den Antwortalternativen „Ja" und „Nein" aufgeteilt. Bei allen Gruppenvergleichen liegt eine Stichprobe von 160 Personen zu Grunde, mit Ausnahme des Gruppenvergleichs zur Häufigkeit des Spielens von Sudoku. Da nicht alle Personen angaben, Sudoku überhaupt zu spielen, fielen hier einige Personen weg. Für alle Auswertungen dieses Abschnitts wurde ein Signifikanzniveau von $\alpha = 0,05$ verwendet. Alle Normalverteilungsüberprüfungen geschahen mittels Kolmogorov-Smirnov-Test. Alle Überprüfungen der Varianzenhomogenität wurden mittels Levene-Test berechnet.

Spielen des Logikrätsels Sudoku:

Normalverteilung konnte nicht angenommen werden ($p < 0,01$), deswegen wurde ein U-Test berechnet. Dieser war hochsignifikant ($U = -3,57$, $p < 0,001$) bei mittleren Rangplätzen von $\bar{x}_1 = 92,29$ und $\bar{x}_0 = 66,09$. Personen, die angaben, Sudoku zu spielen, schnitten also signifikant besser im MAT ab, als Personen, die angaben, Sudoku nicht zu spielen.

Häufigkeit des Spielens von Sudoku:

Hier wurde die Skala „sehr häufig = 1 2 3 4 5 6 = sehr selten“ als Antwortformat benutzt.

Es stand lediglich eine Personenanzahl von 88 zur Verfügung, da nicht alle TeilnehmerInnen angaben, Sudoku zu spielen. Normalverteilung konnte angenommen werden ($p = 0,99$), ebenfalls die Homogenität der Varianzen ($p = 0,56$). Der t-Test zeigte keine Signifikanz ($t = -0,003$, $df = 86$, $p = 0,99$) bei gleichen Mittelwerten ($\bar{x}_1 = 6,89$ und $\bar{x}_0 = 6,89$). Das bedeutet, dass es keinen Unterschied zwischen denjenigen Personen gab, die angaben, häufig bzw. selten Sudoku zu spielen.

Spaß am Umgang mit Zahlen und Mengen:

Hier wurde die Skala „sehr viel = 1 2 3 4 5 6 = sehr wenig“ als Antwortformat benutzt.

Die Überprüfung der Normalverteilung zeigte keine Signifikanz ($p = 0,58$). Das gleiche Ergebnis brachte die Überprüfung der Varianzenhomogenität ($p = 0,24$). Der anschließende t-Test zeigte keine Signifikanz ($t = -1,07$, $df = 158$, $p = 0,29$) bei Mittelwerten von $\bar{x}_1 = 6,13$ und $\bar{x}_0 = 5,40$. Es gab also keinen signifikanten Unterschied zwischen Personen, die berichteten viel bzw. wenig Spaß am Umgang mit Zahlen und Mengen zu haben.

Fähigkeit im Umgang mit Zahlen und Mengen:

Da die Normalverteilungsannahme verletzt war ($p = 0,03$) wurde ein U-Test berechnet. Dieser offenbarte, dass Menschen mit höherer, eingeschätzter Fähigkeit signifikant bessere Leistungen im MAT zeigten als diejenigen mit niedrigerer, eingeschätzter Fähigkeit ($U = -3,07$, $p < 0,01$, $\bar{x}_1 = 88,28$ und $\bar{x}_0 = 64,34$).

Spaß, Rechenaufgaben im Kopf zu lösen:

Hier wurde ebenfalls die Skala „sehr viel = 1 2 3 4 5 6 = sehr wenig“ als Antwortformat benutzt.

Der Kolmogorov-Smirnov-Test zeigte keine signifikante Abweichung von der Normalverteilung ($p = 0,06$). Auch der Levene-Test zeigte keine signifikanten Inhomogenitäten der Varianzen ($p = 0,13$). Deswegen wurde ein t-Test gerechnet. Dieser zeigte, dass sich Personen, die mehr Spaß daran haben, Rechenaufgaben im Kopf zu lösen, signifikant mehr Items des MAT lösen konnten, als Personen, die weniger Spaß daran haben ($t = -2,83$, $df = 158$, $p = 0,01$, $\bar{x}_1 = 6,60$, $\bar{x}_0 = 4,95$).

Fähigkeit, Rechenaufgaben im Kopf zu lösen:

Keine Signifikanz zeigte sich bei der Überprüfung der Normalverteilung (p = 0,17). Ebenfalls keine Signifikanz zeigte der Levene-Test (p = 0,59). Der anschließend berechnete t-Test zeigte Signifikanz (t = -2,14, df = 158, p = 0,03). Die Gruppenmittelwerte lagen bei $\bar{x}_1$ = 6,44 und $\bar{x}_0$ = 5,17. Personen, die angaben, eine höhere Fähigkeit im Kopfrechnen zu haben, schnitten auch besser im MAT ab, als Personen, die von sich behaupteten, sie wären eher schwach im Kopfrechnen.

Mathematisches Faktenwissen:

Die Überprüfung der Normalverteilung zeigte einen nicht signifikanten Wert von p = 0,20. Varianzenhomogenität war ebenfalls gegeben (p = 0,56). Der t-Test zeigte signifikante Unterschiede zwischen TeilnehmerInnen, die ein höheres mathematisches Faktenwissen angaben, und jenen, die ein niedrigeres Faktenwissen angaben (t = -2,02, df = 158, p = 0,045). Die Gruppenmittelwerte zeigten, dass Erstgenannte besser im MAT abschnitten ($\bar{x}_1$ = 6,33, $\bar{x}_0$ = 5,06).

Wissen um mathematische Vorgehensweisen:

Die Normalverteilungsannahme konnte bestätigt werden (p = 0,23). Die Annahme der Homogenität der Varianzen ebenfalls (p = 0,82). Der t-Test zeigte, dass die beiden Gruppen der Personen, die sich als mehr wissend und weniger wissend, in Bezug auf mathematische Vorgehensweisen, eingeschätzt hatten, sich nicht signifikant voneinander unterscheiden (t = -1,37, df = 158, p = 0,17, $\bar{x}_1$ = 6,21, $\bar{x}_0$ = 5,33).

Fähigkeit zur Abschätzung von Größenordnungen:

Die Überprüfung der Normalverteilung ergab deren Annahme (p = 0,65). Die Überprüfung der Varianzenhomogenität brachte das gleiche Ergebnis (p = 0,60). Der t-Test brachte keinen signifikanten Unterschied zum Vorschein (t = -0,60, df = 158, p = 0,55). Bei Mittelwerten von $\bar{x}_1$ = 6,08 und $\bar{x}_0$ = 5,72 unterscheiden sich die TeilnehmerInnen, die sich als fähiger im Abschätzen von Größenordnungen eingeschätzt haben nicht signifikant von denen, die sich als weniger fähig eingeschätzt haben.

Merkfähigkeit in Bezug auf Zahlen:

Normalverteilung konnte angenommen werden (p = 0,14). Ebenso die Homogenität der Varianzen (p = 0,51). Der t-Test zeigte, dass sich die Gruppenmittelwerte signifikant voneinander unterscheiden (t = -2,75, df = 158, p < 0,01). Die Mittelwerte ($\bar{x}_1$ = 6,61, $\bar{x}_0$ = 5,01) zeigten weiter, dass die TeilnehmerInnen, die angaben, sich Zahlen gut merken zu können, einen höheren Mittelwert hatten, als die TeilnehmerInnen, die angaben, sich Zahlen eher schlecht merken zu können.

Allgemeine Konzentrationsfähigkeit:

Der Kolmogorov-Smirnov-Test erlaubte eine Annahme der Normalverteilung (*p* = 0,73). Der Levene-Test erlaubte die Annahme der Varianzenhomogenität (*p* = 0,38). Der anschließende t-Test zeigte keine Signifikanz (*t* = -1,24, *df* = 158, *p* = 0,22). Die Gruppen unterscheiden sich also nicht signifikant voneinander. Die Gruppe derjenigen, die sich als allgemein konzentrationsfähiger einschätzten hatten einen Gruppenmittelwert von $\bar{x}_1$ = 6,15, die Personen der Gruppe, die sich als weniger konzentrationsfähig einschätzten, einen Wert von $\bar{x}_0$ = 5,31.

Konzentrationsfähigkeit in Bezug auf mathematische Aufgaben:

Die Normalverteilungsannahme musste verworfen werden (*p* < 0,01). Deswegen wurde ein U-Test berechnet. Dieser zeigte hohe Signifikanz (*U* = -3,37, *p* < 0,01). Die mittleren Rangplätze ($\bar{x}_1$ = 88,31 und $\bar{x}_0$ = 61,14) zeigten, dass die Personen, die sich eine hohe Konzentrationsfähigkeit in Bezug auf mathematische Aufgaben zuschrieben, auch eine bessere Leistung im MAT zeigten, als die Personen, die sich darin eher schwach einschätzten.

Leistung im Fach Mathematik:

Schließlich wurde hier die Skala „sehr gut = 1 2 3 4 5 6 = sehr schlecht" als Antwortformat benutzt. Auch hier wurde die Normalverteilungsannahme verworfen (*p* = 0,01). Der U-Test zeigte abermals hohe Signifikanz (*U* = -3,80, *p* < 0,001). Und abermals haben diejenigen, die ihre Leistung im Fach Mathematik als besser einschätzten, ebenfalls die bessere Leistung im MAT als diejenigen, die ihre Leistung als eher weniger gut einschätzten ($\bar{x}_1$ = 88,12 und $\bar{x}_0$ = 55,18).

Tabelle 4 fasst die Ergebnisse zusammen.

Tabelle 4: Ergebnisse der Gruppenvergleiche

		n	%	$\bar{x}$ (MAT)	σ	p
Spielen des Logikrätsel Sudoku^{u}		160	100	5,94	3,70	
	1	88	55,0	6,89	3,58	
	0	72	45,0	4,79	3,54	< ,001*
Häufigkeit des Spielens von Sudoku^{t}		88	55,0	6,89	3,58	
	1	18	11,3	6,89	3,92	
	0	70	43,8	6,89	3,51	0,99 n.s.

Spaß am Umgang mit Zahlen und Mengen[t]		160	100	5,94	3,70		
	1	120	75,0	6,13	3,78		
	0	40	25,0	5,40	3,44	0,29	n.s.
Fähigkeit im Umgang mit Zahlen und Mengen[u]		160	100	5,94	3,70		
	1	108	67,5	6,56	3,70		
	0	52	32,5	4,65	3,38	< ,01*	
Spaß, Rechenaufgaben im Kopf zu lösen[t]		160	100	5,94	3,70		
	1	96	60,0	6,60	3,81		
	0	64	40,0	4,95	3,31	,01*	
Fähigkeit, Rechenaufgaben im Kopf zu lösen[t]		160	100	5,94	3,70		
	1	97	60,6	6,44	3,71		
	0	63	39,4	5,17	3,59	,03*	
Mathematisches Faktenwissen[t]		160	100	5,94	3,70		
	1	111	69,4	6,33	3,69		
	0	49	30,6	5,06	3,61	,045*	
Wissen um mathematische Vorgehensweisen[t]		160	100	5,94	3,70		
	1	112	70,0	6,21	3,69		
	0	48	30,0	5,33	3,69	0,17	n.s.
Fähigkeit zur Abschätzung von Größenordnungen[t]		160	100	5,94	3,70		
	1	99	61,9	6,08	3,74		
	0	61	38,1	5,72	3,65	0,55	n.s.
Fähigkeit zum Zahlen merken[t]		160	100	5,94	3,70		
	1	93	58,1	6,61	3,67		
	0	67	41,9	5,01	3,56	,01*	

Allgemeine Konzentrationsfähigkeit[t]		160	100	5,94	3,70		
	1	121	75,6	6,15	3,75		
	0	39	24,4	5,31	3,53	0,22	n.s.
Konzentrationsfähigkeit in Bezug auf mathematische Aufgaben[u]		160	100	5,94	3,70		
	1	114	71,2	6,55	3,72		
	0	46	28,8	4,43	3,22	< ,01*	
Leistung im Fach Mathematik[u]		160	100	5,94	3,70		
	1	123	76,9	6,55	3,67		
	0	37	23,1	3,92	3,04	< ,001*	

[t] = Für die betreffende Variable wurde ein t-Test berechnet; [u] = Für die betreffende Variable wurde ein U-Test berechnet

Eine explorative Faktorenanalyse wurde berechnet um die Ergebnisse der Validitätshinweise zusammenzufassen. Für diese wurden 11 der 13 Variablen zur Matheaffinität in die Analyse aufgenommen. Die Variable „Spielen des Logikrätsels Sudoku“ wurde nicht in die Analyse aufgenommen, da diese dichotome Daten lieferte. Die Variable „Häufigkeit des Spielens von Sudoku“ wurde nicht aufgenommen, weil für diese Variable nur 88 Datensätze vorlagen.
Mit den verbliebenen Variablen wurde eine erste Hauptachsenanalyse (Bühner, 2011, Stewart, Barnes, Cote, Cudeck & Malthouse, 2001) durchgeführt. Diese lieferte drei Faktoren nach dem Kaiser Kriterium (siehe Tabelle 5).

Tabelle 5: Die drei Faktoren der ersten Hauptachsenanalyse inklusive Ladungen

	Faktor		
	1	2	3
Spaß am Umgang mit Zahlen und Mengen	**,750**	-,282	-,281
Fähigkeit im Umgang mit Zahlen und Mengen	**,807**	-,108	,011
Spaß, Rechenaufgaben im Kopf zu lösen	**,787**	-,359	,039
Fähigkeit, Rechenaufgaben im Kopf zu lösen	**,697**	-,112	,179
Mathematisches Faktenwissen	**,625**	,027	,298
Wissen um mathematische Vorgehensweisen	**,672**	-,067	,116
Fähigkeit zur Abschätzung von Größenordnungen	**,563**	-,052	,294
Merkfähigkeit in Bezug auf Zahlen	**,427**	,240	,132
Allgemeine Konzentrationsfähigkeit	**,480**	**,685**	,031
Konzentrationsfähigkeit in Bezug auf mathematische Aufgaben	**,799**	,341	-,211
Leistung im Fach Mathematik	**,582**	,020	**-,491**

Der erste Faktor zeigte einen Eigenwert von 5,25, der zweite einen von 1,23 und der Eigenwert des dritten Faktors lag mit 1,05 knapp über dem Kriterium. Nach dem Scree-Test würde ein Faktor genügen, da die Beugung der Kurve deutlich beim zweiten Faktor beginnt. Die drei Faktoren erklären gemeinsam einen Varianzanteil von 57,78 %. Nur Faktorladungen, die größer als 0,4 sind werden hier als bedeutsam interpretiert (Field, 2009). Sämtliche Variablen luden bedeutsam auf dem ersten Faktor. Gleichzeitig lud nur jeweils eine Variable auf den beiden anderen Faktoren. Nach Rotation nach dem Varimax Kriterium konnte die Interpretierbarkeit nicht verbessert werden. Aufgrund des Scree-Test Kriteriums und aufgrund dessen, dass jeweils nur eine Variable auf den beiden anderen Faktoren bedeutsam lud, wurde eine weitere Hauptachsenanalyse berechnet. Hierbei wurde nun nur ein Faktor extrahiert. Zusätzlich dazu wurde die Variable „Allgemeine Konzentrationsfähigkeit“ aus der Analyse ausgeschlossen, weil sie die einzige Variable war, die inhaltlich nichts mit Mathematik zu tun hatte und weil sie als einzige Variable nicht am höchsten auf dem ersten Faktor lud. Die Faktorladungen zeigt Tabelle 6. Der Eigenwert des Faktors lag bei

5,05 und der erklärte Varianzanteil lag bei 45,57 %. Wieder laden alle Variablen bedeutsam auf dem Faktor. Der Faktor wird als Matheaffinität interpretiert.

Tabelle 6: Der Faktor der zweiten Hauptachsenanalyse inklusive Ladungen

	Faktor
	1
Spaß am Umgang mit Zahlen und Mengen	**,751**
Fähigkeit im Umgang mit Zahlen und Mengen	**,822**
Spaß, Rechenaufgaben im Kopf zu lösen	**,802**
Fähigkeit, Rechenaufgaben im Kopf zu lösen	**,704**
Mathematisches Faktenwissen	**,614**
Wissen um mathematische Vorgehensweisen	**,690**
Fähigkeit zur Abschätzung von Größenordnungen	**,564**
Merkfähigkeit in Bezug auf Zahlen	**,401**
Konzentrationsfähigkeit in Bezug auf mathematische Aufgaben	**,735**
Leistung im Fach Mathematik	**,553**

4.6 Vierte Hypothese: Geschlechtsunterschied im Kopfrechnen

Die Überprüfung der Normalverteilung zeigte Signifikanz (p = 0,04). Deswegen wurde ein U-Test berechnet. Dieser war hochsignifikant (U = -2,67, p < 0,01) bei mittleren Rangplätzen von $\bar{x}_1$ = 66,02 und $\bar{x}_0$ = 87,08. Männliche Teilnehmer schnitten im MAT also besser ab, als weibliche Teilnehmer.

4.7 Fünfte Hypothese: Altersunterschied im Kopfrechnen

Die Normalverteilungsannahme musste verworfen werden (p = 0,04). Deswegen wurde ein U-Test berechnet. Dieser war signifikant (U = -2,48, p = 0,01, $\bar{x}_1$ = 70,97 und $\bar{x}_0$ = 89,12) und zeigte, dass jüngere TeilnehmerInnen im MAT besser abschnitten, als ältere TeilnehmerInnen.

5 Zusammenfassung und Diskussion der Ergebnisse

Die Überprüfung der 20 Mental Arithmetic Test-Items kann als gelungen angesehen werden. Alle Items waren in allen fünf Teilungskriterien Rasch homogen. Lediglich für Item 17 beim Teilungskriterium Mittelwert und für Item 23 beim Teilungskriterium Geschlecht konnten keine Itemparameter aufgrund mangelnder Information geschätzt werden. Dieser Umstand spricht dabei in keinster Weise gegen die Rasch Homogenität dieser Items. Im Gegenteil, es kann vermutet werden, dass diese auch in den betreffenden Teilungskriterien Rasch homogen sind, weil sie in allen anderen Teilungskriterien Rasch homogen waren.
Auch sind nicht schätzbare Itemparameter keineswegs etwas Besonderes. Bei Gabriel (2004) konnte ebenfalls ein Itemparameter aufgrund mangelnder Information nicht geschätzt werden.
Zur Erklärung des Umstands, dass hier alle Items sofort Rasch Homogenität aufwiesen sei hier angeführt, dass die Items des MAT seit Gabriel (2004), Gittler (2008) und Franz (2009) intensiv analysiert und verbessert wurden. Mehr noch, das Itemdesign von Gittler (2008), welches das Ziel hatte Rasch homogene Items zu produzieren lieferte hier ausschließlich Rasch homogene Items. Somit kann das Ziel des Itemdesigns als erfüllt angesehen werden.
Die Überprüfung der diskriminanten Validität kann ebenfalls als gelungen angesehen werden. Mit einer Korrelation von 0,34 zwischen den Leistungen im MAT und den Leistungen in den Analogien wird lediglich ein gemeinsamer Varianzanteil von 11,56 % erklärt. Dieser gemeinsame Varianzanteil kann durch generelle Ähnlichkeiten in Leistungstests erklärt werden. Sowohl der MAT als auch der Test Analogien fordern Leistungen des Gedächtnis, der Lernfähigkeit, der Aufmerksamkeit und Motivation. Dass die Korrelation signifikant ausfiel sagt nur über deren Generalisierbarkeit auf die dahinter stehende Population etwas aus, nicht jedoch über die Bedeutung und Größe des Effekts. Dieser wird mithilfe der Effektgröße interpretiert. Im Falle einer Korrelation ist die Effektgröße gleich dem Korrelationskoeffizienten r, hier also 0,34. Obwohl bei Korrelationen eine Effektgröße ab 0,3 als moderat einzustufen ist (Bortz & Döring, 2006, S.606), gilt eine Korrelation von 0,34 bei Validitätsuntersuchungen als Hinweis auf diskriminante Validität (Hagmann-von Arx et al., 2013). Inhaltlich bedeutet die Korrelation, dass ein schwacher Zusammenhang in der Richtung besteht, dass je besser eine Person im MAT abschnitt, desto besser sie auch in den Analogien abschnitt.

Die ersten Validitätshinweise durch Gruppenunterschiede können ebenfalls als bestätigt angesehen werden. Acht von 13 Gruppenunterschieden waren signifikant und bei jedem Gruppenunterschied hatte die Gruppe, die sich matheaffiner eingeschätzt hatte, den höheren Mittelwert im MAT. Das bedeutet, dass die Personen, die sich als matheaffiner eingeschätzt hatten auch besser im MAT abschneiden, als die Personen, die sich als eher mathefern eingeschätzt hatten. Ausnahme ist da lediglich die Frage nach der Häufigkeit des Sudoku-Spielens. Hier waren die Mittelwerte gleich groß. Die explorative Faktorenanalyse mit 10 der 13 Variablen zur Matheaffinität lieferte einen einzigen Faktor, auf dem alle Variablen bedeutsam luden. Der Faktor wird als Matheaffinität interpretiert. Dabei konnten sieben der Variablen zwischen fähigen und weniger fähigen KopfrechnerInnen zu diskriminieren.

Der berechnete Geschlechtsunterschied lieferte das erwartete Ergebnis, dass Männer im Kopfrechnen besser abschneiden, als Frauen. Dieses Ergebnis ist konform mit der Literatur zum Kopfrechnen (Feingold, 1993, Hyde et al., 1990, Krinzinger et al., 2012, Lynn & Irwing 2008) und wird als Hinweis auf Validität des MAT interpretiert.

Der berechnete Altersunterschied lieferte das Ergebnis, dass jüngere Personen besser im Kopfrechnen sind, als ältere Personen. Dieses Ergebnis ist in keinem der Ergebnisse der Literatur (Allen et al., 1992, Salthouse & Kersten, 1993, Charness & Campbell, 1988) zu finden. Hier gab es entweder keinen Unterschied oder aber ältere Personen schnitten besser ab. Allerdings war das Aufgabenmaterial in den Studien entweder sehr einfach (z. B. einfache Multiplikationsaufgaben, bei Allen et al., 1992) oder es enthielt die Rechenoperation „Quadrieren" (Charness & Campbell, 1988). Die Anforderungen des MAT haben mit den Anforderungen dieser Aufgaben wenig gemein. Geht es bei der Lösung von einfachen Multiplikationen nur um Faktenabruf, wird beim MAT die Arbeitsgedächtnisleistung viel stärker in Anspruch genommen. Weiter wurden in den drei Studien wesentlich distinktere Altersgruppen untersucht als hier (z. B. Personen, die zwischen 18 und 27 Jahren alt waren vs. Personen, die zwischen 60 und 80 Jahren alt waren, bei Salthouse & Kersten, 1993). Hier wurden Personen, die zwischen 17 und 21 Jahren alt waren mit Personen, die zwischen 22 und 61 Jahren alt waren verglichen. Zusätzlich gab es kaum TeilnehmerInnen, die älter als 30 Jahre alt waren. Insofern sind die Studien nicht wirklich vergleichbar. Um einen Altersunterschied in den Anforderungen, die der MAT misst, herauszufinden, sollten mindestens zwei distinkte Altersgruppen untersucht werden. Nichtsdestotrotz bleibt das Ergebnis ein erster Hinweis darauf, dass jüngere Personen besser im „anspruchsvollen" Kopfrechnen sind, als ältere Personen.

Die stärkste Limitation der Arbeit ist, neben der geringen Stichprobengröße, die Art und Weise der Rekrutierung. Diese fand notgedrungen zweigeteilt statt. Im ersten Teil wurde der MAT in ruhigen Räumen appliziert, im zweiten Teil in Aufenthaltsräumen für Studierende. In diesen ist es natürlich weniger ruhig, da auch gesprochen werden darf. Jedoch sprechen die Ergebnisse stark dafür, dass dieser Unterschied in der Testsituation keinen bzw. einen sehr geringen Effekt auf die Leistungen im MAT ausübte. Die Rasch Homogenität der Items spricht dafür, dass eine Fähigkeit für verschiedene Teilstichproben gemessen wird und somit der Test für die Personen fair gewesen ist. Durch die etwas lautere Umgebung scheint also kein Nachteil entstanden zu sein. Dafür spricht ebenfalls, dass lediglich drei Personen ihren Unmut über die zu laute Umgebung äußerten. Für den Rest der Stichprobe schien dies kein limitierendes Problem zu sein.

Jedoch kann diese Art der Rekrutierung durchaus zu verfälschten Ergebnissen im Test Analogien geführt haben. Obwohl dieser ebenfalls Rasch homogen ist, wurde dessen Rasch Homogenität für die hier erhobene Stichprobe nicht überprüft. Andererseits ist es fraglich, weshalb die Art der Rekrutierung beim MAT zu Rasch Homogenität führen sollte und gleichzeitig bei einem bereits Rasch homogenen Test nicht.

Schließlich seien hier nun noch zwei Änderungsmöglichkeiten für den MAT und eine für die Analogien besprochen.

Manchmal fragten TeilnehmerInnen bereits bei der ersten Instruktionsseite des MAT, wie sie denn den MAT bearbeiten sollen, ohne auf den „Weiter-Button“ zu klicken. Sie versuchten sich den Lösungsweg selbst herzuleiten, da Erklärungen auf der Seite fehlten. Um dies zu vermeiden würde ein Hinweis auf der ersten Seite genügen, der darauf verweist, dass der Lösungsweg auf der nächsten Seite erklärt wird. Statt „Jede Aufgabe hat vier Rechenschritte, die nun für das Beispiel vorgeführt werden" könnte z. B. „Jede Aufgabe hat vier Rechenschritte, die nun für das Beispiel auf der nächsten Seite vorgeführt werden" auf der ersten Seite stehen.

Weiter kamen einige TeilnehmerInnen beim dritten Übungsitem nicht weiter. Manchmal fragten sie erst nach ca. 15 Minuten nach, wie sie denn weiterkommen würden. In diesem Fall kostete das den TeilnehmerInnen viele kognitive Ressourcen und es frustrierte sie eventuell, da sie nicht weiterkamen. Hier könnte eine Feedbackschleife eingebaut werden. Diese würde den TeilnehmerInnen, nach einer gewissen Anzahl von Fehlversuchen, rückmelden, wie das Item korrekt zu lösen ist. So ist sichergestellt, dass sich weder emotionale Frustration noch kognitive Erschöpfung auf die Testbearbeitung auswirken. Der zentrale Vorteil liegt hier in der Standardisierung der Testsituation. Durch die Feedbackschleife ist sichergestellt, dass alle TeilnehmerInnen mit vergleichbaren Voraussetzungen in die eigentliche Testsituation einsteigen.

Bei den Analogien verwirrte die erste Instruktionsseite einige TeilnehmerInnen. Dort steht als zweiter Satz: „Wählen Sie dazu das unterste Wort aus der Wortliste auf der linken Seite und das oberste Wort aus der Wortliste auf der rechten Seite – so wie es unten gezeigt ist“. Dieser Satz kann dazu führen, dass vermutet wird, dass dieses Verhalten für jede Aufgabe immer gleich ausgeführt werden muss, obwohl das offensichtlich den Sinn einer Testung unterminieren würde. Zum Beispiel könnte dort stattdessen „Wählen Sie dazu ein Wort aus der Wortliste auf der linken Seite und ein

Wort aus der Wortliste auf der rechten Seite – so wie es unten gezeigt ist“ stehen. So könnte unnötige Verwirrung und möglicherweise die Falschbearbeitung des Tests vermieden werden. Diese Änderungsmöglichkeiten könnten zur Optimierung des MAT und den Analogien beitragen und so deren zukünftige Applikation vorantreiben.

Für diesen Zweck werden wohl weitere MAT-Items auf Rasch Homogenität überprüft werden müssen, um so einen kompletten Itempool für den MAT zu generieren. Geschieht dies mit weiteren Items so könnten Items 17 und 23 als Linkitems dienen, um so diese noch einmal auf Rasch Homogenität zu überprüfen. Item 17 sollte dabei auf Rasch Homogenität im Teilungskriterium Mittelwert und Item 23 auf Rasch Homogenität im Teilungskriterium Geschlecht überprüft werden. Weisen beide Items dann Rasch Homogenität auf, so kann auch dessen Überprüfung als abgeschlossen angesehen werden.
Zwei der 176 StudienteilnehmerInnen lösten alle 14 Items des MAT. Das lässt darauf schließen, dass der Test für diese zwei Personen zu leicht war. Das wiederum lässt darauf schließen, dass das obere Ende des Schwierigkeitsspektrums des MAT mit den hier verwendeten Items noch nicht erreicht ist, zumal hier die meisten StudienteilnehmerInnen Studierende waren. Deswegen scheint es sinnvoll die nächste Version des MAT mit noch ein wenig schwierigeren Items zu gestalten und wenn möglich einer Stichprobe vorzulegen, die a priori als fähiger im Kopfrechnen eingeschätzt wird als Studierende. So werden nicht nur weitere Items auf Rasch Homogenität überprüft, sondern es wird weiter das obere Ende des Schwierigkeitsspektrums des MAT ausgelotet. Die Tatsache, dass acht der 176 StudienteilnehmerInnen keines der 14 Items des MAT lösen konnten zeigt, dass der endgültige Itempool des MAT neben genügend schwierigen Items ebenfalls genügend leichte Items enthalten sollte.

6 Literaturverzeichnis

Allen, P. A., Ashcraft, M. H. & Weber, T. A. (1992). On Mental Multiplication and Age. *Psychology and Aging*, 7(4), 536-545.

Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.

Andersen, E. B. (1973). A Goodness of Fit Test for the Rasch Model. *Psychometrika*, 38(1), 123-140.

Anderson, J. R. (2007). *Kognitive Psychologie*. Berlin: Springer.

Arendasy, M., Horke, L. F., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., Heidinger, C., Herle, M., Körtner, T. (2009). *INSBAT. Intelligenz-Struktur-Batterie*. Wien: Schuhfried.

Aster, M. G. v., Bzufka, M. W. & Horn, R. R. (2009). *ZAREKI-K. Neuropsychologische Testbatterie fuer Zahlenverarbeitung und Rechnen bei Kindern - Kindergartenversion (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Aster, M. G. v., Neubauer, A. & Horn, R. (Hrsg., 2006). *Wechsler-Intelligenztest für Erwachsene. WIE*. Frankfurt am Main: Harcourt.

Aster, M. G. v., Weinhold-Zulauf, M. & Horn, R. (2006). *ZAREKI-R. Neuropsychologische Testbatterie fuer Zahlenverarbeitung und Rechnen bei Kindern (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Ayalon, H. & Livneh, I. (2013). Educational Standardization and Gender Differences in Mathematics Achievement: A Comparative Study. *Social Science Research*, 42, 432-445.

Baddeley, A. D. (1996). The Fractionation of Working Memory. *Proceedings of the National Academy of Sciences of the United States of America*, 9, 13468-13472.

Baddeley, A. D. (2000). The Episodic Buffer: A New Component of Working Memory?. *Trends in Cognitive Science*, 4(11), 417-423.

Baddeley, A. D. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63, 1-29.

Baddeley, A. D., Eysenck, M. W. & Anderson, M. C. (2009). *Memory*. Hove: Psychology Press.

Baddeley, A. D. & Logie, R. H. (1999): Working Memory: The Multiple-Component Model. In A. Myake & P. Shah (Hrsg.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (28-61). Cambridge: Cambridge University Press. Zugriff am 11.02.2014, von scholar.google.com

Beiglböck, W. (2003). Ethische Richtlinien – Österreich. In K.D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 127-130). Weinheim: Beltz.

Binet, A. & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique*, 11, 191-244.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the History of Intelligence Testing. *Journal of Clinical and Experimental Neuropsychology*, 24(3), 383-405.

Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.

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

Bundeszentrale für gesundheitliche Aufklärung (n.d.). *Essstörungen. Wie häufig kommen Essstörungen vor?*. Zugriff am 09.02.2014, http://www.bzga-essstoerungen.de/index.php?id=44

Campbell, J. I. D. & Xue, Q. (2001). Cognitive Arithmetic Across Cultures. *Journal of Experimental Psychology: General*, 130(2), 299-315.

Charness, N. & Campbell, J. I. D. (1988). Acquiring Skill at Mental Calculation in Adulthood: A Task Decomposition. *Journal of Experimental Psychology: General*, 117(2), 115-129.

Deutsche Gesellschaft für Psychologie (2007). *Richtlinien zur Manuskriptgestaltung.* Göttingen: Hogrefe.

Devine, A., Fawcett, K., Szücs, D. & Dowker, A. (2012). Gender Differences in Mathematics Anxiety and the Relation to Mathematics Performancewhile Controlling for Test Anxiety. *Behavioral and Brain Functions*, 8:33.

Eberwein, M. (1991). Einführung in die Intelligenzdiagnostik. In Zentralstelle für Psychologische Information und Dokumentation (ZPID) (Hrsg.), *Intelligenztestverfahren. Eine Spezialbibliographie deutschsprachiger psychologischer Testverfahren*. Trier: ZPID. Zugriff am 26.11.2013 http://www.zpid.de/pub/tests/pt_pub1.pdf

Fehr, T., Weber, J., Willmes, K. & Herrmann, M. (2010). Neural Correlates in Exceptional Mental Arithmetic – About the Neural Architecture of Prodigious Skills. *Neuropsychologia*, 48, 1407-1416.

Feingold, A. (1993). Cognitive Gender DIfferences: A Developmental Perspective. *Sex Roles*, 29, 91-112.

Feng, S., Wang, W., Chen, L. & Liu, H. (2012). The Influence of Depression on Coping Strategies in Mental Arithmetic Stress. *2012 World Congress on Information and Communication Technologies*, 1019-1024.

Field, A. (2009). *Discovering Statistics Using SPSS*. London: SAGE.

Fippinger, F., Steidl, J. & Eher, A. (1972). *AST 4/A. ALLGEMEINER SCHULLEISTUNGSTEST FUER DIE 4. SCHULSTUFE - OESTERREICHISCHE VERSION (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Fischer, G. H. & Molenaar, I. W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications.* New York: Springer.

Fischer, G. H. & Ponocny-Seliger, E. (1998). *Structural Rasch Modeling: Handbook of the Usage of LpcM-Win 1.0*. ProGAMMA.

Franz, N. (2009). *Entwicklung eines Itempools für den "Mental Arithmetic Test" (MAT)*. Unveröffentlichte Diplomarbeit: Universität Wien.

Funke, J. (2006). Alfred Binet (1857 bis 1911) und der erste Intelligenztest der Welt. In: G. Lamberti (Hrsg.), *Intelligenz auf dem Prüfstand. 100 Jahre Psychometrie.* Göttingen: Vandenhoeck & Ruprecht.

Gabriel, M. (2004). *Entwicklung einer Rasch-homogenen und LLTM-kalibrierten Skala zur Erfassung der Fähigkeitsdimension Kopfrechnen*. Unveröffentlichte Diplomarbeit: Universität Wien.

Galton, F. (1879). Psychometric Experiments. *Brain*, 2, 149-162.

Gardner, H. (2002). *Intelligenzen: die Vielfalt des menschlichen Geistes*. Stuttgart: Klett-Cotta.

Geary, D. C., Frensch, P. A. & Wiley, J. G. (1993). Simple and Complex Mental Sutraction: Strategy Choice and Speed-of-Processing Differences in Younger and Older Adults. *Psychology and Aging*, 8(2), 242-256.

Geary, D. C. & Wiley, J. G. (1991). Cognitive Addition: Strategy Choice and Speed-of-Processing Differences in Young and Elderly Adults. *Psychology and Aging*, 6(3), 474-483.

Gittler, G. (2008). *Zum „Mental Arithmetic Test“ (MAT): Generative Regeln für das Erstellen neuer MAT Aufgaben*. Unveröffentlichter Forschungsbericht, Universität Wien.

Gittler, G. (2009). *Computerfassung des „Mental Arithmetic Tests“ (MAT)*. Unveröffentlichtes Computerprogramm (Softwarebasis TNT), Universität Wien.

Hagmann-von Arx, P., Petermann, F. & Grob, A. (2013). Konvergente und diskriminante Validität der WISC-IV und der Intelligence and Development Scales (IDS) bei Kindern mit Migrationshintergrund. *Diagnostica*, 59(4), 170-182.

Hauser, T. U., Rotzer, S., Grabner, R. H., Mèrillat, S. & Jäncke, L. (2013). Enhancing Performance in Numerical Magnitude Processing and Mental Arithmetic using transcranial Direct Current Stimulation (tDCS). *Frontiers in Human Neuroscience*, 7:244.

Heirdsfield, A. M., Cooper, T. J., Mulligan, J. & Irons, C. J., (1999). Children’s mental multiplication and division strategies. In O. Zaslavsky (Hrsg), *Proceedings of the 23rd Psychology of Mathematics Education Conference* (89-96). Haifa: Technion-Israel Institute of Technology.

Heller, K., Gaedike, A.-K- & Weinläder, H. (1985). KFT 4-13+. *Kognitiver Fähigkeits-Test*. Weinheim: Beltz.

Höft, S., Stelling, D. & Maschke, P. (2010). Begutachtung der Eignung als Kopilot einer Fluggesellschaft – Michael H., 28 Jahre. In K. D. Kubinger & T. M. Ortner (Hrsg.), *Psychologische Diagnostik in Fallbeispielen* (142-154). Göttingen: Hogrefe.

Horn, J. L. (1968). Organization of Abilities and the Development of Intelligence. *Psychological Review*, 75(3), 242-259.

Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender Differences in Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, 107(2), 139-155.

Imbo, I. & LeFevre, J.-A. (2009). Cultural Differences in Complex Addition: Efficient Chinese Versus Adaptive Belgians and Canadians. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1465-1476.

Imbo, I. & LeFevre, J.-A. (2010). The Role of Phonological and Visual Working Memory in Complex Arithmetic for Chinese- and Canadian-Educated Adults. *Memory & Cognition*, 38(2), 176-185.

Imbo, I. & Vandierendonck, A. (2010). Instruction and Load-Effects on High-Skill and Low-Skill Individuals: A Study in the Domain of Mental Arithmetic. *European Journal of Cognitive Psychology*, 22(6), 964-989.

Jackson, D. N. & Rushton, J. P. (2006). Males have Greater *g*: Sex differences in General Mental Ability from 100,000 17- to 18-year-olds on the Scholastic Assessment Test. *Intelligence*, 34, 479-486.

Jacobs, C. & Petermann, F. (2005). *RZD 2-6. Rechenfertigkeiten- und Zahlenverarbeitungs-Diagnostikum fuer die 2. bis 6. Klasse (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Kalbe, E., Brand, M. & Kessler, J. (2002). *ZRT. Zahlenverarbeitungs- und Rechentest (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Kaufmann, L., Nuerk, H. C., Graf, M., Krinzinger, H., Delazer, M. & Willmes, K. (2009). *TEDI-MATH. Test zur Erfassung numerisch-rechnerischer Fertigkeiten vom Kindergarten bis zur 3. Klasse (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Kenney-Benson, G. A., Patrick, H., Pomerantz, E. M. & Ryan, A. M. (2006). Sex Differences in Math Performance: The Role of Children's Approach to Schoolwork. *Developmental Psychology*, 42(1), 11-26.

Kersting, M., Althoff, K. & Jäger, K. O. (2008). *Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.

Ketelsen, K. & Welsh, M. (2010). Working Memory and Mental Arithmetic: A Case for Dual Central Executive Resources. *Brain and Cognition*, 74, 203-209.

Klein, E., Moeller, K., Willmes, K., Nuerk, H.-C. & Domahs, F (2011). The Influence of Implicit Hand-Based Representations on Mental Arithmetic. *Frontiers in Psychology*, 2:197.

Krajewski, K., Kuespert, P. & Schneider, W. (2002). *DEMAT 1+. Deutscher Mathematiktest fuer erste Klassen (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Krinzinger, H., Kaufmann, L., Gregoire, J., Desoete, A., Nuerk, H.-C. & Willmes, K. (2012). Gender Differences in the Development of Numerical Skill in Four European Countries. International Journal of Gender, Science and Technology, 4(1), 62-77.

Kubinger, K. D. (2009a). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID 2) samt AID 2-Türkisch*. Göttingen: Beltz.

Kubinger, K. D. (2009b). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens.* (2., überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.

Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2*. Göttingen: Beltz.

Krueger, K., Hylla, E. & Bargmann, R. (1970). *ZR 4+. ZAHLENRECHNEN 4+ (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Landerl, K. & Kaufmann, L. (2008). *Dyskalkulie: Modelle, Diagnostik, Intervention*. München: Ernst Reinhardt.

Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M & Dearing, E. (2013). Spatial Skills as a Predictor of First Grade Girls' Use of Higher Level Arithmetic Strategies. *Learning and Individual Differences*, 23, 123-130.

Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). Manual* (2. erweiterte und überarbeitete Aufl.). Göttingen: Hogrefe.

Logie, R. H., Gilhooly, K. J. & Wynn, V. (1994). Counting on Working Memory in Arithmetic Problem Solving. *Memory & Cognition*, 22(4), 395-410.

Lynn, R. & Irwing, P. (2008). Sex Differences in Mental Arithmetic, Digit Span, and g defined as working memory capacity. *Intelligence*, 36, 226-235.

Mejias, S., Grégoire, J. & Noël, M. (2012). Numerical Estimation in Adults with and without Developmental Dyscalculia. *Learning and Individual Differences*, 22, 164-170.

Molenaar, I. W. (1995). Estimation of Item Parameters. In G. H. Fischer & I. W. Molenaar (Hrsg.), *Rasch Models: Foundations, Recent Developments, and Applications* (S. 39-52). New York: Springer.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., Paulus, W., Hummel, F., Boggio, P. S., Fregni, F. & Pascual-Leone, A. (2008). Trancranial Direct Current Stimulation: State of the Art 2008. *Brain Stimulation*, 1, 206-223.

Penner, A. M. & Paret, M. (2008). Gender Differences in Mathematics Achievement: Exploring the Early Grades and the Extremes. *Social Science Research*, 37, 239-253.

Perham, N., Hodgetts, H. & Banbury, S. (2013). Mental Arithmetic and Non-Speech Office Noise: An Exploration of Interference-by-Content. *Noise & Health*, 15(62), 73-78.

Perleth, C. (2006). Die Münchner Hochbegabungs-Testbatterie (MHBT). In H. Wagner & Thomas-Morus-Akademie Bensberg (Hrsg.), *Intellektuelle Hochbegabung: Aspekte der Diagnostik und Beratung. Tagungsbericht* (56-69). Bad Honnef: Karl Heinrich Bock. Zugriff am 07.02.2014, von scholar.google.com

Petermann, F. & Petermann, U. (Hrsg., 2007). *HAWIK-IV. Hamburg-Wechsler-Intelligenztest für Kinder – IV*. Bern: Huber.

Pivik, R. T., Tennal, K. B., Chapman, S. D. & Gu, Y. (2012). Eating Breakfast Enhances the Efficiency of Neural Networks Engaged During Mental Arithmetic in School-Aged Children. *Physiology & Behavior*, 106, 548-555.

Price, G. R., Mazzocco, M. M. M. & Ansari, D. (2013). Why Mental Arithmetic Counts: Brain Activation During Single Digit Arithmetic Predicts High School Math Scores. *The Journal of Neuroscience*, 33(1), 156-163.

Rasch, G. W. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Kopenhagen: Danmarks Paedagogiske Institut.

Ridgway, A. J. (2009). The Inner Voice. *International Journal of English Studies*, 9(2), 45-58.

Rost, D. H. (2009). *Intelligenz: Fakten und Mythen*. Weinheim: Beltz.

Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Hans Huber.

Rubinsten, O., Bialik, N. & Solar, Y (2012). Exploring the Relationship Between Math Anxiety and Gender Through Implicit Measurement. *Frontiers in Human Neuroscience*, 6:279.

Salthouse, T. A. & Kersten, A. W. (1993). Decomposing Adult Age Differences in Symbol Arithmetic. *Memory & Cognition*, 21(5), 699-710.

Schallberger, U. & Trier, U. P. (1978). *TSF 6-7. TEST FUER SCHULRELEVANTE FAEHIGKEITEN 6-7 (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Schmitt, M. & Altstötter-Gleich, C. (2010). *Differentielle Psychologie und Persönlichkeitspsychologie kompakt*. Weinheim, Basel: Beltz.

Schroeders, U. & Schneider, W. (2008). *TeDDy-PC. Test zur Diagnose von Dyskalkulie (PSYNDEX Tests Review)*. Zugriff am 01.06.2014, von der PSYNDEXplus Tests Datenbank.

Sherin, B. & Fuson, K. (2005). Multiplication Strategies and the Appropriation of Computational Resources. *Journal for Research in Mathematics Education*, 36(4), 347-395.

Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35, 4-28.

Statistik Austria (2014). *Armut und soziale Eingliederung*. Zugriff am 30.01.2014, http://www.statistik.at/web_de/statistiken/soziales/armut_und_soziale_eingliederung/

Stemmler, G., Hagemann, D., Amelang, M. & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung*. (7., vollständig überarbeitete Auflage). Stuttgart: Kohlhammer.

Stewart, D., Barnes, J., Cote, J., Cudeck, R. & Malthouse, E. (2001). Factor Analysis. *Journal of Consumer Psychology*, 10(1&2), 75-82.

Strobl, C. (2012). *Das Rasch-Modell*. München: Rainer Hampp.

Thevenot, C., Castel, C., Danjon, J., Fanget, M. & Fayol, M. (2013). The Use of the Operand-Recognition Paradigm for the Study of Mental Addition in Older Adults. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 68(1), 64-67.

Thompson, I. & Smith, F. (1999). *Mental Calculation Strategies for the Addition and Subtraction of 2-digit Numbers*. Zugriff am 10.02.2014, von scholar.google.com

Universität Wien (n.d.). *Studium an der Universität Wien*. Zugriff am 10.02.2014, http://www.univie.ac.at/studium/

Van der Sluis, S., Posthuma, D., Dolan, C. V., De Geus, E. J. C., Colom, R. & Boomsma, D. I. (2006). Sex Differences on the Dutch WAIS-III. *Intelligence*, 34, 273-289.

Van Impe, A., Coxon, J. P., Goble, D. J., Wenderoth, N. & Swinnen, S. P. (2011). Age-related Changes in Brain Activation Underlying Single- and Dual-Task Performance: Visuomanual Drawing and Mental Arithmetic. *Neuropsychologia*, 49, 2400-2409.

Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54(3), 427-450.

Wiemers, M., Bekkering, H. & Lindemann, O. (in Druck). Spatial Interferences in Mental Arithmetic: Evidence from the Motion-Arithmetic Compatibility Effect. *The Quarterly Journal of Experiment Psychology*.

Wilfinger, G. & Holocher-Ertl, S. (2010). Abklärung einer Intelligenzminderung – Der 15-jährige Yusuf mit Türkisch als Muttersprache. In K. D. Kubinger & T. M. Ortner (Hrsg.), *Psychologische Diagnostik in Fallbeispielen* (142-154). Göttingen: Hogrefe.

7 Anhang

7.1 Zusammenfassung

Kopfrechnen ist eine alltägliche Anforderung in Beruf und Alltag.
Ein Kopfrechentest kann unter anderem in der Leistungs- und Eignungsdiagnostik eingesetzt werden und er kann zum Training der Kopfrechenfähigkeit benutzt werden.
Der Mental Arithmetic Test (MAT) ist ein solcher Test.
20 Items des MAT werden auf Rasch Homogenität überprüft. Das Globalziel ist es dabei, einen Itempool zu generieren, der das gesamte Schwierigkeitsspektrum der Normalbevölkerung abdeckt. Damit soll der Test dann adaptiv und am Computer vorgegeben werden.
Der MAT enthält die vier Grundrechenarten und ein- bis dreistellige Zahlen. Für die Lösung eines MAT-Items ist vornehmlich das Arbeitsgedächtnis nötig, aber auch das Wissen um effiziente Rechenstrategien.
Es werden die 20 MAT-Items auf diskriminante Validität überprüft und es werden erste Validitätshinweise durch Gruppenunterschiede gesammelt.
Geschlechts- und Altersunterschiede im Kopfrechnen werden exploriert.

Mit zwei Versionen und einem inkompletten Design als Methode, zeigen die Ergebnisse Rasch Homogenität für alle Items.
Der MAT ist in Bezug auf den Test Analogien diskriminant valide.
Gruppenunterschiede zeigen, dass Menschen, die sich als fähiger in Mathematik einschätzen auch besser im MAT abschneiden, als Menschen, die sich als weniger fähig in Mathematik einschätzen.
Männer und jüngere Personen schneiden im Kopfrechnen besser ab.
Eine Feedbackschleife würde die Standardisierung der Testsituation des MAT erhöhen. Die Zukunft des MAT wird in der Überprüfung der Rasch Homogenität von weiteren Items gesehen.

7.2 Abstract

Mental arithmetic is commonly used in profession and everyday life.
A mental arithmetic test could be applied in performance and aptitude diagnostics and to train one's mental arithmetic ability.
Such a test is the Mental Arithmetic Test (MAT).
20 MAT-items are tested for Rasch homogeneity. The global goal thereby is to generate an item pool which covers the whole difficulty spectrum of the normal population to then use the pool in adaptive testing on the computer.
The MAT contains the four basic arithmetic operations and one to three-digit numbers. For solving one item one must primarily use skills of working memory and efficient strategies for mental arithmetic.
The 20 MAT items are tested for discriminant validity and group differences are used to provide first hints towards validity.
Gender and age differences in mental arithmetic are explored.

With two used versions of the MAT and an incomplete design, the results speak for Rasch homogeneity of all items.
The MAT achieved discriminant validity in relation to the verbal test Analogien (English: Analogies).
Group differences showed that participants who self-reported to be more competent in mathematics performed better on the MAT than participants who self-reported to be less competent in mathematics.
Men and younger participants performed better in mental arithmetic.
A feedback loop could enhance test standardization of the MAT. Its future is seen in testing more items for Rasch homogeneity to finalize the item pool.

7.3 Lebenslauf

Name:	Voigt
Vorname:	Gregor
Geburtsdatum:	20.06.1988
Geburtsort:	Hamburg
Staatsangehörigkeit:	deutsch
Familienstand:	verheiratet
Adresse:	Donaufelder Straße 54 1210 Wien
Schulausbildung:	1999 – 2008 Abitur am Alexander-von-Humboldt-Gymnasium
Auslandsaufenthalt:	08/2005 – 06/2006 Austauschjahr in Selah, WA, U.S.A.
Studium:	2008 – 2014 Hauptstudium der Psychologie, Universität Wien
Praktikum:	Sechs-Wochen-Praktikum am Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung, Universität Wien und am Initiativkolleg- Empowerment through Human Rights mit dem Schwerpunkt psychometrische Fragebogenanalyse und lineare Strukturgleichungsmodellierung
Sprachkenntnisse:	Englisch in Wort und Schrift, Französisch Grundkenntnisse