

Kognitive Basisfunktionen: Rolle für den Wissenserwerb und Trainierbarkeit



Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Philosophie an der Ludwig-Maximilians-Universität

München

vorgelegt von

Matthias Paul Schwaighofer

München, Dezember 2014

Gutachter: 1. Prof. Dr. Frank Fischer

2. Prof. Dr. Markus Bühner

Datum der mündlichen Prüfung: 22.01.2015

Vorwort

Bei vielen Personen möchte ich mich bedanken.

An erster Stelle bedanke ich mich bei meinen Betreuern Prof. Dr. Frank Fischer und Prof. Dr. Markus Bühner für die Ermöglichung der Promotion und die ausgezeichnete Betreuung. Von ihnen durfte ich viel lernen.

Dank gilt auch PD Dr. Ingo Kollar für die Bereitschaft, als Drittgutachter zur Verfügung zu stehen.

Bei Prof. Dr. Moritz Heene und PD Dr. Christof Wecker bedanke ich mich für die kompetente und wertvolle Beratung bei methodischen Angelegenheiten.

Freydis Vogel und Ulrike Franke danke ich für viele hilfreichen Anregungen beim Schreiben der Dissertation.

Bettina Kieltsch, Christian Meißner, Anja Prinz und Anna Reeb danke ich für die Beteiligung an der Datenerhebung der Studie 1 der vorliegenden Arbeit. Saraswathi Ramasamy danke ich für die Doppelkodierungen bei der Metaanalyse (Studie 2 der vorliegenden Arbeit).

Bei Martina Denz, Rosa Haas, Ruth Maitland, Georg Kuschel und insbesondere Cora Laugs bedanke ich mich für die Hilfe bei organisatorischen Belangen.

Zur besseren Lesbarkeit wird in der vorliegenden Arbeit nur die maskuline Form personenbezogener Substantive verwendet. Es sind jedoch immer männliche und weibliche Personen gemeint, wenn keine separate Bezeichnung erfolgt.

Zusammenfassung

Die Relevanz der kognitiven Basisfunktionen Arbeitsgedächtnis und der Fähigkeit zum Shifting sowie der fluiden Intelligenz für viele kognitive Leistungen wurde wiederholt gezeigt. Insbesondere das Arbeitsgedächtnis ist wichtig für das Lernen und sollte bei Entscheidungen hinsichtlich instruktionaler Unterstützungsmaßnahmen berücksichtigt werden. In der ersten Studie der vorliegenden Arbeit wurde untersucht, ob der Effekt einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung durch Lösungsbeispiele verglichen mit einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) auf den Wissenserwerb von der Arbeitsgedächtniskapazität, der Shifting-Fähigkeit sowie der fluiden Intelligenz moderiert wird. In den beiden Lernumgebungen lernten 76 Studierende aus den Fachrichtungen Pädagogik, Psychologie und Schulpsychologie mit Problemen aus der Domäne Statistik. Die Ergebnisse zeigten, dass die Fähigkeit zum Shifting und die fluide Intelligenz, nicht aber die Arbeitsgedächtniskapazität einen moderierenden Einfluss auf den Erwerb anwendungsorientierten Wissens haben. Die kognitive Belastung war nicht mit der Arbeitsgedächtniskapazität, aber der fluiden Intelligenz, konzeptuellem und anwendungsorientiertem Vorwissen sowie Skalen eines Fragebogens zur Erfassung aktueller Motivation korreliert. Den Ergebnissen der ersten Studie zu Folge könnten die Fähigkeit zum Shifting und die fluide Intelligenz für die Bestimmung des optimalen Grads an instruktionaler Unterstützung wichtig sein. Vor dem Hintergrund der Bedeutung kognitiver Basisfunktionen und einer Vielzahl neuerer Studien zu Arbeitsgedächtnistrainings wurde in der zweiten Studie dieser Arbeit eine Metaanalyse zu Arbeitsgedächtnistrainings durchgeführt. In dieser Metaanalyse wurden mehrere, bisher nicht beachtete Trainingsbedingungen als potentielle Moderatoren berücksichtigt. Es wurden 47 Studien mit 65 Gruppenvergleichen analysiert. Die Ergebnisse zeigten, dass Arbeitsgedächtnistrainings nahe Transfereffekte auf Kurzzeit- und Arbeitsgedächtniskomponenten erzielen, die zum Follow-up erhalten bleiben. Weite Transfereffekte auf andere kognitive Fähigkeiten und Fertigkeiten (nonverbale und verbale Fähigkeit, Fertigkeit zur Wortdekodierung, mathematische Fähigkeiten und Fertigkeiten) waren klein, auf die nonverbale und verbale Fähigkeit beschränkt und blieben nicht zum Follow-up erhalten. Unter den Moderatoren hatten der Interventionstyp, die Trainingsdosis, die Sessiondauer, die Supervision während des Trainings und der Trainingsort einen Einfluss auf nahe und weite Transfereffekte. Den Ergebnissen der zweiten Studie zufolge können Arbeitsgedächtnistrainings robuste, nahe Transfereffekte auf andere Aufgaben erzielen, die Kurzzeit- und Arbeitsgedächtniskomponenten erfassen. Die Bedingungen für Transfereffekte, insbesondere weite Transfereffekte, sind

noch schlecht verstanden. Arbeitsgedächtnistrainings mit optimierten Trainingsbedingungen könnten jedoch zu praktisch relevanten Transfereffekten führen.

Inhaltsverzeichnis

Vorwort	I
Zusammenfassung.....	II
Abkürzungsverzeichnis.....	VII
1 Problemstellung	1
2 Kognitive Basisfunktionen und fluide Intelligenz.....	6
2.1 Begründung für die Auswahl kognitiver Basisfunktionen	6
2.2 Konzeptionalisierung und Erfassung kognitiver Basisfunktionen und fluider Intelligenz	7
2.2.1 Arbeitsgedächtnis	7
2.2.2 Shifting	10
2.2.3 Fluide Intelligenz	12
2.3 Bedeutung kognitiver Basisfunktionen und fluider Intelligenz.....	13
2.3.1 Arbeitsgedächtnis	13
2.3.2 Shifting	15
2.3.3 Fluide Intelligenz	15
2.4 Trainierbarkeit kognitiver Basisfunktionen.....	16
2.4.1 Arbeitsgedächtnistrainings	16
2.4.2 Trainings zur Verbesserung von Shifting.....	23
3 Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung.....	24
4 Forschungsfragen	29
5 Studie 1: Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung (Frage 1)	31
5.1 Ziel und Hypothesen.....	31
5.2 Methode	33
5.2.1 Stichprobe und Design.....	33
5.2.2 Lernumgebung und Versuchsablauf.....	34
5.2.3 Experimentalbedingungen	36
5.2.4 Abhängige Variablen	38
5.2.5 Kontrollvariablen	51
5.2.6 Statistische Analyse	52
5.3 Ergebnisse.....	54

5.3.1	Vorausgehende Analysen	54
5.3.2	Moderatorenrolle kognitiver Basisfunktionen und fluiden Intelligenz (Frage 1)	59
5.4	Diskussion	64
5.4.1	Zusammenfassung der Ergebnisse.....	64
5.4.2	Ergebnisinterpretation.....	65
5.4.3	Limitationen.....	71
5.4.4	Weitere Forschung.....	75
5.4.5	Schlussfolgerungen.....	77
6	Studie 2: Metaanalyse von Arbeitsgedächtnistrainings (Fragen 2 und 3).....	79
6.1	Ziele und Hypothesen.....	79
6.2	Methode.....	82
6.2.1	Literatursuche und Einschlusskriterien.....	82
6.2.2	Codierung von Transfermaßen und Moderatoren.....	83
6.2.3	Effektstärkeberechnungen	88
6.2.4	Random-effects Modelle	89
6.2.5	Durchführung.....	90
6.3	Ergebnisse Studie 2	91
6.3.1	Nahe und weite Transfereffekte (Frage 2).....	92
6.3.2	Moderatoren von Transfereffekten (Frage 3)	107
6.4	Diskussion	114
6.4.1	Zusammenfassung der Ergebnisse.....	114
6.4.2	Ergebnisinterpretation.....	114
6.4.3	Praktische Implikationen	121
6.4.4	Limitationen.....	122
6.4.5	Weitere Forschung.....	123
6.4.6	Schlussfolgerungen.....	125
7	Gesamtdiskussion	126
7.1.1	Zusammenfassung der Ergebnisse der Studien	126
7.1.2	Theoretischer Erkenntnisfortschritt	127
7.1.3	Limitationen und weitere Forschung	133
7.1.4	Praktische Bedeutung	138
	Tabellenverzeichnis	140

Abbildungsverzeichnis	141
Literaturverzeichnis.....	144
Anhang A: Vorwissenstests mit Lösung.....	172
Anhang B: Wissenstests zum Posttest mit Lösung.....	175
Anhang C: Fragebogen zur aktuellen Motivation	179
Anhang D: Statistische Probleme und zugehörige Informationen	181
Anhang E: Lösungsbeispiele	185
Anhang F: Antwortbogen zu den statistischen Problemen	188
Anhang G: Cognitive load Skala.....	189
Anhang H: Korrelationen innerhalb der Aufgaben zur Erfassung der Arbeitsgedächtniskapazität und der Shifting-Fähigkeit.....	190

Abkürzungsverzeichnis

AG	Arbeitsgedächtnis
ALM	Allgemeines Lineares Modell
aospan	automated operation span
arspan	automated reading span
asymspan	automated symmetry span
CL	cognitive load
CLT	cognitive load theory
CSA	complex span Aufgabe
FAM	Fragebogen zur aktuellen Motivation
Gc	Kristalline Intelligenz
Gf	Fluide Intelligenz
INSBAT	Intelligenz-Struktur-Batterie
IQ	Intelligenz
JN	Johnson-Neyman
kBf	Kognitive Basisfunktionen
KZG	Kurzzeitgedächtnis
MC	multiple-choice
PM	primary memory
SM	secondary memory

1 Problemstellung

Unter der Bezeichnung „kognitive Basisfunktionen (kBf)“ lassen sich nach Andersson (2010) grundlegende kognitive Funktionen wie Arbeits- und Kurzzeitgedächtnis, exekutive Funktionen, Langzeitgedächtnis und die Verarbeitungsgeschwindigkeit subsumieren. Insbesondere exekutive Funktionen erhielten in den letzten Jahren Aufmerksamkeit in der Forschung, da sie wichtige Komponenten der Fähigkeit zur Selbstkontrolle (oder Selbstregulation) darstellen (Miyake & Friedman, 2012). Für die Fähigkeit zur Selbstkontrolle wurde gezeigt, dass sie breite Auswirkungen auf das alltägliche Leben hat (Mischel et al., 2011; Moffit et al., 2011). Die drei grundlegenden exekutiven Funktionen, die häufig in der Literatur genannt werden, sind Shifting (die Fähigkeit zum flexiblen Wechsel zwischen Aufgaben oder mentalen Voreinstellungen; Miyake & Friedman, 2012), Updating (Informationsaktualisierung und -überwachung von Repräsentationen im Arbeitsgedächtnis) und Inhibition (Hemmung automatischer, dominanter oder aufdringlicher Antworten wenn nötig) (Best, Miller, & Naglieri, 2011; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). In dieser Arbeit stehen die exekutiven Funktionen Arbeitsgedächtnis (AG) und Shifting im Fokus der Betrachtungen.

Die Bedeutung dieser kBf für viele kognitive Leistungen wurde wiederholt gezeigt (z.B. Yeniad, Malda, Mesman, van Ijzendoorn, & Pieper, 2013; Yuan, Steedle, Shavelson, Alonzo, & Oppezzo, 2006). Insbesondere das AG wird als wichtige Variable für das Lernen angesehen. Die Begründung ist, dass durch das AG ein mentaler Arbeitsbereich bereitgestellt wird, in dem Informationen gehalten werden können während man mental mit anderen bedeutsamen Aktivitäten beschäftigt ist (Gathercole & Alloway, 2007). Dem AG wurde viel Beachtung in Studien zur Optimierung von Lernprozessen durch Gestaltung des Instruktionsdesigns geschenkt. Vor allem im Rahmen der cognitive load theory (CLT) wurde postuliert, dass die Kapazität des AGs bei instruktionalen Maßnahmen beachtet werden sollte (Sweller, 2011). Daher erscheint es sinnvoll, die AG-Kapazität zu berücksichtigen, wenn Lernumgebungen gestaltet werden.

Lernumgebungen unterscheiden sich im Hinblick auf den Grad an instruktionaler Unterstützung. Eine Form von Lernumgebung mit hohem Grad an instruktionaler Unterstützung stellen Lösungsbeispiele dar. Lösungsbeispiele werden oftmals mit Lernumgebungen mit niedrigem Grad an instruktionaler Unterstützung wie beispielsweise Problemlösen verglichen (z.B. Renkl, Gruber, Weber, Lerche, & Schweizer, 2003; Van Gog, Paas, & Van Merriënboer, 2006). Die Wirksamkeit von Lösungsbeispielen könnte durch die AG-Kapazität beeinflusst

werden (Van Gog & Rummel, 2010). Hintergrund ist, dass Anforderungen an die AG-Kapazität in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung hoch sein könnten, beispielsweise bei Problemen, deren Lösung die parallele Speicherung und Verarbeitung mehrerer Informationen erfordert. Lerner mit geringer AG-Kapazität könnten daher Schwierigkeiten in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung haben. Ein hoher Grad an instruktionaler Unterstützung (Lösungsbeispiele) könnte die Anforderungen an die AG-Kapazität reduzieren. Der Vorteil einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) hinsichtlich des Erwerbs von Wissen könnte umso größer sein, je geringer die AG-Kapazität ist. Analog zur AG-Kapazität könnten Anforderungen an Shifting in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung hoch sein, beispielsweise bei Problemen, deren Lösung einen Wechsel zwischen Aufgabenteilen erfordert. Durch einen hohen Grad an instruktionaler Unterstützung (Lösungsbeispiele) könnten die Anforderungen an die Shifting-Fähigkeit verringert werden. Der Vorteil einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) im Hinblick auf den Wissenserwerb könnte umso größer sein, je geringer die Shifting-Fähigkeit ist. Kognitive Basisfunktionen könnten somit im Hinblick auf den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb eine moderierende Rolle spielen. Jedoch wurde die Rolle der AG-Kapazität als auch der Fähigkeit zum Shifting für den Wissenserwerb kaum systematisch berücksichtigt, wenn unterschiedliche Grade an instruktionaler Unterstützung kontrastiert wurden.

Daher stellt sich die Frage, welche Rolle kBf für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung haben. Dieser Frage wird im Rahmen der Studie 1 dieser Arbeit nachgegangen. Neben den kBf Shifting und AG wird auch die Rolle der fluiden Intelligenz (Gf) als ein potentieller Moderator des Wissenserwerbs aus mehreren Gründen untersucht. Fluide Intelligenz kann aufgefasst werden als die Fähigkeit zum schlussfolgernden Denken, um neue Probleme zu lösen (König, Bühner, & Mürling, 2005). Fluide Intelligenz scheint wichtig für Lernen zu sein, da sie den Wissenserwerb fördert (z.B. Schweizer & Koch, 2002; Ziegler, Danay, Heene, Asendorpf, & Bühner, 2012) und prädiktiv für das Lösen neuer und komplexer Probleme ist (z.B. Ackerman & Cianciolo, 2002; Snow, Kyllonen, & Marshalek, 1984). Ähnlich wie es auch für die AG-Kapazität und die Fähigkeit zum Shifting zutrifft, wurde der moderierende Einfluss von Gf auf den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler

Unterstützung noch nicht systematisch untersucht. In einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung könnten die Anforderungen an Gf hoch sein, beispielsweise wenn neue und komplexe Probleme gelöst werden müssen. Das Lösen komplexer Probleme erfordert typischerweise von Lernern zu schlussfolgern, welche Informationen für die Problemlösung relevant sind, was hohe Anforderungen an Gf stellt. Lerner mit geringer Gf könnten folglich Probleme in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung haben. Ein hoher Grad an instruktionaler Unterstützung (Lösungsbeispiele) könnte die Anforderungen an Gf reduzieren. Der Vorteil einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) hinsichtlich des Wissenserwerbs könnte umso größer sein, je geringer die Shifting-Fähigkeit ist. Darüber hinaus weist Gf einen starken Zusammenhang mit der AG-Kapazität auf (z.B. Ackerman, Beier, & Boyle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Oberauer, Süß, Wilhelm & Wittmann, 2008). Der Einschluss von kBf und Gf in die Studie 1 ermöglicht es, für andere kognitive Fähigkeiten zu kontrollieren, wenn der moderierende Einfluss einer spezifischen kognitiven Fähigkeit untersucht wird (z.B. Kontrolle für Gf, wenn der moderierende Einfluss der AG-Kapazität untersucht wird).

In der ersten Studie der vorliegenden Arbeit lernten Studierende der Fachrichtungen Pädagogik, Psychologie und Schulpsychologie mit komplexen Problemen aus der Domäne Statistik entweder in einer Lernumgebung mit einem niedrigen Grad an instruktionaler Unterstützung (Problemlösen) oder einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung durch Lösungsbeispiele. Das Fach Statistik wurde gewählt, da die Überlegenheit von Lösungsbeispielen verglichen mit Problemlösen wiederholt in gut strukturierten Domänen wie Statistik gezeigt werden konnte (z.B. Leppink, Paas, Van Gog, Van der Vleuten, & Van Merriënboer, 2014; Paas, 1992). Um die Rolle von kBf und Gf für den Wissenserwerb in Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung untersuchen zu können, müssen sich die Lernumgebungen im Effekt auf den Wissenserwerb unterscheiden. Des Weiteren ist das Fach Statistik in vielen Berufen und Studiengängen wichtig (Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013). Beispielsweise ist Statistik ein wichtiger und verpflichtender Teil in den Studiengängen Pädagogik, Psychologie und Schulpsychologie. Insgesamt soll mit der Studie 1 die Forschung zur Bedeutsamkeit von kBf und Gf für kognitive Leistungen um die Rolle dieser kognitiven Fähigkeiten für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung erweitert werden.

Vor dem Hintergrund der Forschung zur Bedeutsamkeit von kBf für viele kognitive Leistungen wurden Versuche unternommen, kBf zu trainieren, um neben Verbesserungen in der jeweils trainierten kognitiven Funktion auch Verbesserungen in anderen kognitiven Funktionen zu erzielen. Vor allem für das Training des AGs wuchs das Interesse unter Forschern und Fachleuten. Eine Vielzahl an Studien zu AG-Trainings widmete sich der Frage, ob kognitive Trainings intellektuelle Fähigkeiten verbessern und somit zu Transfereffekten führen können. Die Ergebnisse einer Metaanalyse von Melby-Lervåg und Hulme (2013) lassen jedoch daran zweifeln, ob ein Training des AGs zur Erzielung von Transfereffekten auf andere Maße wie beispielsweise Gf geeignet ist. Seit dieser Metaanalyse wurden viele weitere Studien veröffentlicht, die Transfereffekte von AG-Trainings untersuchten. Zudem wurden viele potentielle Moderatoren von Transfereffekten wie beispielsweise das Feedback zu Leistungen von Trainingsteilnehmern sowie der Trainingsort nicht systematisch berücksichtigt. Folglich besteht die Notwendigkeit, Trainingsbedingungen als potentielle Moderatoren von Transfereffekten zu identifizieren.

Aus diesen Gründen erscheint zum einen die Frage interessant, welche Transfereffekte AG-Trainings unter Berücksichtigung einer Vielzahl neu veröffentlichter Trainingsstudien erzielen. Zum anderen ist fraglich, welche Trainingsbedingungen einen moderierenden Einfluss auf Transfereffekte haben. Diesen Fragen wird im Rahmen der Studie 2 dieser Arbeit nachgegangen. Mit einer Metaanalyse wurden Transfereffekte von AG-Trainings sowie Trainingsbedingungen als potentielle Moderatoren dieser Transfereffekte untersucht. Hinsichtlich der Analysestrategie wurden einige wichtige Änderungen im Vergleich mit der bisherigen Metaanalyse zu AG-Trainings von Melby-Lervåg und Hulme (2013) vorgenommen. Beispielsweise wurde zwischen Transfereffekten auf Kurzzeitgedächtnis- und AG-Komponenten unterschieden.

Vor diesem Hintergrund ist die vorliegende Arbeit in sieben Kapitel gegliedert. Nach der Problemstellung (*Kapitel 1*) wird im *Kapitel 2* („Kognitive Basisfunktionen und fluide Intelligenz“) zunächst die Auswahl von kBf für die vorliegende Arbeit begründet. Im Anschluss daran werden die jeweilige Konzeptionalisierung und Erfassung der kBf AG und Shifting sowie Gf thematisiert. Im Anschluss an einen Abschnitt zur Bedeutung von kBf und Gf für kognitive Leistungen wird auf die Trainierbarkeit von kBf mit einem Schwerpunkt auf AG-Trainings eingegangen.

In *Kapitel 3* („Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung“) werden zunächst Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung beschrieben. Die anschließend dargestellte CLT dient als theoretische

Basis für die Erklärung der, mit einem Instruktionsdesign verbundenen kognitiven Belastung im AG. Anknüpfend an die CLT wird auf die Forschung zu Lösungsbeispielen als eine Form von Lernumgebung mit hohem Grad an instruktionaler Unterstützung eingegangen.

Vor dem theoretischen Hintergrund (*Kapitel 2* und *3*) sind in *Kapitel 4* („Forschungsfragen“) die Fragestellungen der beiden Studien dieser Arbeit beschrieben.

Die erste Studie dieser Arbeit wird in *Kapitel 5* („Studie 1: Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung“) dargestellt. Im Anschluss an das Ziel und die Hypothesen der Studie 1 werden im Rahmen der Methode der Studie 1 Stichprobe und Design, Lernumgebung, Experimentalbedingungen, abhängige Variablen, Kontrollvariablen und statistische Analysen beschrieben. Der Ergebnisteil gliedert sich in einen Abschnitt über vorausgehende Analysen sowie einen Abschnitt über die Moderatorenrolle von kBf und Gf. In der darauf folgenden Diskussion werden die Ergebnisse der Studie zusammenfassend dargestellt, bevor eine detaillierte Diskussion der Ergebnisse der vorausgehenden Analysen und insbesondere der Moderatorenrolle von kBf und Gf erfolgt. Des Weiteren werden Limitationen, Ansatzpunkte für weitere Forschung und Schlussfolgerungen aus der Studie 1 erläutert.

In *Kapitel 6* („Studie 2: Metaanalyse von Arbeitsgedächtnistrainings“) wird die zweite Studie dieser Arbeit dargestellt. Nach den Zielen und Hypothesen der Studie 2 wird die Methode der Metaanalyse beschrieben. Erläutert werden Literatursuche und Einschlusskriterien, die Codierung von Transfermaßen und Moderatoren, Effektstärkeberechnung, random-effects Modelle als Basis für die statistischen Analysen und die Durchführung der Metaanalyse. Danach werden die Ergebnisse getrennt nach nahen und weiten Transfereffekten von AG-Trainings und Moderatoren von Transfereffekten berichtet. Die Ergebnisse werden zusammenfassend am Anfang der Diskussion dargestellt. An die folgende Ergebnisinterpretation schließen sich praktische Implikationen und Limitationen der Studie 2 an. Folglich werden Empfehlungen für zukünftige Forschung gegeben und Schlussfolgerungen aus der Studie 2 gezogen.

In *Kapitel 7* („Gesamtdiskussion“) werden zunächst die Ergebnisse der beiden Studien der vorliegenden Arbeit zusammenfassend dargestellt. Der theoretische Erkenntnisfortschritt der Ergebnisse der beiden Studien dieser Arbeit wird diskutiert und wichtige Limitationen sowie Ansatzpunkte für weitere Forschung beschrieben. Abschließend wird auf die praktische Bedeutung der Ergebnisse der beiden Studien der vorliegenden Arbeit eingegangen.

2 Kognitive Basisfunktionen und fluide Intelligenz

Wie eingangs erwähnt, können grundlegende kognitive Funktionen wie das AG und exekutive Funktionen als kBf betrachtet werden. Im folgenden Abschnitt wird begründet, warum die vorliegende Arbeit auf die kBf AG und Shifting fokussiert. Im darauf folgenden Abschnitt wird auf die Konzeptionalisierung und Erfassung von AG, Shifting und Gf eingegangen.

2.1 Begründung für die Auswahl kognitiver Basisfunktionen

Wie bereits in Kapitel 1 erwähnt, sind Shifting, Updating und Inhibition drei grundlegende exekutive Funktionen, die häufig in der Literatur genannt werden (Best et al., 2011; Miyake et al., 2000). Nachfolgend wird begründet, warum Inhibition in der vorliegenden Arbeit nicht näher betrachtet wird.

Friedman et al. (2008) sowie Friedman, Miyake, Robinson und Hewitt (2011; für einen Überblick siehe Miyake & Friedman, 2012) konnten auf Basis von Strukturgleichungsmodellen zeigen, dass bei Extrahierung eines gemeinsamen Faktors durch die Faktoren für Updating, Shifting und Inhibition für letztere nach Berücksichtigung des gemeinsamen Faktors keine einzigartige Varianz mehr übrig bleibt.

Wie in Abschnitt 2.2.1 noch genauer dargelegt wird, haben sich sogenannte complex span Aufgaben (CSAs) wie beispielsweise die operation span Aufgabe als reliable und valide Indikatoren der AG-Kapazität erwiesen (z.B. Redick, Broadway, et al., 2012). Die operation span Aufgabe scheint von den drei genannten exekutiven Faktoren am meisten durch den Updating-Faktor beeinflusst zu sein (Miyake et al., 2000), was konform mit der Sichtweise von Krumm et al. (2009) ist, nach der Updating nicht von typischen AG-Aufgaben wie komplexen Speicheraufgaben getrennt werden kann.

Folglich könnte Inhibition schon in großem Ausmaß durch die Leistungen in der operation span Aufgabe sowie in Aufgaben zur Erfassung des Aufgabenwechsels reflektiert werden. Des Weiteren wird Inhibition als multidimensionales Konstrukt angesehen (Harnishfeger, 1994; Nigg, 2000; Krumm et al., 2009), da es mehrere, statistisch dissoziierbare, inhibitionsbezogene Fähigkeiten gibt (z.B. Friedman & Miyake, 2004; Hedden & Yoon, 2006). Aus diesen Gründen wird in der vorliegenden Arbeit nicht näher auf Inhibition eingegangen.

2.2 Konzeptionalisierung und Erfassung kognitiver Basisfunktionen und fluider Intelligenz

2.2.1 Arbeitsgedächtnis

Konzeptionalisierung

Nach Baddeley (2007, S. 1) wird davon ausgegangen, dass das AG einen vorübergehenden Speicher darstellt, der unter Aufmerksamkeitskontrolle steht. Dieser Speicher unterstützt die Fähigkeit zum komplexen Denken (Baddeley, 2007, S. 1). Das AG ist kapazitätsbeschränkt und kann Informationen vorübergehend aufrechterhalten und speichern. Zudem unterstützt das AG menschliche Denkprozesse dadurch, dass es eine Schnittstelle zwischen Wahrnehmung, Langzeitgedächtnis und Handlung herstellt (Baddeley, 2003).

Im Gegensatz zum Kurzzeitgedächtnis (KZG) hat das AG aufgrund der Beteiligung von Aufmerksamkeitsprozessen eine Verarbeitungskomponente. Das KZG ist dagegen nur für die temporäre Speicherung von Informationen zuständig (Shipstead, Redick, & Engle, 2012).

Mehrere Arbeitsgedächtnismodelle wurden vorgeschlagen (z.B. Baddeley, Allen, & Hitch, 2011; Cowan, 1995; Oberauer, 2002; Unsworth & Engle, 2007a). Obwohl unterschiedliche Begriffe innerhalb der verschiedenen Modelle verwendet werden, haben diese Modelle einige Gemeinsamkeiten. Dazu zählen die Annahme eines beschränkten Kapazitätssystems sowie die Involvierung einer aufmerksamssteuernden Komponente und des Langzeitgedächtnisses. Aus den oben genannten AG-Modellen soll das Modell von Unsworth und Engle (2007a) näher dargestellt werden, da es validierten und häufig genutzten Aufgaben zur Erfassung der AG-Kapazität zugrunde liegt (siehe unten, Abschnitt „Erfassung der Arbeitsgedächtniskapazität“) und zur Erklärung von Unterschieden in der AG-Kapazität geeignet ist. Zudem ist es die Grundlage aktuell postulierter Mechanismen zur Erklärung von Unterschieden in der AG-Kapazität (Shipstead, Lindsey, Marshall, & Engle, 2014).

Nach Unsworth und Engle (2007a) sind *primary memory* (PM), *secondary memory* (SM) und eine Aufmerksamkeitskomponente die Bestandteile des AGs. In Anlehnung an die Terminologie von James (1890) verwenden Unsworth und Engle (2007a) den Begriff *primary memory* für eine Komponente zur Speicherung einer begrenzten Anzahl an Repräsentationen für die laufende Verarbeitung mittels kontinuierlicher Aufmerksamkeitszuwendung. In Übereinstimmung mit früheren Arbeiten (z.B. Broadbent, 1975; Cowan, 2001) gehen die Autoren davon aus, dass maximal circa vier Items im PM gespeichert werden können. Informationen, die aus dem PM verloren gehen bzw. dessen Kapazität übersteigen, können aus einem große-

ren Langzeitspeicher, dem SM durch Suchprozesse abgerufen werden. Diese Suchprozesse sind abhängig von vorhandenen Hinweisen (z.B. bestimmte Merkmale eines Kontexts). Für die Auswahl von Hinweisen für die Suche sowie die Festlegung, welche Hinweise für eine Suche kombiniert werden sollen, ist Aufmerksamkeit notwendig (Unsworth & Engle, 2007a).

In Anlehnung an das AG-Modell von Unsworth und Engle (2007a) wird davon ausgegangen, dass Unterschiede in der AG-Kapazität primär auf Variationen in der *executive attention* zurückgehen (Engle & Kane, 2004; Kane, Conway, Hambrick, & Engle, 2007). *Executive attention* spiegelt nach Redick, Broadway, et al. (2012) die Fähigkeit wider, zielrelevante Informationen vorübergehend im PM zu halten und Informationen aus dem SM abzurufen. Demnach gehen Unterschiede in der AG-Kapazität auf Unterschieden in der Fähigkeit zurück, Informationen durch kontinuierliche Aufmerksamkeitszuwendung aufrechtzuerhalten und relevante Informationen in der Gegenwart von Interferenz durch andere Reize abzurufen (Unsworth & Engle, 2007a). Die Ergebnisse einer aktuellen Studie von Shipstead et al. (2014) stehen im Einklang mit den Annahmen von Redick, Broadway, et al. (2012) sowie Unsworth und Engle (2007a) über die Mechanismen, die mit Unterschieden in der AG-Kapazität korrespondieren. Diese Mechanismen sind Behalten oder Vergessen im PM, Abruf aus dem SM und Aufmerksamkeitskontrolle (Shipstead et al., 2014). Folgt man dieser Logik, sind domänenübergreifende Unterschiede in Aufmerksamkeitsfähigkeiten wichtig für die Erklärung von Unterschieden in der AG-Kapazität. Diese Konzeptionalisierung findet sich in der Mehrzahl an Studien zum Training des AGs (siehe Abschnitt 2.4.1) wieder (Melby-Lervåg & Hulme, 2013). Nachfolgend wird auf die Erfassung der AG-Kapazität näher eingegangen.

Erfassung der Arbeitsgedächtniskapazität

In Übereinstimmung mit der Trennung zwischen KZG und AG unterscheiden Redick, Broadway, et al. (2012) simple span Aufgaben wie Digit Span und Corsi blocks von CSAs wie operation und symmetry span. Simple span Aufgaben messen lediglich die Kapazität eines kurzfristigen Speichersystems. Complex span Aufgaben erfordern zusätzlich die Verarbeitung von Informationen und erfassen daher die AG-Kapazität (Redick, Broadway, et al., 2012). Ein Beispiel für eine CSA stellt die automated operation span Aufgabe dar, die in Abschnitt 5.2.4 genauer beschrieben wird. In dieser Aufgabe wird zunächst eine mathematische Operation präsentiert und es muss entschieden werden, ob das anschließend dargestellte Ergebnis richtig oder falsch ist. Folglich wird ein Buchstabe präsentiert und nach einer variablen Anzahl an Operation-Buchstabe-Durchgängen müssen die Buchstaben in der korrekten Rei-

henfolge wiedergegeben werden (Unsworth, Heitz, Schrock, & Engle, 2005). Complex span Aufgaben wie die automated operation span Aufgabe haben sich als reliable und valide Indikatoren der AG-Kapazität erwiesen (für Überblicke siehe Redick, Broadway, et al., 2012; Unsworth, Redick, Heitz, Broadway, & Engle, 2009; Unsworth et al., 2005).

Nach Unsworth und Engle (2007b) messen simple und complex span Aufgaben zum Großteil dieselben Teilprozesse wie beispielweise inneres Memorieren der Items zwischen Darbietung und Wiedergabe, Behalten, Aktualisieren und kontrollierte Suche. Simple und complex span Aufgaben unterscheiden sich jedoch im Ausmaß, wie diese Teilprozesse in einer bestimmten Aufgabe operieren (Unsworth & Engle, 2007b). Trotz ähnlicher Basisprozesse ist die von Redick, Broadway, et al. (2012) vorgenommene Trennung zwischen simple und complex span Aufgaben für die Unterscheidung zwischen Maßen des KZG und Maßen des AGs hilfreich.

Shipstead, Redick und Engle (2010) merken an, dass simple span Aufgaben auch AG-Komponenten beanspruchen können. Voraussetzung ist, dass die zu memorierenden Listen die Kapazität des KZG überschreiten (Shipstead et al., 2010). Shipstead et al. (2010) schlussfolgern, dass CSAs bereits bei kurzen Listen prädiktiv für höhere kognitive Fähigkeiten sind, da sie die Verarbeitungskomponente von Beginn der Aufgabe an beinhalten.

Sowohl in simple span als auch in complex span Aufgaben werden unterschiedliche Materialien verwendet. In Abhängigkeit des verwendeten Materials wird die verbale oder visuell-räumliche Modalität des AGs beansprucht. In einigen AG-Trainingsstudien (siehe Abschnitt 2.4.1) wurde eine Unterscheidung vorgenommen zwischen Aufgaben, die die verbale und visuell-räumliche Modalität des KZG und AG beanspruchen.

Insgesamt stellen CSAs bewährte Aufgaben dar, um die AG-Kapazität zu erfassen. Nach Shipstead et al. (2012) wird die AG-Kapazität nicht durch eine Aufgabe allein repräsentiert. Folgt man dem Ansatz von Friedman et al. (2011, 2008), scheinen generell mehrere Aufgaben pro Konstrukt für die Erfassung exekutiver Funktionen empfehlenswert zu sein, da jede Aufgabe spezifische Varianz beinhaltet, die nicht mit exekutiven Funktionen assoziiert ist (Friedman et al., 2011; Miyake & Friedman, 2012). Eine genauere Schätzung für eine bestimmte exekutive Funktion erhält man durch Verwendung mehrerer Aufgaben pro Konstrukt.

2.2.2 Shifting

Konzeptionalisierung

Die Forschung zur kBf Shifting ist intensiv und hat in den letzten Jahren zugenommen. In diesem Abschnitt werden Definition und Charakteristiken von Shifting erläutert, die für das Verständnis dieser kBf hilfreich sind.

Wie in Kapitel 1 erwähnt, versteht man unter Shifting die Fähigkeit zum flexiblen Wechsel zwischen Aufgaben oder mentalen Voreinstellungen (Miyake & Friedman, 2012). Die mentalen Voreinstellungen können durch den Begriff *task set* konkretisiert werden. Darunter versteht man die Organisation kognitiver Prozesse und mentaler Repräsentationen, um den Aufgabenanforderungen zu entsprechen. Konkret muss ein *task set* die Repräsentation aufgabenrelevanter Stimuli und Antworten als auch die korrespondierenden Abbildungen von Stimulus-Antwort Paaren enthalten. Einige Aufgaben enthalten relativ leichte Stimulus-Antwort Paare wie beispielweise die Benennung von Objekten. In anderen Aufgaben ist es schwieriger Stimulus-Antwort Paare zu festigen weil diese Paare willkürlich sind, beispielsweise wenn die Farbe blau mit der linken Antworttaste und die Farbe rot mit der rechten Antworttaste assoziiert werden muss (Kiesel et al., 2010).

Im Kontext eines Aufgabenwechsels können mehrere Phänomene beobachtet werden (für einen Überblick siehe Monsell, 2003).

Durch den Wechsel zwischen Aufgaben bzw. mentalen Voreinstellungen entstehen messbare zeitliche Kosten. Wenn sich Sequenzen von Durchgängen mit gleicher Aufgabe mit Sequenzen von Durchgängen mit anderer, aber gleich bleibender Aufgabe abwechseln (gemischte Durchgänge), sind die Reaktionszeiten höher als bei nicht wechselnden Sequenzen. Diese zeitlichen Kosten werden als *mixing costs* bezeichnet (z.B. Koch, Prinz, & Allport, 2005; Rubin & Meiran, 2005) und stellen nach Kiesel et al. (2010) globale Kosten dar. Kosten in Reaktionszeiten, die durch den Wechsel zwischen *einzelnen* Aufgaben innerhalb von gemischten Durchgängen entstehen, werden als *switch costs* bezeichnet (Korbach & Kray, 2009) und sind lokale Kosten (Kiesel et al., 2010).

Switch costs können durch Vorbereitung auf die auszuführende Aufgabe reduziert werden. Wenn vor einer Aufgabe Wissen über diese vermittelt und Zeit zur Vorbereitung gegeben wird, sind die durchschnittlichen Kosten des Aufgabenwechsels normalerweise reduziert. Dieses Phänomen wird als *preparation effect* bezeichnet (Monsell, 2003). Der *prepara-*

tion effect zeigt sich auch bei Durchgängen ohne Aufgabenwechsel und konnte mehrfach belegt werden (für einen Überblick siehe Kiesel et al., 2010).

Die *switch costs* werden durch diesen Effekt allerdings nicht vollständig reduziert. Es bleiben sogenannte *residual costs* (z.B. Rogers & Monsell, 1995; Kimberg, Aguirre, & D'Esposito, 2000).

Um die mit einem Aufgabenwechsel verbundenen Phänomene zu untersuchen, wurden verschiedene Paradigmen entwickelt. Die Paradigmen spielen im Hinblick auf die Erfassung von Shifting eine Rolle.

Erfassung von Shifting

Zur Operationalisierung von Shifting gibt es eine Vielzahl an Aufgaben. Diese Aufgaben haben einen ähnlichen Aufbau. Allgemein müssen zunächst zwei oder mehrere Aufgaben mit einfachen Stimuli geübt werden. Jede Aufgabe erfordert die Aufmerksamkeit auf einen unterschiedlichen Bestandteil oder ein unterschiedliches Merkmal eines Stimulus und dessen Klassifizierung. Möglich ist auch, dass eine unterschiedliche Eigenschaft des Stimulus aus dem Gedächtnis abgerufen oder errechnet werden muss. Dann wird in jedem Durchgang einer Serie von Durchgängen ein Stimulus präsentiert und eine der Aufgaben muss bearbeitet werden. Welche Aufgabe bearbeitet werden muss, wird durch verschiedene Methoden mitgeteilt. Allen gemeinsam ist, dass die Aufgabe manchmal von einem zum anderen Durchgang wechselt und manchmal nicht (Monsell, 2003).

Die verschiedenen Aufgaben zur Erfassung von Shifting lassen sich bestimmten Paradigmen zuordnen. Ein häufig angewendetes Paradigma stellt das *task-cuing paradigm* dar (für einen Überblick siehe Kiesel et al., 2010). *Switch costs* lassen sich mit diesem Paradigma robust nachweisen (Koch, 2001; Meiran, 1996; Meiran, Chorev, & Sapir, 2000). Beim *task-cuing* Paradigma wird in jedem Durchgang vor oder mit dem Stimulus ein Hinweis präsentiert, der die aktuelle Aufgabe anzeigt. Die Reihenfolge der Aufgaben sowie die Reihenfolge der Aufgabenwechsel und -wiederholungen sind zufällig (Kiesel et al., 2010).

Beispiele für Aufgaben, die nach dem *task-cuing* Paradigma konstruiert wurden, sind die Aufgaben color-shape, number-letter und category-switch. Diese Aufgaben sind detailliert in Abschnitt 5.2.4 beschrieben. Mehrere Gründe sprechen für die Verwendung dieser Aufgaben.

Neuropsychologische Maße von Shifting weisen oft eine geringe Reliabilität auf und beinhalten große Varianzanteile, die nicht mit Shifting in Verbindung stehen (Friedman et al.,

2006). Die Aufgaben color-shape, number-letter und category-switch erwiesen sich in den Studien von Friedman et al. (2011) und Friedman et al. (2006) als reliable Indikatoren von Shifting. Diese Aufgaben sind einfacher aufgebaut und besser analysiert als komplexe neuropsychologische Shifting-Aufgaben. Folglich sind die kognitiven Anforderungen besser verstanden (Friedman et al., 2006).

2.2.3 Fluide Intelligenz

Konzeptionalisierung

Knapp formuliert ist Gf “the ability to reason and to solve novel problems” (König et al., 2005, S. 245). Diesen Autoren zu Folge meint Gf also die Fähigkeit zum schlussfolgernden Denken, um neue Probleme zu lösen. Fluide Intelligenz bezieht sich auf das Lösen von Problemen, wenn Vorerfahrungen eine untergeordnete Rolle spielen (Schneider & McGrew, 2012). Fluide Intelligenz wird in der Intelligenztheorie von Raymond Cattell (1963) von kristalliner Intelligenz (Gc) unterschieden. Kristalline Intelligenz bezieht sich auf die Anwendung vorhandenen Wissens (Cattell, 1963). Das vorhandene Wissen hat eine bedeutsame Beziehung mit Gf, was in Abschnitt 2.3.3 dargelegt wird.

Fluide Intelligenz kann weiterhin in mehrere Subkomponenten ausdifferenziert werden. Schneider und McGrew (2012) unterteilen Gf in die drei Faktoren induktives Denken, deduktives Denken und quantitatives Denken. Induktives Denken meint die Fähigkeit, die dem Auftreten eines Phänomens (z.B. bestimmte Stimuli) zugrundeliegenden Prinzipien oder Regeln zu erkennen. Deduktives Denken ist die Fähigkeit, unter Rückgriff auf bekannte Prinzipien neue Fakten zu erkennen oder Probleme zu lösen. Quantitatives Denken spiegelt die Fähigkeit wider, entweder induktiv oder deduktiv Schlüsse aus Zahlen, mathematischen Beziehungen und Operatoren zu ziehen (Schneider & McGrew, 2012). Diese Ausdifferenzierung von Gf in drei Faktoren ist für die Erfassung von Gf bedeutsam.

Erfassung fluider Intelligenz

Induktives Denken wird mit Analogie-Aufgaben und Aufgaben mit geometrischen Matrizen operationalisiert. Deduktives Denken wird mit Aufgaben erfasst, die deduktives logisches Denken erfordern, wobei Schlussfolgerungen aus Prämissen gezogen werden. Quantitatives Denken wird mit Aufgaben erfasst, die Deduktion und Induktion mit Zahlen und mathematische Operationen erfordern. Für eine Operationalisierung der Faktoren induk-

tives Denken, deduktives Denken und quantitatives Denken erscheint es angemessen, mehrere Maße von Gf zu verwenden. Tatsächlich beinhalten viele Intelligenztestbatterien mehr als einen Test zur Erfassung von Gf (für einen Überblick siehe Primi, 2014).

Intelligenztestbatterien sind oftmals sehr zeitaufwendig und erhöhen die Dauer eines Experiments beträchtlich. Weniger zeitaufwendig sind dagegen adaptive Tests, die an das Leistungsniveau der Testteilnehmer angepasst sind. Verglichen mit Tests, bei denen die Items in einer festen Reihenfolge vorgegeben werden, ist bei adaptiven Tests bei einer geringeren Itemanzahl die Genauigkeit der Ergebnisse höher (Arendasy et al., 2012). Adaptive Tests zur Erfassung von Gf, die eine Einstellung der Messgenauigkeit erlauben, sind die drei Subtests „Numerisch-induktives Denken“, „Figural-induktives Denken“ und „Verbal-deduktives Denken“ der Intelligenz-Struktur-Batterie (INSBAT; Arendasy et al., 2012). Mit diesen Subtests werden die drei von Schneider und McGrew (2012) genannten Subkomponenten von Gf operationalisiert.

2.3 Bedeutung kognitiver Basisfunktionen und fluider Intelligenz

2.3.1 Arbeitsgedächtnis

Wie in Abschnitt 2.2.1 dargestellt, gehen Unterschiede in der AG-Kapazität primär auf Unterschiede in der *executive attention* zurück. Wenig überraschend ist daher, dass Unterschiede in der AG-Kapazität mit Unterschieden in der Leistung in grundlegenden Aufmerksamkeitsmaßen in Verbindung gebracht werden (Unsworth & Engle, 2007a).

Die AG-Kapazität weist des Weiteren Zusammenhänge zu bedeutsamen, kognitiven Leistungen auf. Für die AG-Kapazität wurden Zusammenhänge gefunden mit der Mathematikleistung (Gathercole, Pickering, Knight, & Stegmann, 2004), dem Leseverständnis (Daneman & Merikle, 1996), Leistungen in Chemie (Danili & Reid, 2004; Tsaparlis, 2005) bzw. Aufgaben, die höhere kognitive Fähigkeiten beanspruchen (Engle, 2001). So ist die Arbeitsgedächtniskapazität auch mit Problemlösen (Bühner, Kröner, & Ziegler, 2008; Wittmann & Süß, 1999) und Multitasking assoziiert (Bühner, König, Pick, & Krumm, 2006; König et al., 2005; Hambrick, Oswald, Darowski, Rench, & Brou, 2010). Insbesondere der Zusammenhang zwischen der AG-Kapazität und Gf ist mehrfach belegt (z.B. Ackerman et al., 2002; Bühner, Krumm, Ziegler, & Plücken, 2006; Engle et al., 1999; Oberauer et al., 2008; Redick, Unsworth, Kelly, & Engle, 2012). Beispielsweise fanden Ackerman et al. (2002) eine Korrelation von $r = .48$ zwischen der AG-Kapazität und Gf. Redick, Unsworth, et al. (2012) erhielten einen Wert von $r = .53$. Für die Beziehung zwischen AG-Kapazität und Gf wurden ver-

schiedene Mechanismen verantwortlich gemacht. Beispielsweise konstatieren Unsworth und Engle (2007a), dass Speicherung, Abruf und Aufmerksamkeitsfähigkeiten für Unterschiede in der AG-Kapazität und fluiden Intelligenz von Bedeutung sind. Andere Studien deuten darauf hin, dass insbesondere die Speicherung von Informationen einen Einfluss auf die Beziehung zwischen AG-Kapazität und Gf haben (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Fukuda, Vogel, Mayr, & Awh, 2010). Ungeachtet dieser Mechanismen besteht ein bedeutsamer Zusammenhang zwischen AG-Kapazität und Gf. In der Studie 1 der vorliegenden Arbeit wurden sowohl die AG-Kapazität als auch Gf erfasst, um bei der Analyse der Moderatorenrolle einer der beiden Variablen für den Einfluss der jeweils anderen Variable kontrollieren zu können.

Neben den beschriebenen Zusammenhängen der AG-Kapazität mit mehreren kognitiven Fähigkeiten und Fertigkeiten, ist die wichtige Rolle der AG-Kapazität für Lernprozesse hervorzuheben. Die AG-Kapazität ist beispielsweise mit dem Kategorienlernen (Lewandowsky, Newell, Yang, & Kalish, 2012), dem Lernen einer Programmiersprache (Shute, 1991) und dem Wissenserwerb aus Wikipedia-ähnlichen Lernmaterialien (Banas & Sanchez, 2012) assoziiert. In der Studie von Banas und Sanchez (2012) sollten Studenten in einer Lernphase zwischen Pre- und Posttest Inhalte von einer Wikipedia-ähnlichen Website lesen und parallel dazu Fragen beantworten. Bei den Fragen handelte es sich um einfache Faktenfragen, die in der Lernphase enthalten waren, da die Studenten ähnlich wie in einer authentischen Lernumgebung in mehrere Prozesse durch das Lernen mit multiplen Dokumenten involviert sein sollten. In einem Pre-Post-Vergleich zeigte sich, dass Studenten mit höherer AG-Kapazität (operationalisiert mit der automatisierten Version der operation span Aufgabe) die zugrundeliegende, nicht durch das Material explizierte Struktur komplexer Probleme besser gelernt hatten als Studenten mit geringer AG-Kapazität. Dieser Zusammenhang bestand unabhängig vom Vorwissen der Studenten. Kein Zusammenhang fand sich zwischen der AG-Kapazität und dem Lernen einfacher, expliziter Informationen. Die Ergebnisse ihrer Studie sehen Banas und Sanchez (2012) konform mit mehreren Studien, die nahe legen, dass Personen mit einer höheren AG-Kapazität Informationen besser über Textteile hinweg verbinden können (Chiappe, Hasher, & Siegel, 2000; Daneman & Green, 1986; Masson & Miller, 1983; Singer, Andrusiak, Reisdorf, & Black, 1992; Yuill, Oakhill, & Parkin, 1989). Es wurde geschlossen, dass individuelle Unterschiede einen moderierenden Einfluss auf das Verstehen von, nicht durch das Lernmaterial explizierten Informationen haben. Instruktionale Unterstützungsmaßnahmen wie organisierende Informationen könnten für Personen mit niedriger AG-Kapazität beim Lernen mit mehreren Dokumenten hilfreich sein (Banas & Sanchez,

2012).

Zusammenfassend korrespondieren Unterschiede in der AG-Kapazität mit Unterschieden in einer Vielzahl kognitiver Leistungen. Das AG hat eine wichtige Rolle für das Lernen, da es einen mentalen Arbeitsspeicher bereitstellt, in dem der Lerner Informationen speichern kann, während andere mentale Aktivitäten ausgeführt werden. Lernende mit geringer AG-Kapazität haben Probleme mit dieser Operation und können nicht alle Informationen mental verfügbar halten während eine Aufgabe bearbeitet wird (Gathercole & Alloway, 2007). Wie in Kapitel 3 noch detailliert beschrieben wird, sollte das AG bei Entscheidungen bezüglich des Instruktionsdesigns als eine wichtige Variable berücksichtigt werden. Ob der Einfluss des Grads an instruktionaler Unterstützung auf den Wissenserwerb tatsächlich vom AG abhängt, wurde allerdings bisher kaum systematisch untersucht.

2.3.2 Shifting

Im Vergleich zur Bedeutung der AG-Kapazität für kognitive Leistungen ist die Relevanz von Shifting für kognitive Leistungen weniger intensiv erforscht. Allerdings gibt es auch bedeutsame Befunde zum Zusammenhang zwischen Shifting und akademischen Fertigkeiten.

Schon in frühen Jahren ist die Shifting-Fähigkeit für die erfolgreiche Bewältigung schulischer Aufgaben wichtig (für einen Überblick siehe Best, Miller, & Jones, 2009). Im Speziellen ist Shifting für die Leistung in komplexen, schulischen Aufgaben wichtig, die einen Wechsel zwischen verschiedenen Aspekten eines Problems oder arithmetischen Strategien erfordern (Agostino, Johnson, & Pascual-Leone, 2010; Blair, Knipe, & Gamson, 2008; Van der Sluis, de Jong, & Van der Leij, 2007)

Die Bedeutsamkeit von Shifting für schulische Leistungen wurde von Yeniad et al. (2013) anhand zweier Metaanalysen belegt. Diesen Metaanalysen zu Folge ist Shifting signifikant mit Lese- und Mathematikleistungen sowie der Intelligenz (IQ) assoziiert. Die Autoren schlussfolgern, dass die Zusammenhänge zwischen Shifting, schulischen Fertigkeiten und IQ domänenübergreifend sind (Yeniad et al., 2013).

2.3.3 Fluide Intelligenz

Neben kBf scheint auch Gf wichtig für kognitive Leistungen zu sein. Nach der Investment Theorie von Raymond Cattell (1971) fördert Gf Gc und hat daher einen Einfluss auf die Wissensaneignung. Aktuelle Studien unterstützen diese Theorie (z.B. Bühner, Krumm, et al., 2006; Kvist & Gustafsson, 2008; Ziegler et al., 2012). Die Ergebnisse der Studie von

Ziegler et al. (2012) zeigen, dass Gc im Alter von 23 Jahren durch individuelle Unterschiede in Gf im Alter von 17 Jahren vorhergesagt wird. Fluide Intelligenz hat demnach einen nachhaltigen Einfluss auf den Wissenserwerb.

Der Einfluss von Gf auf Lernen scheint bei neuen und komplexen Aufgaben am offensichtlichsten zu sein. In Übereinstimmung mit dieser Annahme legen Studien nahe, dass Gf prädiktiv für das Lösen neuer und komplexer Probleme ist (z.B. Ackerman & Cianciolo, 2002; Snow et al., 1984). So kann der Lernerfolg in der komplexen Domäne Mathematik durch individuelle Unterschiede in Gf vorhergesagt werden (Primi, Ferrão, & Almeida, 2010).

Zusammenfassend sind kBf und Gf wichtig für kognitive Leistungen in verschiedenen Domänen. Insbesondere die AG-Kapazität wird als wichtige Variable in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung gesehen, worauf in Kapitel 3 näher eingegangen wird. Der folgende Abschnitt thematisiert Beeinflussungsmöglichkeiten der kognitiven Basisfunktionen AG und Shifting.

2.4 Trainierbarkeit kognitiver Basisfunktionen

In diesem Abschnitt wird auf die Beeinflussungsmöglichkeiten der kognitiven Basisfunktionen AG und Shifting durch Trainings eingegangen. Anzumerken ist, dass auch Trainings zur Verbesserung von Gf bzw. im engeren Sinne induktivem Denken vorhanden sind (für eine Metaanalyse siehe Klauer, 2014). Darauf wird jedoch nicht in einem separaten Abschnitt, sondern in der Gesamtdiskussion (Kapitel 7) kurz eingegangen, da der Fokus der Arbeit auf kBf liegt.

2.4.1 Arbeitsgedächtnistrainings

Nachfolgend werden die Idee und Transfereffekte von AG-Trainings sowie Trainingsprogramme zur Verbesserung des AGs thematisiert. Zudem wird auf bereits publizierte Reviews und Unzulänglichkeiten in der bisherigen Forschung zu AG-Trainings eingegangen.

Aufgrund der Bedeutung des AGs für wichtige kognitive Leistungen ist eine typische Idee von AG-Trainings, dass Verbesserungen im AG auf andere kognitive Funktionen generalisiert werden. Berücksichtigt man die Ursachen für individuelle Unterschiede in der AG-Kapazität (siehe Abschnitt 2.2.1), sollte ein AG-Training zu einer Verbesserung in der *executive attention* führen. Die Verbesserung in der domänenübergreifenden Aufmerksamkeitsfähigkeit sollte zu Verbesserungen führen in Aufgaben zur Erfassung der AG-Kapazität

sowie Aufgaben, die andere kognitive Fähigkeiten oder Fertigkeiten erfassen und *executive attention* erfordern (z.B. Raven Matrizen; siehe Melby-Lervåg & Hulme, 2013).

Für eine Systematisierung von Transfereffekten von AG-Trainings ist eine Unterscheidung zwischen nahen und weiten Transfereffekten hilfreich. Nahe Transfereffekte schließen Verbesserungen in KZG- und AG-Aufgaben mit ein. Weite Transfereffekte beziehen sich auf Verbesserungen in Maßen, die über KZG- und AG-Aufgaben hinausgehen. Mehrere Studien zu AG-Trainings wurden veröffentlicht, die nahe und weite Transfereffekte nachweisen konnten. Zum Beispiel wurden nahe Transfereffekte auf das verbale und visuell-räumliche KZG beobachtet (z.B. Brehmer, Westerberg, & Bäckman, 2012; Harrison et al., 2013). Nahe Transfereffekte auf verbales und visuell-räumliches AG wurden von Alloway, Bibile und Lau (2013) sowie Dunning, Holmes und Gathercole (2013) berichtet. Weite Transfereffekte zeigten sich unter anderem auf mathematisches Schlussfolgern (Holmes, Gathercole, & Dunning, 2009), Leseverständnis (Chein & Morrison, 2010) und Gf (z.B. Klingberg et al., 2005; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008, Jaeggi, Studer-Luethi, et al., 2010; Jaušovec & Jaušovec, 2012). Morrison und Chein (2011) stellten heraus, dass sogenannte *Core Trainings* geeigneter als *Strategy Trainings* sind, um weite Transfereffekte zu erzielen. *Strategy Trainings* beinhalten effektive Vorgehensweisen um Enkodierung, Behalten und/oder Erinnerung aus dem AG zu unterstützen. Im Gegensatz dazu zielen *Core Trainings* darauf ab, die Nutzung domänenspezifischer Strategien zu begrenzen und Automatisierung zu minimieren. Bei *Core Trainings* werden Aufgaben verwendet, die adaptiv sind und typischerweise häufiges Updating erfordern (Morrison & Chein, 2011). Bei adaptiven Trainings wird die Trainingsschwierigkeit durchgehend an die aktuelle Leistung der Teilnehmer angepasst (Jaeggi, Studer-Luethi, et al., 2010).

Mehrere Ansätze wurden verfolgt, um mit adaptiven Trainings das AG zu verbessern. In vielen Studien wurden eigens entwickelte Programme wie Training mit dual tasks verwendet (z.B. Jaeggi et al., 2008, 2014; Redick et al., 2013). Ein Beispiel für eine dual task ist das in der Studie von Jaeggi et al. (2008) verwendete n-back Training. Bei diesem Training bekommen Teilnehmer eine Sequenz blauer Quadrate an acht verschiedenen Stellen einer Matrix am Bildschirm präsentiert. Parallel dazu hören die Teilnehmer über Kopfhörer eine Sequenz von Konsonanten. Die Teilnehmer müssen beide Reizströme überwachen und angeben, wann ein Reiz (blaues Quadrat oder Konsonant) mit dem n Positionen vorher präsentierten Reiz übereinstimmt (Jaeggi et al., 2008). Folglich werden durch die n-back Aufgabe die Speicher- und die Verarbeitungskomponente des AGs beansprucht. Neben eigens entwickelten Programmen machen mehrere Trainingsparadigmen Gebrauch von kommerzieller Software.

Eine der am häufigsten benutzten kommerziellen Software stellt Cogmed (<http://cogmed.com/>) dar. Andere kommerzielle Programme, die in AG-Trainings verwendet wurden, sind Jungle Memory (<http://junglememory.com/>) und Cognifit (<https://www.cognifit.com/>). Mehrfach wurde behauptet, dass diese Programme effektiv sind, um Aufmerksamkeit, IQ und Schulnoten zu verbessern (siehe Melby-Lervåg & Hulme, 2013). Die Aufgaben in allen drei kommerziellen Programmen sind adaptiv und beanspruchen sowohl das verbale als auch das visuell-räumliche AG. Eine Limitation dieser Programme betrifft ihre theoretische Fundierung. Es ist unklar, welche Eigenschaften und Aufgabencharakteristiken dieser Programme für Verbesserungen in der AG-Kapazität verantwortlich sind (Melby-Lervåg & Hulme, 2013).

Bisherige Reviews

In ihrem meta-analytischen Review diskutieren Melby-Lervåg und Hulme (2013) einige narrative Reviews über die Effekte von AG-Trainings. Nachfolgend werden einige Hauptaspekte der von Melby-Lervåg und Hulme (2013) diskutierten Reviews beschrieben. Weiterhin werden die Ergebnisse und Schlussfolgerungen der Metaanalyse von Melby-Lervåg und Hulme (2013) thematisiert. Am Ende dieses Abschnitts werden Trainingsbedingungen als potentielle Moderatoren von Transfereffekten von AG-Trainings adressiert, die bisher vernachlässigt wurden.

Hinsichtlich der Effektivität von AG-Trainings wurden optimistische Perspektiven geäußert. Klingberg (2010) folgerte auf der Basis ausgewählter Studien, dass AG-Trainings zu Leistungsverbesserungen in nichttrainierten Aufgaben führen können, die sich auf das AG und die Kontrolle von Aufmerksamkeit stützen. Auf Basis neuerer Studien mit Kindern argumentieren Titz und Karbach (2014), dass ein Training des AGs für schulische Fertigkeiten nützlich sein kann, insbesondere für Sprache, Lesen und in geringerem Ausmaß mathematische Fertigkeiten. Andere Untersuchungen lassen jedoch Bedenken über AG-Trainings aufkommen (Morrison & Chein, 2011; Melby-Lervåg & Hulme, 2013; Shipstead et al., 2012). Morrison und Chein (2011) merken in ihrem Review an, dass Anstrengungs-/Erwartungseffekte auf Ergebnisse (z.B. bessere Ergebnisse der Experimentalgruppe aufgrund von Verbesserungserwartungen der Lerner) nicht kontrolliert werden. Den Autoren zu Folge werden diese und ähnliche Effekte nicht durch Verwendung geeigneter Maße kontrolliert, wie z.B. Selbstberichte und Maße für Motivation sowie Commitment. Um für Erwartungseffekte zu kontrollieren, schlagen Shipstead et al. (2012) den Einschluss einer Kontrollgruppe vor, die ein Training

erhält, welches das AG nicht beansprucht aber ähnlich wie die typischen AG-Trainingsaufgaben adaptiv ist. Eine weitere Limitation betrifft die Erfassung von Transfereffekten. Morrison und Chein (2011) stellen heraus, dass Trainingsparadigmen sowie Aufgaben für die Erfassung von Transfereffekten hoch variabel sind. Diese Autoren äußerten auch Bedenken über den Nachweis von Transfereffekten mit Aufgaben, die nur einen Aspekt eines Konstrukts erfassen (z.B. Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010). Um Bedenken darüber zu minimieren, dass aufgabenspezifische Überlappungen zwischen trainierten Aufgaben und Transferaufgaben für Transfereffekte verantwortlich sind, schlugen Morrison und Chein (2011) vor, mehrere überlappende Aufgaben für die Messung einer latenten, kognitiven Fähigkeit zu verwenden. Übereinstimmend mit dieser Forderung sollten nach Shipstead et al. (2012) die interessierenden Fähigkeiten und Fertigkeiten mit mehreren Instrumenten erfasst werden. Des Weiteren werden nahe Transfereffekte auf AG-Komponenten oftmals mit Aufgaben nachgewiesen, die das KZG erfassen. Validere AG-Aufgaben wie CSAs wurden bisher nicht konsistent genutzt. In einigen Studien konnte ein Transfer zu CSAs nicht festgestellt werden und nahe Transfereffekte sind möglicherweise auf Überlappungen zwischen Trainings- und Transferaufgaben zurückzuführen (Shipstead et al., 2012). Selbst wenn ein Transfer auf CSAs nachgewiesen wurde, sind die dahinter stehenden, spezifischen Mechanismen noch wenig verstanden (Morrison & Chein, 2011; Titz & Karbach, 2014; von Bastian & Oberauer, 2013).

Neben den Mechanismen, die zum Transfer führen, wurden die Trainingsbedingungen, welche einen Einfluss auf Trainingsresultate haben könnten, kaum beachtet (von Bastian & Oberauer, 2013). Klingberg (2010) hebt hervor, dass die Rolle von Variablen wie Dauer und Zeitabstand des Trainings für die Erzielung von Transfereffekten und dauerhaften Verbesserungen noch schlecht verstanden sind. Faktoren, welche auf die Zeitkonfiguration von Trainings bezogen sind (z.B. Dauer einzelner Trainingssessions, Häufigkeit des Trainings pro Woche und das Zeitintervall zwischen einzelnen Sessions), wurden bisher kaum systematisch beachtet. Beispielsweise wurde bezüglich des Zeitintervalls festgestellt, dass verteiltes Lernen vorteilhafter als massiertes Lernen ist (z.B. Baddeley & Longman, 1978; Bloom & Shuell, 1981). Im Hinblick auf AG-Trainings fanden Penner et al. (2012) einen Vorteil des verteilten Trainings (2x pro Woche über 8 Wochen) im Vergleich zum massierten Training (4x pro Woche über 4 Wochen). Ob das Zeitintervall unter Berücksichtigung einer Vielzahl an AG-Trainingsstudien einen moderierenden Einfluss auf Transfereffekte hat, ist unklar. Eine Untersuchung der genannten Variablen, die auf die Zeitkonfiguration von Trainings bezogen sind, könnte dazu beitragen, die Mechanismen besser zu verstehen und das Trainingsdesign zu ver-

bessern, um Transfereffekte zu optimieren. Des Weiteren bedarf es der Untersuchung des Einflusses der durch die Trainingsaufgabe beanspruchten Modalität des AGs (verbal, visuell-räumlich, oder beides) auf diese Effekte (Titz & Karbach, 2014). Studien unterscheiden sich in Hinblick auf die Identifizierung der Modalitäten, die zu größeren Transfereffekten führen, was für die Gestaltung optimaler Trainingsaufgaben hilfreich wäre. Aus theoretischer Perspektive ist von Interesse, in welchem Ausmaß Transfereffekte durch die verbale oder die visuell-räumliche Modalität oder beide Modalitäten moderiert werden.

Transfereffekte von AG-Trainings und einige Moderatoren dieser Effekte wurden von Melby-Lervåg und Hulme (2013) in einer Metaanalyse untersucht, die auf 23 Studien und 30 Gruppenvergleichen basierte. Die Autoren fanden kurzfristige Effekte auf verbales und visuell-räumliches AG und in begrenztem Maß einen Nachweis für andauernde Effekte auf das visuell-räumliche AG zum Follow-up. Die Effektstärke für den kurzfristigen Effekt auf das verbale AG betrug $d_s = 0.79$ und die Effektstärke für den kurzfristigen Transfereffekt auf das visuell-räumliche AG betrug $d_s = 0.52$. Die Effektstärke für den langfristigen Transfereffekt auf das visuell-räumliche AG war $d_s = 0.41$. Melby-Lervåg und Hulme (2013) fanden kleine, kurzfristige Transfereffekte auf die nonverbale Fähigkeit und Aufmerksamkeit (gemessen mit der Stroop Aufgabe), aber keine Transfereffekte auf die verbale Fähigkeit, arithmetische Fähigkeiten und Fertigkeiten und die Fertigkeit zur Wortdekodierung. Die Autoren berichteten eine signifikante Heterogenität zwischen den Studien für kurzfristige und langfristige Transfereffekte auf das verbale AG. Studien unterschieden sich zudem signifikant bezüglich kurzfristiger Transfereffekte auf das visuell-räumliche AG und die nonverbale Fähigkeit.

Melby-Lervåg und Hulme (2013) untersuchten die folgenden Moderatoren: Alter (jüngere Kinder, ältere Kinder, jüngere Erwachsene, ältere Erwachsene); Trainingsdosis (Gesamtdauer des Trainings); Designtyp (randomisierte oder nichtrandomisierte Studien); Art der Kontrollgruppe (aktiv oder passiv); Lernstatus der Teilnehmer (Teilnehmer mit oder ohne Lernschwierigkeiten); und Interventionstyp (Art des Trainingsprogramms). Das Alter war ein signifikanter Moderator von Transfereffekten auf das verbale AG, wobei die Zuwächse im verbalen AG bei jüngeren Kindern größer als bei älteren Kindern waren. Die Effektstärken des Cogmed AG-Trainings (ein kommerzielles Trainingsprogramm) auf das visuell-räumliche AG waren größer als die Effektstärken nicht-kommerzieller Programme. Zwischen dem Cogmed AG-Training und den anderen kommerziellen Programmen gab es keinen Unterschied. Die Art der Kontrollgruppe war ein signifikanter Moderator von Transfereffekten auf die nonverbale Fähigkeit. Bei passiven Kontrollgruppen waren die Zuwächse in der nonverbalen Fähigkeit größer als bei aktiven Kontrollgruppen. Den Autoren zu Folge gibt es bei Maßen

für weite Transfereffekte wahrscheinlich keine systematische Heterogenität zwischen den Studien, die durch Moderatoren erklärt werden könnte. Melby-Lervåg und Hulme (2013) kommen zu dem Schluss, dass die Idee der Verbesserung kognitiver Fähigkeiten bei Kindern und gesunden Erwachsenen durch AG-Trainings zumindest zweifelhaft ist.

Es bleibt fraglich, ob die Schlussfolgerungen von Melby-Lervåg und Hulme (2013) vor dem Hintergrund mehrerer, kürzlich veröffentlichter Studien immer noch angemessen sind. Was die Bedingungen von AG-Trainings betrifft, könnten weitere Trainingsbedingungen einen Einfluss auf Transfereffekte von AG-Trainings haben. Diese Trainingsbedingungen wurden allerdings noch nicht systematisch berücksichtigt. Im Hinblick auf die trainierte Aufgabe variiert neben der trainierten Modalität auch das Feedback über die Ergebnisse. In einigen Studien erhielten die Teilnehmer nur Informationen über die Richtigkeit ihrer Antworten, d.h., reines Ergebniswissen (z.B. Heinzl et al., 2014; Van der Molen, Van Luit, Van der Molen, Klugkist, & Jongmans, 2010). In anderen Studien erhielten die Teilnehmer elaborierteres Feedback, beispielsweise über die individuelle Verbesserung (Alloway et al., 2013; Egeland, Aarli, & Saunes, 2013). Die Berücksichtigung des Feedbacks über die trainierte Aufgabe scheint wichtig zu sein, da die Art des Feedbacks differenzielle Effekte auf Lernen und Transfer haben kann (Hattie & Timperley, 2007). Andere Variablen betreffen den Trainingsprozess vom Beginn bis zum Ende des Trainings. Eine offensichtlich wichtige Variable könnte Supervision sein, d.h., ob die Trainingsaktivitäten der Teilnehmer durch jemand anderes überwacht oder gelenkt wurden. In einigen Studien überwachten Experimentatoren oder andere Personen wie Eltern die Teilnehmer, ob sie richtig trainierten (z.B. Borella, Carretti, Riboldi, & De Beni, 2010; Holmes & Gathercole, 2013). In anderen Untersuchungen waren Experimentatoren nur anwesend, ohne das Training der Teilnehmer zu überwachen oder einzugreifen (z.B. Horowitz-Kraus & Breznitz, 2009; Salminen, Strobach, & Schubert, 2012). Der Einfluss der bloßen Anwesenheit von jemand anderem auf die Aufgabenleistung ist mehrfach belegt (z.B. Aiello & Douthitt, 2001; Guerin, 1986). Die Leistung in komplexen Aufgaben wird durch die bloße Anwesenheit anderer Personen negativ beeinflusst (für einen Überblick siehe Aiello & Douthitt, 2001). Eines der Anliegen der, in Kapitel 6 dargestellten Metaanalyse ist daher, den Effekt der Supervision zu untersuchen, d.h., ob jemand anderes während des Trainings anwesend ist *und* das Training überwacht sowie gewährleistet, dass die Teilnehmer der Trainingsinstruktion richtig folgen.

Eine andere Prozessvariable, die Beachtung verdient, ist die instruktionale Unterstützung, d.h., ob Teilnehmer zusätzliche Erklärungen und Hilfe während des Trainings bekommen. In einigen Studien erhielten Teilnehmer nur *am Anfang* des Trainings Instruktionen dar-

über, wie sie Aufgaben ausführen sollen (z.B. Jaeggi et al., 2008; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009). Dagegen erhielten Teilnehmer in anderen Studien zusätzlich *während* des Trainings Instruktionen über die Trainingsaufgaben (z.B. Heinzl et al., 2014; Jaušovec & Jaušovec, 2012). Zusätzliche Instruktionen während des Trainings könnten Teilnehmern helfen, die Aufgabe richtig zu bearbeiten und die Effektivität des Trainings verbessern. In der, in Kapitel 6 beschriebenen Metaanalyse wurde die instruktionale Unterstützung als Moderator mit eingeschlossen, um zu untersuchen, ob sie einen Einfluss auf Transfereffekte von AG-Trainings hat.

Eine weitere Variable, die bisher noch nicht systematisch berücksichtigt wurde, ist der Trainingsort. Die meisten AG-Trainings fanden im Labor statt, aber in einigen Studien trainierten Teilnehmer an einem anderen Ort, wie beispielsweise zu Hause (Brehmer et al., 2011, 2012). Es ist plausibel, dass im Labor weniger Stör- und Ablenkungsquellen als an anderen Orten vorhanden sind. Der Trainingsort wurde in der, in Kapitel 6 dargestellten Metaanalyse mit eingeschlossen, um zu untersuchen, ob Trainings in oder außerhalb des Labors die Trainingsergebnisse unterschiedlich beeinflussen.

Die dargestellten Trainingsbedingungen könnten bedeutsame Anteile in der Variabilität der Transfereffekte von Trainingsstudien erklären. Somit könnte die Identifizierung von Moderatoren unter den Trainingsbedingungen zur verbesserten Trainingsgestaltung für optimierte Ergebnisse beitragen.

Zusammenfassend kann festgestellt werden, dass zwar Analysen über die Effektivität von AG-Trainings als Intervention zur Verbesserung kognitiver Funktionen durchgeführt wurden (z.B. Klingberg, 2010; Titz & Karbach, 2014), in der bisherigen Forschung jedoch einige Unzulänglichkeiten existieren. In die Metaanalyse von Melby-Lervåg und Hulme (2013) wurden Studien bis zum 5.11.2011 eingeschlossen. Seit diesem Datum wurden mehrere neue Studien veröffentlicht, weshalb sich die Frage stellt, ob die übergreifenden Schlussfolgerungen zu Transfereffekten von AG-Trainings in früheren Studien im Lichte neuerer empirischer Evidenz immer noch valide sind. Des Weiteren betrachteten Melby-Lervåg und Hulme (2013) nicht den Einfluss mehrerer Trainingsbedingungen auf Transfereffekte von AG-Trainings. Plausibel ist, dass die spezifischen Bedingungen unter denen AG-Trainings stattfinden, die Trainingseffektivität und Transfereffekte wesentlich beeinflussen. Allerdings wurden die Effekte der Modalität der trainierten Aufgabe, von Supervision, Trainingsort und anderen Trainingsbedingungen noch nicht systematisch untersucht. Die in Kapitel 6 dargestellte Metaanalyse fokussiert daher auf Trainingsbedingungen als weitere, potentielle Moderatoren.

2.4.2 Trainings zur Verbesserung von Shifting

Wie bereits in Abschnitt 2.3.2 erwähnt, ist die Forschung zur Relevanz von Shifting für kognitive Leistungen verglichen mit der Forschung zur Bedeutung der AG-Kapazität für kognitive Leistungen weniger stark ausgeprägt. Zudem gibt es weitaus weniger Studien zur Trainierbarkeit von Shifting und Transfereffekten von Shifting-Trainings. Darauf wird in diesem Abschnitt eingegangen.

Einige Studien konnten zeigen, dass Shifting trainierbar ist. In diesen Studien konnte eine Reduktion von *switch costs* durch das Training mit Shifting-Aufgaben nachgewiesen werden (z.B. Cepeda, Kramer, & Gonzalez de Sather, 2001; Karbach & Kray, 2009; Strobach, Liepelt, Schubert, & Kiesel, 2012; für einen Überblick siehe Strobach, Salminen, Karbach, & Schubert, 2014). *Switch costs* konnten jedoch oftmals nicht vollständig eliminiert werden, d.h., es blieben *residual costs*, sowohl bei einer geringen Anzahl an Trainingssessions (Kray & Lindenberger, 2000; Strobach et al., 2012) als auch einer hohen Anzahl an Trainingssessions (Berryhill & Hughes, 2009).

Eine geringe Anzahl an Studien untersuchte Transfereffekte von Shifting-Trainings auf andere kognitive Fähigkeiten und Fertigkeiten. Beispielhaft werden zwei Studien aufgeführt. Karbach und Kray (2009) konnten positive Effekte einer Intervention zur Verbesserung von Shifting bei Kindern und Erwachsenen auf die Fähigkeit zum Shifting, die Kontrolle von Interferenz, sowie auf das verbale und visuelle AG und Gf beobachten. In einer neueren Untersuchung von Kray, Karbach, Haenig und Freitag (2012) wurden Kinder mit der Diagnose ADHS in der Fähigkeit zum Shifting in vier Sessions trainiert. Die Kinder verbesserten sich in Folge des Shifting-Trainings im Shifting selbst, im verbalen AG sowie der inhibitorischen Kontrolle. Im Gegensatz zur Studie von Karbach und Kray (2009) gab es keinen Transfereffekt des Shifting-Trainings auf Gf. Unklar bleibt, ob größere Trainingseffekte durch adaptive oder intensivere Trainings erreicht werden können und ob die Effekte des Trainings über einen längeren Zeitraum aufrechterhalten werden können (Kray et al., 2012).

Insgesamt gibt es vermehrt Hinweise darauf, dass die Fähigkeit zum Shifting trainierbar ist. Wenig ist über Transfereffekte von Shifting-Trainings auf andere kognitive Fähigkeiten und Fertigkeiten bekannt.

3 Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung

In diesem Kapitel werden Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung beschrieben. Als Hintergrund für die Beanspruchung des AGs in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung wird auf die CLT eingegangen. Unter Berücksichtigung der CLT werden im Anschluss daran Lösungsbeispiele als eine effektive instruktionale Unterstützungsmaßnahme thematisiert. Kapitel 3 schließt mit einer Betrachtung der möglichen Moderatorenrolle von kBf und Gf in Lernumgebungen mit hohem vs. niedrigem Grad an instruktionaler Unterstützung.

Bezogen auf den Unterricht bestehen Lernumgebungen aus einem Arrangement von Unterrichtsmethoden, Unterrichtstechniken, Lernmaterialien und Medien (Reinmann & Mandl, 2006) sowie der organisierten Interaktion zwischen Lehrenden und Lernenden. In der vorliegenden Arbeit wird von einem weiten Verständnis von Lernumgebungen ausgegangen, welches sämtliche Merkmale eines Kontexts in Bezug auf die Unterstützung von Lernprozessen (also z.B. auch verschiedene Instruktionsansätze) mit einschließt. Lernumgebungen unterscheiden sich stark darin, welchen Komplexitätsgrad die Aufgaben aufweisen, die die Lernenden bearbeiten sollen und welcher Grad an instruktionaler Unterstützung zur lernförderlichen Bearbeitung der Aufgaben angeboten wird. Lernumgebungen, in denen ein hoher Grad an instruktionaler Unterstützung implementiert ist, wie beispielsweise die Direkte Instruktion (z.B. Klahr & Nigam, 2004) könnten kognitive Funktionen in ganz anderer Weise beanspruchen als Lernumgebungen mit einem niedrigem Grad an instruktionaler Unterstützung wie beispielsweise Problemlösen. Beim Problemlösen müssen Lerner generelle Strategien oder andere oberflächliche Strategien für die Lösung des Problems anwenden, insbesondere wenn Domänenprinzipien noch schlecht verstanden werden (Renkl & Atkinson, 2007). Es wird eine Belastung im AG induziert, die den Wissenserwerb beeinträchtigt. Die Belastung kann zu einer kognitiven Überlastung führen oder zumindest produktive Lernprozesse wie Selbsterklärungen verhindern (Renkl, Hilbert, & Schworm, 2009). Theoretische Basis für die Erklärung dieser Phänomene bildet die CLT (z.B. Sweller, Van Merriënboer, & Paas, 1998).

Cognitive load theory

Die CLT fußt auf einer kognitiven Architektur, die aus einem begrenzten AG mit teilweise unabhängig voneinander operierenden Verarbeitungseinheiten für visuelle/räumliche und auditive/verbale Informationen besteht. Das AG interagiert mit einem Langzeitgedäch-

nis, dessen Kapazität unbeschränkt ist. Zentral ist die Idee, dass bei Instruktionsdesigns ein Hauptaugenmerk auf die Architektur des AGs und seine Begrenzungen gelegt werden sollte. Die wichtigsten Lernprozesse in diesem Kontext sind Schemakonstruktion und – automatisierung (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Ein Schema stellt im Rahmen der CLT eine kognitive Struktur dar, die es Problemlösern erlaubt, Probleme als zugehörig zu bestimmten Kategorien zu erkennen, die bestimmte Operationen für die Lösungsfindung erfordern (Paas & Van Merriënboër, 1994a). Die Schemakonstruktion erfolgt, indem einzelne Informationselemente in einem größeren Element bzw. Schema zusammengefasst werden. Schemata werden durch Hinzufügen neuer Informationselemente elaboriert und können durch intensives sowie gleichmäßiges Üben automatisiert, d.h. ohne kontrollierte Verarbeitung ausgeführt werden (Van Gog & Paas, 2008). Schemakonstruktion und – automatisierung stellen im Kontext der CLT die primären Instruktionsziele dar (Paas et al., 2003). Optimiert werden können diese Lernprozesse aus Sicht der CLT durch eine Minimierung bestimmter Aspekte der kognitiven Belastung des AGs.

Klassischerweise werden drei Arten von cognitive load (CL) unterschieden. Der intrinsic CL bezieht sich auf die Komplexität des Lernmaterials und wird durch die Anzahl interagierender Informationselemente im Lernmaterial bestimmt (Sweller, 2010). Wenn die Anzahl interagierender Informationselemente hoch ist, wird ein hoher intrinsic CL induziert, da die Informationselemente gleichzeitig verarbeitet werden müssen (Sweller et al., 1998). Der intrinsic CL hängt vom Vorwissen des Lerners über das Lernmaterial ab. Lerner mit einem hohen Vorwissen haben mehrere oder alle, der durch das Lernmaterial präsentierten Informationselemente in einem Schema integriert. Dieses Schema kann als einzelnes Element im AG betrachtet werden und reduziert den intrinsic CL (Van Gog & Paas, 2008).

Während der intrinsic CL durch die Komplexität des Lernmaterials bedingt ist, hängt der extraneous CL von der Präsentation der Informationen ab. Instruktionsmethoden, bei denen Lerner *unnötigerweise* viele Informationselemente verarbeiten müssen, erzeugen einen hohen extraneous CL, der Lernen beeinträchtigt (Sweller, 2011).

Als lernförderlich wird dagegen der germane CL angesehen. Dieser CL bezieht sich nach einer aktuellen Konzeptionalisierung von Sweller (2010) auf die AG-Ressourcen, die ein Lerner für die Auseinandersetzung mit dem intrinsic CL aufwendet. Im Gegensatz zu intrinsic und extraneous CL, stellt der germane CL also keine unabhängige Ursache für den CL dar (Kalyuga, 2011; Sweller, 2010). Der germane CL wird durch die Gestaltung von Instruktionen beeinflusst. Falls Instruktionen derart organisiert sind, dass AG-Ressourcen für die Auseinandersetzung mit den Elementen die einen intrinsic CL erzeugen verfügbar sind, wird der ger-

mane CL und folglich Lernen optimiert. Dagegen wird Lernen bei einem hohen extraneous CL reduziert, da AG-Ressourcen für die Auseinandersetzung mit dem extraneous CL aufgewendet werden müssen und weniger AG-Ressourcen für die Auseinandersetzung mit dem intrinsic CL übrig bleiben (Sweller, 2010). Im Falle des Problemlösens müssen Novizen Unterziele speichern und verschiedene Lösungsoptionen beachten. Dies kann einen hohen extraneous CL erzeugen (Renkl, 2014). Eine mögliche instruktionale Unterstützungsmaßnahme zur Verringerung des extraneous CL sind Lösungsbeispiele.

Lösungsbeispiele

Lösungsbeispiele bestehen aus einer Problemformulierung, Lösungsschritten und der endgültigen Lösung (Schwonke et al., 2009). Es wird angenommen, dass Lösungsbeispiele zur Reduzierung des extraneous CL beitragen können, weil keine Suche nach spezifischen Lösungen notwendig ist (Renkl, 2014) und die Anzahl an interagierenden Elementen reduziert wird (Sweller, 2011). Durch die freien AG-Ressourcen kann ein Schema für späteres Problemlösen erzeugt werden (Renkl, 2014).

Nach Renkl (2014) bestehen wenig Zweifel darüber, dass beispielbasiertes Lernen zu einer der am besten untersuchten Lernmethoden gehört. Für Lösungsbeispiele wurde gezeigt, dass sie in verschiedenen Domänen effektiver als Problemlösen (niedriger Grad an instruktionaler Unterstützung) sind, beispielsweise in der Mathematik (Carroll, 1994), der Fehlersuche bei elektrischen Kreisen (Van Gog et al., 2006) und Statistik (Leppink et al., 2014; Paas, 1992).

In diesen und anderen Studien zur Effektivität von Lösungsbeispielen wurde der Lernerfolg mit unterschiedlichen Tests erfasst. Diese Tests erfassen in Abhängigkeit der gelernten Inhalte unterschiedliche Wissensarten. Beispielsweise wurden in den Studien von Salden, Alevén, Schwonke und Renkl (2010) sowie Schwonke, Renkl, Salden und Alevén (2011) Inhalte aus dem Bereich Geometrie gelernt und der Lernerfolg mit Tests erfasst, die prozedurales und konzeptuelles Wissen erfassen. Während prozedurales Wissen „Wissen über eine angemessene Ausführungsweise einer Handlung“ (Woolfolk, 2008, S. 319) meint, ist konzeptuelles Wissen durch ein Netzwerk charakterisiert, in dem Informationen miteinander verbunden sind (Hiebert & Lefevre, 1984, S. 3f.). Konzeptuelles Wissen kann auch bei der Lösung bestimmter Aufgaben aus der Domäne Statistik eine wichtige Rolle spielen, da für die Lösung oftmals miteinander verbundene Inhalte abgerufen werden müssen. Ein Beispiel für eine Aufgabe aus der Domäne Statistik, deren Lösung konzeptuelles Wissen erfordert, ist in Abschnitt 5.2.4

aufgeführt. Für die Lösung komplexer statistischer Probleme kann dagegen in erster Linie anwendungsorientiertes Wissen erforderlich sein, da für die Lösung wesentliche Problemaspekte identifiziert und theoretisches Wissen auf diese Aspekte angewendet werden müssen. Ein Beispiel für ein komplexes statistisches Problem aus der Domäne Statistik, dessen Lösung anwendungsorientiertes Wissen erfordert, ist in Abschnitt 5.2.4 aufgeführt.

Lösungsbeispiele sind effektiver, wenn das Vorwissen niedrig ist. Für Lerner mit hohem Vorwissen ist Problemlösen effektiver. Dieses Phänomen wird *expertise reversal effect* genannt (für einen Überblick siehe Kalyuga, 2007). Die Erklärung für den *expertise reversal effect* ist, dass Lösungsbeispiele den extraneous CL für Lerner mit niedrigem Vorwissen reduzieren, da die Lerner nicht nach den Problemlösungen suchen müssen. Für erfahrenere Lerner können Lösungsbeispiele redundante Informationen enthalten, die unnötigerweise verarbeitet werden müssen und dadurch den CL erhöhen (Sweller, 2011).

Zusammenfassend sind Lösungsbeispiele eine effektive Form instruktionaler Unterstützung für den anfänglichen Wissenserwerb. Als Begründung wird angenommen, dass Lösungsbeispiele den extraneous CL reduzieren und so mehr Raum für Lernprozesse zur Verfügung steht. In der Tat konnten einige Studien zeigen, dass Lösungsbeispiele verglichen mit einer Problemlösebedingung den CL verringern (z.B. Paas & Van Merriënboer, 1994a; Renkl et al., 2003; Van Gog et al., 2006). Anzumerken ist, dass bei diesen Studien nur der *gesamte* CL gemessen wurde, also keine Unterscheidung zwischen den verschiedenen Arten von CL vorgenommen wurde. Unterschiede zwischen Lernumgebungen im gesamten CL auf Unterschiede in einer Art oder mehreren Arten von CL zurückzuführen, ist bei diesen Untersuchungen spekulativ.

Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung können sich also in der kognitiven Belastung des AGs unterscheiden. Unbekannt ist jedoch, ob eine Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) zur Reduzierung des CL insbesondere dann hinsichtlich des Wissenserwerbs effektiver als eine Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) und hohem CL ist, wenn die AG-Kapazität gering ist. Bei Lernern mit geringer AG-Kapazität könnten Lösungsbeispiele noch wirksamer sein (Van Gog & Rummel, 2010). Erweitert man die Argumentation von Van Gog und Rummel (2010), könnten Lösungsbeispiele verglichen mit einer Problemlösebedingung *weniger* effektiv für Lerner mit einer hohen AG-Kapazität sein, da ein hoher CL in der Problemlösebedingung aufgrund der hohen AG-Kapazität weniger problematisch für diese Lerner ist. Somit könnte die AG-Kapazität beim Vergleich zwischen Lösungsbeispielen und Problemlösen einen moderierenden Einfluss auf den Wissenserwerb

haben. Um diesen moderierenden Einfluss zu untersuchen, muss die AG-Kapazität erfasst werden. Allerdings wurde die AG-Kapazität in der Forschung zum CL kaum gemessen (de Jong, 2010). Dies bedeutet, dass der AG-Kapazität mit Bezug auf den CL eine wichtige Rolle in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung zugeschrieben wurde, sie jedoch selten erfasst wurde.

Vielleicht noch offensichtlicher ist, dass Shifting und Gf trotz der nachgewiesenen Wichtigkeit für das Lernen (siehe Abschnitte 2.3.2 und 2.3.3) in Studien, die Lösungsbeispiele und Problemlösen verglichen, nicht berücksichtigt wurden. Das Lösen eines komplexen Problems erfordert möglicherweise den Wechsel zwischen bestimmten Aspekten des Problems oder unterschiedlichen Dokumenten mit relevanten Informationen sowie den Wechsel zwischen Problemen und den Informationen zur Problemlösung. Lerner mit einer geringen Shifting-Fähigkeit könnten in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung Schwierigkeiten haben. Bei einer instruktionalen Unterstützung durch Lösungsbeispiele könnten die Shifting-Anforderungen reduziert werden, da die Lösungsbeispiele Lösungsschritte und die endgültige Lösung beinhalten. Lösungsbeispiele könnten daher im Hinblick auf den Wissenserwerb vor allem dann dem Problemlösen überlegen sein, wenn Lerner über eine geringe Shifting-Fähigkeit verfügen. Bei Lernern mit hoher Shifting-Fähigkeit könnte die instruktionale Unterstützung durch Lösungsbeispiele verglichen mit einer Problemlösebedingung *weniger* effektiv sein, da möglicherweise hohe Shifting-Anforderungen in der Problemlösebedingung für diese Lerner weniger problematisch sind.

In einer analogen Weise könnte auch Gf die Effekte von Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung auf den Wissenserwerb moderieren. Lerner müssen für die Lösung komplexer Probleme typischerweise schlussfolgern, welche Informationen für die Lösung notwendig sind. Lernern mit geringer Gf fällt diese Selektion von relevanten Informationen möglicherweise schwer. Schlussfolgerungen über notwendige Informationen zur Lösung komplexer Probleme müssen bei Lösungsbeispielen nicht gezogen werden. Lösungsbeispiele könnten demnach hinsichtlich des Wissenserwerbs dem Problemlösen insbesondere dann überlegen sein, wenn die Gf der Lerner niedrig ist. Falls die Gf der Lerner hoch ist, könnten Lösungsbeispiele verglichen mit einer Problemlösebedingung *weniger* effektiv sein, da möglicherweise hohe Anforderungen an Gf in der Problemlösebedingung für Lerner mit hoher Gf weniger problematisch sind.

4 Forschungsfragen

Mit der vorliegenden Arbeit soll zum einen die Rolle von kBf und Gf für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung untersucht werden. Zum anderen sollen nahe und weite Transfereffekte von AG-Trainings und der Einfluss von Trainingsbedingungen als potentielle Moderatoren auf diese Transfereffekte untersucht werden.

KBf und Gf sind für viele kognitive Leistungen relevant (siehe Abschnitt 2.3). Insbesondere dem AG bzw. der AG-Kapazität wird eine wichtige Rolle für Lernprozesse zugeschrieben. Auch im Rahmen von Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung wurde der AG-Kapazität basierend auf der CLT eine zentrale Rolle für den Wissenserwerb eingeräumt (siehe Kapitel 3). Allerdings wurde in der Forschung zum Vergleich von Lernumgebungen mit hohem vs. niedrigem Grad an instruktionaler Unterstützung die AG-Kapazität kaum und Shifting sowie Gf überhaupt nicht erfasst. Unterschiede in kBf und Gf könnten relevant sein für die Effektivität von Lösungsbeispielen (hoher Grad an instruktionaler Unterstützung) im Vergleich zu Problemlösen (niedriger Grad an instruktionaler Unterstützung). Es ist jedoch unklar, welche Bedeutung kBf und Gf für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung haben. Daher ergibt sich die erste Forschungsfrage:

Frage 1:

Welche Rolle spielen kognitive Basisfunktionen und fluide Intelligenz für den Wissenserwerb in unterschiedlich instruktional unterstützten (Lösungsbeispiele/hoher Grad an instruktionaler Unterstützung vs. Problemlösen/niedriger Grad an instruktionaler Unterstützung) Lernumgebungen?

Vor dem Hintergrund der Bedeutsamkeit von kBf für kognitive Leistungen wurden Versuche unternommen, kBf zu trainieren um einen Transfer auf andere kognitive Fähigkeiten und Fertigkeiten zu erzielen. Eine große Anzahl an Studien widmete sich dem Training des AGs. Eine bisherige Metaanalyse von Melby-Lervåg und Hulme (2013) fand nahe Transfereffekte auf AG-Komponenten, die teilweise über mehrere Monate bestehen blieben. Es zeigte sich nur ein weiter Transfereffekt auf die nonverbale Fähigkeit. Dieser Effekt war klein und kurzfristig. Seit der Veröffentlichung der Metaanalyse von Melby-Lervåg und Hulme (2013) wurde eine Vielzahl neuerer Studien zum Training des AGs publiziert. Unklar ist, wie die Be-

fundlage zu nahen und weiten Transfereffekten angesichts dieser neueren Studien ist. Aus diesem Grund ergibt sich die zweite Forschungsfrage:

Frage 2:

Welche nahen und weiten Transfereffekte folgen auf Arbeitsgedächtnistrainings?

In der Metaanalyse von Melby-Lervåg und Hulme (2013) wurde nur eine geringe Anzahl an Moderatoren untersucht, die möglicherweise einen Einfluss auf Transfereffekte von AG-Trainings haben. Das Alter, der Interventionstyp und die Art der Kontrollgruppe erwiesen sich dabei als signifikante Moderatoren bestimmter Transfereffekte. Die Rolle dieser Moderatoren ist vor dem Hintergrund neuerer Trainingsstudien unbekannt. Zudem wurde eine Vielzahl anderer Trainingsbedingungen wie z.B. die trainierte Modalität des AGs und die Supervision des Trainings vernachlässigt. Diese und weitere Trainingsbedingungen könnten aus theoretischer Sicht einen Einfluss auf Transfereffekte haben, wurden aber nicht untersucht. Folglich ergibt sich die dritte Forschungsfrage:

Frage 3:

Welche Trainingsbedingungen haben einen moderierenden Einfluss auf Transfereffekte von Arbeitsgedächtnistrainings?

In den folgenden zwei Kapiteln (Kapitel 5 und 6) werden zwei Studien dargestellt, die der Beantwortung der Forschungsfragen dienen. Studie 1 dient der Beantwortung der Forschungsfrage 1. Mit der zweiten Studie sollen die Forschungsfragen 2 und 3 beantwortet werden. Im letzten Kapitel (Kapitel 7) werden die Ergebnisse beider Untersuchungen unter einer übergeordneten Perspektive diskutiert.

5 Studie 1: Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung (Frage 1)

5.1 Ziel und Hypothesen

Mit Studie 1 soll die Forschungsfrage 1 beantwortet werden.

Frage 1:

Welche Rolle spielen kognitive Basisfunktionen und fluide Intelligenz für den Wissenserwerb in unterschiedlich instruktional unterstützten (Lösungsbeispiele/hoher Grad an instruktionaler Unterstützung vs. Problemlösen/niedriger Grad an instruktionaler Unterstützung) Lernumgebungen?

Vor der Formulierung der geeigneten Haupthypothesen für diese Forschungsfrage wird kurz der theoretische Hintergrund für Hypothesen gegeben, die für vorausgehende Analysen formuliert werden.

Der Forschung zur Wirksamkeit von Lösungsbeispielen (siehe Kapitel 3) zu Folge sollte in einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) der CL niedriger und der Wissenserwerb höher ausfallen. Der Effekt der Lösungsbeispiele verglichen mit Problemlösen auf den Wissenserwerb sollte umso geringer sein, je höher das Vorwissen ist. Das heißt, das Vorwissen sollte eine moderierende Rolle im Hinblick auf den Wissenserwerb haben. Diese Hypothese lässt sich mit dem *expertise reversal effect* begründen, demzufolge die Effektivität von Lösungsbeispielen mit wachsendem Vorwissen abnimmt (Kalyuga, 2007). Aus diesen Gründen werden Hypothesen für vorausgehende Analysen formuliert, die den Haupthypothesen über die moderierende Rolle von kBf und Gf vorangestellt werden. Hinsichtlich des Wissenserwerbs wurde zwischen dem Erwerb von konzeptuellem und dem Erwerb von anwendungsorientiertem Wissen unterschieden. Es werden separate Hypothesen für die verschiedenen Arten des Wissenserwerbs von Pre- zu Post-test formuliert.

Hypothesen für vorausgehende Analysen

1. Die kognitive Belastung ist in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) höher als in einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele).
2. a. Der Erwerb von konzeptuellem Wissen ist in einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) höher als in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen).
b. Der Erwerb von anwendungsorientiertem Wissen ist in einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) höher als in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen).
3. Der Effekt einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) auf den Wissenserwerb ist umso geringer, je höher das Vorwissen ist.

Die Fragestellung 1 und der theoretische Hintergrund (Abschnitt 2.3 und Kapitel 3) führen zu folgenden Haupthypothesen, die sich auf die Moderatorenrolle von kBf und Gf beziehen:

Haupthypothesen

4. a. Der Vorteil einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) gegenüber einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) im Hinblick auf den Wissenserwerb ist umso größer, je geringer die AG-Kapazität ist.
b. Der Vorteil einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) gegenüber einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) im Hinblick auf den Wissenserwerb ist umso größer, je geringer die Shifting-Fähigkeit ist.
c. Der Vorteil einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) gegenüber einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) im Hinblick auf den Wissenserwerb ist umso größer, je geringer Gf ist.

5.2 Methode

5.2.1 Stichprobe und Design

Über Aushänge in der Ludwig-Maximilians-Universität München, den Infodienst der Ludwig-Maximilians-Universität sowie Werbung in Seminaren an der Ludwig-Maximilians-Universität wurden Teilnehmer für die Studie rekrutiert. Die Stichprobe bestand aus 76 Studierenden der Fachrichtungen Pädagogik, Psychologie und Schulpsychologie an der Ludwig-Maximilians-Universität. Studierende der Pädagogik und Psychologie durften ab dem 2. bis zum 4. Semester an der Studie teilnehmen. Studierende der Schulpsychologie konnten ab dem 2. Semester bis vor dem Schreiben ihrer Abschlussarbeit teilnehmen. Gründe für diese Auswahlkriterien waren, dass ein niedriges Vorwissen in Statistik aufgrund des Besuchens von Statistikveranstaltungen vorhanden sein sollte, um die statistischen Probleme in der Lernphase verstehen zu können. Zudem sind Lösungsbeispiele insbesondere bei niedrigem Vorwissen effektiv (Kalyuga, 2007). Studierende in den Bachelorstudiengängen Pädagogik und Psychologie erfüllen die Auswahlkriterien bis zum 4. Semester, da im Anschluss daran erneut statistische Methoden gelehrt und die Bachelorarbeit angefertigt wird. Mit dem Schreiben einer Abschlussarbeit ist i.d.R. die Auseinandersetzung mit statistischen Methoden verbunden, weshalb Studierende der Schulpsychologie bis vor dem Beginn des Schreibens der Abschlussarbeit an der Studie teilnehmen durften.

Das Durchschnittsalter der Teilnehmer betrug 23.83 Jahre ($SD = 5.70$). Von den Teilnehmern waren 67 weiblich und 9 männlich. Tabelle 1 zeigt eine Übersicht der Stichprobenzusammensetzung.

Tabelle 1

Stichprobenzusammensetzung

Studiengang	Geschlecht		Alter			Semesterzahl
	weiblich	männlich	$M (SD)$	Min	Max	$M (SD)$
Pädagogik	24	5	22.66 (3.20)	19.08	31.42	2.83 (1.00)
Psychologie	21	3	25.03 (7.71)	18.08	44.83	2.25 (0.68)
Schulpsychologie	22	1	24.03 (5.69)	19.17	44.75	4.39 (2.50)

Anmerkungen. M = Mittelwert; SD = Standardabweichung; Min = Minimum; Max = Maximum.

Studie 1 ist als einfaktorielles Design (siehe Tabelle 2) mit der unabhängigen Variablen „Grad an instruktionaler Unterstützung“ (Lösungsbeispiele/hoher Grad an instruktionaler Unterstützung vs. Problemlösen/niedriger Grad an instruktionaler Unterstützung) konzipiert.

Tabelle 2

Design von Studie 1

Grad an instruktionaler Unterstützung	
Niedrig (Problemlösen)	Hoch (Lösungsbeispiele)
$n = 38$	$n = 38$

Anmerkung. n = Stichprobengröße.

Die Teilnehmer wurden den Experimentalbedingungen (siehe Tabelle 2) zufällig zugewiesen.

Für die Teilnahme an der gesamten Studie gab es wahlweise 40 Euro oder 5 Versuchspersonenstunden, die in Psychologie-Studiengängen benötigt werden.

5.2.2 Lernumgebung und Versuchsablauf

In einer computerbasierten Lernumgebung bearbeiteten Teilnehmer komplexe Probleme aus der Domäne Statistik. Die statistischen Probleme und Informationen zur Aufgabebearbeitung wurden auf Laptops mit PowerPoint-Folien präsentiert, durch welche die Teilnehmer blättern konnten. In beiden Experimentalbedingungen bzw. Lernumgebungen konnten die Teilnehmer Informationen zur Aufgabebearbeitung aufrufen. In der Experimentalbedingung mit Lösungsbeispielen erhielten die Teilnehmer zusätzlich Folien mit Lösungsschritten zu den statischen Problemen. Die Studie fand in Laborräumen der Ludwig-Maximilians-Universität statt.

Studie 1 erstreckte sich über ein Zeitfenster von etwa 4 Stunden. In einem Pretest wurden demografische Daten, Vorwissen in Statistik, die AG-Kapazität, Shifting sowie Gf erhoben. Zunächst wurden die Teilnehmer begrüßt und der Ablauf des Pretests erklärt. Im Anschluss daran bearbeiteten die Teilnehmer einen Wissenstest zur Erfassung konzeptuellen Vorwissens und einen Wissenstest zur Erfassung anwendungsorientierten Vorwissens. Beide Vorwissenstests waren zusammengeheftet In die Kopfzeile der ersten Seite des Wissenstests zur Erfassung konzeptuellen Vorwissens mussten demografische Daten eingetragen werden. Durch die Angabe eines individuellen Codes auf dem Wissenstests zur Erfassung konzeptuel-

len Vorwissens war die Anonymisierung der Teilnehmer gewährleistet. Instruktionen zur Beantwortung der Items zur Erfassung konzeptuellen Vorwissens sind auf der ersten Seite des Wissenstests zur Erfassung konzeptuellen Vorwissens abgedruckt (siehe Anhang A). Die Bearbeitung der Vorwissenstests dauerte ca. 20 Minuten. Nach den Vorwissenstests bearbeiteten die Teilnehmer an Laptops drei Aufgaben zur Erfassung der AG-Kapazität und an MacBooks drei Aufgaben zur Erfassung der Shifting-Fähigkeit. Der Versuchsleiter richtete die jeweiligen Aufgaben ein. Standardisierte Instruktionen für die Aufgabenbearbeitung erhielten die Teilnehmer vor der jeweiligen Aufgabe am Laptop bzw. MacBook. Die Instruktionen zu den AG-Aufgaben waren in englischer Sprache, da sie nicht verändert werden konnten. Es gab jedoch kaum Verständnisschwierigkeiten bei den Teilnehmern. Falls doch, wurden diese durch den Versuchsleiter behoben. Die Aufgaben zur Erfassung der AG-Kapazität und Shifting nahmen jeweils ca. 45 Minuten (15 Minuten pro Aufgabe) in Anspruch. Am Ende des Pretests bearbeiteten die Teilnehmer einen Test zur Erfassung von Gf am Laptop. Standardisierte Instruktionen hierfür erhielten die Teilnehmer ebenfalls am Laptop. Der Intelligenztest dauerte ca. 48 Minuten.

Zu einem zweiten Messzeitpunkt fanden die Interventionsphase sowie der Posttest statt. In der Interventionsphase bekamen die Teilnehmer zunächst Informationen über den Ablauf der Interventionsphase und des Posttests. Die Aufgabenstellung für das erste zu bearbeitende statistische Problem wurde den Teilnehmern vorgelesen und im Anschluss daran Interesse sowie Motivation zur Bearbeitung derartiger Probleme mittels eines Fragebogens erfasst. Das Ausfüllen des Fragebogens zur Erfassung von Interesse und Motivation nahm ca. 5 Minuten in Anspruch. Im Anschluss daran bearbeiteten die Teilnehmer drei komplexe, statistische Probleme in Abhängigkeit der Experimentalbedingung mit oder ohne Lösungsbeispiele. Auf der ersten PowerPoint-Folie wurden Instruktionen zu den Aufgaben mit statistischen Problemen gegeben. Nach der Bearbeitung jedes Problems gaben die Teilnehmer an, wie hoch die kognitive Belastung bei der Aufgabenbearbeitung war. Die Dauer der Interventionsphase betrug ca. 60 Minuten. Im Anschluss an die Bearbeitung aller statistischen Probleme wurde in einem Posttest das Wissen in Statistik erneut erfasst. Der Wissenstest zum Posttest nahm ca. 20 Minuten in Anspruch. Am Ende der Studie erhielten die Teilnehmer 40 Euro oder 5 Versuchspersonenstunden.

5.2.3 Experimentalbedingungen

Die Teilnehmer bearbeiteten individuell drei statistische Probleme (siehe Anhang D) aus dem Themenbereich ALM (Allgemeines Lineares Modell), je nach Experimentalbedingung mit hohem (Lösungsbeispiele) oder niedrigem (Problemlösen) Grad an instruktionaler Unterstützung. Die Aufgabenbearbeitung erfolgte schriftlich auf einem Antwortbogen (Anhang F). In beiden Experimentalbedingungen bzw. Lernumgebungen wurden Selbsterklärungsaufforderungen (Aleven & Koedinger, 2002) verwendet, da der Vorteil von Lösungsbeispielen gegenüber Problemlösen verschwinden könnte, wenn nur in der Bedingung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) Selbsterklärungsaufforderungen vorhanden sind (Schwonke et al., 2009) Die Begründung ist, dass Selbsterklärungen den Lernerfolg erhöhen können (für einen Überblick siehe Renkl, 2014). Die Selbsterklärung in der Studie 1 dieser Arbeit bestand darin, dass die Teilnehmer ihre Antworten begründen mussten.

In der Bedingung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) konnten die Teilnehmer nach jedem statistischen Problem Folien mit Informationen zur Problemlösung aufrufen. Abbildung 1 zeigt das erste zu bearbeitende statistische Problem mit der Aufgabenstellung und der Selbsterklärungsaufforderung, die Antworten wenn möglich zu begründen. Ein Teil der dazugehörigen Informationen ist in Abbildung 2 dargestellt.

Problem 1

Der Kognitionsforscher Max Musterhirn möchte herausfinden, ob die Anwesenheit einer Person einen Einfluss auf die Effektivität eines kognitiven Trainings hat. Als Probanden hat er Studierende der Pädagogik und Psychologie (N = 120) akquiriert, die er zufällig zwei Gruppen zuwies. Eine Gruppe absolvierte ein kognitives Training über einen Zeitraum von 4 Wochen allein in einem Labor, die andere Gruppe absolvierte dasselbe kognitive Training über den gleichen Zeitraum, allerdings unter der Aufsicht eines Experimentators. Die Trainingsleistung (intervallskaliertes Merkmal) hat Max Musterhirn zu Beginn, nach zwei Wochen Training und am Ende des Trainings erhoben. Da er noch keine Untersuchung mit dem kognitiven Training an einer studentischen Stichprobe durchgeführt hat, ist Max Musterhirn daran interessiert, ob sich die Trainingsleistung der Versuchspersonen zwischen den drei Messzeitpunkten signifikant unterscheidet. Er vermutet zudem, dass der Effekt der Anwesenheit eines Experimentators nicht sofort erkennbar ist, sondern von der Dauer des Trainings abhängt. Um in Hinblick auf die statistische Auswertung korrekt vorzugehen, möchte Max Musterhirn auch die Voraussetzungen für seine Analyse berücksichtigen. *Wie kann die Fragestellung von Max Musterhirn statistisch beantwortet werden? Bitte begründen Sie ihre Antworten wenn möglich.*

Abbildung 1. Erstes statistisches Problem mit Aufgabenstellung.

Diese Informationen waren ähnlich wie ein Lehrtext aufgebaut. Die Probleme und Informationen wurden so gestaltet, dass sie eine eher hohe Komplexität aufweisen. Um eine

hohe Komplexität der statistischen Probleme und der zugehörigen Informationen zu erreichen, enthielten die statistischen Probleme und die zugehörigen Informationen auch Inhalte, die irrelevant für die Aufgabenbearbeitung waren.

Informationen zu Problem 1

Für den Vergleich von Mittelwerten aus mehr als zwei Stichproben gibt es verschiedene statistische Verfahren. Von Bedeutung sind die Anzahl zu vergleichender Stichproben und ob diese unabhängig voneinander sind. Für den Vergleich mehrerer unabhängiger Stichproben, die Realisationen eines Faktors bzw. einer unabhängigen Variablen sind, ist eine einfaktorielles Varianzanalyse ohne Messwiederholung geeignet. Dabei liegt ein einfaktorielles Design vor, da nur Mittelwerte aus den Stufen eines Faktors bzw. einer unabhängigen Variable betrachtet werden. Wird eine Stichprobe zu mehreren Zeitpunkten untersucht, sind die Werte in der abhängigen Variablen zu den verschiedenen Zeitpunkten abhängig voneinander. In diesem Fall liegt ebenfalls ein einfaktorielles Design mit einem Messwiederholungsfaktor (unabhängige Variable) vor. Vergleicht man dagegen Mittelwerte aus Kombinationen von Faktorstufen, liegt ein mehrfaktorielles Design vor. Hat man bspw. drei unabhängige Variablen bzw. Faktoren mit jeweils 3 Stufen, liegt ein 3x3x3-faktorielles Design vor. Die Auswertung erfolgt dann mit einer dreifaktoriellen Varianzanalyse ohne Messwiederholung. Einen Spezialfall stellt die Kombination eines Messwiederholungsfaktors und einem oder mehreren anderen Faktoren dar. In diesem Fall kann man auch von einem gemischten Design sprechen. Eine Auswertung kann mit einer dreifaktoriellen Varianzanalyse mit Messwiederholung erfolgen. Wäre in dem Beispiel mit drei Faktoren ein Faktor ein Messwiederholungsfaktor (mit drei Stufen), könnte man Unterschiede zwischen den Mittelwerten aus den Faktorstufenkombinationen mit einer dreifaktoriellen Varianzanalyse mit Messwiederholung feststellen.

Abbildung 2. Teil der Informationen für die Lösung des ersten statistischen Problems.

In der Bedingung mit hohem Grad an instruktionaler Unterstützung bekamen die Teilnehmer dieselben statistischen Probleme und dieselben zugehörigen Informationen wie in der Bedingung mit niedrigem Grad an instruktionaler Unterstützung präsentiert. Zusätzlich konnten die Teilnehmer zu jedem Problem vor den zugehörigen Informationen ein Lösungsbeispiel aufrufen (siehe Anhang E). Das Lösungsbeispiel beinhaltete die notwendigen Schritte zur Lösung des statistischen Problems. Den Teilnehmern wurden zwischen den statistischen Problemen und den Informationen zur Aufgabenbearbeitung PowerPoint-Folien präsentiert, auf denen jeweils drei ausgearbeitete Lösungsschritte für das jeweilige statistische Problem dargestellt waren. *Abbildung 3* zeigt die Folie des ersten Lösungsschritts für das erste statistische Problem. Das Schema der Lösungsschritte war für alle drei statistischen Probleme sehr ähnlich. Im ersten Lösungsschritt wurden das Design und/oder die unabhängige/n sowie die abhängige Variable dargestellt. Der zweite Lösungsschritt zeigte die geeignete statistische Auswertungsmethode bzw. das geeignete Testverfahren für das jeweilige statistische Problem. Im dritten Schritt wurden Voraussetzungen zur Anwendung einer statistischen Auswertungsmethode (Probleme 1 und 3) sowie die Berücksichtigung der Kausalität der Ergebnisse (Problem 2) aufgeführt.

Lösung Problem 1

1. Schritt: Bestimmung des Designs und der unabhängigen Variablen sowie der abhängigen Variable

Es handelt sich um ein zweifaktorielles (2x3 bzw. 3x2-faktorielles) Design mit den unabhängigen Variablen (UVs) bzw. den Faktoren „Anwesenheit einer Person während des Trainings“ (zweistufig, Ja vs. Nein) sowie Messzeitpunkt (dreistufig). Die abhängige Variable (AV) ist die Trainingsleistung.

Abbildung 3. Erster Lösungsschritt für das erste statistische Problem.

Selbsterklärungsaufforderungen sollten die die Teilnehmer dazu anregen, die Lösungsschritte zu begründen. Für die Begründung der Antworten war ein Rückgriff auf die zugehörigen Informationen (auf separaten Folien; für ein Beispiel siehe Abbildung 2) zur Aufgabenbearbeitung erforderlich. Ohne Selbsterklärungsaufforderungen hätten die Teilnehmer in der Bedingung mit Lösungsbeispielen keine Antworten geben und nicht von den Informationen zur Aufgabenbearbeitung Gebrauch machen müssen.

5.2.4 Abhängige Variablen

Abhängige Variablen waren die AG-Kapazität, Shifting, Gf, der CL sowie der Wissenserwerb vom Pre- zum Posttest.

AG-Kapazität. Die AG-Kapazität wurde auf Konstruktebene durch die automatisierten Versionen der complex span Aufgaben (CSAs) automated operation span (aospan), automated symmetry span (asymspan) und automated reading span (arspan) operationalisiert (Redick, Broadway, et al., 2012; Unsworth et al., 2009; Unsworth et al., 2005). Diese erlauben eine schnelle Handhabung, sind vollständig computerisiert, mit der Maus zu bedienen und Leistungen werden automatisch erfasst (Redick, Broadway, et al., 2012). Als Software für die AG-Aufgaben wurde E-Prime (Version 2.0.8.22) verwendet. Alle drei verwendeten CSAs lassen sich in zwei Phasen unterteilen. In der ersten Phase müssen abwechselnd Informationen verarbeitet und memoriert werden. In der zweiten Phase müssen die zu memorierenden Informationen wiedergegeben werden. Vor den eigentlichen Durchgängen mussten in allen drei CSAs drei Übungsblöcke (nur Memorieren, nur Verarbeiten, Memorieren und Verarbeiten) durchlaufen werden. Die abhängige Variable in den drei CSAs war die Anzahl korrekt erinnelter Elemente wie z.B. Buchstaben bei der aospan Aufgabe (für Methoden zur Berechnung der

abhängigen Variable bei CSAs siehe Conway et al., 2005). Die Aufgaben werden nachfolgend detaillierter beschrieben.

Im ersten Übungsblock der aospa Aufgabe wurde nur das Memorieren von Buchstaben geübt. In jedem Durchgang erschienen mehrere Buchstaben einzeln nacheinander auf dem Bildschirm. Die möglichen Buchstaben waren F, H, J, K, L, N, P, Q, R, S, T und standen zueinander nicht in Beziehung. Nach der Präsentationsphase sahen die Teilnehmer Buchstaben mit Kästchen zur Auswahl und mussten die vorher präsentierten Buchstaben in der korrekten Reihenfolge durch Anklicken von Feldern wiedergeben. Falls Buchstaben vergessen wurden, konnte dies durch Anklicken eines Felds angegeben werden. Im Anschluss daran bekamen die Teilnehmer Feedback über die Anzahl korrekt erinnerter Buchstaben. Im zweiten Übungsblock wurde das zeitabhängige Lösen einfacher mathematischer Operationen geübt. Teilnehmer bekamen in den Durchgängen zunächst eine einfache mathematische Operation präsentiert (z.B. $(1*2) + 1 = ?$). Dann erschien eine Zahl und die Teilnehmer mussten durch Anklicken eines true- oder eines false-Felds so schnell wie möglich angeben, ob diese Zahl die Lösung der vorherigen Operation war oder nicht. Nach jedem Durchgang bekamen die Teilnehmer Feedback über die Richtigkeit der Antwort. Im Anschluss an die Übungsdurchgänge mit mathematischen Gleichungen wurde vom Programm für jede Person die mittlere Zeit für das Lösen der mathematischen Gleichungen berechnet. Diese Zeit plus 2.5 Standardabweichungen der Zeit wurde im dritten Übungsblock und in den eigentlichen Durchgängen als Limit für die Zeit zum Lösen der mathematischen Gleichungen verwendet. Benötigte ein Teilnehmer in diesen Durchgängen für das Lösen einer mathematischen Operation länger als sein individuelles Zeitlimit, machte das Programm mit der Präsentation des nächsten Buchstabens weiter und wertete die mathematische Gleichung als nicht gelöst. Durch das Fortfahren sollte verhindert werden, dass die Teilnehmer die Buchstaben wiederholt memorieren. Im dritten Übungsblock mussten die Teilnehmer wie in den eigentlichen Durchgängen in Folgen abwechselnd mathematische Gleichungen lösen und Buchstaben memorieren. Die Anzahl an zu memorierenden Buchstaben im dritten Übungsblock betrug 2. Wie im ersten Übungsblock mussten die Teilnehmer am Ende jeden Durchgangs die präsentierten Buchstaben anklicken. Danach erhielten die Teilnehmer Feedback über die Anzahl korrekt erinnerter Buchstaben und die Anzahl an Fehlern bei den mathematischen Operationen. Die tatsächlichen Übungsdurchgänge waren mit den Übungsdurchgängen im dritten Übungsblock bezüglich des Ablaufs identisch. Allerdings variierte die Anzahl an zu memorierenden Buchstaben mit einer Listenlänge von 3 bis 7. Jede Listenlänge kam insgesamt dreimal vor. Folglich erga-

ben sich insgesamt 75 zu lösende mathematische Operationen und 75 zu memorierende Buchstaben (für eine Aufgabenbeschreibung siehe Unsworth et al., 2005). Die abhängige Variable bei der aospa Aufgabe war die Anzahl korrekt erinnerter Buchstaben geteilt durch die Gesamtanzahl präsentierter Buchstaben.

Bei der arspa Aufgabe mussten im ersten Übungsblock in jedem Durchgang Folgen von Buchstaben erinnert werden. Dieser Block war mit dem ersten Übungsblock der aospa Aufgabe identisch. Im zweiten Übungsblock lasen die Teilnehmer in den Durchgängen Sätze am Bildschirm. Die Sätze waren in englischer Sprache, da sie aufgrund der Aufgabenprogrammierung nicht übersetzt werden konnten (Beispielsatz: “Andy was stopped by the policeman because he crossed the yellow heaven”). Von den 10-15 Wörtern eines Satzes entschied eines über die Sinnhaftigkeit des Satzes. Nach dem Lesen eines Satzes mussten die Teilnehmer durch Anklicken eines true- oder eines false-Felds so schnell wie möglich angeben, ob der zuvor präsentierte Satz Sinn macht oder nicht. Die Teilnehmer erhielten nach jedem Satz ein Feedback über die Richtigkeit der Antwort. Im Anschluss an die Übungsdurchgänge mit Sinnhaftigkeitsentscheidungen wurde vom Programm analog zur aospa Aufgabe für jede Person ein Zeitlimit für die Beurteilung der Sinnhaftigkeit eines Satzes im dritten Übungsblock und den eigentlichen Durchgängen berechnet. Benötigte ein Teilnehmer in diesen Durchgängen für die Beurteilung der Sinnhaftigkeit eines Satzes länger als sein individuelles Zeitlimit, machte das Programm mit der Präsentation des nächsten Buchstabens weiter und wertete die Antwort bei der Sinnhaftigkeitsbeurteilung als falsch. Das Memorieren von Buchstaben während der Präsentation von Sätzen sollte somit verhindert werden. Durch die Berechnung von individuellen Zeitgrenzen für die Beurteilung der Sinnhaftigkeit von Sätzen wurden interindividuelle Unterschiede im Leseverständnis berücksichtigt. Im dritten Übungsblock mussten die Teilnehmer wie in den eigentlichen Durchgängen in Folgen abwechselnd Sätze im Hinblick auf deren Sinnhaftigkeit beurteilen und Buchstaben memorieren. Die Anzahl an zu memorierenden Buchstaben im dritten Übungsblock betrug 2. Wie im ersten Übungsblock mussten die Teilnehmer am Ende jeden Durchgangs die präsentierten Buchstaben anklicken. Danach erhielten die Teilnehmer Feedback über die Anzahl korrekt erinnerter Buchstaben und die Anzahl an Fehlern bei den mathematischen Operationen. Die tatsächlichen Übungsdurchgänge waren mit den Übungsdurchgängen im dritten Übungsblock bezüglich des Ablaufs identisch. Allerdings variierte die Anzahl an zu memorierenden Buchstaben mit einer Listenlänge von 3 bis 7. Jede Listenlänge kam insgesamt dreimal vor. Folglich ergaben sich insgesamt 75 zu beurteilende Sätze und 75 zu memorierende Buchstaben (für eine

Aufgabenbeschreibung siehe Unsworth et al., 2009). Die abhängige Variable bei der arspan Aufgabe war wie bei der aospan Aufgabe die Anzahl korrekt erinnelter Buchstaben geteilt durch die Gesamtanzahl präsentierter Buchstaben.

Im ersten Übungsblock der asymspan Aufgabe wurde den Teilnehmern in jedem Durchgang eine Sequenz rot aufleuchtender Quadrate in einer 4x4-Matrix präsentiert. Diese Sequenz musste durch Anklicken von weißen Quadraten in einer 4x4-Matrix in der korrekten Reihenfolge wiedergegeben werden. Im zweiten Übungsblock der asymspan Aufgabe sahen Teilnehmer in den Durchgängen zunächst ein Muster schwarzer Quadrate auf einer 8x8-Matrix. Danach mussten die Teilnehmer durch Anklicken eines yes- oder eines no-Felds entscheiden, ob dieses Muster bezüglich einer gedachten vertikalen Linie in der Mitte der 8x8-Matrix symmetrisch war. Analog der aospan Aufgabe und der arspan Aufgabe wurde bei der asymspan Aufgabe ein Zeitlimit für die Symmetrieentscheidungen beim dritten Übungsblock und den eigentlichen Durchgängen anhand der durchschnittlichen Dauer für die Symmetrieentscheidung berechnet. Benötigte ein Teilnehmer in diesen Durchgängen für die Symmetrieentscheidung länger als sein individuelles Zeitlimit, machte das Programm mit der Präsentation des nächsten Musters von roten Quadraten weiter und wertete die Antwort bei der Symmetrieentscheidung als falsch. Im dritten Übungsblock mussten die Teilnehmer wie in den eigentlichen Durchgängen in Folgen abwechselnd Symmetrieentscheidungen treffen oder rot aufleuchtende Quadrate memorieren. Die Anzahl an zu memorierenden roten Quadraten im dritten Übungsblock betrug 2. Wie im ersten Übungsblock mussten die Teilnehmer am Ende jeden Durchgangs die präsentierten Quadrate in einer 4x4-Matrix anklicken. Danach erhielten die Teilnehmer Feedback über die Anzahl korrekt erinnelter roter Quadrate und die Anzahl an Fehlern bei den Symmetrieentscheidungen. Die tatsächlichen Übungsdurchgänge waren mit den Übungsdurchgängen im dritten Übungsblock bezüglich des Ablaufs identisch. Allerdings variierte die Anzahl an zu memorierenden roten Quadraten von 2 bis 5. Folglich ergaben sich 42 Symmetrieentscheidungen und 42 zu memorierende Positionen der roten Quadrate (für eine Aufgabenbeschreibung siehe Unsworth et al., 2009). Bei der asymspan Aufgabe war die abhängige Variable die Anzahl korrekt erinnelter Quadrate geteilt durch die Gesamtanzahl präsentierter Quadrate.

Die internen Konsistenzen (Cronbach's Alpha) für jede einzelne CSA (jeweils über die einzelnen Durchgänge) wurden mit der Methode von Kane et al. (2004) berechnet. Es ergaben sich folgende Werte: $\alpha = .87$ (aospan Aufgabe), $\alpha = .71$ (asymspan Aufgabe) und $\alpha = .89$ (ar-

span Aufgabe). Für die statistischen Analysen wurde der Mittelwert der Aufgaben aospanspan und arspan als Maß für die AG-Kapazität verwendet.

Shifting. Shifting wurde wie die AG-Kapazität auf Konstruktebene mit den computerbasierten Aufgaben color-shape, number-letter und category-switch (siehe Friedman et al., 2011, 2008; Miyake & Friedman, 2012) erfasst. Für diese Aufgaben wurden die von Friedman et al. (2014, submitted) angewandten Methoden verwendet. Die Aufgabenbearbeitung erfolgte tastaturgesteuert auf MacBooks mit dem Programm PsyScope X B51. Die Tasten D und L wurden mit je einem Etikett mit der Aufschrift „Links“ (Taste D) bzw. „Rechts“ (Taste L) beklebt. Diese Tasten mussten von den Teilnehmern für die Aufgabenbearbeitung mit dem linken („Links“-Taste) bzw. rechten („Rechts“-Taste) Zeigefinger gedrückt werden. Mit der Leertaste erfolgte die weitere Bedienung der Aufgaben (z.B. Aufgabenbeginn). Alle drei Shifting-Aufgaben beinhalteten zum einen no-switch-Durchgänge, bei denen dieselbe Aufgabenregel wie beim jeweils vorausgehenden Durchgang angewendet werden musste. Zum anderen beinhalteten die Shifting-Aufgaben switch-Durchgänge, bei denen eine andere Aufgabenregel wie beim jeweils vorausgehenden Durchgang angewendet werden musste.

Bei der color-shape Aufgabe (Miyake, Emerson, Padilla, & Ahn, 2004) erschien in jedem Durchgang ein Hinweisbuchstabe (C oder S) über einem farbigem Symbol (Kreis oder Dreieck). Die Farben waren rot oder grün. Wenn ein C erschien, musste die Farbe des Symbols durch Klicken der „Links“-Taste bei der Farbe Rot oder durch Klicken der „Rechts“-Taste bei der Farbe Grün angegeben werden. Wenn ein S erschien, musste die Form des Symbols (Kreis oder Dreieck) durch Klicken der „Links“-Taste bei einem Kreis oder der „Rechts“-Taste bei einem Dreieck angegeben werden. Abbildung 4 zeigt Beispiele von Durchgängen mit der Klassifizierung von Farbe bzw. Form.



Abbildung 4. Beispiele für Durchgänge bei der color-shape Aufgabe mit der Klassifizierung von Farbe (links) bzw. Form (rechts).

Vor den eigentlichen Durchgängen absolvierten die Teilnehmer jeweils 12 Übungsdurchgänge nur für die Reaktion auf Farben (es erschien immer ein C neben dem Symbol) und Symbole (es erschien immer ein S neben dem Symbol). Nach den jeweiligen Übungsdurchgängen folgten die eigentlichen Durchgänge, jeweils 26 nur für eine Kategorisierungs-

art. Die eigentlichen Durchgänge waren mit den jeweiligen Übungsdurchgängen bezüglich des Ablaufs gleich. Nach den eigentlichen Durchgängen mit jeweils nur C oder S neben einem Symbol absolvierten die Teilnehmer 24 gemischte Übungsdurchgänge, bei denen sowohl ein C oder ein S neben einem Symbol erscheinen konnte. In diesen Übungsdurchgängen gab es no-switch- und switch-Durchgänge. Bei no-switch-Durchgängen war die Kategorisierungsart identisch zum jeweils vorausgehenden Durchgang (z.B. musste wie beim vorausgehenden Durchgang die Form kategorisiert werden). Bei switch-Durchgängen wechselte die Kategorisierungsart im Vergleich zum *jeweils* vorausgehenden Durchgang (z.B. musste die Form kategorisiert werden und beim vorausgehenden Durchgang die Farbe). Es folgten zwei Blöcke mit jeweils 60 eigentlichen, gemischten Durchgängen, die mit den Übungsdurchgängen hinsichtlich des Ablaufs gleich waren.

In der number-letter Aufgabe (adaptiert von Rogers & Monsell, 1995) sahen die Teilnehmer in jedem Durchgang ein Zahl-Buchstabe Paar wie z.B. 7G in einem Quadranten einer 2x2-Matrix. Die 2x2-Matrix war unten rechts auf dem Bildschirm positioniert. In der Mitte des Bildschirms befand sich ein schwarz umrahmtes Quadrat. Die Zahlen konnten gerade (2, 4, 6, 8) oder ungerade (3, 5, 7, 9) sein. Die Buchstaben waren entweder Konsonanten (G, K, M, R) oder Vokale (A, E, I, U). Wenn das Zahl-Buchstabe-Paar in einem der beiden oberen Quadranten erschien, mussten die Teilnehmer die „Links“-Taste bei einer ungeraden Zahl und die „Rechts“-Taste bei einer geraden Zahl drücken. Wenn das Zahl-Buchstabe-Paar in einem der beiden unteren Quadranten erschien, mussten die Teilnehmer die „Links“-Taste bei einem Konsonanten und die „Rechts“-Taste bei einem Vokal drücken. In Abbildung 5 sind Durchgänge als Beispiele für die Klassifikation einer Zahl bzw. eines Buchstabens dargestellt.

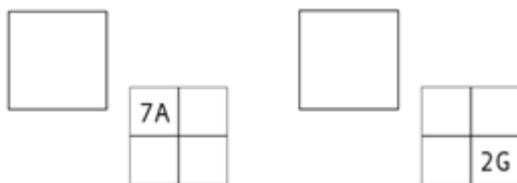


Abbildung 5. Beispiele von Durchgängen bei der number-letter Aufgabe mit der Klassifikation einer Zahl (links) bzw. eines Buchstabens (rechts).

Zunächst absolvierten die Teilnehmer jeweils 12 Übungsdurchgänge nur für Klassifikationen der Zahlen (das Zahl-Buchstabe Paar erschien immer in einem der beiden oberen Quadranten) und Klassifikationen der Buchstaben (das Zahl-Buchstabe-Paar erschien immer

in einem der beiden unteren Quadranten). Anschließend bearbeiteten die Teilnehmer die eigentlichen Durchgänge, jeweils 34 nur für Klassifikationen von Zahlen und Buchstaben. Die eigentlichen Durchgänge stimmten mit den Übungsdurchgängen bezüglich des Ablaufs überein. Nach den eigentlichen Durchgängen mit jeweils nur Zahlen oder Buchstaben absolvierten die Teilnehmer 12 gemischte Übungsdurchgänge, bei denen Klassifikationen von Buchstaben und Zahlen innerhalb der Durchgänge möglich waren (das Zahl-Buchstabe-Paar konnte in einem der oberen oder einem der unteren Quadranten erscheinen). In den gemischten Übungsdurchgängen gab es no-switch-Durchgänge und switch-Durchgänge. Bei no-switch-Durchgängen musste dieselbe Klassifikation wie in den jeweils vorausgehenden Durchgängen angewendet werden (z.B. musste wie im vorausgehenden Durchgang die Zahl klassifiziert werden). Bei switch-Durchgängen musste eine andere Klassifikation wie in den *jeweils* vorausgehenden Durchgängen angewendet werden (z.B. musste eine Zahl klassifiziert werden und im vorausgehenden Durchgang ein Buchstabe). Das Zahl-Buchstabe-Paar erschien bei diesen gemischten Durchgängen an einer vorhersagbaren Stelle, da es im Uhrzeigersinn in den Quadranten erschien. Es folgten zwei Blöcke mit jeweils 68 eigentlichen, gemischten Durchgängen, die mit den Übungsdurchgängen hinsichtlich des Ablaufs übereinstimmten. Nach den gemischten Durchgängen mit vorhersagbarem Erscheinen des Zahl-Buchstabe-Paars folgte ein Übungsblock mit 24 gemischten Durchgängen, in welchen das Zahl-Buchstabe-Paar in einem der Quadranten zufällig erschien. Kurz vor dem Erscheinen des Zahl-Buchstabe-Paars färbte sich der Rahmen des betreffenden Quadranten schwarz. Bei diesen Durchgängen gab es analog zu den gemischten Blöcken mit vorhersagbarer Reihenfolge no-switch- und switch-Durchgänge. Nach dem Übungsblock mit zufälligem Erscheinen des Zahl-Buchstabe-Paars folgten zwei Blöcke mit jeweils 68 eigentlichen, gemischten Durchgängen, die hinsichtlich des Ablaufs mit den Durchgängen des Übungsblocks übereinstimmten.

Bei der category-switch Aufgabe (adaptiert von Mayr & Kliegl, 2000) sahen die Teilnehmer in jedem Durchgang ein Wort, welches entweder etwas beschreibt, das lebendig/nicht lebendig oder kleiner/größer als ein Fußball ist. Die Wörter in den einzelnen Durchgängen waren in englischer Sprache, da sie nicht verändert werden konnten. Den Teilnehmern wurden vor Aufgabenbeginn als Hilfestellung die deutschen Wörter in Klammern hinter den englischen Wörtern gezeigt. In der Anfangsinstruktion war ein Hinweis, dass sich die Teilnehmer Zeit nehmen sollten, sich die englischen Wörter und deren Bedeutungen zu merken. Folgende Wörter wurden präsentiert: table, bicycle, coat, cloud, pebble, knob, marble, snowflake, shark,

lion, oak, alligator, mushroom, sparrow, goldfish und lizard. Ein Symbol (Herz oder Kreuz) unter dem Wort gab einen Hinweis darauf, welche Kategorisierung (lebendig/nicht lebendig vs. kleiner/größer als ein Fußball) angewendet werden musste. Wenn ein Herz unter dem Wort erschien, musste die „Links“-Taste für die Kategorisierung „nichtlebendig“ und die „Rechts“-Taste für die Kategorisierung „lebendig“ gedrückt werden. Wenn ein Kreuz unter dem Wort erschien, musste die „Links“-Taste für die Kategorisierung „kleiner als ein Fußball“ und die „Rechts“-Taste für die Kategorisierung „größer als ein Fußball“ gedrückt werden. Abbildung 6 zeigt Beispiele für Durchgänge mit der Kategorisierung als „lebendig/nicht lebendig“ bzw. „kleiner/größer als ein Fußball“.



Abbildung 6. Beispiele für Durchgänge bei der category-switch Aufgabe mit der Kategorisierung als „lebendig/nicht lebendig“ (links) bzw. „kleiner/größer als ein Fußball“ (rechts).

Vor den eigentlichen Durchgängen absolvierten die Teilnehmer jeweils 12 Übungsdurchgänge für die Kategorisierung „nichtlebendig/lebendig“ (es erschien immer ein Herz unter dem Wort) und die Kategorisierung „kleiner/größer als ein Fußball“ (es erschien immer ein Kreuz unter dem Wort). Die jeweils darauf folgenden, eigentlichen Durchgänge (34 pro Kategorisierungsart) für die beiden Kategorisierungsarten waren mit den Übungsdurchgängen bezüglich des Ablaufs gleich. Nach den eigentlichen Durchgängen mit nur einer Kategorisierungsart (entweder nichtlebendig/lebendig oder kleiner/größer als ein Fußball) absolvierten die Teilnehmer 24 gemischte Übungsdurchgänge, bei denen beide Kategorisierungsarten wechseln konnten (es konnte ein Herz oder ein Kreuz unter dem Wort erscheinen). In diesen Durchgängen gab es no-switch und switch-Durchgänge. Bei den no-switch-Durchgängen musste dieselbe Kategorisierungsart wie bei den jeweils vorausgehenden Durchgängen angewendet werden (z.B. musste wie im vorausgehenden Durchgang entschieden werden, ob durch ein Wort etwas beschrieben wird, das lebendig/nicht lebendig ist). Bei den switch-Durchgängen musste eine andere Kategorisierungsart wie bei den *jeweils* vorausgehenden Durchgängen angewendet werden (z.B. musste entschieden werden, ob ein Wort etwas beschreibt, das lebendig/nicht lebendig ist und beim vorausgehenden Durchgang, ob ein Wort etwas beschreibt, das kleiner/größer als ein Fußball ist). Es folgten zwei Blöcke mit jeweils 68 eigentlichen, gemischten Durchgängen, die mit den Übungsdurchgängen hinsichtlich des Ablaufs gleich waren.

Die abhängige Variable bei allen Shifting-Aufgaben war die Differenz in den mittleren Reaktionszeiten zur Antworteingabe zwischen den korrekten switch- und no-switch-Durchgängen. Durchgänge, die auf Fehler folgten, wurden von der Analyse ausgeschlossen. Ebenso gab es trash-Durchgänge, die ebenfalls nicht mit in die Analyse einfließen. Ein Wert in der Number letter Aufgabe wich mehr als drei Standardabweichungen vom Mittelwert ab und wurde durch den Wert ersetzt, der drei Standardabweichungen vom Mittelwert abwich. Diese Transformation wurde vorgenommen, um den Einfluss von Extremwerten zu reduzieren und eine Normalverteilung der Werte besser zu erreichen (Friedmann et al., 2008).

Für jede der drei Shifting-Aufgaben wurde die Split-Half-Reliabilität (Guttman) durch Verwendung von switch-costs für zwei Testhälften berechnet. Es ergaben sich folgende Werte: $r_{tt} = .91$ (color-shape Aufgabe), $r_{tt} = .90$ (number-letter Aufgabe), $r_{tt} = .86$ (category-switch Aufgabe). Für die statistischen Analysen wurde der Mittelwert der Aufgaben color-shape, number-letter und category-switch als Maß für die Shifting-Fähigkeit verwendet.

Fluide Intelligenz. Für die Messung von Gf wurden die drei Subtests „Numerisch-induktives Denken“, „Figural-induktives Denken“ und „Verbal-deduktives Denken“ der Intelligenz-Struktur-Batterie (INSBAT; Arendasy et al., 2012) verwendet. Bei allen Subtests war die Aufgabenanzahl variabel, da sie von den Leistungen der Teilnehmer abhängig war (adapatives Testen; Arendasy et al., 2012). Die Dauer der Aufgabenbearbeitung betrug ca. 48 Minuten.

Im Subtest „Numerisch-induktives Denken“ bekamen Teilnehmer bei jedem Item eine Folge von Zahlen präsentiert. Dieser Folge von Zahlen lag eine Regel zugrunde, die von den Teilnehmern erkannt werden musste. Nach dem Erschließen der Regel musste die korrekte Zahl zur Ergänzung der Folge eingegeben werden. Abbildung 7 zeigt ein Beispielitem aus dem Subtest „Numerisch-induktives Denken“.

Kapitel 5: Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung

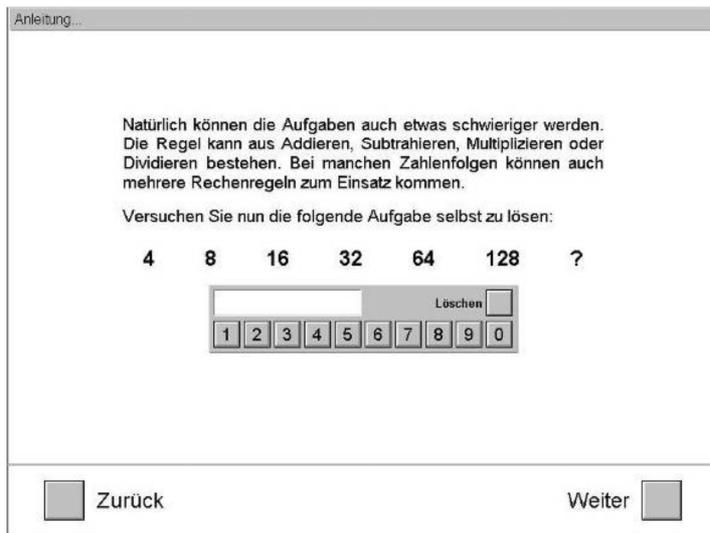


Abbildung 7. Beispielitem aus dem Subtest „Numerisch-induktives Denken“ der INSBAT (unverändert übernommen aus Arendasy et al., 2012).

Im Subtest „Figural-induktives Denken“ mussten Teilnehmer bei jedem Durchgang das nicht ausgefüllte Feld einer 3x3-Matrix mit einem Symbol aus anderen Feldern ergänzen. Die Symbole in den acht ausgefüllten Feldern folgten einer Regel, die der Teilnehmer erkennen musste. Es standen acht mögliche Antwortkategorien bzw. Symbole zur Auswahl. Abbildung 8 zeigt ein Beispielitem des Subtests „Figural-induktives Denken“.

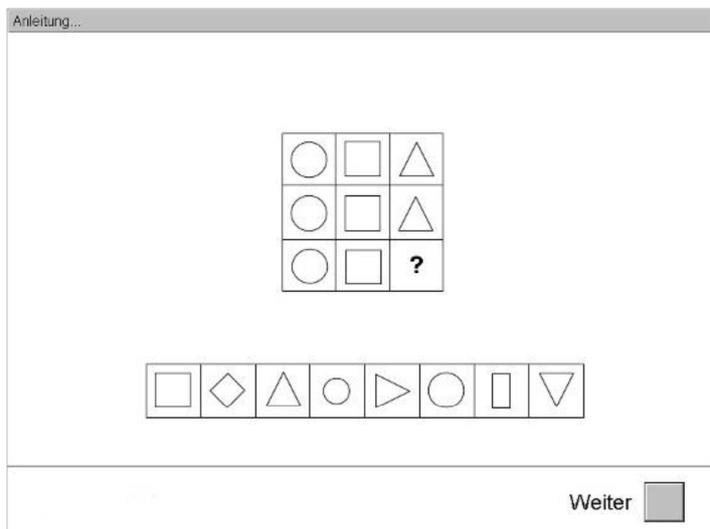


Abbildung 8. Beispielitem aus dem Subtest „Figural-induktives Denken“ der INSBAT (unverändert übernommen aus Arendasy et al., 2012).

Beim Subtest „Verbal-deduktives Denken“ bekamen Teilnehmer Items mit jeweils zwei Aussagen präsentiert. Die Aussagen galten als Voraussetzungen und es musste eine

Schlussfolgerung daraus gezogen werden. Es standen fünf Antwortmöglichkeiten zur Auswahl, von denen eine innerhalb von 45 Sekunden ausgewählt werden musste. Die Bearbeitungsdauer und die noch verfügbare Zeit wurden durch einen Balken angezeigt. Dieser Balken färbte sich kurz vor Ablauf der übrigen Zeit rot. Abbildung 9 zeigt ein Beispielitem des Subtests „Verbal-deduktives Denken“.

Anleitung...

Noch ein Beispiel:
Welche Schlussfolgerung muss bei dieser Aufgabe immer gelten?

**Alle Bienen sind keine Wespen.
Einige Insekten sind Bienen.**

Alle Insekten sind Wespen.
 Einige Insekten sind Wespen.
 Alle Insekten sind keine Wespen.
 Einige Insekten sind keine Wespen.
 Keine der Schlussfolgerungen ist richtig.

Abbildung 9. Beispielitem aus dem Subtest „Verbal-deduktives Denken“ der INSBAT (unverändert übernommen aus Arendasy et al., 2012).

Die Ergebnisse der drei Subtests wurden in einem Rohwert abgebildet, der als Schätzung für Gf verwendet wurde. Die Reliabilitäten der drei Subtests der INSBAT konnten voreingestellt werden und betragen jeweils $\alpha = .70$ (Arendasy et al., 2012).

Cognitive load. Das Niveau der kognitiven Belastung wurde mittels der, aus einem Item bestehenden Ratingskala von Paas (1992) operationalisiert, die unmittelbar nach der Bearbeitung jedes statistischen Problems eingesetzt wurde. Nach Brünken, Plass und Leutner (2003) scheint diese Skala den subjektiv wahrgenommenen Aufwand reliabel erfassen zu können. Die CL Skala (siehe Anhang G) war neunstufig und wurde ins Deutsche übersetzt. Die Ankerpunkte der Skala reichten von sehr, sehr niedriger kognitiver Belastung (1) bis zu sehr, sehr hoher kognitiver Belastung (9). Für eindimensionale Ratingskalen konnte gezeigt werden, dass diese sensitiv in Bezug auf Unterschiede in der kognitiven Belastung sowie valide, reliabel und nicht intrusiv sind (z.B. Gimino, 2002; Paas, Van Merriënboer, & Adam, 1994). In der vorliegenden Studie wurde als Reliabilitätsmaß für den CL die interne Konsis-

tenz über die drei Messwerte im CL berechnet. Die interne Konsistenz betrug $\alpha = .80$. Für die statistischen Analysen wurde der Mittelwert des CL aus den drei Messungen verwendet.

Wissenserwerb. Der Erwerb konzeptuellen Wissens und der Erwerb anwendungsorientierten Wissens wurden jeweils durch die Punktedifferenz zwischen einem Vorwissenstest und einem Wissenstest zum Posttest berechnet. Die Wissenstests zum Posttest (siehe Anhang B) stellen eine Parallelförmigkeit der Vorwissenstests (siehe Anhang A) dar, die beim Pretest bearbeitet wurden. Die Wissenstests wurden auf Basis vorhandener Fragen (Übungs- und Klausurfragen) am Lehrstuhl für psychologische Methodenlehre und Diagnostik der Ludwig-Maximilians-Universität München konstruiert, um eine hohe Inhaltsvalidität der Wissenstests zu erzielen. Die Wissenstests zur Erfassung konzeptuellen Wissens beinhalten eine multiple-choice (MC)-Frage zu Skalenniveaus als grundlegenden statistischen Inhalt. Die restlichen Fragen sind aus dem Themenbereich ALM, weisen jedoch oftmals einen inhaltlichen Bezug zu Skalenniveaus auf. Die Wissenstests zur Erfassung konzeptuellen Wissens bestehen aus jeweils vier MC-Fragen. Bei jeder MC-Frage gibt es fünf Antwortmöglichkeiten, von denen mehrere richtig sein können. Die MC-Fragen erfassen in erster Linie konzeptuelles Wissen, da zur Beantwortung der Fragen Informationen abgerufen werden müssen, die üblicherweise miteinander verbunden sind. Abbildung 10 zeigt beispielhaft die zweite MC-Frage aus dem Wissenstest zur Erfassung konzeptuellen Vorwissens mit Lösung.

2. Kreuzen Sie die Voraussetzungen an, die bei einer zweifaktoriellen Varianzanalyse ohne Messwiederholung gegeben sein müssen:

- a) Die abhängige Variable ist mindestens nominal skaliert.
- b) Die Interaktion zwischen den beiden Faktoren ist ordinal.
- c) *Die abhängige Variable ist in der Population in allen Zellen der Faktorstufenkombinationen normalverteilt.*
- d) *Die Varianzen müssen in jeder Teilpopulation gleich sein.*
- e) Die unabhängige Variable muss mindestens ordinalskaliert sein.

Abbildung 10. Zweite MC-Frage aus dem Wissenstest zur Erfassung konzeptuellen Vorwissens. Richtige Antwortmöglichkeiten sind kursiv gedruckt.

Zur Beantwortung dieser Frage müssen Informationen zu den Voraussetzungen einer zweifaktoriellen Varianzanalyse ohne Messwiederholung abgerufen werden. Diese Informati-

onen können typischerweise im Rahmen des Konzepts der zweifaktoriellen Varianzanalyse ohne Messwiederholung miteinander verbunden sein. Beispielsweise weisen die Unterfragen 2.a und 2.b aufgrund des Wissens über Skalenniveaus Beziehungen zueinander auf und beide können unter die Voraussetzungen der zweifaktoriellen Varianzanalyse ohne Messwiederholung subsumiert werden.

Für jede der 5 Antwortmöglichkeiten bei jeder der 4 Fragen gab es bei richtiger Beantwortung einen Punkt, bei falscher Beantwortung keinen Punkt. Die maximale Punktzahl bei den MC-Fragen betrug somit 20 Punkte.

Die Wissenstests zur Erfassung anwendungsorientierten Wissens bestehen aus jeweils zwei offenen Fragen. Die offenen Fragen erfassen in erster Linie anwendungsorientiertes Wissen, da zur Beantwortung der Fragen wesentliche Aspekte eines statistischen Problems identifiziert und die zur Lösung des Problems notwendigen Informationen abgerufen werden müssen. In Abbildung 11 ist beispielhaft die erste offene Frage des Wissenstests zur Erfassung anwendungsorientierten Vorwissens dargestellt.

1. In einer Untersuchung soll eine neue Fruchtjoghurtsorte hinsichtlich der Beurteilung des Geschmacks (intervallskaliertes Merkmal) mit einer bereits auf dem Markt befindlichen Fruchtjoghurtsorte und einem Joghurt ohne Fruchtzubereitung verglichen werden. Zusätzlich sollen unterschiedliche Geschmackspräferenzen von Männern und Frauen berücksichtigt werden. Von Interesse ist darüber hinaus auch, ob die Präferenz einer Joghurtsorte vom Geschlecht abhängt. *Um welche Art von Design handelt es sich bei der vorliegenden Untersuchung? Wie können die Fragestellungen statistisch beantwortet werden? Bitte begründen Sie ihre Antworten kurz.*

Abbildung 11. Erste offene Frage aus dem Wissenstest zur Erfassung anwendungsorientierten Vorwissens.

Um beispielweise das Design der Untersuchung in dieser Frage zu bestimmen, müssen die Aspekte, die das Design bestimmen, identifiziert werden. Unter Rückgriff auf theoretisches Wissen kann dann das Design benannt werden.

Für die Bewertung der offenen Fragen wurde ein Auswertungsschema entwickelt. Jede Frage wurde in drei Teilfragen unterteilt, für die es bei richtiger Beantwortung einen, bei falscher Beantwortung keinen Punkt gab. Die maximale Punktzahl bei den offenen Fragen betrug somit 6. In der ersten offenen Frage des Wissenstests zur Erfassung anwendungsorientierten Vorwissens (siehe Abbildung 11) gab es einen Punkt für die Nennung des Designs (2x3-

oder 3x2-Design), einen Punkt für die Nennung der unabhängigen Variablen bzw. Faktoren Geschlecht (zweistufig) sowie Art des Joghurts (dreistufig) und einen Punkt für die Nennung der Auswertungsmethode zur Beantwortung der statistischen Fragestellung (zweifaktorielle Varianzanalyse (ohne Messwiederholung) oder mehrfaktorielle Varianzanalyse mit einem zwei- und einem dreistufigen Faktor).

Die Wissenstests unterscheiden sich im Antwortformat und dem erfassten Wissen (MC-Fragen zur Erfassung konzeptuellen Wissens vs. offene Fragen zur Erfassung anwendungsorientierten Wissens). Daher erfolgte eine getrennte Auswertung für MC- und offene Fragen. Der Wissenserwerb wurde jeweils getrennt für die beiden Frage- bzw. Wissensarten betrachtet.

Für die Wissenstests wurde getrennt nach Wissensarten jeweils der Kuder-Richardson-20-Koeffizient als Reliabilitätsmaß berechnet. Es ergaben sich folgende Werte: $r_{tt} = .47$ (konzeptuelles Vorwissen), $r_{tt} = .49$ (anwendungsorientiertes Vorwissen), $r_{tt} = .47$ (konzeptuelles Wissen zum Posttest) und $r_{tt} = .44$ (anwendungsorientiertes Wissen zum Posttest).

Für die Quantifizierung des Grads der Übereinstimmung von zwei Ratern (dem Autor der Arbeit sowie dem zusammengesetzten Urteil von weiblichen Studierenden für separate Aufgaben) bezüglich über 40% der Itemantworten wurde Cohen's Kappa für MC- und offene Fragen getrennt für die Vorwissenstests und die Wissenstests zum Posttest berechnet. Für die Berechnung wurden alle Antwortmöglichkeiten bei den MC-Fragen und die Teilfragen bei den offenen Fragen als dichotome Items behandelt. Bei den Vorwissenstests betragen die durchschnittlichen Werte von Cohen's Kappa $\kappa = .96$ für die MC-Fragen und $\kappa = .95$ für die offenen Fragen. Bei den Wissenstests zum Posttest betragen die durchschnittlichen Werte von Cohen's Kappa $\kappa = .97$ für die MC-Fragen und $\kappa = .92$ für die offenen Fragen. Bei Nichtübereinstimmung wurden die betreffenden Antworten erneut überprüft und Unstimmigkeiten behoben.

5.2.5 Kontrollvariablen

Kontrollvariablen waren demografische Daten (Alter, Geschlecht), Semesterzahl sowie Interesse und Motivation.

Interesse und Motivation wurden mit dem Fragebogen zur aktuellen Motivation in Lern- und Leistungssituationen (FAM; Rheinberg, Vollmeyer, & Burns, 2001) operationalisiert. Die siebenstufigen Items des FAM erfassen die Skalen Interesse, Misserfolgsbefürch-

tung, Erfolgswahrscheinlichkeit und Herausforderung. Für jede Skala wurde als Reliabilitätsmaß Cronbach's Alpha berechnet. Es ergaben sich folgende Werte: $\alpha = .80$ (Skala Interesse), $\alpha = .82$ (Skala Misserfolgsbefürchtung), $\alpha = .89$ (Skala Erfolgswahrscheinlichkeit) und $\alpha = .52$ (Skala Herausforderung). Der FAM ist in Anhang C aufgeführt.

5.2.6 Statistische Analyse

Für alle statistischen Tests wurden die Voraussetzungen überprüft (für Voraussetzungen zu t-Tests, Varianzanalysen, Korrelationen und Regressionen siehe Bühner & Ziegler, 2009; für Voraussetzungen von Kovarianzanalysen siehe Bortz & Schuster, 2010). Für eine zweifaktorielle Varianzanalyse mit Messwiederholung sowie Kovarianzanalysen wurde die Normalverteilungsannahme der abhängigen Variable in den beiden Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung mit Kolmogoroff-Smirnov Tests und Histogrammen überprüft. Für die Berechnung von Korrelationen wurden die Normalverteilung der jeweiligen Variablen ebenfalls mit Kolmogoroff-Smirnov Tests und Histogrammen sowie die Linearität der Zusammenhänge von jeweils zwei Variablen und das Vorliegen von Ausreißerwerten mit Streudiagrammen überprüft. Zur Berechnung der zweifaktorielle Varianzanalyse mit Messwiederholung und der Kovarianzanalyse wurde die Sphärizitätsannahme mit dem Mauchly-Test sowie im Falle der Kovarianzanalyse die Homogenität der Steigungen in den einzelnen Stichproben mittels Interaktionsdiagrammen überprüft. Bei den Moderatoranalysen wurden Voraussetzungen für Regressionsanalysen überprüft. Diese Voraussetzungen sind Linearität (mit Streudiagrammen überprüft), keine Kollinearität (mit Konditionsindizes und Tolerance überprüft), Normalverteilung der Residuen (mit Kolmogoroff-Smirnov Tests sowie Histogrammen überprüft) und Homoskedastizität (mit Streudiagrammen untersucht). In den statistischen Analysen konnten keine merklichen Voraussetzungsverletzungen mit Ausnahme der Verletzung der Normalverteilungsannahme festgestellt werden. Einzelne Variablen wie beispielsweise konzeptuelles und anwendungsorientiertes Vorwissen waren nicht normalverteilt. Allerdings sind die Varianzanalyse und die Kovarianzanalyse stabil gegen Voraussetzungsverletzungen (Bortz & Schuster, 2010, S. 214, S.311; Glass, Peckham & Sanders, 1972). Zudem wird bei einem $n > 30$ häufig davon ausgegangen, dass die Verteilung der Mittelwerte normal ist (Bortz & Schuster, 2010, S. 87). Diese Anforderung an die Stichprobengrößen war in der vorliegenden Studie gegeben.

Mit t-Tests für unabhängige Stichproben wurde auf Unterschiede in den abhängigen

Variablen und Kontrollvariablen zwischen den beiden Lernumgebungen zum Pretest getestet. Für die Beurteilung der Größe von Hedge's g wurden folgende Richtlinien von Cohen (1988) verwendet: $g = 0.20$ (kleiner Effekt), $g = 0.50$ (moderater Effekt), $g = 0.80$ (starker Effekt).

Mittels punktbiserialer Korrelationen und Pearson-Korrelationen wurde auf Zusammenhänge zwischen Kontrollvariablen und abhängigen Variablen getestet. Für die Beurteilung der Größe einer Korrelation wurden folgende Richtlinien von Cohen (1988) verwendet: $r = .10$ (kleiner Effekt); $r = .30$ (mittlerer Effekt); $r = .50$ (starker Effekt).

Mit einer einfaktoriellen, univariaten Kovarianzanalyse wurde überprüft, ob sich die Lernumgebungen im CL unter Kontrolle von konzeptuellem und anwendungsorientiertem Vorwissen, FAM Skala Misserfolgsbefürchtung, FAM Skala Erfolgswahrscheinlichkeit und FAM Skala Herausforderung unterscheiden.

Mit einer zweifaktoriellen, univariaten Varianzanalyse mit Messwiederholung wurde überprüft, ob Unterschiede im konzeptuellen Wissen zwischen den Messzeitpunkten vom Grad an instruktionaler Unterstützung abhängen.

Mit einer zweifaktoriellen, univariaten Kovarianzanalyse mit Messwiederholung wurde überprüft, ob Unterschiede im anwendungsorientierten Wissen zwischen den Messzeitpunkten unter Kontrolle von Alter und Semesterzahl vom Grad an instruktionaler Unterstützung abhängen.

Für die Beurteilung der Größe von η^2 wurden folgende Konventionen von Murphy und Myers (2004) verwendet: $\eta^2 = .01$ (kleiner Effekt), $\eta^2 = .06$ (mittlerer Effekt), $\eta^2 = .14$ (großer Effekt).

Die Moderationshypothesen 3 sowie 4.a, 4.b und 4.c wurden mit dem SPSS-Macro PROCESS (Hayes, 2013) getestet. Im Rahmen von Moderatoranalysen mit PROCESS wurde jeweils regressionsanalytisch überprüft, ob eine signifikante Interaktion zwischen der unabhängigen Variable (Grad an instruktionaler Unterstützung) und den angenommenen Moderatoren Vorwissen (Hypothese 3), AG-Kapazität (Hypothese 4.a), Shifting (Hypothese 4.b) und Gf (Hypothese 4.c) vorliegt. Zudem wurden Kontrollvariablen berücksichtigt. Wie von Hayes (2012) vorgeschlagen, wurden Standardfehler geschätzt, die konsistent gegenüber Heteroskedastizität sind. Die Johnson-Neyman (JN) Technik (Hayes, 2013) wurde angewandt, um den Effekt der unabhängigen Variable (Grad an instruktionaler Unterstützung) auf die abhängige Variable (Wissenserwerb) für unterschiedliche Ausprägungen der jeweiligen Moderatorvariable (Vorwissen, AG-Kapazität, Shifting oder Gf) zu quantifizieren. Mit dieser Technik kann

festgestellt werden, ob bei einer bestimmten Ausprägung der Moderatorvariable die unabhängige Variable einen signifikanten Effekt auf die abhängige Variable hat (Hayes, 2013).

Das Signifikanzniveau wurde für alle statistischen Analysen auf 5% festgelegt. Für die unstandardisierten Regressionskoeffizienten und den bedingten Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb, jeweils für unterschiedliche Werte von AG-Kapazität, Shifting und Gf werden die 95%-Konfidenzintervalle angegeben.

5.3 Ergebnisse

5.3.1 Vorausgehende Analysen

Bedeutsame Korrelationen zwischen abhängigen Variablen und Kontrollvariablen

Die Korrelationen innerhalb der Aufgaben zur Erfassung der AG-Kapazität und der Shifting-Fähigkeit sind in Anhang H aufgeführt.

Die AG-Kapazität korrelierte nur mit Gf signifikant, $r(76) = .40$ (mittlerer bis starker Effekt), $p < .001$. Es ist anzumerken, dass die AG-Kapazität nicht signifikant mit dem CL zusammenhing, $r(76) = -.10$ (kleiner Effekt), $p = .38$.

Shifting war mit keiner der abhängigen Variablen korreliert.

Außer dem signifikanten Zusammenhang mit der AG-Kapazität korrelierte Gf auch mit dem CL, $r(76) = -.30$ (mittlerer Effekt), $p < .01$. Die Korrelation zwischen Gf und CL blieb auch dann signifikant, wenn der Einfluss der AG-Kapazität auf beide Variablen auspartialisiert wurde, $r_{GfCL.WMC}(76) = -.28$ (kleiner bis mittlerer Effekt), $p < .05$.

Neben dem signifikanten Zusammenhang des CL mit Gf, war der CL auch mit folgenden Variablen korreliert: konzeptuelles Vorwissen ($r(76) = -.27$ (kleiner bis mittlerer Effekt), $p < .05$), anwendungsorientiertes Vorwissen ($r(76) = -.33$ (mittlerer bis starker Effekt), $p < .01$), FAM Skala Misserfolgsbefürchtung ($r(76) = .25$ (kleiner bis mittlerer Effekt), $p < .05$), FAM Skala Erfolgswahrscheinlichkeit ($r(76) = -.36$ (mittlerer bis starker Effekt), $p < .01$), FAM Skala Herausforderung ($r(76) = .25$ (kleiner bis mittlerer Effekt), $p < .05$). Unter diesen Variablen war das konzeptuelle Vorwissen mit der FAM Skala Erfolgswahrscheinlichkeit korreliert ($r(76) = .26$ (kleiner bis mittlerer Effekt), $p < .05$). Das anwendungsorientierte Vorwissen korrelierte ebenfalls mit der FAM Skala Erfolgswahrscheinlichkeit ($r(76) = .35$ (mittlerer bis starker Effekt), $p < .01$) sowie mit der FAM Skala Herausforderung ($r(76) = -.27$ (kleiner bis mittlerer Effekt), $p < .05$). Darüber hinaus war das anwendungsorientierte Vorwissen auch signifikant mit Gf korreliert ($r(76) = .28$ (kleiner bis mittle-

rer Effekt), $p < .05$). Die Variablen, die mit dem CL korreliert waren, wurden in der Kovarianzanalyse zur Überprüfung des Unterschieds im CL zwischen den Lernumgebungen berücksichtigt.

Unterschiede zum Pretest und Korrelationen mit dem Wissenserwerb

Nur die Semesterzahl unterschied sich signifikant zwischen den Experimentalbedingungen bzw. Lernumgebungen, $t(74) = -2.79$, $p < .05$, $g = 0.57$ (moderater bis starker Effekt). Die Semesterzahl war in der Bedingung mit hohem Grad an instruktionaler Unterstützung ($M = 3.61$, $SD = 2.13$) höher als in der Bedingung mit niedrigem Grad an instruktionaler Unterstützung ($M = 2.63$, $SD = 1.15$).

Nur das konzeptuelle Vorwissen und das konzeptuelle Wissen zum Posttest korrelierten signifikant mit dem Erwerb von konzeptuellem Wissen ($r(76) = -.70$ (starker Effekt), $p < .001$ für das konzeptuelle Vorwissen; $r(76) = .53$ (starker Effekt), $p < .001$ für das konzeptuelle Wissen zum Posttest).

Das anwendungsorientierte Vorwissen und das anwendungsorientierte Wissen zum Posttest korrelierten signifikant mit dem Erwerb von anwendungsorientiertem Wissen ($r(76) = -.49$ (mittlerer bis starker Effekt), $p < .001$ für das anwendungsorientierte Vorwissen; $r(76) = .63$ (starker Effekt), $p < .001$ für das anwendungsorientierte Wissen zum Posttest). Außerdem waren das Alter und die Semesterzahl signifikant mit dem Erwerb von anwendungsorientiertem Wissen korreliert ($r(76) = .23$ (kleiner bis mittlerer Effekt), $p < .05$ für das Alter; $r(76) = .25$ (kleiner bis mittlerer Effekt), $p < .05$ für die Semesterzahl). Folglich wurden in der Kovarianzanalyse für die Testung der Hypothese 2.b sowie den Moderatoranalysen zur Testung der Hypothesen 3 und 4 für das Alter und die Semesterzahl kontrolliert.

Unterschiede im cognitive load und dem Wissenserwerb (Hypothesen 1 und 2)

Deskriptiv war der CL in der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung ($M = 6.05$, $SD = 1.11$) höher als in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung ($M = 5.85$, $SD = 1.48$). Im Gegensatz zur Hypothese 1 war dieser Unterschied in einer einfaktoriellen Kovarianzanalyse mit dem Faktor Grad an instruktionaler Unterstützung (niedrig vs. hoch) und den Kontrollvariablen konzeptuelles und anwendungsorientiertes Vorwissen, FAM Skala Misserfolgsbefürchtung, FAM Skala Erfolgswahrschein-

lichkeit und FAM Skala Herausforderung jedoch nicht signifikant, $F(1, 68) = 0.93$, $p = .34$, partielles $\eta^2 = .01$ (kleiner Effekt).

Eine zweifaktorielle Varianzanalyse mit Messwiederholung und den Faktoren Grad an instruktionaler Unterstützung (niedrig vs. hoch) und Messzeitpunkt (Pretest vs. Posttest) wurde für die abhängige Variable konzeptuelles Wissen durchgeführt. Es gab keinen signifikanten Haupteffekt für den Grad an instruktionaler Unterstützung ($F(1, 74) = 0.17$, $p = .68$, partielles $\eta^2 = .002$), aber einen signifikanten Haupteffekt für den Messzeitpunkt ($F(1, 74) = 33.16$, $p < .001$, partielles $\eta^2 = .31$ (großer Effekt)). Im Gegensatz zur Hypothese 2.a war der Wissenserwerb unabhängig vom Grad an instruktionaler Unterstützung, $F(1, 74) = 0.56$, $p = .46$, partielles $\eta^2 = .01$ (kleiner Effekt). Tabelle 3 zeigt die Mittelwerte und Standardabweichungen der Pretest- und Posttestwerte des konzeptuellen Wissens getrennt nach Experimentalbedingungen bzw. Lernumgebungen.

Eine zweifaktorielle Kovarianzanalyse mit Messwiederholung und den Faktoren Grad an instruktionaler Unterstützung (niedrig vs. hoch) und Messzeitpunkt (Pretest vs. Posttest) sowie den Kontrollvariablen Alter und Semesterzahl wurde für die abhängige Variable anwendungsorientiertes Wissen durchgeführt. Es gab keinen signifikanten Haupteffekt für den Grad an instruktionaler Unterstützung ($F(1, 72) = 1.38$, $p = .24$, $\eta^2 = .02$ (kleiner Effekt)) und keinen signifikanten Haupteffekt für den Messzeitpunkt ($F(1, 72) = 0.20$, $p = .65$, partielles $\eta^2 = .003$). Konform mit Hypothese 2.b war der Wissenserwerb vom Grad an instruktionaler Unterstützung abhängig, $F(1, 72) = 5.75$, $p < .05$, partielles $\eta^2 = .07$ (mittlerer bis großer Effekt). Der Erwerb von anwendungsorientiertem Wissen (Differenz zwischen dem Wissenstest beim Posttest und dem Vorwissenstest) war bei der Lernumgebung mit hohem Grad an instruktionaler Unterstützung höher ($M_{diff} = 1.53$) als bei der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung ($M_{diff} = 0.53$). Tabelle 4 zeigt die Mittelwerte und Standardabweichungen der Pretest- und Posttestwerte des anwendungsorientierten Wissens getrennt nach Experimentalbedingungen.

Kapitel 5: Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung

Tabelle 3

Mittelwerte und Standardabweichungen für Pretest- und Posttestwerte des konzeptuellen Wissens getrennt nach Experimentalbedingungen

	Grad an instruktionaler Unterstützung	<i>M</i>	<i>SD</i>
Vorwissen (konzeptuell)	Gering (Problemlösen)	12.29	2.42
	Hoch (Lösungsbeispiele)	11.84	2.94
Wissen zum Posttest (konzeptuell)	Gering (Problemlösen)	14.05	2.34
	Hoch (Lösungsbeispiele)	14.13	2.23

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung.

Tabelle 4

Mittelwerte und Standardabweichungen für Pretest- und Posttestwerte des anwendungsorientierten Wissens getrennt nach Experimentalbedingungen

	Grad an instruktionaler Unterstützung	<i>M</i>	<i>SD</i>
Vorwissen (anwendungsorientiert)	Gering (Problemlösen)	1.68	1.07
	Hoch (Lösungsbeispiele)	1.47	1.35
Wissen zum Posttest (anwendungsorientiert)	Gering (Problemlösen)	2.21	1.36
	Hoch (Lösungsbeispiele)	3.00	1.27

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung.

Moderierende Rolle des Vorwissens (Hypothese 3)

Es gab keinen Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von konzeptuellem Wissen (siehe oben, Abschnitt „*Unterschiede im cognitive load und dem Wissenserwerb (Hypothesen 1 und 2)*“). Zudem ergab eine Häufigkeitsanalyse für das konzeptuelle Vorwissen, dass bei einer Spanne von 0 bis 20 Punkten im Gesamtwert der Fragen zur Erfassung des konzeptuellen Vorwissens 98.68% der Teilnehmer ein konzeptuelles Vorwissen von mindestens 8 Punkten aufwiesen. Es gab nur einen Teilnehmer mit einem geringen Vorwissen von 5 Punkten. Da folglich der moderierende Einfluss des Grads an instruktionaler Unterstützung im Bereich geringen konzeptuellen Vorwissens kaum nachgewiesen werden kann und sich die Lernumgebungen nicht im Erwerb konzeptuellen Wissens unterschieden, wurde nur der moderierende Einfluss des anwendungsorientierten Vorwissens für den Erwerb von anwendungsorientiertem Wissen untersucht.

In der Moderatoranalyse für den Erwerb von anwendungsorientiertem Wissen wurden das Alter und die Semesterzahl als Kovariaten berücksichtigt. Der Grund hierfür war, dass das Alter und die Semesterzahl einen signifikanten Einfluss auf den Erwerb von anwendungsorientiertem Wissen hatten (siehe oben, Abschnitt „Unterschiede zum Pretest und Korrelationen mit dem Wissenserwerb“). Konträr zur Hypothese 3 hatte das anwendungsorientierte Vorwissen global keinen moderierenden Einfluss auf den Erwerb anwendungsorientierten Wissens, $b = -0.02$, 95% KI $[-0.53, 0.49]$, $p = .93$. Die Anwendung der JN Technik ergab, dass es Werte des anwendungsorientierten Vorwissens im unteren Bereich der theoretisch möglichen Werte gab, in dem der Grad der instruktionalen Unterstützung einen signifikanten Effekt auf den Erwerb von anwendungsorientiertem Wissen hatte. In diesem Bereich war der Effekt des hohen Grads an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit dem niedrigen Grad an instruktionaler Unterstützung (Problemlösen) auf den Erwerb von anwendungsorientiertem Wissen umso geringer, je höher das anwendungsorientierte Vorwissen war. Im übrigen Bereich war der Effekt des hohen Grads an instruktionaler Unterstützung im Vergleich mit dem niedrigen Grad an instruktionaler Unterstützung auf den Wissenserwerb ebenfalls umso geringer, je höher das anwendungsorientierte Vorwissen war. Allerdings war der Effekt in diesem Bereich nicht signifikant (siehe Tabelle 5). Eine Häufigkeitsanalyse für die Werte des anwendungsorientierten Vorwissens ergab, dass bei einer Spanne von 0 bis 6 Punkten im Gesamtwert der Fragen zur Erfassung des anwendungsorientierten Vorwissens 52.63% der Teilnehmer ein anwendungsorientiertes Vorwissen von 1 Punkt aufwiesen. 90.79% der Teilnehmer hatten ein anwendungsorientiertes Vorwissen von *höchstens* 3. Fünf Teilnehmer hatten einen Gesamtwert von 4 Punkten, jeweils ein Teilnehmer hatte 5 bzw. 6 Punkte im Gesamtwert der Fragen zur Erfassung des anwendungsorientierten Vorwissens.

Tabelle 5

Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte des anwendungsorientierten Vorwissens

Anwendungs-orientiertes Vorwissen	Effekt des Grads an instruktionaler Unterstützung	SE	t	p	95% KI	
					UG	OG
0.0000	0.9134	0.4721	1.9348	.0569	-0.0277	1.8544
0.0713	0.9121	0.4575	1.9935	.0500	0.0000	1.8241
0.3000	0.9080	0.4128	2.1994	.0311	0.0850	1.7309
0.6000	0.9026	0.3602	2.5060	.0145	0.1846	1.6206
0.9000	0.8972	0.3175	2.8262	.0061	0.2643	1.5300
1.2000	0.8918	0.2891	3.0847	.0029	0.3155	1.4681
1.5000	0.8864	0.2795	3.1713	.0022	0.3292	1.4436
1.8000	0.8810	0.2905	3.0322	.0034	0.3018	1.4602
2.1000	0.8756	0.3201	2.7355	.0078	0.2375	1.5137
2.4000	0.8702	0.3637	2.3929	.0193	0.1453	1.5952
2.7000	0.8648	0.4169	2.0745	.0416	0.0338	1.6958
2.7842	0.8633	0.4331	1.9935	.0500	0.0000	1.7266
3.0000	0.8594	0.4765	1.8036	.0755	-0.0905	1.8093
3.3000	0.8540	0.5405	1.5802	.1184	-0.2234	1.9314
3.6000	0.8486	0.6074	1.3973	.1666	-0.3621	2.0594
3.9000	0.8432	0.6763	1.2468	.2165	-0.5050	2.1915
4.2000	0.8379	0.7468	1.1220	.2656	-0.6508	2.3265
4.5000	0.8325	0.8183	1.0173	.3124	-0.7989	2.4638
4.8000	0.8271	0.8908	0.9285	.3563	-0.9486	2.6028
5.1000	0.8217	0.9638	0.8525	.3968	-1.0997	2.7430
5.4000	0.8163	1.0374	0.7868	.4340	-1.2518	2.8843
5.7000	0.8109	1.1114	0.7296	.4680	-1.4047	3.0265
6.0000	0.8055	1.1857	0.6793	.4991	-1.5582	3.1692

Anmerkungen. Theoretische Werte des anwendungsorientierten Vorwissens liegen zwischen 0 und 6. Höhere Werte bedeuten ein höheres anwendungsorientiertes Vorwissen. Werte zwischen 0 und 1 stellen theoretische Werte des anwendungsorientierten Vorwissens dar, für die auf Basis der Stichprobendaten jeweils der Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb berechnet wurde. In der Stichprobe wurden nur ganzzahlige Werte im anwendungsorientierten Vorwissen zwischen 0 und 6 realisiert. SE = Standardfehler; UG = Untere Grenze des Konfidenzintervalls; OG = Obere Grenze des Konfidenzintervalls.

5.3.2 Moderatorenrolle kognitiver Basisfunktionen und fluider Intelligenz (Frage 1)

Bei einem Teilnehmer fehlte der Wert für den Subtest „Verbal-deduktives Denken“ der INSBAT. In den Analysen wurde der Rohwert für Gf aus den Subtests „Numerisch-induktives Denken“ und „Figural-induktives Denken“ verwendet. Bei einem anderen Teilnehmer fehlte der Wert für die category-switch Aufgabe. Für die Berechnung der Fähigkeit zum Shifting

wurde der Mittelwert aus der color-shape Aufgabe und der number-letter Aufgabe verwendet. Der Rohwert in Gf von nur zwei Subtests der INSBAT und der Mittelwert von color-shape und number-letter Aufgabe wurden bei je einem Teilnehmer verwendet, um eine Verringerung der Stichprobengröße bei den statistischen Analysen mit Gf und Shifting zu verhindern.

Analog zur Untersuchung der moderierenden Rolle des Vorwissens wurde der moderierende Einfluss von kBf und Gf nur für den Erwerb von anwendungsorientiertem Wissen untersucht, da es keinen Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von konzeptuellem Wissen gab. Neben den Kovariaten Alter und Semesterzahl wurde Gf als eine Kovariate in der Moderatoranalyse für die AG-Kapazität und die AG-Kapazität in der Moderatoranalyse für Gf als Kovariate berücksichtigt, da die AG-Kapazität und Gf miteinander korrelierten (siehe Abschnitt 5.3.1).

Die AG-Kapazität hatte global keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung (niedrig vs. hoch) auf den Erwerb von anwendungsorientiertem Wissen unter Kontrolle von Gf, Alter und Semesterzahl, $b = -0.47$, 95% KI [-4.07, 3.14], $p = .80$. Die Anwendung der JN Technik ergab, dass es wenige Werte für die AG-Kapazität im mittleren Bereich der theoretisch möglichen Werte gab, für die der Grad an instruktionaler Unterstützung einen signifikanten Effekt auf den Erwerb von anwendungsorientiertem Wissen hatte. In diesem Bereich war der Effekt der hohen instruktionalen Unterstützung verglichen mit der niedrigen instruktionalen Unterstützung auf den Erwerb von anwendungsorientiertem Wissen umso größer, je geringer die AG-Kapazität war (siehe Tabelle 6).

Kapitel 5: Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung

Tabelle 6

Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte der AG-Kapazität

AG-Kapazität	Effekt des Grads an instruktionaler Unterstützung	SE	t	p	95% KI	
					UG	OG
.2633	0.9858	0.8216	1.1998	.2343	-0.6533	2.6249
.2972	0.9700	0.7668	1.2650	.2101	-0.5598	2.4997
.3310	0.9542	0.7130	1.3383	.1852	-0.4682	2.3767
.3648	0.9384	0.6605	1.4208	.1599	-0.3792	2.2561
.3987	0.9227	0.6096	1.5136	.1347	-0.2934	2.1387
.4325	0.9069	0.5607	1.6175	.1103	-0.2117	2.0254
.4663	0.8911	0.5144	1.7323	.0877	-0.1351	1.9174
.5002	0.8754	0.4715	1.8564	.0677	-0.0653	1.8160
.5340	0.8596	0.4330	1.9851	.0511	-0.0043	1.7234
.5366	0.8584	0.4303	1.9950	.0500	0.0000	1.7167
.5678	0.8438	0.4002	2.1085	.0386	0.0455	1.6422
.6017	0.8280	0.3745	2.2111	.0303	0.0809	1.5751
.6355	0.8123	0.3575	2.2720	.0262	0.0990	1.5255
.6693	0.7965	0.3505	2.2725	.0262	0.0973	1.4957
.7032	0.7807	0.3540	2.2053	.0308	0.0745	1.4870
.7370	0.7649	0.3678	2.0798	.0413	0.0312	1.4987
.7553	0.7564	0.3792	1.9950	.0500	0.0000	1.5128
.7708	0.7492	0.3908	1.9172	.0594	-0.0304	1.5287
.8047	0.7334	0.4214	1.7404	.0862	-0.1073	1.5741
.8385	0.7176	0.4582	1.5663	.1218	-0.1964	1.6316
.8723	0.7019	0.4997	1.4045	.1647	-0.2951	1.6988
.9062	0.6861	0.5450	1.2590	.2123	-0.4011	1.7733
.9400	0.6703	0.5931	1.1303	.2623	-0.5128	1.8534

Anmerkungen. Theoretische Werte der AG-Kapazität liegen zwischen 0 und 1. Höhere Werte bedeuten eine höhere AG-Kapazität. Werte in der AG-Kapazität unter .26 und über .94 wurden in der Stichprobe nicht realisiert. Werte zwischen .26 und .94 stellen theoretische Werte der AG-Kapazität dar, für die auf Basis der Stichprobendaten jeweils der Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb berechnet wurde. SE = Standardfehler; UG = Untere Grenze des Konfidenzintervalls; OG = Obere Grenze des Konfidenzintervalls.

Shifting moderierte den Effekt des Grads an instruktionaler Unterstützung (niedrig vs. hoch) auf den Erwerb von anwendungsorientiertem Wissen unter Kontrolle von Alter und Semesterzahl, $b = 0.004$, 95% KI [0.001, 0.007], $p < .01$. Die Anwendung der JN Technik ergab, dass es einen breiten Bereich der Werte für Shifting gab, in dem der Grad an instruktionaler Unterstützung einen signifikanten Einfluss auf den Erwerb von anwendungsorientiertem Wissen hatte. Der Effekt der hohen instruktionalen Unterstützung verglichen mit der niedrigen instruktionalen Unterstützung auf den Erwerb von anwendungsorientiertem

Wissen war umso größer, je geringer die Shifting-Fähigkeit ausgeprägt war, d.h., je höher die Reaktionszeiten waren (siehe Tabelle 7).

Tabelle 7

Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte von Shifting

Shifting (ms)	Effekt des Grads an instruktionaler Unterstützung	SE	t	p	95% KI	
					UG	OG
53.1839	-0.0832	0.4491	-0.1853	.8535	-0.9789	0.8124
89.7237	0.0710	0.4163	0.1706	.8650	-0.7593	0.9013
126.2635	0.2252	0.3876	0.5811	.5631	-0.5479	0.9984
162.8033	0.3795	0.3640	1.0425	.3008	-0.3465	1.1054
199.3431	0.5337	0.3464	1.5406	.1279	-0.1572	1.2246
231.9738	0.6714	0.3367	1.9944	.0500	0.0000	1.3429
235.8829	0.6879	0.3359	2.0481	.0443	0.0180	1.3578
272.4226	0.8422	0.3330	2.5287	.0137	0.1779	1.5064
308.9624	0.9964	0.3381	2.9472	.0044	0.3221	1.6707
345.5022	1.1506	0.3507	3.2813	.0016	0.4512	1.8500
382.0420	1.3048	0.3700	3.5263	.0007	0.5668	2.0429
418.5818	1.4591	0.3952	3.6921	.0004	0.6709	2.2472
455.1216	1.6133	0.4251	3.7952	.0003	0.7655	2.4611
491.6614	1.7675	0.4588	3.8522	.0003	0.8524	2.6826
528.2012	1.9218	0.4956	3.8775	.0002	0.9333	2.9102
564.7410	2.0760	0.5348	3.8816	.0002	1.0093	3.1427
601.2808	2.2302	0.5760	3.8722	.0002	1.0815	3.3789
637.8205	2.3844	0.6186	3.8544	.0003	1.1506	3.6183
674.3603	2.5387	0.6625	3.8317	.0003	1.2173	3.8601
710.9001	2.6929	0.7075	3.8064	.0003	1.2819	4.1039
747.4399	2.8471	0.7532	3.7799	.0003	1.3449	4.3494
783.9797	3.0014	0.7997	3.7533	.0004	1.4065	4.5963

Anmerkungen. Bei den Shifting-Werten handelt es sich um Reaktionszeiten (in ms). Höhere Reaktionszeiten bedeuten eine geringere Shifting-Fähigkeit. Shifting-Werte unter 53.18 ms und über 783.98 ms wurden in der Stichprobe nicht realisiert. Werte zwischen 53.18 ms und 783.98 ms stellen theoretische Werte der Shifting-Fähigkeit dar, für die auf Basis der Stichprobendaten jeweils der Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb berechnet wurde. SE = Standardfehler; UG = Untere Grenze des Konfidenzintervalls; OG = Obere Grenze des Konfidenzintervalls.

Fluide Intelligenz moderierte den Effekt des Grads an instruktionaler Unterstützung (niedrig vs. hoch) auf den Erwerb von anwendungsorientiertem Wissen unter Kontrolle von AG-Kapazität, Alter und Semesterzahl, $b = -0.83$, 95% KI [-1.57, -0.08], $p < .05$. Die Anwendung der JN Technik ergab, dass es einen breiten Bereich der Stichprobenwerte für Gf gab, in

dem der Grad an instruktionaler Unterstützung einen signifikanten Einfluss auf den Erwerb von anwendungsorientiertem Wissen hatte. Der Effekt der hohen instruktionalen Unterstützung verglichen mit der niedrigen instruktionalen Unterstützung auf den Erwerb von anwendungsorientiertem Wissen war umso größer, je geringer die Werte von Gf waren (siehe Tabelle 8).

Tabelle 8

Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte von Gf

Gf (Rohwert)	Effekt des Grads an instruktionaler Unterstützung	SE	t	p	95% KI	
					UG	OG
-1.5700	2.5127	0.8040	3.1253	.0026	0.9088	4.1165
-1.3350	2.3181	0.7253	3.1958	.0021	0.8711	3.7651
-1.1000	2.1235	0.6490	3.2719	.0017	0.8288	3.4183
-0.8650	1.9289	0.5759	3.3494	.0013	0.7800	3.0778
-0.6300	1.7343	0.5074	3.4182	.0011	0.7221	2.7466
-0.3950	1.5398	0.4456	3.4555	.0009	0.6508	2.4287
-0.1600	1.3452	0.3937	3.4165	.0011	0.5597	2.1307
0.0750	1.1506	0.3561	3.2309	.0019	0.4402	1.8611
0.3100	0.9560	0.3376	2.8318	.0061	0.2825	1.6296
0.5450	0.7615	0.3413	2.2312	.0289	0.0806	1.4423
0.6268	0.6938	0.3478	1.9950	.0500	0.0000	1.3875
0.7800	0.5669	0.3665	1.5469	.1265	-0.1642	1.2980
1.0150	0.3723	0.4092	0.9098	.3661	-0.4441	1.1887
1.2500	0.1777	0.4647	0.3824	.7033	-0.7494	1.1049
1.4850	-0.0168	0.5290	-0.0318	.9747	-1.0722	1.0385
1.7200	-0.2114	0.5992	-0.3528	.7253	-1.4068	0.9840
1.9550	-0.4060	0.6735	-0.6028	.5486	-1.7496	0.9376
2.1900	-0.6006	0.7506	-0.8001	.4264	-2.0981	0.8969
2.4250	-0.7951	0.8299	-0.9581	.3413	-2.4507	0.8604
2.6600	-0.9897	0.9106	-1.0868	.2809	-2.8064	0.8270
2.8950	-1.1843	0.9926	-1.1932	.2369	-3.1644	0.7958
3.1300	-1.3789	1.0754	-1.2822	.2041	-3.5242	0.7665

Anmerkungen. Höhere Rohwerte bedeuten eine höhere Gf. Werte der Gf unter -1.57 und über 3.13 wurden in der Stichprobe nicht realisiert. Die Werte zwischen -1.57 und 3.13 stellen bis auf den Wert -0.16 (in der Stichprobe realisiert) theoretische Werte der Gf dar, für die auf Basis der Stichprobendaten jeweils der Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb berechnet wurde. SE = Standardfehler; UG = Untere Grenze des Konfidenzintervalls; OG = Obere Grenze des Konfidenzintervalls.

5.4 Diskussion

5.4.1 Zusammenfassung der Ergebnisse

Die vorausgehenden Analysen der Studie 1 zeigten, dass sich die Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung nicht signifikant im CL unterscheiden. Der CL korrelierte nicht signifikant mit der AG-Kapazität, aber mit Gf, konzeptuellem Vorwissen, anwendungsorientiertem Vorwissen sowie den Skalen Misserfolgsbefürchtung, Erfolgswahrscheinlichkeit und Herausforderung.

Teilnehmer erwarben in beiden Lernumgebungen konzeptuelles Wissen. Allerdings hing der Erwerb konzeptuellen Wissens nicht vom Grad der instruktionalen Unterstützung ab. Teilnehmer erwarben in der Lernumgebung mit einem hohen Grad an instruktionaler Unterstützung (Lösungsbeispiele) mehr anwendungsorientiertes Wissen als Teilnehmer in der Lernumgebung mit einem niedrigen Grad an instruktionaler Unterstützung (Problemlösen).

Das anwendungsorientierte Vorwissen hatte global keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens.

Das Hauptergebnis der Studie 1 dieser Arbeit ist, dass Shifting und Gf, jedoch nicht die AG-Kapazität den Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen moderierten. Je geringer die Fähigkeit zum Shifting und Gf der Teilnehmer waren, desto größer war der Effekt der hohen instruktionalen Unterstützung verglichen mit der niedrigen instruktionalen Unterstützung auf den Erwerb von anwendungsorientiertem Wissen.

5.4.2 Ergebnisinterpretation

Unterschiede im cognitive load (Hypothese 1)

Im Gegensatz zur Hypothese 1 unterschieden sich die Lernumgebungen nicht signifikant im CL. Dieses Ergebnis ist konträr zu Studien, die einen geringeren CL in einer Bedingung mit Lösungsbeispielen verglichen mit einer Problemlösebedingung fanden (z.B. Renkl et al., 2003; Schwonke et al., 2009). Möglicherweise war der intrinsic CL in beiden Bedingungen aufgrund der Komplexität des Lernmaterials sehr hoch. In beiden Bedingungen sollten Teilnehmer komplexe statistische Probleme und umfangreiche Informationen lesen, um die Probleme zu lösen (Problemlösebedingung) oder Gründe für die Lösungsschritte (Bedingung mit Lösungsbeispielen) zu geben. Wahrscheinlich war es für Teilnehmer in der Bedingung mit Lösungsbeispielen leichter, nach Informationen zur Begründung der Lösungen zu suchen, da die Lösungen präsentiert wurden und Teilnehmer gezielter nach Informationen suchen konnten. Dennoch könnte der intrinsic CL in dieser Bedingung aufgrund der Komplexität der statistische Probleme mit den zugehörigen Informationen bzw. einer hohen Elementinteraktivität so hoch gewesen sein, dass ein Unterschied im gesamten CL zwischen den Lernumgebungen nicht auftrat. Allerdings merkt de Jong (2010) an, dass solche post-hoc Erklärungen mit der CLT problematisch sind, da die Theorie kaum falsifiziert werden kann. Zudem

differenziert die in der Studie 1 verwendete Ratingskala nicht zwischen verschiedenen Arten des CL. Daher ist unklar, welche Art von CL hauptsächlich für den, von den Teilnehmern hoch eingeschätzten, gesamten CL in beiden Lernumgebungen verantwortlich war. Nach einer Diskussion der Ergebnisse zum Wissenserwerb in den Lernumgebungen wird bei der Diskussion der Moderatorenrolle der AG-Kapazität u.a. auf Gründe für fehlende Unterschiede in der kognitiven Belastung der AG-Kapazität *ohne* einen Bezug zur Ratingskala eingegangen. Eine Interpretation der kognitiven Belastung, die über den Bezug zu verschiedenen Arten des CL hinausgeht, erfolgt in der Gesamtdiskussion (siehe Abschnitt 7.1.2).

Unterschiede im Wissenserwerb (Hypothese 2)

Wie erwartet, zeigten sich Effekte des Grads an instruktionaler Unterstützung auf den Wissenserwerb. Im Gegensatz zu Hypothese 2.a unterschieden sich die Lernumgebungen nicht hinsichtlich des Erwerbs von konzeptuellem Wissen. Übereinstimmend mit der Hypothese 2.b erwarben Teilnehmer in der Bedingung mit Lösungsbeispielen mehr anwendungsorientiertes Wissen als Lerner in der Problemlösebedingung. Ein Grund für den fehlenden Unterschied im Erwerb konzeptuellen Wissens zwischen den Lernumgebungen könnte das relativ hohe konzeptuelle Vorwissen der Teilnehmer (siehe Tabelle 3) sein, da Lösungsbeispiele in erster Linie für Lerner mit niedrigem Vorwissen effektiv sind (Kalyuga, 2007). In beiden Lernumgebungen wurden zusätzliche Informationen zur Aufgabenbearbeitung gegeben, die eine inhaltliche Überlappung zu den Fragen zur Erfassung konzeptuellen Wissens aufwiesen. Diese inhaltliche Überlappung könnte ein Grund dafür sein, dass Teilnehmer in beiden Lernumgebungen konzeptuelles Wissen erwarben. Im Gegensatz zum konzeptuellen Vorwissen war das anwendungsorientierte Vorwissen eher gering (siehe Tabelle 4). Des Weiteren wurde post-hoc (d.h. im Anschluss an die Studie) festgestellt, dass die inhaltliche Überlappung der Fragen zur Erfassung des konzeptuellen Wissens und den Lösungsbeispielen geringer war als die inhaltliche Überlappung zwischen den Fragen zur Erfassung des anwendungsorientierten Wissens und den Lösungsbeispielen. Mit anderen Worten waren die durch die Lösungsbeispiele gelernten Inhalte geeigneter, um anwendungsorientiertes Wissen zu erwerben. Folglich förderten die Lösungsbeispiele den Erwerb von anwendungsorientiertem Wissen mehr als den Erwerb von konzeptuellem Wissen.

Moderierende Rolle des Vorwissens (Hypothese 3)

Aufgrund der Befunde zum *expertise reversal effect* (Kalyuga, 2007) war zu erwarten, dass der Effekt der hohen instruktionalen Unterstützung (Lösungsbeispiele) verglichen mit der niedrigen instruktionalen Unterstützung (Problemlösen) auf den Wissenserwerb umso geringer ist, je höher das Vorwissen ist. Im Gegensatz zur Hypothese 3 hatte das anwendungsorientierte Vorwissen global keinen moderierenden Einfluss auf den Erwerb anwendungsorientierten Wissens. Allerdings gab es einen theoretischen Wertebereich, in dem der Effekt der hohen instruktionalen Unterstützung verglichen mit der niedrigen instruktionalen Unterstützung hypothesenkonform auf den Wissenserwerb umso geringer war, je höher das anwendungsorientierte Vorwissen war (siehe Tabelle 5). Eine mögliche Erklärung für diese Befunde ist, dass die Varianz im anwendungsorientierten Vorwissen stark eingeschränkt war. In der Stichprobe gab es nur ganzzahlige Werte von 0 bis 6 (Gesamtwerte der offenen Fragen) für das anwendungsorientierte Vorwissen. Die meisten Teilnehmer hatten ein ähnlich niedriges anwendungsorientiertes Vorwissen, d.h., im Bereich niedrigen anwendungsorientierten Vorwissens gab es eine geringe Varianz. Ein sehr geringer Anteil an Teilnehmern hatte ein mittleres bis hohes Vorwissen. Aufgrund der geringen Varianz, insbesondere im Bereich niedrigen anwendungsorientierten Vorwissens, konnte wahrscheinlich eine moderierende Rolle über den gesamten Wertebereich des anwendungsorientierten Vorwissens nicht festgestellt werden. Diese Befundlage ist mit dem Vorteil der hohen instruktionalen Unterstützung verglichen mit der niedrigen instruktionalen Unterstützung im Hinblick auf den Erwerb anwendungsorientierten Wissens (siehe vorheriger Abschnitt) vereinbar. Der Grund hierfür ist, dass das anwendungsorientierte Wissen in der Stichprobe im Mittel gering war und in der Kovarianzanalyse Unterschiede zwischen den Mittelwerten der beiden Lernumgebungen im anwendungsorientierten Wissen festgestellt wurden. Lösungsbeispiele scheinen den Ergebnissen der Studie 1 dieser Arbeit zu Folge somit für vorwissensschwache Lerner hypothesenkonform effektiv zu sein. Allerdings konnte vermutlich aufgrund der starken Varianzeinschränkung im niedrigen Bereich des anwendungsorientierten Vorwissens global kein *expertise-reversal effect* (Kalyuga, 2007) mit steigendem Vorwissen festgestellt werden.

Moderierende Rolle kognitiver Basisfunktionen und fluider Intelligenz (Hypothese 4)

Es wurde davon ausgegangen, dass die AG-Kapazität, die Fähigkeit zum Shifting und Gf einen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb haben.

In Widerspruch zu Hypothese 4.a hatte die AG-Kapazität global keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen. In einem kleinen Wertebereich war die instruktionale Unterstützung jedoch hypothesenkonform umso wirksamer, je geringer die AG-Kapazität war (siehe Tabelle 6). Für die Hypothesenbildung wurden die Annahmen von Van Gog und Rummel (2010) erweitert und davon ausgegangen, dass ein hoher Grad an instruktionaler Unterstützung (Lösungsbeispiele) für Lerner mit einer hohen AG-Kapazität weniger wichtig ist, da der CL deren AG-Kapazität nicht überschreitet. Der CL unterschied sich allerdings nicht signifikant zwischen den Lernumgebungen. Demnach könnte ein hoher Grad an instruktionaler Unterstützung nicht zu einer signifikanten Reduzierung des CL im AG geführt haben. Aufgrund der fehlenden, signifikanten Reduzierung des CL hatten Lerner mit geringer AG-Kapazität möglicherweise keinen Vorteil durch den hohen Grad an instruktionaler Unterstützung. Daher könnte die AG-Kapazität keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb gehabt haben. Diese Annahme ist jedoch nicht plausibel, da die AG-Kapazität und der CL nicht miteinander korrelierten, obwohl eine höhere AG-Kapazität mit einer relativ geringeren kognitiven Belastung im AG einhergehen sollte. Der fehlende Zusammenhang zwischen AG-Kapazität und CL lässt an der Validität der verwendeten Ratingskala zur Erfassung des CL im AG zweifeln. Tatsächlich wurde die Validität der Ratingskala nie in Frage gestellt (de Jong, 2010). Des Weiteren kann bezweifelt werden, ob die Ratingskala für den Grad an instruktionaler Unterstützung beachtet werden muss, da ein hoher Grad an instruktionaler Unterstützung (Lösungsbeispiele) den Erwerb von anwendungsorientiertem Wissen trotz fehlender Unterschiede im CL zwischen den Lernumgebungen förderte. Diese Befundlage ist in Übereinstimmung mit den Ergebnissen mehrerer Studien, die zeigten, dass ein hoher Grad an instruktionaler Unterstützung (Lösungsbeispiele) effektiv sein kann, auch wenn sich Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung nicht im CL unterscheiden (z.B. de Koning, Tabbers, Rikers, & Paas, 2009; Lusk & Atkinson, 2007; Wouters, Paas, & Van Merriënboër, 2009).

Ohne eine Bezugnahme auf den, mit der Ratingskala gemessenen CL, könnten die Lösungsbeispiele die Anforderungen an die AG-Kapazität nicht signifikant reduziert haben. Obwohl die Teilnehmer in der Lernumgebung mit einem hohen Grad an instruktionaler Unterstützung Lösungsschritte präsentiert bekamen, mussten sie diese Schritte im AG behalten, wenn sie nach den richtigen Informationen für die Begründung der Lösungsschritte suchten. In Anlehnung an Ayres und Sweller (2005) könnte die Notwendigkeit der Integration getrennter Informationen (Lösungsschritte und Informationen zu deren Begründung waren auf unterschiedlichen Folien) eine hohe Anforderung an die AG-Kapazität gestellt haben. Die Anforderungen an die AG-Kapazität könnten daher in der Bedingung mit Lösungsbeispielen ähnlich hoch wie in der Problemlösebedingung gewesen sein. In der Problemlösebedingung mussten die Teilnehmer Informationen in den statistischen Problemen und die zugehörigen Informationen verarbeiten und speichern. Die hohen Anforderungen an die AG-Kapazität aufgrund der Notwendigkeit zur mentalen Integration getrennter Informationen in beiden Lernumgebungen wirkte sich jedoch anscheinend nicht oder nur teilweise hinderlich auf das Lernen in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung aus, da Teilnehmer in dieser Bedingung mehr anwendungsorientiertes Wissen erwarben als in der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung. Schließlich bleibt unklar, ob die AG-Kapazität einen moderierenden Einfluss auf den Wissenserwerb hat, wenn sich Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung stärker in den kognitiven Anforderungen an das AG unterscheiden. Da zumindest in einem kleinen Wertebereich größere Effekte des Grads an instruktionaler Unterstützung auf den Wissenserwerb bei geringerer AG-Kapazität festzustellen waren, gibt es Hinweise auf eine moderierende Rolle der AG-Kapazität—speziell für Lerner mit einer sehr geringen AG-Kapazität.

Neben dem AG könnten andere kognitive Fähigkeiten und Fertigkeiten für die Moderation des Effekts des Grads an instruktionaler Unterstützung auf den Wissenserwerb wichtig sein. In dieser Studie waren die Fähigkeit zum Shifting und Gf unabhängig von der AG-Kapazität Moderatoren des Erwerbs von anwendungsorientiertem Wissen. Mit anderen Worten war die moderierende Rolle von Shifting und Gf einzigartig.

Der moderierende Einfluss von Shifting könnte durch unterschiedliche Anforderungen an die Fähigkeit zum Shifting in den Lernumgebungen erklärt werden. In der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung erforderte das Lösen der komplexen, statistischen Probleme höchstwahrscheinlich den Wechsel zwischen gewissen Aspekten des jeweiligen Problems, insbesondere den Wechsel zwischen dem Problem und Informationen

zur Problemlösung sowie den Wechsel zwischen diesen Informationen. Daher sollte die Fähigkeit zum Shifting insbesondere in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung wichtig für den Wissenserwerb sein. In Übereinstimmung mit dieser Annahme wurde gefolgert, dass die Fähigkeit zum Shifting insbesondere für komplexe Aufgaben relevant sein sollte, die den Wechsel zwischen gewissen Aspekten eines Problems erfordern (Blair et al., 2008; Van der Sluis et al., 2007). Lerner mit einer hohen Shifting-Fähigkeit könnten weniger Probleme in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung haben, da sie mit den Shifting-Anforderungen zurechtkommen. Lerner mit geringer Shifting-Fähigkeit sollten dagegen Schwierigkeiten in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung haben und von einem hohen Grad an instruktionaler Unterstützung profitieren. Tatsächlich war konform mit der Hypothese 4.b der Effekt des hohen Grads an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit dem niedrigen Grad an instruktionaler Unterstützung (Problemlösen) auf den Erwerb von anwendungsorientiertem Wissen in der Studie 1 dieser Arbeit für Teilnehmer mit einer geringen Shifting-Fähigkeit am größten (siehe Tabelle 7). Dieses Ergebnis scheint plausibel zu sein, da Teilnehmer in der Bedingung mit hohem Grad an instruktionaler Unterstützung die Lösungsschritte sahen und nicht zwischen den statistischen Problemen und den zugehörigen Informationen wechseln mussten. Folglich könnte die Notwendigkeit zum Wechsel zwischen dem Problem und den Informationen zur Problemlösung bedeutsam reduziert worden sein. Unterschiede in der Fähigkeit zum Shifting könnten deshalb in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung eine untergeordnete Rolle gespielt haben. Der Vorteil der Lösungsbeispiele verglichen mit der Problemlösebedingung scheint umso geringer zu sein, je höher die Shifting-Fähigkeit ist.

Ähnlich zur Interpretation der moderierenden Rolle von Shifting könnte der moderierende Einfluss von Gf durch Unterschiede in den Anforderungen an Gf zwischen den Lernumgebungen erklärt werden. In der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) mussten Teilnehmer höchstwahrscheinlich schlussfolgern, welche Informationen für die Lösung des jeweiligen statistischen Problems notwendig sind. Die Aufgabenkomplexität war in der Problemlösebedingung bedeutsam höher als in der Bedingung mit Lösungsbeispielen. Fluide Intelligenz ist insbesondere bei neuen *und* komplexen kognitiven Aufgaben wichtig (Primi et al., 2010). Lerner mit hoher Gf könnten die komplexen Aufgaben auch dann lösen, wenn die instruktionale Unterstützung gering ist, da sie die notwendige Fähigkeit zum schlussfolgernden Denken besitzen. Im Gegensatz dazu könnten Ler-

ner mit geringer Gf in einer Lernumgebung mit einem niedrigen Grad an instruktionaler Unterstützung Schwierigkeiten haben und besonders von der instruktionalen Unterstützung durch Lösungsbeispiele profitieren. In Übereinstimmung mit dieser Annahme und konform mit der Hypothese 4.c war der Vorteil der Lösungsbeispiele im Vergleich mit der Problemlösebedingung umso größer, je geringer Gf war (siehe Tabelle 8). Teilnehmer mit geringer Gf könnten in der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung Schwierigkeiten gehabt haben, weil es ihnen womöglich schwerfiel zu schlussfolgern, welche Informationen für die Lösung eines Problems notwendig sind. Die Lösungsschritte der Lösungsbeispiele könnten solchen Teilnehmern in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung geholfen haben, da es nicht erforderlich war zu schlussfolgern, welche Informationen für die Lösung eines statistischen Problems notwendig sind. Der Vorteil der Lernumgebung mit hohem Grad an instruktionaler Unterstützung scheint jedoch umso geringer zu sein, je höher Gf ist.

5.4.3 Limitationen

Eine Limitation betrifft Unterschiede im Vorwissen zwischen den Wissensarten der Teilnehmer. Die Teilnehmer der Studie 1 dieser Arbeit hatten ein relativ hohes konzeptuelles Vorwissen und ein relativ geringes anwendungsorientiertes Vorwissen. Diese Unterschiede im Vorwissen zwischen den Wissensarten wurden bei der Ergebnisinterpretation (siehe Abschnitt 5.4.2) berücksichtigt und schränken die Generalisierbarkeit der Ergebnisse ein. Eine Generalisierung auf Studierende mit geringem konzeptuellem Vorwissen und/oder hohem anwendungsorientiertem Vorwissen ist kaum möglich, zumal Lösungsbeispiele vor allem für Lerner mit niedrigem Vorwissen effektiv sind (Kalyuga, 2007).

Hinsichtlich der Wissenstests muss einschränkend eingeräumt werden, dass diese spezifisch für die Studie entwickelt wurden und konzeptuelles Wissen sowie anwendungsorientiertes Wissen erfassen. Offen ist, ob eine moderierende Rolle von kBf und Gf mit Wissenstests nachgewiesen werden kann, die andere Wissensarten erfassen (z.B. prozedurales Wissen). Die Reliabilitäten (internen Konsistenzen) der Wissenstests (getrennt nach Wissensarten) waren niedrig. Obwohl zwischen verschiedenen Wissensarten differenziert wurde, unterschieden sich die Inhalte der Fragen zur Erfassung der jeweiligen Wissensart inhaltlich deutlich voneinander. Grund für die inhaltlichen Unterschiede zwischen den Fragen und der wahrscheinlich damit verbundenen niedrigen Reliabilität der Wissenstests war, dass ein breites

Spektrum an Wissen über statistische Inhalte erfasst werden sollte. Die Fragen der Wissenstests wurden auf Basis vorhandener Fragen (Übungs- und Klausurfragen) am Lehrstuhl für psychologische Methodenlehre und Diagnostik der Ludwig-Maximilians-Universität München konstruiert, um eine hohe Inhaltsvalidität der Wissenstests zu erzielen. Ein geringes bzw. hohes Wissen in einem bestimmten Inhaltsbereich geht nicht notwendigerweise mit einem geringen bzw. hohem Wissen in einem anderen Inhaltsbereich einher, was die niedrigen Reliabilitäten der Wissenstests erklären könnte. Konform mit diesen Ausführungen merken Wecker et al. (2013) an, dass es für Wissenstests, die einen breiten Inhaltsbereich abdecken nicht gänzlich untypisch ist, dass die, durch einzelne Items erfassten Wissens Elemente nicht durch enge Beziehungen verbunden sind, die es Lernern erlauben, sich diese Wissens Elemente voneinander zu erschließen. Trotz der niedrigen Reliabilitäten der Wissenstests gab es einen Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens. Dieser Effekt spricht für die Trennung zwischen verschiedenen Wissensarten, da es keinen Effekt der Lösungsbeispiele (hoher Grad an instruktionaler Unterstützung) auf den Erwerb konzeptuellen Wissens gab. Zudem konnte nachgewiesen werden, dass es in beiden Lernumgebungen einen signifikanten Erwerb konzeptuellen Wissens von Pre- zu Posttest gab. Die Reliabilitäten der Wissenstests waren demnach ausreichend, um trotz eines hohen Fehleranteils in der Varianz einen Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens und einen Effekt des Messzeitpunkts (Pre- vs. Posttest) auf den Erwerb konzeptuellen Wissens festzustellen. Durch Verwendung einer größeren Anzahl inhaltsähnlicher Fragen zur Erfassung des konzeptuellen Wissens könnte eine reliablere Erfassung dieses Wissens erzielt werden. Der Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens hätte bei Verwendung inhaltlich ähnlicherer Fragen zur Erfassung des anwendungsorientierten Wissens, die zu einer höheren Reliabilität führen sollten, wahrscheinlich genauer geschätzt werden können.

Bezüglich der Wissenstests muss weiterhin angemerkt werden, dass nur 2 offene Fragen mit je drei Unterfragen zur Erfassung des anwendungsorientierten Wissens verwendet wurden. In der Studie von Leppink et al. (2014) zur Wirksamkeit von Lösungsbeispielen in der Domäne Statistik wurden ebenfalls lediglich 6 offene Fragen genutzt. Mehrere offene Fragen, die zudem inhaltsähnlicher sind, hätten aber wahrscheinlich zu reliableren Schätzungen des anwendungsorientierten Vorwissens der Teilnehmer geführt und in der Folge die Ergebnisse der vorliegenden Studie stärker untermauert. Zudem war die inhaltliche Überlappung der MC-Fragen zur Messung von konzeptuellem Wissen und den Lösungsbeispielen geringer

als die inhaltliche Überlappung der offenen Fragen mit den Lösungsbeispielen. Vor dem Hintergrund des hohen konzeptuellen Vorwissens der Teilnehmer bleibt unklar, ob ein Vorteil der Lernumgebung mit hohem Grad an instruktionalen Unterstützung (Lösungsbeispiele) gegenüber der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) im Hinblick auf den Erwerb konzeptuellen Wissens nachgewiesen werden kann, wenn das konzeptuelle Vorwissen niedrig ist und die inhaltliche Überlappung der MC-Fragen und den Lösungsbeispielen hoch ist.

In der Studie 1 dieser Arbeit war die Analyse der moderierenden Rolle des konzeptuellen Vorwissens nicht möglich, da sich die Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung nicht im Erwerb konzeptuellen Wissens unterschieden. Zudem wies nur 1 Teilnehmer ein niedriges Vorwissen auf. Fraglich ist daher, ob das konzeptuelle Vorwissen den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb moderiert, wenn sich die Lernumgebungen im Erwerb konzeptuellen Wissens unterscheiden und die Varianz im Bereich geringen konzeptuellen Vorwissens höher ist. Im Gegensatz zum konzeptuellen Vorwissen hatten viele Teilnehmer ein geringes anwendungsorientiertes Vorwissen. Insbesondere im Bereich niedrigen anwendungsorientierten Vorwissens gab es jedoch eine geringe Varianz in diesem Wissen zwischen den Teilnehmern. Diese geringe Varianz im anwendungsorientierten Vorwissen ist eine mögliche Ursache für die ausbleibende moderierende Rolle des anwendungsorientierten Vorwissens (siehe Abschnitt 5.4.2). Daher ist unklar, ob das anwendungsorientierte Vorwissen bei einer höheren Varianz, insbesondere im Bereich niedrigen anwendungsorientierten Vorwissens den Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens moderiert.

Eine weitere Limitation betrifft die statistischen Probleme, mit denen in der Interventionsphase gelernt wurde. Die Inhalte und Aufgabenstellungen der Probleme waren hochspezifisch und aus dem Themenbereich ALM. Eine offene Frage ist, ob ähnliche Ergebnisse bei anderen statistischen Problemen oder Problemen aus anderen Domänen gefunden werden können. Vor allem im Hinblick auf die moderierende Rolle von Shifting und Gf muss eingeräumt werden, dass in der Studie 1 dieser Arbeit die Lösung der statistischen Probleme vermutlich eine hohe Shifting-Fähigkeit und eine hohe Gf erforderten. In anderen Studien zur Effektivität von Lösungsbeispielen verglichen mit Problemlösen wurden andere Probleme verwendet, deren Lösung möglicherweise geringere Anforderungen an die Shifting-Fähigkeit und Gf stellen. Beispielweise wurden den Teilnehmern in einer Studie von Renkl et al. (2003) Probleme aus dem Themengebiet der Wahrscheinlichkeitsrechnung gezeigt, das auch für die

Domäne Statistik von Bedeutung ist. Beispielsweise musste bei diesen Problemen die Wahrscheinlichkeit bestimmt werden, beim zweimaligen Ziehen einer Kugel ohne Zurücklegen aus einer Urne mit drei roten und zwei weißen Kugeln beim ersten Ziehen eine rote und beim zweiten Ziehen eine weiße Kugel zu ziehen. Die Probleme wurden auf einem Bildschirm gezeigt und es gab keine zusätzlichen Informationsquellen (Renkl et al., 2003). In der Studie 1 dieser Arbeit wurden dagegen Probleme und die zugehörigen Informationen zur Problemlösung auf separaten Folien präsentiert, zwischen denen die Teilnehmer in der Problemlösebedingung wechseln mussten. Die Probleme in der vorliegenden Studie waren zudem weitaus umfangreicher als in der Studie von Renkl et al. (2003). Die Probleme in der Studie von Renkl et al. (2003) stellten vermutlich weitaus geringere Anforderungen an die Fähigkeit zum Shifting. So war bezogen auf das Urnenbeispiel zur Problemlösung mutmaßlich nur ein Wechsel zwischen wenigen Aspekten des kurzen Problems nötig. In größerem Ausmaß als Shifting wurde aber möglicherweise Gf beansprucht, da auf Basis gegebener und/oder berechneter Wahrscheinlichkeiten geschlussfolgert werden musste, wie sich Zwischen- und/oder Endergebnisse errechnen lassen. Aufgrund der wahrscheinlich geringen Shifting-Anforderungen in der Problemlösebedingung (niedriger Grad an instruktionaler Unterstützung) könnten die Lösungsbeispiele (hoher Grad an instruktionaler Unterstützung) bei Problemen wie in der Studie von Renkl et al. (2003) nicht zu einer bedeutsamen Reduzierung der Shifting-Anforderungen führen. Interindividuelle Unterschiede in der Shifting-Fähigkeit spielen daher wahrscheinlich eine untergeordnete Rolle, weshalb die Fähigkeit zum Shifting keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb haben könnte. Neben den Problemen in der Studie von Renkl et al. (2003) wurden auch in anderen Studien weitaus weniger umfangreiche Probleme verwendet als in der Studie 1 dieser Arbeit (z.B. Leppink et al., 2014; Paas, 1992). Daher muss bei der Interpretation der Ergebnisse zur Moderatorenrolle der Shifting-Fähigkeit und Gf in der Studie 1 berücksichtigt werden, dass speziell in dieser Studie statistische Probleme verwendet wurden, deren Lösung wahrscheinlich hohe Anforderungen an diese kognitiven Fähigkeiten stellen. Bei anderen statistischen Problemen oder Problemen aus anderen Domänen, deren Lösung möglicherweise unterschiedliche Anforderungen an kBf und Gf stellt, könnte sich ein abweichendes Ergebnismuster zeigen.

Eine spezifische Limitation betrifft das dritte statistische Problem. Der, mit der Ratingskala eingeschätzte CL war für dieses Problem etwas höher als für die anderen Probleme. Eine post-hoc Analyse des Problems zeigte, dass es komplexer als die beiden anderen statisti-

schen Probleme war. Der Grund hierfür ist, dass nicht klar ersichtlich die statistische Methode zur Problemlösung nahegelegt wird, die im zugehörigen Lösungsbeispiel und Informationen zur Lösung des Problems erwähnt wird. Allerdings äußerte kein Teilnehmer Bedenken hinsichtlich des dritten statistischen Problems und dem zugehörigen Lösungsbeispiel sowie den Informationen zur Lösung des Problems. Nur die dritte MC-Frage wies einen direkten Bezug zu dem dritten statistischen Problem bzw. dem zugehörigen Lösungsbeispiel und Informationen zur Lösung des Problems auf. Für diese Frage konnte jedoch ein signifikanter Erwerb von konzeptuellem Wissen in beiden Lernumgebungen festgestellt werden (die Ergebnisse dieser post-hoc Analyse sind in der vorliegenden Arbeit nicht dargestellt). Demnach hatte die höhere Komplexität des dritten statistischen Problems keine negativen Auswirkungen auf den Wissenserwerb.

Im Hinblick auf die ausbleibende Korrelation zwischen dem CL und der AG-Kapazität kann einschränkend eingeräumt werden, dass mit der Ratingskala womöglich der CL im AG nicht valide erfasst wurde, zumal die Validität der Ratingskala bisher kaum in Frage gestellt wurde (de Jong, 2010). Der CL hätte auch mit anderen Verfahren erfasst werden können. Ein Beispiel hierfür sind Zweitaufgaben, die parallel zu den Primäraufgaben ausgeführt werden müssen. Unterschiede im CL durch eine Primäraufgabe können durch Unterschiede in Leistungen in einer Zweitaufgabe festgestellt werden. Zu beachten ist dabei jedoch, dass die Leistung in der Primäraufgabe durch die Zweitaufgabe beeinträchtigt werden kann (Brünken et al., 2003). Zudem ist denkbar, dass sich Unterschiede in der Zweitaufgabe bzw. im CL ergeben, wenn die Primäraufgabe beispielsweise hohe Anforderungen an die Shifting-Fähigkeit und/oder Gf stellt.

5.4.4 Weitere Forschung

Die Ergebnisse der Studie 1 dieser Arbeit sollten mit anderen Stichproben und insbesondere anderen statistischen Problemen sowie Problemen aus anderen Domänen repliziert werden.

Lösungsbeispiele sind insbesondere für Lerner mit niedrigem Vorwissen effektiv (Kalyuga, 2007). Bei Themengebieten, zu denen Teilnehmer ein niedriges konzeptuelles Vorwissen haben, könnte ein Vorteil der Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) hinsichtlich des Erwerbs konzeptuellen Wissens

festgestellt werden, insbesondere bei einer hohen inhaltlichen Überlappung zwischen den Fragen zur Erfassung des konzeptuellen Wissens und den Lösungsbeispielen. Falls ein Unterschied zwischen Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung im Erwerb konzeptuellen Wissens feststellbar ist und Teilnehmer im Bereich niedrigen konzeptuellen Vorwissens eine hohe Variabilität aufweisen, könnte zudem die moderierende Rolle des konzeptuellen Vorwissens genauer untersucht werden. Eine hohe Variabilität des konzeptuellen Vorwissens im unteren Wertebereich könnte mit Fragen erreicht werden, die im unteren Wertebereich des konzeptuellen Vorwissens gut zwischen den Teilnehmern differenzieren. Analog dazu könnten auch für die Erfassung des anwendungsorientierten Vorwissens Fragen verwendet werden, die im unteren Wertebereich gut differenzieren. Eine moderierende Rolle des anwendungsorientierten Vorwissens könnte folglich festgestellt werden. Des Weiteren könnte die Verwendung einer größeren Anzahl an Fragen zur Messung des anwendungsorientierten Wissens reliablere Schätzungen des anwendungsorientierten Wissens der Teilnehmer liefern und somit die Ergebnisse zukünftiger Studien untermauern. Zudem könnten reliablere Schätzungen von anwendungsorientiertem sowie konzeptuellem Wissen erreicht werden, wenn inhaltsähnlichere bzw. homogenere Fragen zur Erfassung dieser Wissensarten verwendet werden. Bezüglich des konzeptuellen Wissens könnten weitere Studien die moderierende Rolle von kBf und Gf untersuchen, wenn sich die Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung im Erwerb konzeptuellen Wissens unterscheiden.

Zukünftige Studien sollten zudem die mögliche moderierende Rolle der AG-Kapazität in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung genauer untersuchen. Eine moderierende Rolle der AG-Kapazität könnte aufgedeckt werden, falls es deutliche Unterschiede in den Anforderungen an die AG-Kapazität zwischen den Lernumgebungen gibt. Unterschiede in den Anforderungen an die AG-Kapazität zwischen Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung könnten beispielsweise beobachtet werden, wenn die instruktionale Unterstützung durch die Lösungsbeispiele optimiert wird. Ein Format, in dem die physikalisch oder zeitlich getrennten Informationen integriert sind, könnte nach Ayres und Sweller (2005) verhindern, dass mehrere Informationsquellen mental integriert werden müssen. Mit Bezug auf die Lernumgebung mit hohem Grad an instruktionaler Unterstützung könnte die Integration von Lösungsbeispielen und den Informationen zur Lösung statistischer Probleme in einem Format (z.B. eine Grafik) die Anforderungen an die AG-Kapazität in dieser Bedingung reduzieren. Der Grund hierfür ist,

dass die Notwendigkeit zur mentalen Integration von Lösungsbeispielen und den Informationen zur Lösung statistischer Probleme im AG durch das integrierte Format höchstwahrscheinlich reduziert ist. Dadurch könnte sich der Unterschied zwischen den Lernumgebungen in den Anforderungen an die AG-Kapazität erhöhen und eine moderierende Rolle der AG-Kapazität möglicherweise aufgedeckt werden.

Die Ergebnisse der Studie 1 dieser Arbeit liefern Evidenz für die Ansicht, dass der selbstberichtete CL von Lernern ein Prädiktor mit geringerer Validität für Lernerfolgsmaße in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung ist als objektivere Maße von kBf. Zusätzlich lässt die nichtsignifikante Korrelation zwischen CL und AG-Kapazität stark an der Konstruktvalidität des CL zweifeln. Sicherlich könnten die Ergebnisse der Studie 1 dahingehend betrachtet werden, dass sie zu den 20% der erwarteten Fälle gehören, die bei Studien mit einer Teststärke von 80% einen nachgewiesenen Effekt nicht finden können. Um diese Erklärung ausschließen zu können, sind Replikationen der Studie 1 dieser Arbeit mit unterschiedlichen und möglicherweise größeren Stichproben erforderlich. Wird jedoch davon ausgegangen, dass der Effekt in dem Setting und für die Stichprobe der vorliegenden Studie nicht existiert, ist die Situation anders. Das Ergebnismuster deutet vor dem Hintergrund des Wissens über kBf und deren objektiven Erfassung offensichtlich darauf hin, den CL zu rekonzeptualisieren. Obwohl es für die Gestalter von Instruktionsdesigns womöglich unangenehm erscheint, könnte sich herausstellen, dass der CL eher ein multidimensionales als ein eindimensionales Konstrukt ist. Denkbar wäre, dass in Abhängigkeit der Charakteristiken von zu lösenden Problemen unterschiedliche Anforderungen an kBf und nicht nur an die AG-Kapazität gestellt werden. Die Fähigkeit zum Shifting könnte eine kBf sein, die für Probleme relevant ist, deren Lösung ein häufiges Wechseln der Aufmerksamkeit zwischen unterschiedlichen Arten von Informationen erfordern. Um diese Informationen zu integrieren und zu einem Schluss hinsichtlich der Lösung von Problemen zu kommen (z.B. durch Auswahl der für eine Problemlösung relevanten Informationen) könnte über kBf und domänenspezifisches Wissen hinaus eine generelle Kapazität zum schlussfolgernden Denken von Bedeutung sein. Diese Kapazität könnte durch Maße für Gf erfasst werden.

5.4.5 Schlussfolgerungen

Auf Basis der Ergebnisse der Studie 1 dieser Arbeit wird geschlussfolgert, dass die kBf Shifting sowie Gf für die Bestimmung des optimalen Grads an instruktionaler Unterstüt-

zung wichtig sein könnten. Ein hoher Grad an instruktionaler Unterstützung scheint umso wichtiger zu sein, je geringer die Shifting-Fähigkeit und Gf sind.

Die AG-Kapazität war in der vorliegenden Studie kein signifikanter Moderator für den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb. Ein moderierender Einfluss der AG-Kapazität könnte aber aufgedeckt werden, wenn sich Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung signifikant in den Anforderungen an die AG-Kapazität unterscheiden. Obwohl die AG-Kapazität als wichtige Variable für Entscheidungen hinsichtlich eines optimalen Grades an instruktionaler Unterstützung angesehen wird, legen die Ergebnisse der vorliegenden Studie nahe, Shifting und Gf zu berücksichtigen, wenn Lernumgebungen mit geringem und hohem Grad an instruktionaler Unterstützung hinsichtlich des Wissenserwerbs kontrastiert werden. Für Entscheidungen über den optimalen Grad an instruktionaler Unterstützung sind möglicherweise interindividuelle Unterschiede in der Fähigkeit zum Shifting und Gf bedeutsam.

Obwohl das Ergebnismuster der Studie einige theoretische Implikationen hat, gibt es direkte praktische Schlussfolgerungen: Konfrontiert man Lerner in frühen Phasen des Fertigkeitserwerbs mit offenem Problemlösen, könnte das vorzugweise Lehrenden helfen, zwischen Lernern mit unterschiedlicher Shifting-Fähigkeit als eine kBf und Gf zu differenzieren. Das Geben zusätzlicher instruktionaler Unterstützung könnte dazu beitragen, Lernern mit unterschiedlichen Voraussetzungen ähnliche Möglichkeiten zum Lernen zu geben.

Allerdings sind Replikationsstudien notwendig, um den moderierenden Einfluss von Shifting und Gf zu bestätigen und die Rolle der AG-Kapazität in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung genauer zu untersuchen.

6 Studie 2: Metaanalyse von Arbeitsgedächtnistrainings (Fragen 2 und 3)

6.1 Ziele und Hypothesen

Mit Studie 2 sollen die Forschungsfragen 2 und 3 beantwortet werden.

Frage 2:

Welche nahen und weiten Transfereffekte folgen auf Arbeitsgedächtnistrainings?

Die zugehörigen Hypothesen zur zweiten Forschungsfrage wurden aus den Ergebnissen der Metaanalyse von Melby-Lervåg und Hulme (2013) abgeleitet (siehe Abschnitt 2.4.1). Für eine bessere Vergleichbarkeit der Metaanalyse in der vorliegenden Arbeit mit der Metaanalyse von Melby-Lervåg und Hulme (2013) wurde deren Analysestrategie angewendet, jedoch einige wichtige Änderungen vorgenommen. Neben methodischen Änderungen, die in den folgenden Abschnitten erwähnt werden, wurde zwischen nahen Transfereffekten auf das KZG und AG unterschieden. Melby-Lervåg und Hulme (2013) nahmen diese Unterscheidung bei ihrer Analyse nicht vor und untersuchten nur nahe Transfereffekte auf das AG. Wie in Abschnitt 2.2.1 erwähnt, unterscheiden sich KZG- und AG-Aufgaben nur in dem Ausmaß, wie grundlegende Teilprozesse wie inneres Memorieren der Items zwischen Darbietung und Wiedergabe, Behalten, Aktualisieren und kontrollierte Suche in einer bestimmten Aufgabe operieren (Unsworth & Engle, 2007b). Daher sollten Transfereffekte auf das AG mit Transfereffekten auf das KZG und umgekehrt Transfereffekte auf das KZG mit Transfereffekten auf das AG einhergehen. Folgende Hypothesen bezüglich Transfereffekten von AG-Trainings sind der zweiten Forschungsfrage zugeordnet:

1. a. Es gibt nahe Transfereffekte von AG-Trainings auf KZG- und AG-Komponenten (verbale und visuell-räumliche Modalität) zum Posttest.
b. Es gibt keine Transfereffekte auf die verbale Modalität von KZG und AG zum Follow-up.
c. Es gibt Transfereffekte auf die visuell-räumliche Modalität von KZG und AG zum Follow-up.
2. a. Es gibt einen kleinen, weiten Transfereffekt auf die nonverbale Fähigkeit zum Posttest, der zum Follow-up nicht erhalten bleibt.
b. Es gibt keine Transfereffekte auf die verbale Fähigkeit, die Fertigkeit zur Wort-

dekodierung und mathematische Fähigkeiten und Fertigkeiten.

Die dritte Forschungsfrage bezieht sich auf potentielle Moderatoren der Transfereffekte von AG-Trainings.

Frage 3:

Welche Trainingsbedingungen haben einen moderierenden Einfluss auf Transfereffekte von Arbeitsgedächtnistrainings?

Bezüglich der signifikanten Moderatoren in der Metaanalyse von Melby-Lervåg und Hulme (2013) werden folgende Hypothesen formuliert:

3. a. Das Alter ist ein Moderator von Transfereffekten von AG-Trainings auf das verbale AG. Jüngere Teilnehmer sollten größere Verbesserungen im verbalen AG erzielen als ältere Teilnehmer.
- b. Der Interventionstyp (d.h., die Art des Trainingsprogramms) ist ein Moderator von Transfereffekten auf das visuell-räumliche AG. Die Effekte des Cogmed Trainingsprogramms auf das visuell-räumliche AG sind größer als die Effekte anderer, nicht-kommerzieller Programme. Zwischen den kommerziellen Programmen gibt es keine Unterschiede.
- c. Die Art der Kontrollgruppe ist ein Moderator von Transfereffekten auf die non-verbale Fähigkeit. Der mittlere Effekt auf die nonverbale Fähigkeit ist bei einem Vergleich von Trainingsgruppen mit passiven Kontrollgruppen größer als beim Vergleich von Trainingsgruppen mit aktiven Kontrollgruppen.
- d. Die Trainingsdosis (Gesamtdauer des Trainings) ist kein Moderator von Transfereffekten von AG-Trainings.

Der Hauptunterschied zur Metaanalyse von Melby-Lervåg und Hulme (2013) war die Berücksichtigung einer Vielzahl weiterer, möglicher Moderatoren von Transfereffekten von AG-Trainings (siehe Abschnitt 2.4.1). Zusätzlich zur Trainingsdosis werden folgende Hypothesen bezüglich der Zeitkonfiguration von AG-Trainings formuliert:

4. a. Die Dauer einzelner Trainingssessions ist ein Moderator von Transfereffekten von AG-Trainings. Es wird angenommen, dass längere Trainingssessions zu größeren Effekten führen.

b. Die Häufigkeit des Trainings pro Woche ist ein Moderator von Transfereffekten von AG-Trainings. Es wird angenommen, dass mehrere Trainingssessions pro Woche zu größeren Effekten führen.

c. Das Trainingsintervall ist ein Moderator von Transfereffekten von AG-Trainings. Das Trainingsintervall lag in den meisten der Trainingsstudien bei 1 oder 2 Tagen. Es wird angenommen, dass ein Trainingsintervall von 2 Tagen zu größeren Effekten führt als ein Trainingsintervall von 1 Tag.

Des Weiteren werden die folgenden Hypothesen formuliert, die sich auf die Charakteristiken von Trainingsbedingungen und der Implementation von Trainings beziehen:

5. a. Die trainierte Modalität ist ein Moderator von Transfereffekten von AG-Trainings. Transfermaße für KZG und AG beanspruchen unterschiedliche Modalitäten des KZG und AG (verbal und visuell-räumlich). Aus diesem Grund wird angenommen, dass Transfereffekte auf verbales KZG und AG bei Interventionen größer sind, in denen das verbale AG trainiert wurde. In analoger Weise sollten Transfereffekte auf visuell-räumliches KZG und AG bei Interventionen größer sein, in denen das visuell-räumliche AG trainiert wurde. Maße für weiten Transfer unterscheiden sich im Ausmaß, in dem bestimmte Komponenten des AGs beansprucht werden. Daher wird keine gerichtete Hypothese hinsichtlich des moderierenden Einflusses der trainierten Modalität auf die Variabilität zwischen Effektstärken bei Maßen für den weiten Transfer formuliert.

b. Supervision ist ein Moderator von Transfereffekten von AG-Trainings. Die Effekte sind bei supervidierten Trainings größer als bei Trainings ohne die Anwesenheit anderer Personen und Trainings, die in der Anwesenheit anderer Personen stattfinden. Die Effekte sind bei Trainings ohne die Anwesenheit anderer Personen größer als bei Trainings, die in der Anwesenheit anderer Personen stattfinden.

c. Die instruktionale Unterstützung ist ein Moderator von Transfereffekten von AG-Trainings. Transfereffekte von Trainings mit zusätzlicher instruktionaler Unterstützung über die Erklärungen zu Beginn des Trainings hinaus sind größer als Transfereffekte von Trainings ohne diese zusätzlichen Erklärungen.

d. Feedback ist ein Moderator von Transfereffekten von AG-Trainings. Trainings mit Feedback über reines Ergebniswissen hinaus erzielen größere Effekte als Trainings, die Feedback in Form reinen Ergebniswissens mit einschließen.

e. Der Trainingsort ist ein Moderator von Transfereffekten von AG-Trainings. Die

Transfereffekte von Trainings im Labor sind größer als die Transfereffekte von Trainings in der Schule oder zu Hause.

6.2 Methode

6.2.1 Literatursuche und Einschlusskriterien

In den elektronischen Datenbanken PsycINFO, PsycARTICLES, ERIC und Medline wurde nach Studien über AG-Trainings gesucht, die im Zeitraum von 1981 bis Dezember 2013 veröffentlicht wurden. Schlüsselwörter waren „working memory training“, die als eine Zeichenkette in einer allgemeinen Suche (d.h., ohne Eingrenzungen, z.B. auf Schlagwörter) verwendet wurden. Des Weiteren wurden Literaturverzeichnisse von Artikeln durchsucht. Die Autoren der in die Metaanalyse eingeschlossenen Studien wurden darum gebeten, andere (auch nicht publizierte) Artikel bereitzustellen. Die Datenbanksuche lieferte 579 Einträge nachdem Duplikate entfernt wurden. Zusammen mit Artikeln aus anderen Quellen (z.B. Reviews, Anfragen an Autoren) wurden etwa 600 Einträge gefunden.

Es wurden die Einschlusskriterien von Melby-Lervåg und Hulme (2013) verwendet, um die Analyse diesbezüglich vergleichbar zu halten und die von Melby-Lervåg und Hulme (2013) analysierten Studien einzuschließen. Nachfolgend sind die Einschlusskriterien aufgeführt:

1. Bei den Studien musste es sich um randomisierte, kontrollierte Studien oder Quasi-Experimente handeln. Die Studien mussten eine Trainingsgruppe sowie entweder eine aktive oder passive Kontrollgruppe beinhalten, die zu einem Pre- und einem Posttest getestet wurden.
2. Die Trainingsgruppe musste für mindestens zwei Wochen eine Intervention in Form eines adaptiven, computerisierten Programms erhalten, das darauf abzielt, AG-Modalitäten (verbal, visuell-räumlich, oder beide) zu trainieren.
3. Der sprachliche Hintergrund und Lernstatus der Teilnehmer (mit oder ohne Lernschwierigkeiten) spielten keine Rolle. Personen mit einem Alter über 75 Jahre wurden jedoch ausgeschlossen.
4. Bei den Studien mussten Daten vorhanden sein, um eine Effektstärke für die Transfermaße berechnen zu können (Melby-Lervåg & Hulme, 2013).

Bezüglich des letzten Punkts in der Aufzählung wurden die Autoren der Studien darum gebeten, Daten für die Berechnung von Effektstärken zur Verfügung zu stellen, falls die Daten nicht in den Artikeln berichtet wurden.

Insgesamt flossen 47 Studien mit 65 Gruppenvergleichen aus Zeitschriften in die Analyse ein. Dreiundzwanzig dieser Studien mit 30 Gruppenvergleichen wurden aus der Analyse von Melby-Lervåg und Hulme (2013) übernommen. Zwei Limitationen in deren Metaanalyse betreffen die Studien von Schmiedek, Lövdén und Lindenberger (2010) und Klingberg, Forsberg und Westerberg (2002). In der Studie von Schmiedek et al. (2010) trainierten Teilnehmer mit Aufgaben, die AG, aber auch Wahrnehmungsgeschwindigkeit und episodisches Gedächtnis spezifisch (d.h., separate Aufgaben in Abhängigkeit des trainierten Konstrukts) beanspruchen. In der Studie von Klingberg et al. (2002) trainierten Teilnehmer mit AG-Aufgaben und zusätzlich mit einer Mischung aus einer Reaktionszeit- und einer Inhibitionsaufgabe. Die vorliegende Metaanalyse sollte in erster Linie auf AG-Trainings fokussieren aber die Vergleichbarkeit zu der Metaanalyse von Melby-Lervåg und Hulme (2013) gegeben sein. Deshalb wurde die Analyse mit und ohne die Studien von Schmiedek et al. (2010) und Klingberg et al. (2002) gerechnet. Der Ausschluss dieser beiden Studien beeinflusste nahe Transfereffekte auf verbales und visuell-räumliches KZG und AG sowie weite Transfereffekte auf nonverbale und verbale Fähigkeiten.

6.2.2 Codierung von Transfermaßen und Moderatoren

Codierung von Transfermaßen

Es wurden Maße kodiert für KZG und AG (jeweils verbal und visuell-räumlich), nonverbale und verbale Fähigkeiten, die Fertigkeit zur Wortdekodierung und mathematische Fähigkeiten und Fertigkeiten. Wie bereits unter Abschnitt 2.2.1 erwähnt, unterscheiden Redick, Broadway, et al. (2012) KZG-Aufgaben, die nur die vorübergehende Speicherung von Informationen erfordern, von AG-Aufgaben, welche zusätzlich die Verarbeitung von Informationen erfordern. Obwohl Melby-Lervåg und Hulme (2013) erwähnen, dass sie eine Unterscheidung zwischen diesen zwei Aufgabentypen vornehmen, wurde diese Unterscheidung tatsächlich nicht konsistent vorgenommen. Der (nahe) Transfer auf das AG wurde von mehreren Autoren von Trainingsstudien mit KZG-Aufgaben erfasst (siehe hierzu auch Shipstead et al., 2012). Die fehlende Unterscheidung zwischen KZG- und AG-Aufgaben in den einzelnen Studien wurde teilweise von Melby-Lervåg und Hulme (2013) übernommen. Beispielsweise kodierten Melby-Lervåg und Hulme (2013) Maße für visuell-räumliches KZG wie die span-

board Aufgabe (Klingberg et al., 2005) und die grid Aufgabe (Nutley et al., 2011) als Maße für visuell-räumliches AG. Bei der grid Aufgabe, welche eine Adaptation der corsi blocks Aufgabe darstellt, muss die Reihenfolge einer Serie von Punkten erinnert werden, die auf einem Gitter präsentiert werden (Alloway, Gathercole, & Pickering, 2006). Folglich haben die grid Aufgabe und die corsi blocks Aufgabe nur eine Speicherkomponente und werden als Beispiele für KZG-Aufgaben angesehen. Neben den Empfehlungen von Redick, Broadway, et al. (2012) für eine Trennung von KZG- und AG-Aufgaben wurde in anderen Studien berücksichtigt, dass simple span Aufgaben mit visuell-räumlichem Material Maße für visuell-räumliches KZG darstellen (z.B. Alloway et al., 2006; Kessels, Berg, & Brands, 2008; Miyake, Friedman, Shah, Rettinger, & Hegarty, 2001). In der vorliegenden Metaanalyse wurden simple span Aufgaben mit visuell-räumlichem Material wie die grid Aufgabe als Maße für visuell-räumliches KZG kodiert. Maße mit einer zusätzlichen Verarbeitungskomponente wie der shape recall Test (verwendet in der Studie von Alloway et al., 2013) wurden als Maße für visuell-räumliches AG kodiert.

Die Kodierung bei der Differenzierung zwischen Maßen von verbalem KZG und verbalem AG war etwas komplexer als in der Metaanalyse von Melby-Lervåg und Hulme (2013). Melby-Lervåg und Hulme (2013) kodierten die digit span backward Aufgabe in den Studien von Borella et al. (2010) und Van der Molen et al. (2010; zusammen mit listening recall, eine CSA) als Maße für verbales AG. In der digit span Aufgabe muss eine Serie von gesprochenen Zahlen in der gleichen (digit span forward) oder umgekehrten (digit span backward) Reihenfolge wiedergegeben werden (Alloway et al., 2006). Die Ergebnisse der Studie von St. Clair-Thompson (2010) legen nahe, dass es sich bei der digit span backward Aufgabe um ein Maß für (verbales) AG bei Kindern und um ein Maß für (verbales) KZG bei Erwachsenen handelt. Diese Ergebnisse sind konform mit Ergebnissen aus anderen Untersuchungen (z.B. Colom, Abad, Rebollo, & Chun Shih, 2005; Gathercole, Pickering, Ambridge, & Wearing, 2004; St Clair-Thompson & Allen, 2013). Folglich wurden die digit span backward Aufgabe bzw. dazu analoge Aufgaben als Maße für verbales AG bei Kindern und als Maße für verbales KZG bei Erwachsenen kodiert. Simple span Maße wie die Aufgaben digit span forward und word recall (z.B. in der Studie von Dunning, Holmes, & Gathercole, 2013 verwendet) wurden als Maße für (verbales) KZG kodiert. CSAs bzw. Maße mit einer zusätzlichen Verarbeitungskomponente wie die reading span Aufgabe (z.B. verwendet in der Studie von Richmond, Morrison, Chein, & Olson, 2011) wurden unabhängig vom Alter als Maße für (verbales) AG kodiert. Melby-Lervåg und Hulme (2013) nutzten die in der Studie von Jaeggie, Studer-Luethi, et al. (2010) berichteten Werte in der n-back Aufgabe um einen Transfer

zum verbalen AG zu erfassen. In der vorliegenden Metaanalyse wurde stattdessen die operation span Aufgabe für die Erfassung des Transfers zum verbalen AG verwendet, da die Übereinstimmungsvalidität der n-back Aufgabe als Maß für das verbale AG allgemein noch nicht akzeptiert ist (für einen Überblick siehe Jaeggi, Buschkuhl, Perrig, & Meier, 2010). Im Gegensatz zur n-back Aufgabe wurde für die operation span Aufgabe gezeigt, dass sie ein reliables und valides Maß für die AG-Kapazität darstellt (Redick, Broadway, et al., 2012).

Die Kodierung von weiten Transfermaßen war identisch zum Vorgehen in der Metaanalyse von Melby-Lervåg und Hulme (2013), außer für Inhibition (in deren Analyse als Aufmerksamkeit benannt und mit der Stroop-Aufgabe gemessen). Wie bereits in Abschnitt 2.1 erwähnt, stellt Inhibition ein multidimensionales Konstrukt dar und könnte in bedeutendem Ausmaß durch CSAs erfasst werden. Daher wurde die Stroop Aufgabe nicht als Maß für Inhibition bzw. Aufmerksamkeit in die vorliegende Analyse eingeschlossen.

Die kodierten Maße für die nonverbale Fähigkeit erfordern das Lösen von Problemen ohne verbale Informationen. Tests, die hauptsächlich verbale Informationen für das Lösen von Problemen erfordern, wurden als Maße für die verbale Fähigkeit kodiert. Die kodierten Maße für die Fertigkeit zur Wortdekodierung beinhalten Geschwindigkeit und Qualität des Lesens von Wörtern und Nichtwörtern, jedoch nicht das Leseverständnis (diese Tests wurden als Maße für die verbale Fähigkeit kodiert). Eine Ausnahme hinsichtlich des Leseverständnisses als ein Aspekt der verbalen Fähigkeit ergab sich in der Studie von Dunning et al. (2013). Ein zusammengesetzter Wert für den verbalen IQ wurde gegenüber dem Maß für Leseverständnis bevorzugt. Der zusammengesetzte Wert wurde ausgewählt, weil davon ausgegangen wird, dass er mehr Gewicht in der Erfassung der verbalen Fähigkeit hat als ein einzelner Wert. Tests, die das Lösen mathematischer Probleme erfordern (z.B. die Anwendung arithmetischer Operationen) wurden als Maße für mathematische Fähigkeiten und Fertigkeiten kodiert (siehe Melby-Lervåg & Hulme, 2013). Der Term "arithmetisch", wie von Melby-Lervåg und Hulme (2013) benutzt, wurde nicht gewählt. Der Grund hierfür war, dass auch andere Aspekte mathematischer Fähigkeiten und Fertigkeiten abgedeckt werden sollten, die durch Tests in einigen Studien erfasst wurden (z.B. der Vergleich von Zahlen in der Studie von Karbach, Strobach, & Schubert, 2014). Allerdings erfordern die meisten mathematischen Tests, die in den eingeschlossenen Studien verwendet wurden, arithmetische Fertigkeiten (z.B. Addition, Division).

Codierung von Moderatoren

Zunächst wurden die signifikanten Moderatoren (Alter, Interventionstyp und Art der Kontrollgruppe) in der Analyse von Melby-Lervåg und Hulme (2013) eingeschlossen und im Anschluss daran Modifikationen vorgenommen. Das Alter über war über die Stichproben der Studien hinweg nicht normalverteilt, jedoch der dekadische Logarithmus des Alters. Daher wurde der dekadische Logarithmus des Alters in die Analyse eingeschlossen, um einen Informationsverlust zu vermeiden, der mit der Kategorisierung von kontinuierlichen Moderatoren in der Analyse von Melby-Lervåg und Hulme (2013) assoziiert war. Obwohl die Trainingsdosis kein signifikanter Moderator in der Metaanalyse von Melby-Lervåg und Hulme (2013) war, wurde diese Variable in die Analyse eingeschlossen. Grund hierfür war, dass die Trainingsdosis als ein wichtiger Faktor mit bedeutender Variation zwischen den Studien angesehen wird und in der Metaanalyse von Melby-Lervåg und Hulme (2013) als dichotomer Moderator eingeschlossen wurde. Die Trainingsdosis (Gesamtdauer des Trainings in Studien) wurde als kontinuierlicher Moderator in die vorliegende Analyse eingeschlossen. Da die Trainingsdosis nicht normalverteilt war, wurde der dekadische Logarithmus der Trainingsdosis (der normalverteilt war) in die Moderatorenanalyse eingeschlossen.

Falls es eine aktive und eine passive Kontrollgruppe in einer Studie gab, wurde die Trainingsgruppe mit der aktiven und der passiven Kontrollgruppe verglichen. Dies steht im Gegensatz zum Vorgehen von Melby-Lervåg und Hulme (2013), die Vergleiche zwischen der aktiven Kontrollgruppe und der Trainingsgruppe mit der passiven Kontrollgruppe einschlossen. Eine Ausnahme betrifft die Studie von Alloway, Bibile und Lau (2013), für die die aktive Kontrollgruppe als zweite Trainingsgruppe eingeschlossen wurde, da diese Trainingsgruppe ein adaptives AG-Training mit einer geringeren Häufigkeit absolvierte. Die zweite Trainingsgruppe und die erste Trainingsgruppe (AG-Training mit hoher Häufigkeit) wurden beide mit der passiven Kontrollgruppe verglichen. Aktive Kontrollgruppen erhielten ein Training, das das AG nicht beanspruchte oder nicht adaptiv war (konstante Schwierigkeit) und deshalb höchstwahrscheinlich nicht dem Training des AGs diente. Es wurde ein anderer Vergleich als von Melby-Lervåg und Hulme (2013) vorgenommen, da keine aktiven Kontrollgruppen ohne adaptives AG-Training als Trainingsgruppen eingeschlossen werden sollten (siehe Einschlusskriterien in Abschnitt 6.2.1). Auch wenn die Vergleiche mehrerer Trainingsgruppen mit einer Kontrollgruppe problematisch sind, wurde diese Prozedur gewählt, um Informationen zu bekommen, die bei anderen Vorgehensweisen verloren gehen könnten (für eine ähnliche Vorgehensweise siehe Wecker & Fischer, 2014).

Andere Moderatoren in der vorliegenden Analyse waren

1. Sessiondauer: Dauer einzelner Trainingssessions (in Minuten).
2. Häufigkeit des Trainings pro Woche (in Tagen). Es konnten nicht immer exakte Informationen über die Anzahl an Trainingstagen pro Woche (z.B. 1 oder 2 Trainingssessions pro Woche) erhalten werden, weshalb drei Kategorien für die Häufigkeit des Trainings pro Woche erstellt wurden (1 oder 2 Tage, 3 Tage, 4 bis 6 Tage).
3. Trainingsintervall: Zeitintervall zwischen einzelnen Trainingssessions (in Tagen; Wochenenden ausgeschlossen). Das Trainingsintervall betrug in den meisten Trainings 1 oder 2 Tage, weshalb diese Variable dichotomisiert wurde (Trainingsintervall von 1 Tag vs. Trainingsintervall von 2 Tagen).

Der Einfluss dieser Variablen wurde untersucht, um detaillierte Informationen über die Rolle der Zeitkonfiguration für Transfereffekte von AG-Trainings zu bekommen.

Die folgenden Moderatoren wurden ausgewählt, um den Einfluss von Variablen zu untersuchen, die für Trainingsparadigmen und die Implementierung von Trainings charakteristisch sind:

1. Modalität: Es wurde kodiert, ob die verbale oder visuell-räumliche Modalität oder beide Modalitäten durch das Trainingsprogramm trainiert wurden.
2. Supervision: Es wurde jede Art von Supervision kodiert, die über die Nichtanwesenheit oder Anwesenheit einer Person wie dem Experimentator hinausging (d.h., keine Anwesenheit einer Person vs. Anwesenheit einer Person vs. Supervision)
3. Instruktionale Unterstützung: Es wurde kodiert, ob es irgendeine Form instruktionaler Unterstützung gab, die über Erklärungen vor dem oder zum Beginn des Trainings hinausging (d.h., keine instruktionale Unterstützung vs. instruktionale Unterstützung).
4. Feedback: Es wurde die Art des Feedbacks über die Ergebnisse der Teilnehmer kodiert. Reines Ergebniswissen (d.h., Richtigkeit der gegebenen Antworten) wurde kontrastiert mit dem Feedback, das darüber hinausging. In den verschiedenen Studien wurde das Feedback entweder nach dem Beenden eines Durchgangs innerhalb einer Aufgabe, am Ende eines Trainingstags, wöchentlich oder nach dem Beenden des kompletten Trainings gegeben. Feedback, das Teilnehmer am Ende des kompletten Trainings erhielten, wurde nicht in die

Analyse eingeschlossen. Bei den meisten, in die vorliegende Analyse eingeschlossenen Trainingsparadigmen erhielten Teilnehmer wenigstens Feedback über die Richtigkeit ihrer Antworten. Daher lag das Interesse auf den Effekten von Feedback, das über das Feedback in Form von reinem Ergebniswissen hinausging (auch Feedback zur Aufgabe genannt; siehe Hattie & Timperley, 2007).

5. Trainingsort: Training im Labor wurde kontrastiert mit Training in der Schule oder zu Hause.

6.2.3 Effektstärkeberechnungen

Hedge's g mit Bias-Korrektur (Hedges & Olkin, 1985) wurde als Effektstärke für jedes Maß berechnet. Um für Unterschiede zwischen Trainings- und Kontrollgruppen zum Pretest zu kontrollieren, wurde die Differenz zwischen den Zuwächsen von Pre- zu Posttest sowie Pretest zu Follow-up berechnet und diese Differenz als Zähler in der Berechnung von Hedge's g berechnet. Die gepoolte Standardabweichung der Standardabweichungen der jeweiligen Trainings- und Kontrollgruppe zum Pretest wurde als Nenner in der Berechnung von Hedge's g verwendet. Morris (2008) favorisiert diese Art der Berechnung von Hedge's g in Pretest-Posttest-Kontrollgruppen Designs. Melby-Lervåg und Hulme (2013) erhielten ihre Effektstärkeschätzungen auf ähnliche Weise, außer dass sie den Mittelwert der Standardabweichungen von Trainings- und Kontrollgruppe zum Pretest im Zähler von Hedge's g verwendeten. Dieser Unterschied führte manchmal zu leicht unterschiedlichen Ergebnissen (z.B. 1.10 für die Effektstärke beim verbalen AG in der vorliegenden Analyse verglichen mit 1.09 in der Analyse von Melby-Lervåg und Hulme (2013) bei der Studie von Thorell et al., (2009).

Die Kontrolle von Differenzen zwischen Trainings- und Kontrollgruppen zum Pretest kann auch dann signifikante, positive Effektstärken erzeugen, wenn sich die Trainingsgruppe nicht verbessert aber sich die Kontrollgruppe verschlechtert. Dies war der Fall für den Transfereffekt auf die nonverbale Fähigkeit bei der Studie von Harrison et al. (2013). In ähnlicher Weise kann ein negativer Effekt resultieren, wenn sich die Kontrollgruppe verbessert und die Trainingsgruppe verschlechtert. Dies war der Fall für die Transfereffekte auf das verbale KZG und die nonverbale Fähigkeit in der Studie von Richmond et al. (2011). Folglich wurden die Analysen für nonverbale Fähigkeit und verbales KZG mit und ohne die Effektstärken von Harrison et al. (2013) bzw. Richmond et al. (2011) gerechnet. Falls Studien mehrere Maße für ein Konstrukt beinhalteten, wurden die Mittelwerte und Standardabweichungen für die Be-

rechnung einer Effektstärke gemittelt. Dieses Vorgehen wurde auch von Melby-Lervåg und Hulme (2013) berichtet.

Große Unterschiede in den Effektstärken ergaben sich, wenn verschiedene Ergebnismaße für die Berechnungen verwendet wurden. Beispielsweise nutzten Melby-Lervåg und Hulme (2013) die letter memory Aufgabe als ein Maß für verbales AG in der Studie von Dahlin, Nyberg, Bäckman und Neely (2008). Diese Aufgabe war jedoch eine der Trainingsaufgaben in der Studie von Dahlin, Nyberg, et al. (2008). In der vorliegenden Analyse sollte der Transfer auf das verbale AG mit CSAs wie der computation span Aufgabe erfasst werden, die auch in der Studie von Dahlin, Nyberg, et al. (2008) verwendet wurde. Als Folge unterschieden sich die Effektstärken stark.

Die Studien wurden vom Autor der Arbeit kodiert und alle Effektstärken wurden durch eine studentische Hilfskraft doppelkodiert. Falls Unterschiede in den Ergebnissen auftraten, wurden diese durch den Autor der vorliegenden Arbeit neu berechnet. Danach betrug die Intraklassenkorrelation zwischen den Ratern 1. Der Autor der vorliegenden Arbeit kodierte und doppelkodierte auch die Moderatoren. Die Moderatoren von 33 Gruppenvergleichen wurden dann nochmals von einer studentischen Hilfskraft kodiert. Die Interrater-Reliabilität für kategoriale Daten wurde mittels Cohen's Kappa berechnet und betrug $\kappa = 1$ mit einer Übereinstimmungsrate von 100%. Die Interrater-Reliabilität für kontinuierliche Moderatoren wurde mittels Intraklassenkorrelation berechnet und betrug 1 mit einer Übereinstimmungsrate von 100%.

6.2.4 Random-effects Modelle

Ähnlich wie bei Melby-Lervåg und Hulme (2013) wurde ein random-effects Modell ausgewählt, um die globale (mittlere) Effektstärke auf Signifikanz zu testen. Die globale Effektstärke selbst wurde durch Gewichtung der Effektstärken der einzelnen Studien mit der inversen Varianz berechnet (Lipsey & Wilson, 2001). Random-effects Modelle werden aktuell für Metaanalysen favorisiert (Schmidt & Hunter, 2003). Fixed-effects Modelle scheinen nicht empfehlenswert zu sein, insbesondere wenn Effektstärken zwischen Studien variieren (Mosteller & Colditz, 1996; Overton, 1998). Die Wahl eines random-effects Modells für die vorliegende Analyse impliziert, dass die eingeschlossenen Effektstärken vom Populationsmittelwert aufgrund eines Stichprobenfehlers *und* einer zufälligen Varianzkomponente abweichen. Generell sind random-effects Modelle konservativer als fixed-effects Modelle (Lipsey & Wilson, 2001). Die Analysen wurden in SPSS mit den Makros für Metaanalysen von Wil-

son (2005) durchgeführt.

6.2.5 Durchführung

Die Hypothesen über die Transfereffekte von AG-Trainings wurden getestet, indem überprüft wurde, ob die mittlere Effektstärke für jedes Transfermaß signifikant von Null abwich. Für die mittleren Effektstärken wird jeweils das 95%-Konfidenzintervall angegeben. Um auf signifikante Heterogenität in den Effektstärken zwischen den Studien zu testen, wurde die Q -Statistik für jedes Transfermaß berechnet (Hedges & Olkin, 1985). Zusätzlich wurde I^2 als ein Maß berechnet, das über den Grad der Heterogenität zwischen den Studien informiert bzw. wie viel Prozent der gesamten Varianz in den Effektstärken durch wahre Varianz zwischen den Studien erklärt werden kann (Huedo-Medina, Sanchez-Meca, Martin-Martinez, & Botella, 2006). Falls die Q -Statistik signifikant war, wurden Moderatorenanalysen durchgeführt, um die Heterogenität zwischen den Studien zu erklären.

Für alle Analysen der Moderatoren wurden mixed-effects Modelle basierend auf der Methode der Momente verwendet. Mixed-effects Modelle implizieren, dass ein gewisser Anteil der Varianz zwischen den Studien durch einen Moderator erklärt werden kann, aber signifikante Varianz zwischen den Studien übrig bleibt (Lipsey & Wilson, 2001). Ein mixed-effects Modell wurde gewählt, da davon ausgegangen wurde, dass ein einzelner Moderator einen gewissen Anteil der Varianz zwischen den Studien aufklären kann, aber ein signifikanter Restanteil der Varianz übrig bleibt. Kategoriale Moderatoren wurden durch die Verwendung spezieller Varianzanalysen für Metaanalysen analysiert, die einfachen Varianzanalysen ähnlich sind (Lipsey & Wilson, 2001). Falls signifikante Gruppenunterschiede zwischen den Kategorien eines Moderators mit mehreren Kategorien bestanden, wurde eine Varianzanalyse für jeden paarweisen Vergleich mit einer Bonferroni-Korrektur des Signifikanzlevels durchgeführt, um für multiples Testen zu kontrollieren. Der Einfluss kontinuierlicher Moderatoren auf Effektstärken wurde durch Verwendung von Meta-Regressionen untersucht. Moderatoranalysen wurden für jeden einzelnen Moderator durchgeführt. Nicht verwendet werden konnten multiple Meta-Regressionen, die alle Moderatoren beinhalteten, um Informationen über den Einfluss eines einzelnen Moderators unter Kontrolle der anderen Moderatoren zu bekommen. Grund hierfür war, dass bei einzelnen Moderatoren Daten fehlten, was zu großen Verringerungen in den Stichprobengrößen führte, wenn alle Moderatoren in die Analysen einbezogen wurden. Allerdings konnten multiple Meta-Regressionen mit Moderatoren durchgeführt werden, die in einzelnen Meta-Regressionen signifikant waren. Bei den aufgeführten

Meta-Regressionen werden die standardisierten Regressionsgewichte berichtet.

Forest plots wurden erstellt, um die Heterogenität der Effektstärken mit ihren Konfidenzintervallen zu visualisieren (Walker, Hernandez, & Kattan, 2008) und Ausreißer zu identifizieren. Der Einfluss von Ausreißern wurde durch Sensitivitätsanalysen untersucht, in denen die mittleren Effektstärken nach Ausschluss der Ausreißer geschätzt wurden.

Um das Vorliegen eines Publikationsbias zu überprüfen, wurden funnel plots für jeden Transfereffekt erstellt. In diesen funnel plots wurden die entsprechende Effektstärke auf der x-Achse und ein Präzisionsmaß (Standardfehler) auf der y-Achse aufgetragen. Falls die Effektstärken nicht symmetrisch um die globale Effektstärke verteilt waren, deutete dies auf das Vorliegen eines Publikationsbias hin. Da die Interpretation eines Funnel-Plots subjektiv ist, wurde eine leicht modifizierte Version des Egger-Tests (Egger, Davey Smith, Schneider, & Minder, 1997; Sterne & Egger, 2005) angewandt. In der Originalversion des Egger-Tests wird die Effektstärke geteilt durch den Standardfehler durch den inversen Standardfehler vorhergesagt. Falls der Schnittpunkt der Regressionsgeraden mit der y-Achse von Null abweicht, ist dies ein Hinweis für einen Bias (Egger et al., 1997; Rothstein, 2008). Wenn die Effektstärke in einer gewichteten Regression durch den Standardfehler vorhergesagt wird, ist die Steigung der Regressionsgeraden das Pendant zum Schnittpunkt der Regressionsgeraden mit der y-Achse beim originalen Egger-Test (Rothstein, 2008). In der vorliegenden Analyse wurde die Effektstärke geteilt durch deren Varianz als Kriterium und der Standardfehler als Prädiktor genutzt (Sterne & Egger, 2005), um den Egger-Test mit einer (gewichteten) Meta-Regression unter Verwendung des entsprechenden Macros von Wilson (2005) durchzuführen.

Zusätzlich wurde die trim-and-fill Analyse (Duval & Tweedie, 2000a, 2000b) für random-effects Modelle durchgeführt, wenn funnel plot und Egger-Test auf einen potentiellen Publikationsbias hinwiesen. Bei der trim-and-fill Methode (Duval & Tweedie, 2000a, 2000b) wird mittels verschiedener Schätzer die Anzahl fehlender Studien geschätzt, um den funnel plot symmetrisch zu machen. Einer dieser Schätzer ist L_0 , der sich als robuster als andere Schätzer erwies (Talebi, 2013). Daher wurde L_0 für die trim-and-fill Analyse verwendet.

Hinsichtlich der Behandlung fehlender Werte wurde dasselbe Vorgehen wie von Melby-Lervåg und Hulme (2013) verwendet. Studien wurden in alle Analysen eingeschlossen, für die genügend Daten zur Verfügung standen.

6.3 Ergebnisse Studie 2

Für die Beurteilung der Größe von Hedge's g wurden folgende Richtlinien von Cohen

(1988) verwendet: $g = 0.20$ (kleiner Effekt), $g = 0.50$ (moderater Effekt), $g = 0.80$ (starker Effekt).

6.3.1 Nahe und weite Transfereffekte (Frage 2)

Um zu untersuchen, welche Transfereffekte auf AG-Trainings folgen (Frage 2), wurden Metaanalysen für jeden nahen und jeden weiten Transfereffekt durchgeführt.

Verbales KZG. Es wurde erwartet, dass es einen Transfer von AG-Trainings auf das verbale KZG zum Posttest (Hypothese 1.a), jedoch nicht zum Follow-up (Hypothese 1.b) gibt. Ein forest plot für 32 Effektstärken für den Vergleich von Zuwächsen im verbalen KZG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 12 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf das verbale KZG war klein bis moderat ($g = 0.37$, 95% KI [0.19, 0.56], $p < .001$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(31) = 93.40$, $p < .001$, $I^2 = 66.81\%$. Nach der Entfernung von Ausreißern reichte die Effektstärke von $g = 0.25$, 95% KI [0.11, 0.39] bis $g = 0.42$, 95% KI [0.25, 0.58]. Nachdem die Effektstärke der Studie von Richmond et al. (2011) ausgeschlossen wurde (in dieser Studie verschlechterte sich die Trainingsgruppe und die Kontrollgruppe verbesserte sich), betrug die mittlere Effektstärke $g = 0.42$, 95% KI [0.25, 0.58]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 1.a einen kurzfristigen Transfereffekt von AG-Trainings auf das verbale KZG.

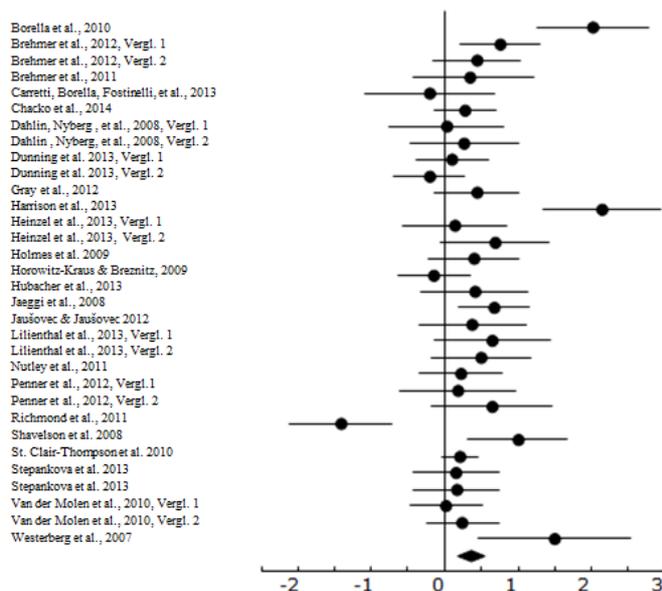


Abbildung 12. Forest plot für kurzfristige Transfereffekte auf das verbale KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Im Hinblick auf langfristige Transfereffekte (Hypothese 1.b) ist ein forest plot für 9 Effektstärken für den Vergleich von Zuwächsen im verbalen KZG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up in Abbildung 13 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf das verbale KZG war klein bis moderat ($g = 0.22$, 95% KI [0.02, 0.42], $p = .03$). Der Follow-up fand im Mittel 8.11 Monate nach dem Posttest statt. Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(8) = 4.57$, $p = .80$. Nach der Entfernung von Ausreißern reichte die mittlere Effektstärke von $g = 0.15$, 95% KI [-0.07, 0.36] bis $g = 0.27$, 95% KI [0.05, 0.47]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es im Gegensatz zur Hypothese 1.b einen langfristigen Transfereffekt von AG-Trainings auf das verbale KZG.

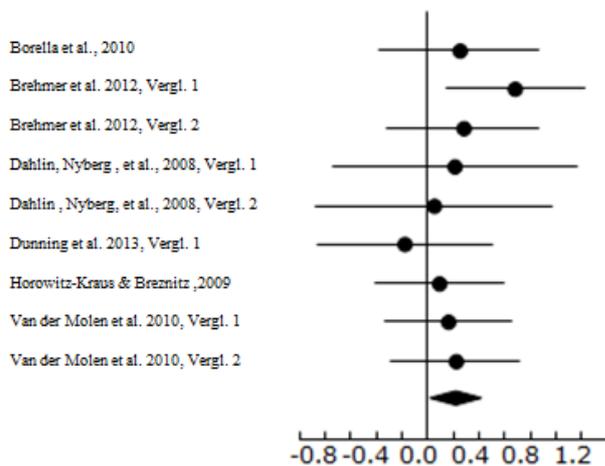


Abbildung 13. Forest plot für langfristige Transfereffekte auf das verbale KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Visuell-räumliches KZG. Es wurde erwartet, dass es einen Transfer von AG-Trainings auf das visuell-räumliche KZG zum Posttest (Hypothese 1.a) und zum Follow-up (Hypothese 1.c) gibt. Ein forest plot für 25 Effektstärken für den Vergleich von Zuwächsen im visuell-räumlichen KZG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 14 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf das visuell-räumliche KZG war moderat bis groß ($g = 0.72$, 95% KI [0.56, 0.92], $p < .001$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(24) = 68.20$, $p < .001$, $I^2 = 64.82\%$. Nachdem die Effektstärke der Studie von Klingberg et al. (2002) entfernt wurde (Teilnehmer in dieser Studie trainierten zusätzlich zu einer AG-Aufgabe mit einer Mischung aus einer Reaktionszeitaufgabe und einer Inhibitionsaufgabe), war der mittlere Effekt moderat bis groß ($g = 0.70$, 95% KI [0.50, 0.91]). Nach der Entfernung von Ausreißern reichte die

mittlere Effektstärke von $g = 0.61$, 95% KI [0.43, 0.80] bis $g = 0.76$, 95% KI [0.55, 0.97]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 1.a einen kurzfristigen Transfereffekt von AG-Trainings auf das visuell-räumliche KZG.

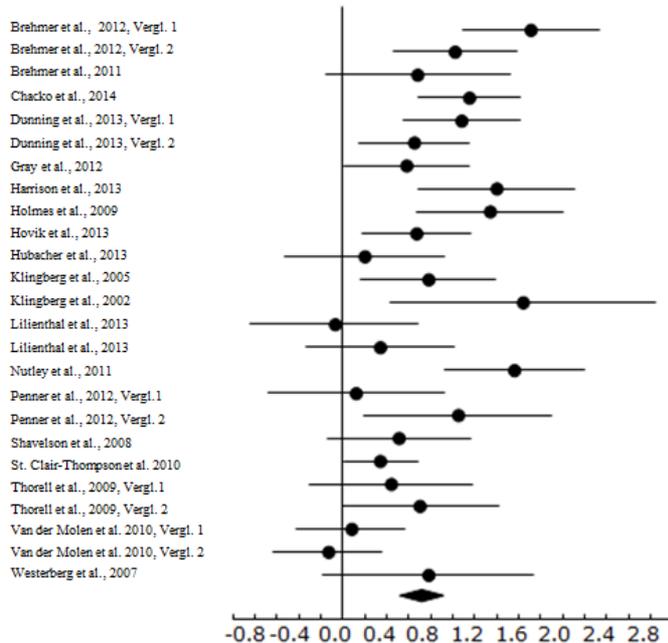


Abbildung 14. Forest plot für kurzfristige Transfereffekte auf das visuell-räumliche KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Ein forest plot für 7 Effektstärken für den Vergleich von Zuwächsen im visuell-räumlichen KZG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up ist in Abbildung 15 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf das visuell-räumliche KZG war moderat bis groß ($g = 0.78$, 95% KI [0.33, 1.23], $p < .001$). Der Follow-up fand im Mittel 4.86 Monate nach dem Posttest statt. Die Effektstärken variierten signifikant zwischen den Studien, $Q(6) = 25.68$, $p < .001$, $I^2 = 76.64\%$. Nach der Entfernung von Ausreißern reichte die mittlere Effektstärke von $g = 0.63$, 95% KI [0.22, 1.03] bis $g = 0.91$, 95% KI [0.47, 1.35]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 1.c einen langfristigen Transfereffekt von AG-Trainings auf das visuell-räumliche KZG.

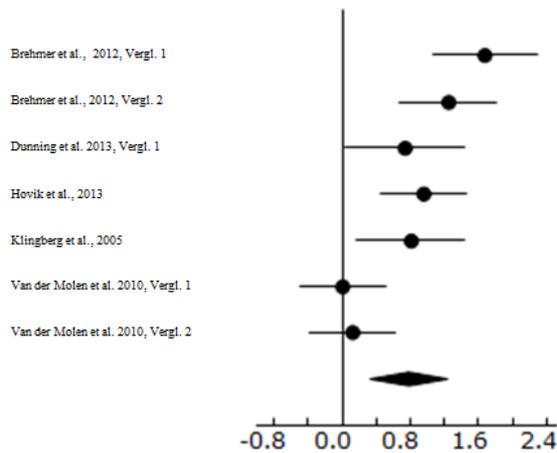


Abbildung 15. Forest plot für langfristige Transfereffekte auf das visuell-räumliche KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Verbales AG. Es wurde erwartet, dass es einen Transfer von AG-Trainings auf das verbale AG zum Posttest (Hypothese 1.a), jedoch nicht zum Follow-up (Hypothese 1.b) gibt. Ein forest plot für 42 Effektstärken für den Vergleich von Zuwächsen im verbalen AG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 16 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf das verbale AG war moderat bis groß ($g = 0.55$, 95% KI [0.33, 0.78], $p < .001$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(41) = 252.00$, $p < .001$, $I^2 = 83.73\%$. Nachdem die Effektstärken der Studien von Klingberg et al. (2002) und Schmiedek et al. (2010) entfernt wurden, war der mittlere Effekt klein bis moderat ($g = 0.58$, 95% KI [0.34, 0.82], $p < .001$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(39) = 242.20$, $p < .001$, $I^2 = 83.90\%$. Nach der Entfernung von Ausreißern reichte die mittlere Effektstärke von $g = 0.30$, 95% KI [0.14, 0.46] bis $g = 0.62$, 95% KI [0.39, 0.85]. Der funnel plot gab einen Hinweis auf einen leichten, potentiellen Bias, wobei Studien auf der linken Seite des Mittelwerts fehlten. Der Egger-Test deutete auch auf einen Publikationsbias hin, $b = 2.50$, 95% KI [0.67, 4.32], $p < .01$. Kein Hinweis auf fehlende Studien wurde in einer trim-and-fill Analyse gefunden, weshalb keine Studie imputiert wurde. Demzufolge gab es in Übereinstimmung mit der Hypothese 1.a einen kurzfristigen Transfereffekt von AG-Trainings auf das verbale AG.

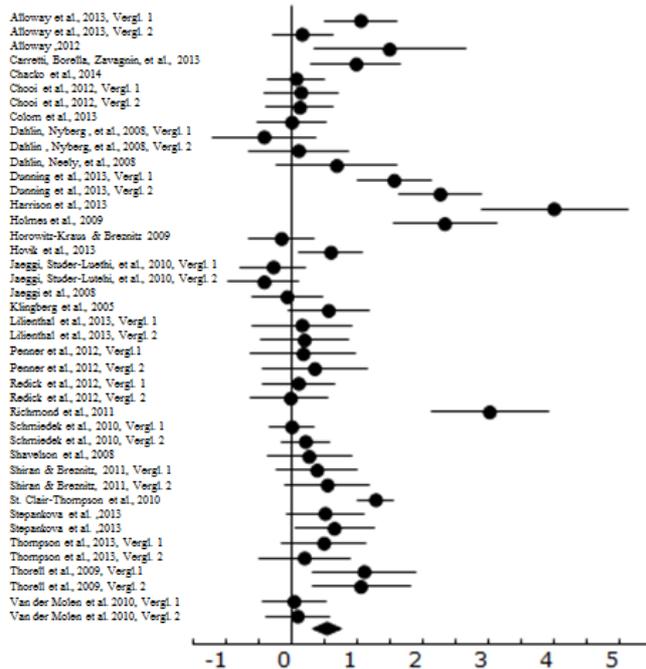


Abbildung 16. Forest plot für kurzfristige Transfereffekte auf das verbale AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Ein forest plot für 11 Effektstärken für den Vergleich von Zuwächsen im verbalen AG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up ist in Abbildung 17 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf das verbale AG war klein bis moderat ($g = 0.35$, 95% KI [0.17, 0.68], $p < .05$). Der Follow-up fand im Mittel 8.36 Monate nach dem Posttest statt. Die Effektstärken variierten signifikant zwischen den Studien, $Q(10) = 20.29$, $p < .001$, $I^2 = 66.98\%$. Nach der Entfernung von Ausreißern reichte die mittlere Effektstärke von $g = 0.16$, 95% KI [-0.12, 0.47] bis $g = 0.44$, 95% KI [0.14, 0.75]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es im Gegensatz mit der Hypothese 1.b einen langfristigen Transfereffekt von AG-Trainings auf das verbale AG.

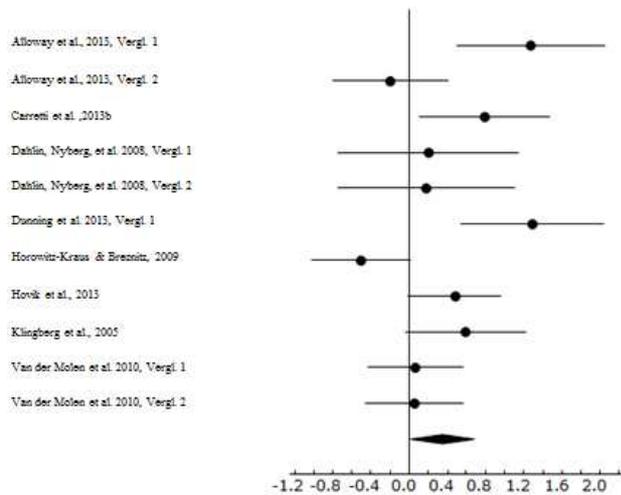


Abbildung 17. Forest plot für langfristige Transfereffekte auf das verbale AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Visuell-räumliches AG. Es wurde erwartet, dass es einen Transfer von AG-Trainings auf das visuell-räumliche AG zum Posttest (Hypothese 1.a) und zum Follow-up (Hypothese 1.c) gibt. Ein forest plot für 19 Effektstärken für den Vergleich von Zuwächsen im visuell-räumlichen AG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 18 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf das visuell-räumliche AG war moderat bis groß ($g = 0.63$, 95% KI [0.35, 0.90], $p < .001$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(18) = 87.69$, $p < .001$, $I^2 = 79.47\%$. Nachdem die Effektstärke der Studie von Schmiedek et al. (2010) entfernt wurde (Teilnehmer in dieser Studie trainierten zusätzlich zu AG-Aufgaben mit Aufgaben, die Wahrnehmungsgeschwindigkeit und episodisches Gedächtnis erfassen), war die mittlere Effektstärke weiterhin moderat bis groß ($g = 0.69$, 95% KI [0.37, 1.00], $p < .001$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(16) = 80.48$, $p < .001$, $I^2 = 81.46\%$. Nach der Entfernung eines Ausreißers betrug die Effektstärke $g = 0.49$, 95% KI [0.30, 0.67]. Der funnel plot gab einen Hinweis auf einen potentiellen Bias, wobei Studien auf der linken Seite des Mittelwerts fehlten. Der Egger-Test deutete auch auf einen Publikationsbias hin, $b = 7.26$, 95% KI [4.70, 9.83], $p < .001$. Nachdem der einzelne Ausreißer entfernt wurde, gaben funnel plot und Egger-Test keine Hinweise auf einen Publikationsbias. Aus diesem Grund wurden keine Studien durch eine trim-and-fill Analyse imputiert. Demzufolge gab es in Übereinstimmung mit der Hypothese 1.a einen kurzfristigen Transfereffekt von AG-Trainings auf das visuell-räumliche AG.

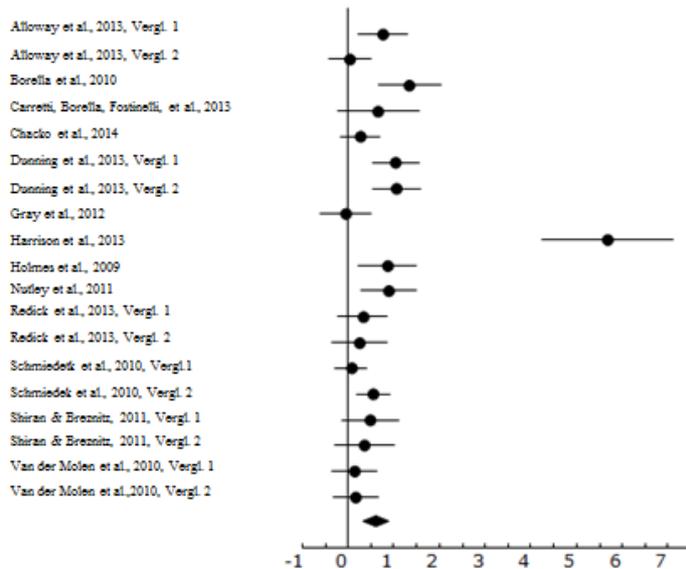


Abbildung 18. Forest plot für kurzfristige Transfereffekte auf das visuell-räumliche AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Hinsichtlich langfristiger Transfereffekte ist ein forest plot für 6 Effektstärken für den Vergleich von Zuwächsen im visuell-räumlichen AG zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up in Abbildung 19 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf das visuell-räumliche AG war klein bis moderat ($g = 0.41$, 95% KI [0.00, 0.81], $p < .05$). Der Follow-up fand im Mittel 6.83 Monate nach dem Posttest statt. Die Effektstärken variierten signifikant zwischen den Studien, $Q(5) = 13.11$, $p < .05$, $I^2 = 61.87\%$. Nachdem ein Ausreißer entfernt wurde, betrug die Effektstärke $g = 0.21$, 95% KI [-0.04, 0.47]. Der funnel plot gab einen Hinweis auf einen potentiellen Bias, wobei Studien auf der linken Seite des Mittelwerts fehlten. Der Egger-Test deutete auch auf einen Publikationsbias hin, $b = 7.06$, 95% KI [0.87, 13.26], $p < .05$. Nachdem der einzelne Ausreißer entfernt wurde, gaben funnel plot und Egger-Test keine Hinweise auf einen Publikationsbias. Aus diesem Grund wurden keine Studien durch eine trim-and-fill Analyse imputiert. Demzufolge gab es in Übereinstimmung mit der Hypothese 1.c einen langfristigen Transfereffekt von AG-Trainings auf das visuell-räumliche AG.

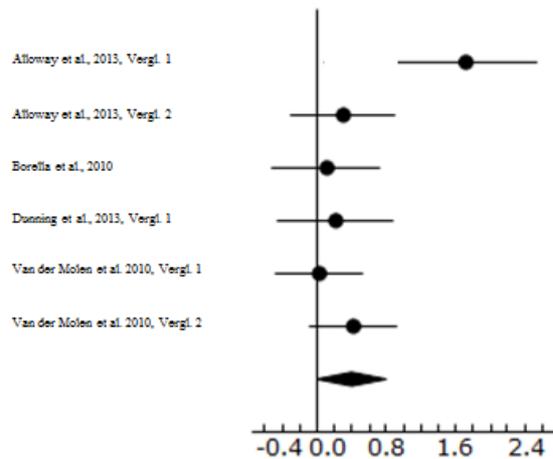


Abbildung 19. Forest plot für langfristige Transfereffekte auf das visuell-räumliche AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Nonverbale Fähigkeit. Es wurde erwartet, dass es einen Transfer von AG-Trainings auf die nonverbale Fähigkeit zum Posttest (Hypothese 2.a), jedoch nicht zum Follow-up gibt (Hypothese 2.a). Ein forest plot für 45 Effektstärken für den Vergleich von Zuwächsen in der nonverbalen Fähigkeit zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 20 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf die nonverbale Fähigkeit war klein ($g = 0.14$, 95% KI [0.01, 0.27], $p < .05$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(44) = 95.44$, $p < .001$, $I^2 = 53.90\%$. Nachdem die Effektstärken der Studien von Klingberg et al. (2002) und Schmiedek et al. (2010) entfernt wurden, war der Effekt klein und nicht signifikant ($g = 0.01$, 95% KI [-0.03, 0.23], $p = .13$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(41) = 79.97$, $p < .001$, $I^2 = 48.73\%$. Nachdem die Effektstärke der Studie von Harrison et al. (2013) entfernt wurde (in dieser Studie verbesserte sich die Trainingsgruppe nicht, aber die Kontrollgruppe verschlechterte sich), war der mittlere Effekt mit allen anderen Studien weiterhin klein und nicht signifikant ($g = 0.13$, KI [-0.01, 0.26], $p = .06$). Wenn die Effektstärken der Studien von Klingberg et al. (2002) und Schmiedek et al. (2010) zusätzlich entfernt wurden, war der mittlere Effekt noch kleiner ($g = 0.00$, KI [-0.05, 0.21], $p = .20$). Der zusätzliche Ausschluss der Effektstärke der Studie von Richmond et al. (2011) resultierte in einem kleinen mittleren Effekt ($g = 0.12$, 95% KI [0.01, 0.22], $p < .05$). Nach der Entfernung von Ausreißern reichte die mittlere Effektstärke von $g = 0.08$, KI [-0.04, 0.20] bis $g = 0.17$, KI [0.06, 0.28]. Nachdem Ausreißer und die Effektstärken der Studien von Klingberg et al. (2002) und Schmiedek et al. (2010) entfernt wurden, war der mittlere Effekt klein und nicht signifikant ($g = 0.08$, KI [-0.01, 0.18], $p = .08$). Funnel plot und Egger-Test lieferten keine Hinweise auf

einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.a einen kurzfristigen, kleinen Transfereffekt von AG-Trainings auf die nonverbale Fähigkeit.

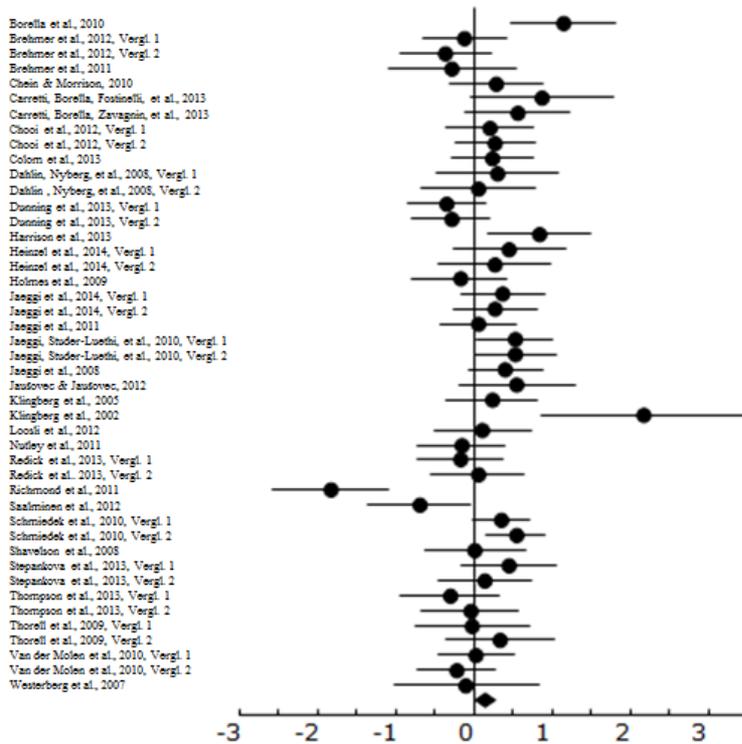


Abbildung 20. Forest plot für kurzfristige Transfereffekte auf die nonverbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Im Hinblick auf langfristige Transfereffekte wurde erwartet, dass es keinen Transfer von AG-Trainings auf die nonverbale Fähigkeit zum Posttest gibt (Hypothese 2.a). Ein forest plot für 11 Effektstärken für den Vergleich von Zuwächsen in der nonverbalen Fähigkeit zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up ist in Abbildung 21 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf die nonverbale Fähigkeit war klein und nicht signifikant ($g = 0.02$, 95% KI [-0.17, 0.20], $p = .85$). Der Follow-up fand im Mittel 6.54 Monate nach dem Posttest statt. Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(10) = 10.14$, $p = .43$. Nach der Entfernung von Ausreißern betrug die mittlere Effektstärke $g = -0.12$, KI [-0.32, 0.09]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.a keinen langfristigen Transfereffekt von AG-Trainings auf die nonverbale Fähigkeit.

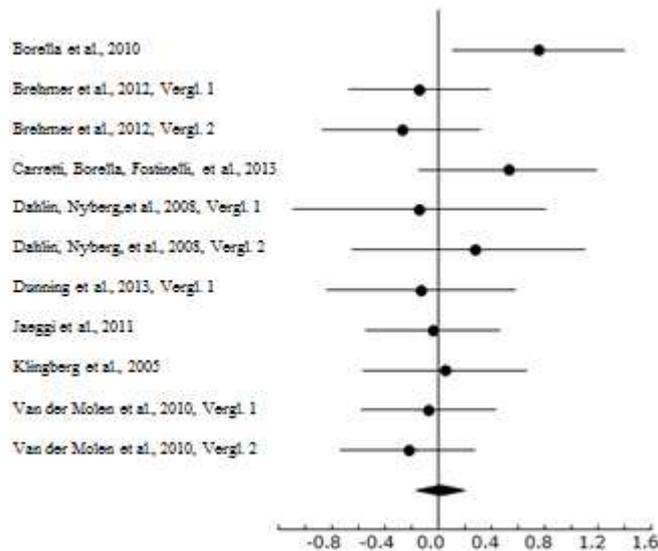


Abbildung 21. Forest plot für langfristige Transfereffekte auf die nonverbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Verbale Fähigkeit. Es wurde erwartet, dass es keinen Transfer von AG-Trainings auf die verbale Fähigkeit gibt (Hypothese 2.b). Ein forest plot für 29 Effektstärken für den Vergleich von Zuwächsen in der verbalen Fähigkeit zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 22 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf die verbale Fähigkeit war klein ($g = 0.16$, 95% KI [0.05, 0.27], $p < .01$). Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(28) = 30.69$, $p = .33$. Nachdem die Effektstärken der Studie von Schmiedek et al. (2010) entfernt wurden, war der mittlere Transfereffekt klein ($g = 0.18$, 95% KI [0.06, 0.31], $p < .05$). Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(26) = 29.77$, $p = .28$. Nach der Entfernung von Ausreißern reichte die mittlere Effektstärke von $g = 0.09$, KI [-0.02, 0.19] bis $g = 0.20$, KI [0.08, 0.31]. Der funnel plot gab einen Hinweis auf einen potentiellen Bias, wobei Studien auf der linken Seite des Mittelwerts fehlten. Der Egger-Test deutete auch auf einen Publikationsbias hin, $b = -0.13$, KI [-0.24, -0.01], $p < .05$. Nachdem eine Studie in der trim-and-fill Analyse imputiert wurde, betrug die mittlere Effektstärke $g = 0.14$, KI [0.04, 0.25], $p < .05$. Demzufolge gab es im Gegensatz zur Hypothese 2.b einen kurzfristigen Transfereffekt von AG-Trainings auf die verbale Fähigkeit.

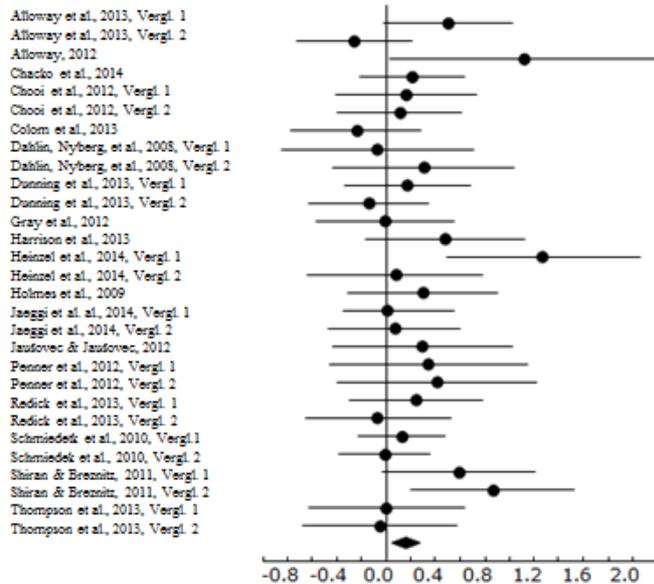


Abbildung 22. Forest plot für kurzfristige Transfereffekte auf die verbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Ein forest plot für 5 Effektstärken für den Vergleich von Zuwächsen in der verbalen Fähigkeit zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up ist in Abbildung 23 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf die verbale Fähigkeit war klein und nicht signifikant ($g = 0.26$, 95% KI $[-0.33, 0.86]$, $p = .39$). Der Follow-up fand im Mittel 12.8 Monate nach dem Posttest statt. Die Effektstärken variierten signifikant zwischen den Studien, $Q(4) = 11.71$, $p < .05$, $I^2 = 65.84\%$. Nachdem ein Ausreißer entfernt wurde, betrug die mittlere Effektstärke $g = -0.06$, KI $[-0.44, 0.32]$. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.b keinen langfristigen Transfereffekt von AG-Trainings auf die verbale Fähigkeit.

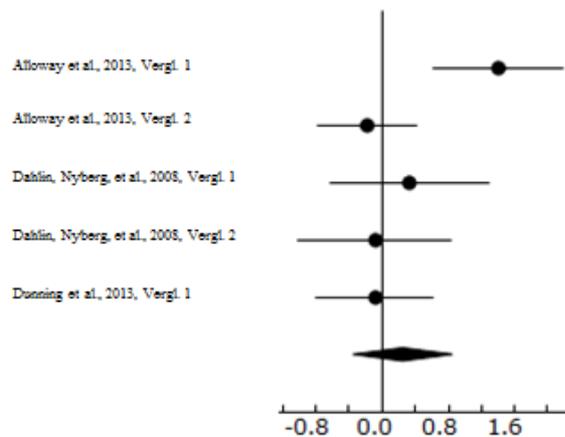


Abbildung 23. Forest plot für langfristige Transfereffekte auf die verbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Fertigkeit zur Wortdekodierung. Es wurde erwartet, dass es keinen Transfer von AG-Trainings auf die Fertigkeit zur Wortdekodierung gibt (Hypothese 2.b). Ein forest plot für 14 Effektstärken für den Vergleich von Zuwächsen in der Fertigkeit zur Wortdekodierung zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 24 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf die Fertigkeit zur Wortdekodierung war klein und nicht signifikant ($g = 0.08$, 95% KI [-0.06, 0.23], $p = .25$). Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(13) = 8.85$, $p = .78$. Nachdem Ausreißer entfernt wurden, reichte die mittlere Effektstärke von $g = 0.04$, KI [-0.11, 0.19] bis $g = 0.09$, KI [-0.05, 0.24]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.b keinen kurzfristigen Transfereffekt von AG-Trainings auf die Fertigkeit zur Wortdekodierung.

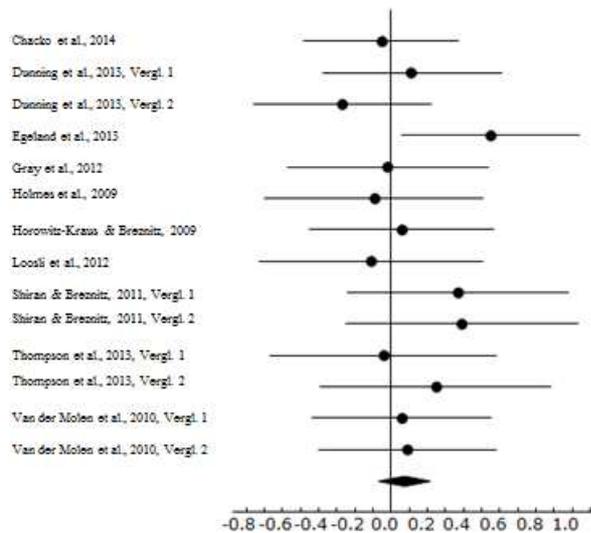


Abbildung 24. Forest plot für kurzfristige Transfereffekte auf die Fertigkeit zur Wortdekodierung. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Hinsichtlich langfristiger Transfereffekte ist ein forest plot für 5 Effektstärken für den Vergleich von Zuwächsen in der Fertigkeit zur Wortdekodierung zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up in Abbildung 25 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf die Fertigkeit zur Wortdekodierung war klein und nicht signifikant ($g = 0.21$, 95% KI $[-0.22, 0.45]$, $p = .08$). Der Follow-up fand im Mittel 6.2 Monate nach dem Posttest statt. Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(4) = 3.52$, $p = .47$. Nachdem ein Ausreißer entfernt wurde, betrug die mittlere Effektstärke $g = 0.09$, KI $[-0.18, 0.36]$. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.b keinen langfristigen Transfereffekt von AG-Trainings auf die Fertigkeit zur Wortdekodierung.

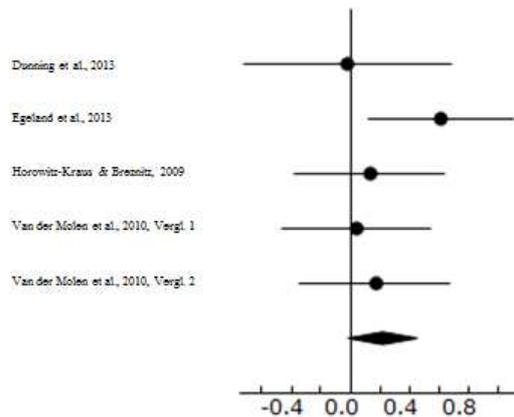


Abbildung 25. Forest plot für langfristige Transfereffekte auf die Fertigkeit zur Wortdekodierung. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Mathematische Fähigkeiten und Fertigkeiten. Es wurde erwartet, dass es keinen Transfer von AG-Trainings auf mathematische Fähigkeiten und Fertigkeiten gibt (Hypothese 2.b). Ein forest plot für 15 Effektstärken für den Vergleich von Zuwächsen in mathematischen Fähigkeiten und Fertigkeiten zwischen AG-Trainingsgruppen und Kontrollgruppen von Pre- zu Posttest ist in Abbildung 26 dargestellt. Der kurzfristige, mittlere Transfereffekt von AG-Trainings auf mathematische Fähigkeiten und Fertigkeiten war klein und nicht signifikant ($g = 0.09$, 95% KI [-0.09, 0.27], $p = .34$). Die Effektstärken variierten signifikant zwischen den Studien, $Q(14) = 27.81$, $p < .05$, $I^2 = 49.65\%$. Nachdem Ausreißer entfernt wurden, reichte die mittlere Effektstärke von $g = 0.04$, KI [-0.11, 0.20] bis $g = 0.13$, KI [-0.03, 0.30]. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.b keinen kurzfristigen Transfereffekt von AG-Trainings auf mathematische Fähigkeiten und Fertigkeiten.

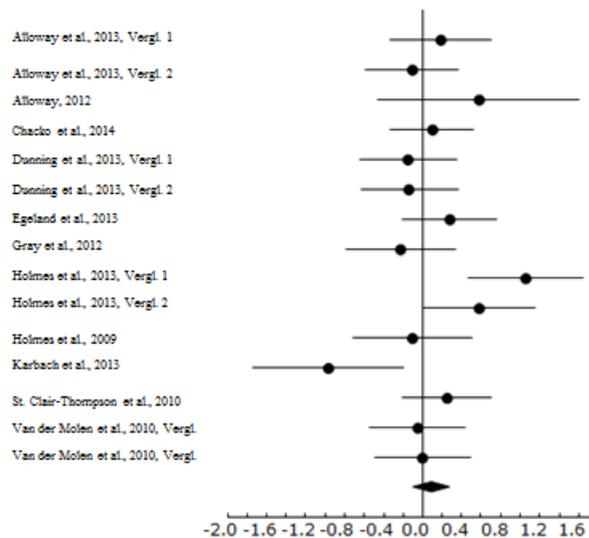


Abbildung 26. Forest plot für kurzfristige Transfereffekte auf mathematische Fähigkeiten und Fertigkeiten. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

Ein forest plot für 8 Effektstärken für den Vergleich von Zuwächsen in mathematischen Fähigkeiten und Fertigkeiten zwischen AG-Trainingsgruppen und Kontrollgruppen von Pretest zu Follow-up ist in Abbildung 27 dargestellt. Der langfristige, mittlere Transfereffekt von AG-Trainings auf mathematische Fähigkeiten und Fertigkeiten war klein und nicht signifikant ($g = 0.08$, 95% KI $[-0.12, 0.28]$, $p = .44$). Der Follow-up fand im Mittel 6.13 Monate nach dem Posttest statt. Die Effektstärken variierten nicht signifikant zwischen den Studien, $Q(7) = 3.22$, $p = .86$. Nachdem Ausreißer entfernt wurden, reichte die mittlere Effektstärke von $g = 0.05$, KI $[-0.15, 0.26]$ bis $g = 0.11$, KI $[-0.10, 0.32]$. Funnel plot und Egger-Test lieferten keine Hinweise auf einen Publikationsbias. Demzufolge gab es in Übereinstimmung mit der Hypothese 2.b keinen langfristigen Transfereffekt von AG-Trainings auf mathematische Fähigkeiten und Fertigkeiten.

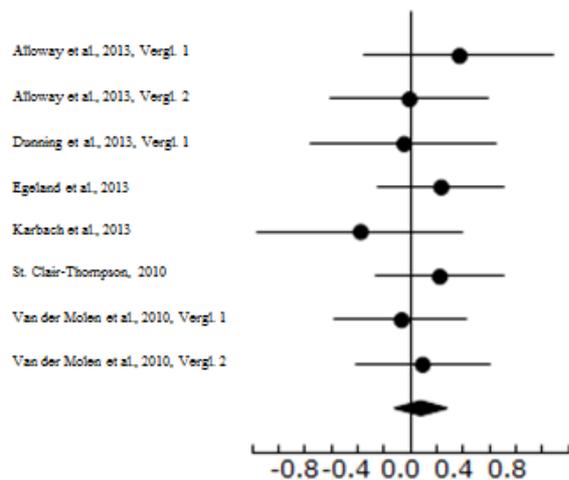


Abbildung 27. Forest plot für langfristige Transfereffekte auf mathematische Fähigkeiten und Fertigkeiten. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.

6.3.2 Moderatoren von Transfereffekten (Frage 3)

Um zu untersuchen, welche Trainingsbedingungen einen Einfluss auf Transfereffekte von AG-Trainings haben (Frage 3), wurden Moderatoranalysen für folgende Transfereffekte durchgeführt: kurzfristige Transfereffekte auf verbales KZG und verbales AG, kurzfristige Transfereffekte auf visuell-räumliches KZG und visuell-räumliches AG und kurzfristige Transfereffekte auf die nonverbale Fähigkeit und mathematische Fähigkeiten und Fertigkeiten. Für die anderen Transfereffekte mit einer signifikanten Q -Statistik waren nicht genügend Studien vorhanden, um Moderatoranalysen sinnvoll durchzuführen. Die untersuchten Moderatoren und die erklärte Varianz in der Heterogenität zwischen Effektstärken für nahe und weite Transfereffekte sind in Tabelle 9 dargestellt. Darauf folgen detaillierte Analysen für jeden Moderator in der Reihenfolge der anfänglichen Hypothesen. Die in der Tabelle 9 dargestellten Effektstärken werden hinsichtlich Signifikanz und Effektstärkekonventionen für η^2 und R^2 klassifiziert. Für die Beurteilung der Größe von η^2 wurden die Richtlinien von Murphy und Myers (2004) verwendet: $\eta^2 = .01$ (kleiner Effekt), $\eta^2 = .06$ (mittlerer Effekt), $\eta^2 = .14$ (großer Effekt). Für die Beurteilung der Größe von R^2 wurden folgende Konventionen verwendet (Murphy & Myers, 2004): $R^2 = .01$ (kleiner Effekt), $R^2 = .06$ (mittlerer Effekt) und $R^2 = .15$ (großer Effekt).

Diese Konventionen wurden zusätzlich zur Signifikanz der Ergebnisse betrachtet, da Effektstärken aufgrund ungenügender Teststärke nicht signifikant sein können.

Tabelle 9

Effektstärken η^2 und R^2 für Moderatoren von nahen und weiten Transfereffekten

	Ver- bales KZG	Visuell- räumli- ches KZG	Ver- bales AG	Visuell- räumli- ches AG	Non- verbale Fähig- keit	Mathema- tische Fä- higkeiten und Fer- tigkeiten
Alter	.00	.00	.04	.00	.01	.00
Interventions- typ	.01	.28**	.18*	.00	.12*	.01
Art der Kon- trollgruppe	.00	.09	.06	.02	.02	.24*
Trainingsdo- sis	.01	.30**	.01	.00	.00	.05
Sessiondauer	.10*	.05	.01	.04	.07	.03
Häufigkeit des Trainings pro Woche	.04	.00	.04	.06	.10	.11
Trainingsin- tervall	.02	-	.11	.00	.09	-
Modalität	.02	.01	.01	.03	.11	.15
Supervision	.06	.17	.17**	.16*	.01	.06
Instruktionale Unterstüt- zung	.01	.11	.01	.02	.00	.00
Feedback	.00	.05	.00	.00	.02	.11
Trainingsort	.08	.24*	.19**	.04	.20**	-

Anmerkungen. Die Effektstärke η^2 wurde für kategoriale Moderatoren berechnet. Die Effektstärke R^2 wurde für kontinuierliche Moderatoren berechnet. Gedankenstriche zeigen an, dass eine Moderatoranalyse aufgrund ungenügender Daten nicht möglich war.

* $p < .05$. ** $p < .01$.

Alter. Es wurde erwartet, dass das Alter ein Moderator von Transfereffekten auf das verbale AG ist (Hypothese 3.a). Jüngere Teilnehmer sollten im Vergleich zu älteren Teilnehmern größere Trainingsverbesserungen zeigen. Im Gegensatz zur Hypothese 3.a war das Alter von keinem Transfereffekt von AG-Trainings ein signifikanter Moderator. Die Werte für die Effektstärken (R^2) waren klein und nahe null.

Interventionstyp. Es wurde erwartet, dass der Interventionstyp ein Moderator von Transfereffekten auf das visuell-räumliche AG ist. Es wurde angenommen, dass die Effekte des Cogmed Trainingsprogramms auf das visuell-räumliche AG größer sind als die Effekte anderer kommerzieller und nichtkommerzieller Trainingsprogramme. Es sollte keinen Unterschied zwischen den kommerziellen Programmen geben (Hypothese 3.b). Wie aus Tabelle 9 hervorgeht, erklärte der Interventionstyp Variabilität in Transfereffekten auf visuell-

räumliches KZG, verbales AG und nonverbale Fähigkeit. Der Interventionstyp war ein signifikanter Moderator von Transfereffekten auf das visuell-räumliche KZG, $Q(1) = 9.68$, $p < .01$. Eta^2 war groß (.28). Die Residualvarianz war nicht signifikant ($p = .33$), d.h., nachdem der Interventionstyp als Moderator für Transfereffekte auf das visuell-räumliche KZG berücksichtigt wurde, blieb keine signifikante Variabilität in den Effektstärken zwischen Studien übrig. Cogmed Training als ein kommerzielles Trainingsprogramm erzielte eine größere, mittlere Effektstärke als nichtkommerzielle Trainingsprogramme. Da mehrere Moderatoren in einzelnen Regressionsanalysen jeder Variable signifikant wurden, wurde eine multiple Regression mit Interventionstyp, dekadischem Logarithmus der Trainingsdosis und zwei dummy-codierten Variablen für den Trainingsort als Prädiktoren für Transfereffekte auf das visuell-räumliche KZG durchgeführt. Das Regressionsmodell war signifikant, $Q(4) = 23.49$, $p < .001$, k (Anzahl an Gruppenvergleichen) = 23, $R^2 = .55$. Die Residualvarianz war nicht signifikant ($p = .38$). Keine einzelne Variable war ein signifikanter Prädiktor für Transfereffekte auf das visuell-räumliche KZG unter Kontrolle der anderen Prädiktoren. Allerdings waren der Interventionstyp (die Art des Trainingsprogramms; in diesem Fall Cogmed Training vs. anderes Training) und die Trainingsdosis korreliert, $r_{pb} = -.48$, $p < .05$.

Der Interventionstyp war zudem ein signifikanter Moderator von Transfereffekten auf das verbale AG, $Q(3) = 10.58$, $p < .05$. Eta^2 war groß (.18). Die Residualvarianz war nicht signifikant ($p = .11$). Paarweise Vergleiche zeigten, dass Cogmed Training einen größeren, mittleren Effekt auf das verbale AG erzielte als n-back Training, $Q(1) = 15.02$, $p < .001$. Jungle memory Training erzielte nur ohne Bonferroni-Korrektur eine größere, mittlere Effektstärke als n-back Training, $Q(1) = 5.94$, $p = .015$. Da mehrere Moderatoren in einzelnen Analysen signifikant waren, wurde eine multiple Regression mit 7 dummy-codierten Variablen (3 für den Interventionstyp, je 2 für Supervision und Trainingsort) als Prädiktoren für Transfereffekte auf das verbale AG durchgeführt. Das Regressionsmodell war signifikant, $Q(7) = 25.22$, $p < .001$, $k = 38$, $R^2 = .39$. Die Residualvarianz war nicht signifikant ($p = .11$). Keine dummy-codierte Variable für den Interventionstyp war unter Kontrolle der anderen Prädiktoren signifikant. Der Interventionstyp und der Trainingsort waren hoch korreliert, Cramer's $V = .59$, $p < .001$.

Der Interventionstyp war des Weiteren ein signifikanter Moderator von Transfereffekten auf die nonverbale Fähigkeit, $Q(2) = 6.44$, $p < .05$. Eta^2 war mittel (.12). Die Residualvarianz war nicht signifikant ($p = .23$). Paarweise Vergleiche zeigten, dass Cogmed Training einen kleineren, mittleren Effekt als n-back Training erzielte, $Q(1) = 8.02$, $p < .01$. Da mehrere Moderatoren in einzelnen Regressionen signifikant wurden, wurde eine multiple Regressi-

on mit 4 dummy-codierten Variablen (jeweils 2 für Interventionstyp und Trainingsort) als Prädiktoren für Transfereffekte auf die nonverbale Fähigkeit durchgeführt. Das Regressionsmodell war signifikant, $Q(4) = 11.66, p < .05, k = 40, R^2 = .23$. Die Residualvarianz war nicht signifikant ($p = .30$). Kein einzelner Prädiktor war unter Kontrolle der anderen Prädiktoren signifikant. Der Interventionstyp und der Trainingsort waren korreliert, Cramer's $V = .44, p < .01$.

Zusammenfassend war der Interventionstyp ein signifikanter Moderator für (nahe) Transfereffekte auf das visuell-räumliche KZG, das verbale AG und die nonverbale Fähigkeit. Die Werte der Effektstärken (η^2) waren groß für Transfereffekte auf das verbale AG und das visuell-räumliche KZG und mittel für den Transfereffekt auf die nonverbale Fähigkeit. Im Gegensatz zur Hypothese 3.b war der Interventionstyp kein signifikanter Moderator von Transfereffekten auf das visuell-räumliche AG. Cogmed Training erzielte verglichen mit anderen kommerziellen und nichtkommerziellen Trainingsprogrammen keinen größeren, mittleren Effekt. Die (nichtsignifikanten) Effektstärken (η^2) waren gering und nahe null für das visuell-räumliche AG und die übrigen Konstrukte.

Art der Kontrollgruppe. Es wurde erwartet, dass die Art der Kontrollgruppe ein Moderator von Transfereffekten auf die nonverbale Fähigkeit ist. Der mittlere Effekt auf die nonverbale Fähigkeit für den Vergleich von Trainingsgruppen mit passiven Kontrollgruppen sollte größer sein als für den Vergleich von Trainingsgruppen mit aktiven Kontrollgruppen (Hypothese 3.c). Wie aus Tabelle 9 hervorgeht, erklärte die Art der Kontrollgruppe Variabilität in Transfereffekten auf mathematische Fähigkeiten und Fertigkeiten. Die Art der Kontrollgruppe war der einzige signifikante Moderator von Transfereffekten auf mathematische Fähigkeiten und Fertigkeiten, $Q(1) = 4.58, p < .05$. η^2 war groß (.24). Die Residualvarianz war nicht signifikant ($p = .32$). Die mittlere Effektstärke für Vergleiche von Trainingsgruppen mit passiven Kontrollgruppen war signifikant größer als für Vergleiche von Trainingsgruppen mit aktiven Kontrollgruppen. Im Gegensatz zur Hypothese 3.c war die Art der Kontrollgruppe kein signifikanter Moderator von Transfereffekten auf die nonverbale Fähigkeit. Die Werte für die Effektstärken (η^2) waren gering bis mittel für Transfereffekte, auf die die Art der Kontrollgruppe keinen signifikanten Einfluss hatte.

Trainingsdosis. Es wurde erwartet, dass die Trainingsdosis kein Moderator von Transfereffekten von AG-Trainings ist (Hypothese 3.d). Um eine konservative Hypothesentestung zu gewährleisten, wurde ein Alpha-Level von .20 angewendet, um die Hypothese leichter abzulehnen. Wie aus Tabelle 9 hervorgeht, erklärte die Trainingsdosis konträr zur Hypothese 3.d Variabilität in den Transfereffekten auf das visuell-räumliche KZG, $\beta = .55, p < .01$,

$k = 25$, $R^2 = .30$ (großer Effekt). Die Residualvarianz war nicht signifikant ($p = .41$). Trainings mit einer höheren Trainingsdosis (angezeigt durch den dekadischen Logarithmus der Trainingsdosis) erzielten einen größeren, mittleren Effekt auf das visuell-räumliche KZG als Trainings mit einem geringeren dekadischen Logarithmus der Trainingsdosis. Der dekadische Logarithmus der Trainingsdosis war kein signifikanter Moderator von Transfereffekten auf das visuell-räumliche KZG unter Kontrolle des Interventionstyp und zwei dummy-codierten Variablen für den Trainingsort.

Sessiondauer. Es wurde erwartet, dass die Dauer einzelner Sessions ein Moderator von Transfereffekten von AG-Trainings ist. Je höher die Sessiondauer, desto größer sollten die Effektstärken sein (Hypothese 4.a). Wie aus Tabelle 9 hervorgeht, erklärte die Sessiondauer Variabilität in den Transfereffekten auf das verbale KZG. Die Sessiondauer war der einzige signifikante Moderator von Transfereffekten auf das verbale KZG, $\beta = .32$, $p < .05$, $k = 32$, $R^2 = .10$ (mittlerer Effekt). Trainings mit längeren Trainingssessions erzielten größere Effekte als Trainings mit kürzeren Trainingssessions. Die Residualvarianz war nicht signifikant ($p = .17$). Im Gegensatz zur Hypothese 4.a war die Sessiondauer kein signifikanter Moderator von anderen Transfereffekten von AG-Trainings. Die Werte der Effektstärken (R^2) waren gering bis mittel für Transfereffekte, auf die die Sessiondauer keinen signifikanten Einfluss hatte.

Häufigkeit des Trainings pro Woche. Es wurde erwartet, dass die Häufigkeit des Trainings pro Woche ein Moderator von Transfereffekten von AG-Trainings ist. Je mehr Trainingssessions pro Woche, desto größer sollten die Effektstärken sein (Hypothese 4.b). Im Gegensatz zur Hypothese 4.b war die Häufigkeit des Trainings pro Woche für keinen Transfereffekt von AG-Trainings ein signifikanter Moderator. Die Werte der Effektstärken (η^2) waren gering bis mittel.

Trainingsintervall. Es wurde erwartet, dass das Trainingsintervall ein Moderator von Transfereffekten von AG-Trainings ist. Ein Trainingsintervall von 2 Tagen sollte zu größeren Effekten führen als ein Trainingsintervall von 1 Tag (Hypothese 4.c). Im Gegensatz zur Hypothese 4.c war das Trainingsintervall von keinem Transfereffekt von AG-Trainings ein signifikanter Moderator. Die Werte der Effektstärken (η^2) waren klein bis mittel.

Modalität. Es wurde erwartet, dass die trainierte Modalität ein Moderator von Transfereffekten von AG-Trainings ist. Transfereffekte auf verbales KZG und AG sollten für Interventionen größer sein, in denen das verbale AG trainiert wurde. Transfereffekte auf visuell-räumliches KZG und AG sollten für Interventionen größer sein, in denen das visuell-räumliche AG trainiert wurde. Es gab keine *gerichtete* Hypothese hinsichtlich des moderie-

renden Einflusses der trainierten Modalität auf die Variabilität zwischen Effektstärken von weiten Transfereffekten (Hypothese 5.a). Im Gegensatz zur Hypothese 5.a war die trainierte Modalität von keinem Transfereffekt von AG-Trainings ein signifikanter Moderator. Die Werte für die Effektstärken (η^2) waren allerdings klein bis groß.

Supervision. Es wurde davon ausgegangen, dass Supervision ein Moderator von Transfereffekten von AG-Trainings ist. Transfereffekte bei supervidierten Trainings sollten größer sein als bei Trainings ohne die Anwesenheit von Personen sowie Trainings in der bloßen Anwesenheit anderer Personen. Transfereffekte bei Trainings ohne die Anwesenheit anderer Personen sollten größer sein als bei Trainings in der bloßen Anwesenheit anderer Personen (Hypothese 5.b). Wie aus Tabelle 9 hervorgeht, erklärte Supervision Variabilität in Transfereffekten auf verbales und visuell-räumliches AG. Supervision war ein signifikanter Moderator von Transfereffekten auf das verbale AG, $Q(2) = 9.53, p < .01$. η^2 war groß (.17). Die Residualvarianz war nicht signifikant ($p = .20$). Paarweise Vergleiche zeigten, dass supervidiertes Training einen größeren, mittleren Transfereffekt erzielte als Training in der bloßen Anwesenheit anderer Personen, $Q(1) = 7.75, p < .01$. In der multiplen Regression mit Interventionstyp, Supervision und Trainingsort als Prädiktoren von Transfereffekten auf das verbale AG erzielte das Training in der bloßen Anwesenheit anderer Personen kleinere Effekte als das Training ohne die Anwesenheit anderer Personen (Referenzkategorie), $\beta = -.35, p < .05$.

Supervision war der einzige signifikante Moderator von Transfereffekten auf das visuell-räumliche AG, $Q(2) = 6.53, p < .05$. η^2 war groß (.16). Die Residualvarianz war signifikant ($p < .01$). Paarweise Vergleiche zeigten, dass die mittlere Effektstärke für supervidiertes Training größer war als die mittlere Effektstärke für das Training in der bloßen Anwesenheit anderer Personen auf einem Alpha-Niveau von .05, $Q(1) = 4.66, p = .03$. Der Unterschied zwischen diesen beiden Gruppen war jedoch bei einer Bonferroni-Korrektur des Alpha-Niveaus (korrigiertes Alpha-Level von .017) nicht signifikant. Daher konnte die Hypothese 5.b nur teilweise bestätigt werden. Die Werte der Effektstärken (η^2) waren klein bis mittel für weite Transfereffekte, auf die Supervision keinen Einfluss hatte.

Instruktionale Unterstützung. Es wurde davon ausgegangen, dass die instruktionale Unterstützung ein Moderator für Transfereffekte von AG-Trainings ist. Transfereffekte für Trainings mit zusätzlichen Erklärungen über die zu Beginn des Trainings gegebenen Erklärungen hinaus sollten größer sein als für Trainings ohne diese zusätzlichen Erklärungen (Hypothese 5.c). Im Gegensatz zur Hypothese 5.c war die instruktionale Unterstützung von keinem Transfereffekt von AG-Trainings ein signifikanter Moderator. Die Werte für die Effektstärken (η^2) waren klein bis mittel.

Feedback. Es wurde davon ausgegangen, dass das Feedback ein Moderator für Transfereffekte von AG-Trainings ist. Trainings mit Feedback über reines Ergebniswissen hinaus sollten größere Effektstärken erzielen als Trainings, die nur Feedback in Form reinen Ergebniswissens beinhalten (Hypothese 5.d). Im Gegensatz zur Hypothese 5.d war das Feedback für keinen Transfereffekt von AG-Trainings ein signifikanter Moderator. Die Werte für die Effektstärken (η^2) waren klein bis groß.

Trainingsort. Es wurde davon ausgegangen, dass der Trainingsort ein Moderator für Transfereffekte von AG-Trainings ist. Transfereffekte von Trainings im Labor sollten größer sein als Transfereffekte in der Schule oder zu Hause (Hypothese 5.e). Der Trainingsort war ein signifikanter Moderator für Transfereffekte auf das visuell-räumliche KZG, $Q(2) = 6.93$, $p < .05$. η^2 war groß ($\eta^2 = .24$). Die Residualvarianz war nicht signifikant ($p = .37$). Paarweise Vergleiche zeigten, dass das Training zu Hause einen größeren, mittleren Transfereffekt als das Training in der Schule erzielte, $Q(1) = 6.16$, $p = .013$. In der multiplen Regression mit Interventionstyp, Supervision und Trainingsort als Prädiktoren von Transfereffekten auf das visuell-räumliche KZG war keine dummy-codierte Variable für den Trainingsort ein signifikanter Moderator unter Kontrolle der anderen Prädiktoren.

Der Trainingsort war zudem ein signifikanter Moderator für Transfereffekte auf das verbale AG, $Q(2) = 12.14$, $p < .01$. η^2 war groß (.19). Die Residualvarianz war signifikant ($p < .05$). Paarweise Vergleiche zeigten, dass das Training in der Schule eine größere, mittlere Effektstärke erzielte als das Training im Labor, $Q(1) = 13.40$, $p < .001$. In der multiplen Regression mit Interventionstyp, Supervision und Trainingsort als Prädiktoren von Transfereffekten auf das verbale AG war keine dummy-codierte Variable für den Trainingsort ein signifikanter Moderator unter Kontrolle der anderen Prädiktoren.

Zusätzlich war der Trainingsort ein signifikanter Moderator für Transfereffekte auf die nonverbale Fähigkeit, $Q(2) = 10.04$, $p < .01$. η^2 war groß (.20). Die Residualvarianz war nicht signifikant ($p = .31$). Paarweise Vergleiche zeigten, dass das Training im Labor eine größere, mittlere Effektstärke erzielte als das Training in der Schule, $Q(1) = 8.13$, $p < .01$, sowie das Training zu Hause, $Q(1) = 6.43$, $p = .01$. In der multiplen Regression mit Interventionstyp und Trainingsort als Prädiktoren von Transfereffekten auf die nonverbale Fähigkeit war der Trainingsort kein signifikanter Moderator unter Kontrolle des Interventionstyps.

Zusammenfassend war der Trainingsort ein signifikanter Moderator für Transfereffekte auf visuell-räumliches KZG, verbales AG und die nonverbale Fähigkeit. Das Einflussmuster des Trainingsorts auf Transfereffekte war für diese Konstrukte inkonsistent. Training im Labor war nur für Transfereffekte auf die nonverbale Fähigkeit besser als Training in der

Schule und Training zu Hause. Folglich gab es nur teilweise Evidenz für die Hypothese 5.e. Die Werte für die Effektstärken (η^2) waren klein bis mittel für Transfereffekte, auf die der Trainingsort keinen signifikanten Einfluss hatte.

6.4 Diskussion

6.4.1 Zusammenfassung der Ergebnisse

Die Ergebnisse der Metaanalyse (Studie 2) zeigten, dass AG-Trainings langfristige, nahe Transfereffekte auf KZG- und AG-Komponenten erzielen. Kurzfristige, kleine Transfereffekte wurden auf die nonverbale und verbale Fähigkeit gefunden. Interventionstyp, Art der Kontrollgruppe, Trainingsdosis, Sessiondauer, Supervision und Trainingsort waren signifikante Moderatoren für verschiedene Transfereffekte.

6.4.2 Ergebnisinterpretation

Beziehungen zu bisheriger Metaanalyse

Einige der Hypothesen (siehe Abschnitt 6.1) basierten auf den Ergebnissen der Metaanalyse von Melby-Lervåg und Hulme (2013). Das Muster naher Transfereffekte (Hypothesen 1. a, b, c) ist nur teilweise mit den Ergebnissen der früheren Metaanalyse vergleichbar. Wie in der Analyse von Melby-Lervåg und Hulme (2013) wurden kurzfristige Transfereffekte von AG-Trainings auf das verbale AG gefunden. Allerdings war der Gesamteffekt in der bisherigen Metaanalyse etwas größer als der in der vorliegenden Analyse erhaltene Effekt ($d_s = 0.79$, verglichen mit $g = 0.55$ in der vorliegenden Analyse). Eine mögliche Erklärung für diesen Unterschied könnte sein, dass im Gegensatz zur Metaanalyse von Melby-Lervåg und Hulme (2013) in der vorliegenden Analyse zwischen Transfereffekten auf das verbale KZG und Transfereffekten auf das verbale AG unterschieden wurde. Da diese beiden Transfereffekte jedoch kleiner als der Transfereffekt von Melby-Lervåg und Hulme (2013) waren, ist es wahrscheinlicher, dass die Effekte auf verbales KZG und verbales AG in der Population kleiner sind als der von Melby-Lervåg und Hulme (2013) gefundene Transfereffekt auf das verbale AG. Aufgrund größerer Stichproben in der vorliegenden Studie sind die Schätzungen für die Transfereffekte genauer als in der Analyse von Melby-Lervåg und Hulme (2013).

Die Transfereffekte auf verbales KZG und AG waren unterschiedlich, was durch eine Überlappung zwischen trainierten AG-Aufgaben und AG-Aufgaben zur Erfassung von nahen Transfereffekten erklärt werden kann (siehe auch Harrison et al., 2013). Falls trainierte AG-

Aufgaben und AG-Aufgaben zur Erfassung von Transfereffekten Gemeinsamkeiten aufweisen (z.B. ähnliche Operationen zur Lösung der Aufgabe), könnten die Effekte auf AG-Aufgaben größer als auf KZG-Aufgaben (oder andere Transferaufgaben) sein. Beispielsweise trainierten Teilnehmer in der Studie von Richmond et al. (2011) mit einer verbalen und einer räumlichen CSA. Ein Transfer zeigte sich auf die reading span Aufgabe (eine verbale CSA, die der trainierten Aufgabe ähnlich ist), aber nicht auf die Aufgaben backward und forward digit span (KZG-Aufgaben). Diese Ergebnisse spiegeln sich auch in den bedeutenden Unterschieden zwischen den Effektstärken für verbales KZG und AG in dieser Studie wider.

Im Gegensatz zur Metaanalyse von Melby-Lervåg und Hulme (2013) konnte in der vorliegenden Analyse ein Transfereffekt auf das verbale AG auch noch zum Follow-up festgestellt werden. Melby-Lervåg und Hulme (2013) schlossen in ihre Analyse der langfristigen Transfereffekte auf das verbale AG 6 Effektstärken ein, in der vorliegenden Analyse waren es 11. Da beide globalen Effektstärken in ihrer Größe vergleichbar sind, kann der nichtsignifikante Transfereffekt auf das verbale AG in der Analyse von Melby-Lervåg und Hulme (2013) auf die geringere Anzahl an Effektstärken (d.h., eine geringe statistische Teststärke) in ihrer Analyse zurückgeführt werden. Die Anzahl an Effektstärken in der vorliegenden Analyse ist noch als gering zu beurteilen, weshalb der langfristige Transfereffekt auf das verbale AG mit Vorsicht zu interpretieren ist.

Hinsichtlich des visuell-räumlichen AG wurden wie in der Analyse von Melby-Lervåg und Hulme (2013) kurzfristige und langfristige Transfereffekte gefunden. Der kurzfristige Transfereffekt war etwas größer ($g = 0.63$) als in deren Analyse ($d_s = 0.52$). Wenn die Effektstärke der Studie von Harrison et al. (2013) aufgrund methodischer Erwägungen ausgeschlossen wurde (siehe Abschnitt 6.2.3), war die Effektstärke in der vorliegenden Analyse ähnlicher ($g = 0.49$). Der langfristige Transfereffekt, der auf Basis von 6 Gruppenvergleichen gefunden wurde, ist vergleichbar mit dem Transfereffekt in der Analyse von Melby-Lervåg und Hulme (2013) auf Basis von 4 Gruppenvergleichen ($d_s = 0.41$ verglichen mit $g = 0.41$ in der vorliegenden Analyse). Daher änderten sich die Ergebnisse durch Einschluss von zwei weiteren Studien nicht bedeutend. Vor dem Hintergrund der (kurz- und langfristigen) Transfereffekte auf das verbale KZG (z.B. kurzfristiger Transfereffekt $g = 0.37$) und das verbale AG (z.B. kurzfristiger Transfereffekt $g = 0.55$), ist es überraschend, dass die (kurz- und langfristigen) Transfereffekte auf das visuell-räumliche KZG (z.B. kurzfristiger Transfereffekt $g = 0.72$) deskriptiv größer sind als die Transfereffekte auf das visuell-räumliche AG (z.B. kurzfristiger Transfereffekt $g = 0.63$). Noch überraschender ist, dass die Effekte auf KZG und AG über die Modalitäten (verbal, visuell-räumlich) hinweg nicht konsistent sind. Zusammenfassend wur-

den Transfereffekte auf KZG und AG für verbale und visuell-räumliche Aufgabeninhalte gefunden. Eine plausible Erklärung könnte sein, dass das AG-Training die *executive attention* verbessert, wodurch wiederum Transfereffekte ermöglicht werden (siehe von Bastian und Oberauer, 2014). Den Ergebnissen zu Folge sind Transfereffekte auf visuell-räumliches KZG und AG größer als Transfereffekte auf verbales KZG und AG. Über eine Erklärung für diese Effekte kann nur spekuliert werden. Denkbar ist, dass sich Personen im Alltag mehr mit verbalen Informationen auseinandersetzen als mit visuell-räumlichen Informationen. Daher könnte das verbale System bereits stärker trainiert sein und Verbesserungen in diesem System sind schwieriger zu erreichen. Mögliches Resultat sind geringere Transfereffekte. Allerdings sind weitere Studien nötig, um ein besseres Verständnis für die Unterschiede in den Transfer-effekten zwischen visuell-räumlichen und verbalen KZG- und AG-Komponenten zu bekommen.

In Übereinstimmung mit den Ergebnissen von Melby-Lervåg und Hulme (2013) wurden keine weiten Transfereffekte auf die Fertigkeit zur Wortdekodierung und mathematische Fähigkeiten und Fertigkeiten gefunden. Die Anzahl an Vergleichen in der vorliegenden Analyse war mindestens doppelt so hoch wie in der Metaanalyse von Melby-Lervåg und Hulme (2013) (14 für die Fertigkeit zur Wortdekodierung und 15 für mathematische Fähigkeiten und Fertigkeiten in der vorliegenden Analyse vs. 7 für die Fertigkeit zur Wortdekodierung und 7 für arithmetische Fähigkeiten und Fertigkeiten in deren Analyse). Aus diesem Grund wäre eine einfache Interpretation, dass AG-Trainings keine Effekte auf die Fertigkeit zur Wortdekodierung und mathematische Fähigkeiten und Fertigkeiten haben. Es wurde ein kleiner, aber signifikanter, kurzfristiger Effekt auf die nonverbale Fähigkeit gefunden ($g = 0.14$), ähnlich wie der von Melby-Lervåg und Hulme (2013) gefundene Effekt ($d_s = 0.19$) aber auf der Basis von über doppelt so vielen Gruppenvergleichen ($k = 45$ in der vorliegenden Analyse vs. $k = 22$ in deren Analyse). Der von Melby-Lervåg und Hulme (2013) gefundene Effekt scheint sich tendenziell zu verringern, wenn mehrere Studien in die Analyse eingeschlossen werden. Nachdem Ausreißer ausgeschlossen wurden, war kein signifikanter Transfereffekt mehr zu beobachten. Morrison und Chein (2011) zu Folge wurden in vielen Studien zu AG-Trainings Transfereffekte nur mit einer Aufgabe erfasst. Dies trifft auch auf die, in die vorliegende Metaanalyse eingeschlossenen Studien zu. Matrizentests für schlussfolgerndes Denken (insbesondere Raven) wurden häufig als einzelne Transfermaße für die nonverbale Fähigkeit verwendet. Nur in wenigen Studien wurden mehrere Maße für die nonverbale Fähigkeit verwendet. In der Folge wäre eine weniger optimistische Interpretation der vorliegenden Ergebnisse, dass es höchstens kleine, kurzfristige Transfereffekte auf die nonverbale Fähigkeit gibt und

diese auf bestimmte Aspekte der nonverbalen Fähigkeit beschränkt sind.

Im Falle der verbalen Fähigkeit wurde ein kurzfristiger, weiter Transfereffekt gefunden, der klein war ($g = 0.16$). Melby-Lervåg und Hulme (2013) fanden einen ähnlichen, kurzfristigen Transfereffekt ($d_s = 0.13$), der allerdings nicht signifikant war. Diese Autoren schlossen nur 8 Effektstärken in ihre Analyse kurzfristiger, weiter Transfereffekte auf die verbale Fähigkeit ein. In die vorliegende Analyse wurden 29 Effektstärken eingeschlossen. Somit war der kurzfristige Transfereffekt auf die verbale Fähigkeit in der Analyse von Melby-Lervåg und Hulme (2013) möglicherweise aufgrund einer zu geringen Teststärke nicht signifikant. Nach der gleichen Logik könnte der nichtsignifikante, langfristige Effekt auf die verbale Fähigkeit ($g = 0.26$) in der vorliegenden Metaanalyse auf die kleine Anzahl an Gruppenvergleichen ($k = 5$) zurückgeführt werden.

Für beide signifikanten, weiten Transfereffekte auf nonverbale und verbale Fähigkeit sind die für den Transfer verantwortlichen Mechanismen unklar. Dieser Befund ist konkordant mit den Schlussfolgerungen, die in aktuellen Reviews gezogen werden (Morrison & Chein, 2011; Titz & Karbach, 2014; Von Bastian & Oberauer, 2013).

Melby-Lervåg und Hulme (2013) schlossen einige Trainingsbedingungen als Moderatoren in ihre Analyse ein. Die signifikanten Moderatoren in deren Analyse wurden erneut untersucht und zahlreiche, weitere Trainingsbedingungen als potentielle Moderatoren eingeschlossen. In den Ergebnissen der Moderatoranalysen der vorliegenden Metaanalyse ergaben sich größere Änderungen zu den Ergebnissen von Melby-Lervåg und Hulme (2013). Im Gegensatz zu der früheren Metaanalyse war das Alter für keinen Transfereffekt ein Moderator. In der vorliegenden Analyse wurde konträr zur bisherigen Metaanalyse das Alter nicht kategorisiert sondern der dekadische Logarithmus des Alters genutzt. Dadurch wurde ein Informationsverlust vermieden, der mit der Kategorisierung kontinuierlicher Moderatoren verbunden ist. Innerhalb der Altersspanne der, in die vorliegende Analyse eingeschlossenen Studien scheinen Transfereffekte nicht altersabhängig zu sein.

Die Art der Kontrollgruppe scheint in der vorliegenden Analyse weniger wichtig zu sein als in der früheren Metaanalyse. Obwohl Trainingsgruppen mit aktiven oder passiven Kontrollgruppen (d.h., keine aktiven Kontrollgruppen mit passiven Kontrollgruppen wie in der früheren Metaanalyse) verglichen wurden, scheint die Art der Kontrollgruppe nur teilweise einen Unterschied hinsichtlich der Transfereffekte von AG-Trainings zu machen. Transfereffekte auf mathematische Fähigkeiten und Fertigkeiten waren für passive Kontrollgruppen größer als für aktive Kontrollgruppen. Dieses Ergebnis bestätigt teilweise die Hypothese, dass aktive Kontrollgruppen besser zur Kontrolle konfundierender Variablen geeignet sind als pas-

sive Kontrollgruppen (Morrison & Chein, 2011).

Der Interventionstyp (d.h., die Art des Trainingsprogramms wie z.B. Cogmed Training und n-back Training) war lediglich ein signifikanter Moderator von Transfereffekten auf das visuell-räumliche AG in der Analyse von Melby-Lervåg und Hulme (2013). Im Gegensatz zu deren Ergebnissen war der Interventionstyp in der vorliegenden Analyse kein Moderator von Transfereffekten auf das visuell-räumliche AG, aber von Transfereffekten auf visuell-räumliches KZG, verbales AG und nonverbale Fähigkeit. Der Einfluss des Interventionstyps auf Transfereffekte auf diese Konstrukte war nicht konsistent. Im Gegensatz zu Melby-Lervåg und Hulme (2013) wurde zwischen Maßen für das visuell-räumliche KZG und das visuell-räumliche AG differenziert, da Aufgaben zur Erfassung des KZG von Aufgaben zur Erfassung des AGs getrennt werden sollten (siehe Abschnitt 2.2.1). Folglich wurden einige Aufgaben, die zur Messung des visuell-räumlichen AG in der Analyse von Melby-Lervåg und Hulme (2013) verwendet wurden in der vorliegenden Analyse als Maße für visuell-räumliches KZG kodiert. Eine Konsequenz dieser Trennung ist ein differenziertes Ergebnismuster im Hinblick auf Transfereffekte auf KZG- und AG-Komponenten. Als Folge davon wurde die moderierende Rolle des Interventionstyps für mehrere Transfereffekte untersucht. Zusammen mit der größeren Anzahl an Studien in der vorliegenden Metaanalyse könnte das differenziertere Ergebnismuster für die Unterschiede in der moderierenden Rolle des Interventionstyps von Transfereffekten auf visuell-räumliches AG, verbales AG und nonverbale Fähigkeit verantwortlich sein.

Diskussion weiterer Ergebnisse

Die vorliegende Metaanalyse ging insofern über bereits existierende Metaanalysen hinaus, als dass der Einfluss mehrerer, bisher nicht berücksichtigter Trainingsbedingungen untersucht wurde, die möglicherweise einen moderierenden Einfluss auf nahe und weite Transfereffekte von AG-Trainings haben. Hinsichtlich der Zeitkonfiguration des Trainings wurden einige offensichtliche Effekte der Trainingsdosis und Sessiondauer auf nahe, kurzfristige Transfereffekte auf das verbale und visuell-räumliche KZG gefunden. Somit scheint die Zeitkonfiguration im Hinblick auf nahe Transfereffekte eine Rolle zu spielen. Bezüglich weiterer Transfereffekte sind die Ergebnisse unschlüssig. Es wurden mittlere Effektstärken gefunden, die wahrscheinlich aufgrund einer geringen Anzahl verfügbarer Gruppenvergleiche nicht signifikant waren. Als Fazit kann festgehalten werden, dass trotz des offensichtlichen Einflusses der Zeitkonfiguration auf nahe, kurzfristige Transfereffekte die Ergebnisse hinsichtlich

weiter, langfristiger Effekte wahrscheinlich in erster Linie aufgrund geringer Teststärken un-schlüssig sind.

In Bezug auf die anderen Moderatoren hatte Supervision einen signifikanten Einfluss auf nahe, kurzfristige Transfereffekte. Im Falle von verbalem und visuell-räumlichem AG erzielte supervidiertes Training eine größere, mittlere Effektstärke als Training in der bloßen Anwesenheit anderer Personen. Dies ist konform mit Befunden die zeigen, dass sich die bloße Anwesenheit anderer Personen nachteilig auf die Leistung in komplexen Aufgaben auswirkt (für einen Überblick siehe Aiello & Douthitt, 2001). Die Effektstärken (η^2) waren groß für Transfereffekte auf verbales und visuell-räumliches AG, was eine wichtige Rolle der Supervision auf die Variabilität der Transfereffekte nahelegt. Supervision könnte auch einen Einfluss auf die Variabilität anderer Transfereffekte haben. Ungenügende Teststärken könnten eine mögliche Ursache sein für nichtsignifikante, mittlere bis große Effekte von Supervision auf die Variabilität einiger Transfereffekte (z.B. verbales KZG, mathematische Fähigkeiten und Fertigkeiten, visuell-räumliches KZG).

Des Weiteren war der Trainingsort ein signifikanter Moderator mit großen Effektstärken. Allerdings war die Moderation in ihrer Richtung inkonsistent. Konkret war der Einfluss des Trainingsorts auf Transfereffekte auf visuell-räumliches KZG, verbales AG und nonverbale Fähigkeit nicht eindeutig. Beispielsweise erzielte nur das Training zu Hause einen größeren Transfereffekt auf das visuell-räumliche KZG als das Training im Labor. Nur das Training in der Schule erzielte einen größeren Transfereffekt auf das verbale AG als das Training im Labor. Allerdings können sich die spezifischen Bedingungen innerhalb einer Kategorie des Trainingsorts stark unterscheiden. Für das Training zu Hause sind wichtige Trainingsbedingungen unbekannt, beispielsweise ob das Training in einem leisen oder lauten Raum mit Ablenkung durch Fernsehen stattfand. Zudem könnten Trainingsbedingungen mit anderen Moderatoren konfundiert sein, wie etwa der Trainingsort und der Interventionstyp. Es wurden multiple Regressionen durchgeführt, um konfundierte Moderatoren zu berücksichtigen. Mit einer Ausnahme zeigten die multiplen Regressionen mit dem Trainingsort und anderen Moderatoren, dass es einen bedeutsamen Anteil erklärbarer Varianz aber keine signifikanten Moderatoren gibt. Dieses Ergebnismuster könnte aufgrund der beträchtlichen Überlappung zwischen den Moderatoren zustande gekommen sein. Beispielsweise waren im Falle des Transfereffekts auf das verbale AG der Trainingsort und der Interventionstyp signifikante Moderatoren in einzelnen Regressionen, jedoch nicht in der multiplen Regression mit Trainingsort, Interventionstyp und Supervision. Zwei Gründe könnten diese nichtsignifikanten Einflüsse der Prädiktoren erklären: der eine ist, dass Trainingsort und Interventionstyp hoch korreliert waren,

der andere ist eine ungenügende Teststärke.

Die trainierte Modalität des AGs (d.h., verbal, visuell-räumlich, oder beide), instruktionale Unterstützung (d.h., ob es zusätzliche Erklärungen über die Instruktionen zu Beginn des Trainings hinaus gab) und Feedback (d.h., ob es zusätzliches Feedback über reines Ergebniswissen hinaus gab) hatten keinen signifikanten Effekt auf die Variabilität von Transfereffekten. Hinsichtlich der Transfereffekte von AG-Trainings scheint es keinen Unterschied zu machen, ob die verbale oder visuell-räumliche Modalität oder beide Modalitäten trainiert werden. Der nichtsignifikante Einfluss der trainierten Modalität könnte auf eine geringe Teststärke zurückzuführen sein, da die mittleren Effektstärken für Transfereffekte auf die nonverbale Fähigkeit und mathematische Fähigkeiten und Fertigkeiten mittel bis groß waren. Zu beachten ist auch, dass in den meisten der, in die Analyse eingeschlossenen Studien beide Modalitäten trainiert wurden. Der Einschluss mehrerer Studien, in denen nur eine Modalität des AGs trainiert wird, könnte zu einer anderen Interpretation der Rolle der trainierten Modalität für Transfereffekte führen.

Der nichtsignifikante Einfluss der instruktionalen Unterstützung und des Feedbacks muss vorsichtig interpretiert werden. Wie auch für den Trainingsort zutreffend, wurden die Kategorien für instruktionale Unterstützung und Feedback sehr breit definiert, da es hinsichtlich der Charakteristiken dieser potentiellen Moderatoren eine große Varianz zwischen den Studien gab. Für die Bildung elaborierterer Kategorien waren keine Informationen in den Beschreibungen der Studien vorhanden. Daher mussten qualitativ unterschiedliche instruktionale Unterstützungsmaßnahmen in verschiedenen Studien mit der gleichen Zahl kodiert werden. Beispielsweise wurde das Vorhandensein einer instruktionalen Unterstützung kodiert, wenn während des Trainings vor jeder Aufgabe Instruktionen auf einem Bildschirm präsentiert wurden und der Experimentator Fragen beantwortete (Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013). Das Vorhandensein instruktionaler Unterstützung wurde auch dann kodiert, falls Instruktionen für die trainierte Aufgabe gegeben wurden, wenn die Teilnehmer überprüfen wollten, was ihre Aufgabe war (Dahlin, Nyberg, et al., 2008). Somit ist unklar, ob elaboriertere Kategorien der instruktionalen Unterstützung eine moderierende Rolle dieser Variable aufdecken würden.

Das Feedback hatte generell keinen Einfluss auf Transfereffekte von AG-Trainings. Zunächst muss festgehalten werden, dass verschiedene Arten von (elaboriertem) Feedback über reines Ergebniswissen hinaus existieren und unterschiedliche Effekte auf Lernen haben (für einen Überblick siehe Hattie & Timperley, 2007). Derartiges Feedback konnte motivational sein (z.B. Lob für das Erreichen eines neuen Schwierigkeitslevels, d.h. anstrengungsbezo-

genes Lob (siehe Thompson et al., 2013)) oder tägliches Feedback in Form von Graphen, welche den Fortschritt von Teilnehmern darstellten (Horowitz-Kraus & Breznitz, 2009). Eine mögliche Erklärung für die fehlende Effektivität elaborierter Arten von Feedback könnte sein, dass das reine Ergebniswissen im Hinblick auf das Training des AGs bereits ausreichend effektiv ist. Feedback in Form reinen Ergebniswissens ist sehr effektiv für Lernen, insbesondere wenn neue Fertigkeiten oder Aufgaben gelernt werden (Hattie & Timperley, 2007). Die Ergebnisse scheinen nahezu legen, dass das Feedback über reines Ergebniswissen hinaus (z.B. prozessbezogenes Feedback) keinen, zusätzlich förderlichen Effekt haben könnte. Dies erscheint plausibel, da wenig darüber bekannt ist, welche anderen Arten von Feedback hilfreich für die Verbesserung von kBf sind. Primärstudien sind notwendig, die diese Hypothese direkt durch systematische Variation unterschiedlicher Arten von Feedback testen. In der vorliegenden Analyse war es nicht möglich, zwischen verschiedenen Arten von Feedback zu differenzieren, die über reines Ergebniswissen hinausgehen. Somit kann nicht ausgeschlossen werden, dass diese Arten von Feedback unterschiedliche Einflüsse (positive oder negative) auf Transfereffekte von AG-Trainings haben, die sich ausgleichen und in einem nichtsignifikanten Effekt resultieren.

6.4.3 Praktische Implikationen

Nach aktueller Befundlage gibt es kurzfristige Transfereffekte von AG-Trainings auf KZG- und AG-Komponenten, die zum Follow-up erhalten bleiben. Weite Transfereffekte sind klein, auf die nonverbale und die verbale Fähigkeit beschränkt und bleiben nicht zum Follow-up erhalten. Daher stellt sich die Frage, in welchem Ausmaß es einen praktischen Nutzen von AG-Trainings gibt. Der kleine Transfereffekt auf die nonverbale Fähigkeit verschwand nach Ausschluss zweier Studien, die nicht den Einschlusskriterien genügten. Zudem gab es keinen Transfereffekt auf mathematische Fähigkeiten und Fertigkeiten, weshalb Zweifel darüber bestehen, ob schulrelevante Fähigkeiten und Fertigkeiten von den derzeit gestalteten und implementierten AG-Trainings profitieren können.

Die Moderatoranalysen legen allerdings nahe, dass bestimmte Trainingsbedingungen für Transfereffekte wichtig sein könnten. Deshalb bleibt es möglich, dass Trainings mit optimierten Trainingsbedingungen zu praktisch relevanten Transfereffekten von AG-Trainings führen.

6.4.4 Limitationen

Einige der Limitationen der Metaanalyse von Melby-Lervåg und Hulme (2013) treffen auch auf die vorliegende Analyse zu, weshalb sie kurz erwähnt werden. Eines dieser Probleme betrifft einen potentiellen Publikationsbias, der aus einer überwiegenden Anzahl an veröffentlichten Studien in der vorliegenden Metaanalyse resultierte. Es konnte nur eine Studie eingeschlossen werden (Karbach et al., 2014), die im Zeitraum der Literatursuche nicht publiziert wurde, obwohl zusätzlich zur Literatursuche eine Vielzahl an Autoren von AG-Trainingsstudien darum gebeten wurde, nicht veröffentlichte Arbeiten bereitzustellen. Nicht veröffentlichte Studien und graue Literatur könnten keine signifikanten Transfereffekte von AG-Trainings gefunden haben. Ein Publikationsbias im dem Sinn, dass positive Ergebnisse wahrscheinlicher als negative publiziert werden, könnte zu Überschätzungen der mittleren Effektstärken führen. Allerdings wurden leichte Anzeichen für einen Publikationsbias bei nur drei Transfereffekten gefunden (verbales und visuell-räumliches AG sowie verbale Fähigkeit).

Obwohl mehr Studien als in der Analyse von Melby-Lervåg und Hulme (2013) eingeschlossen wurden, trat dasselbe Problem bezüglich der Zusammenfassung einer Vielzahl unterschiedlicher Stichproben auf—nämlich, dass Gruppengrößen zu klein waren um Analysen für separate Untergruppen durchzuführen.

Eine weitere, in Metaanalysen häufig gesehene Limitation ist die Heterogenität in der Erfassung von Ergebnissen (Walker et al., 2008). Aufgrund von Unterschieden in Validitäten und Reliabilitäten in den Ergebnismaßen könnten Transfereffekte unterschiedliche Bedeutungen haben, da diese Unterschiede in methodischen Qualitätskriterien in der Berechnung des mittleren Effekts nicht berücksichtigt werden.

Wie bereits in Abschnitt 6.2.3 erwähnt, besteht ein spezifischeres Problem der vorliegenden Analyse in der Berechnung einzelner Effektstärken durch die Differenzen der Zuwächse zwischen Trainings- und Kontrollgruppen zur Verwendung als Zähler in der Berechnung von Hedge's g . Dadurch können auch dann signifikante Effektstärken entstehen, wenn die Trainingsgruppe keine Verbesserung zeigte, aber sich die Kontrollgruppe verschlechterte. Zudem kann eine signifikante Effektstärke resultieren, wenn sich die Kontrollgruppe verbessert und die Trainingsgruppe verschlechtert. Diese Probleme traten allerdings nur bei zwei Studien auf und wurden in den Metaanalysen beachtet (Effektberechnungen mit und ohne diese Studien).

In Bezug auf langfristige Transfereffekte waren teilweise nur wenige Effektstärken für

die Analysen vorhanden (z.B. für die Analyse des langfristigen Transfers auf die verbale Fähigkeit). Aus diesem Grund war die Teststärke für diese Analysen reduziert und die Aussagekraft in Bezug auf langfristige Transfereffekte ist eingeschränkt. Teststärkeprobleme traten vermutlich auch bei den Moderatoranalysen auf. Bei kategorialen Moderatoren waren einzelne Kategorien nur mit wenigen Studien besetzt, was die Aussagekraft des Einflusses der Moderatoren aufgrund einer geringen Teststärke einschränkt.

Weitere Limitationen betreffen die einzelnen Moderatoren, die in der Analyse verwendet wurden. Wie unter 6.4.2 erwähnt, wurden die Kategorien für instruktionale Unterstützung, Feedback und Trainingsort sehr breit definiert, um eine grobe Schätzung des Einflusses dieser Variablen zu bekommen. Damit verbunden ist eine mangelnde Differenzierung (d.h., differenziertere Kategorien für die einzelnen Moderatoren). Dieses Problem trifft auch auf die potentiellen Moderatoren Häufigkeit des Trainings pro Woche und Trainingsintervall zu. Für diese Variablen mussten Kategorisierungen angewendet werden, da nicht genug Informationen verfügbar waren, um sie als kontinuierliche Moderatoren zu untersuchen. Im Falle des Trainingsintervalls fehlten zudem detailliertere Informationen. Beispielsweise war unklar, ob eine Pause von einem Tag zwischen zwei Trainingssessions meinte, dass das Training jeden Tag zur selben Uhrzeit stattfand. Schließlich können andere, potentielle Moderatoren definiert werden. Beispielsweise wurde der Einfluss der Motivation auf Trainingsergebnisse noch kaum systematisch untersucht. Informationen für den Einschluss dieses Moderators fehlen jedoch gegenwärtig in den meisten der, in der vorliegenden Arbeit betrachteten Studien.

6.4.5 Weitere Forschung

Basierend auf den Ergebnissen der vorliegenden Analyse können einige Aspekte durch zukünftige Studien adressiert werden. Obwohl in einigen der, in die Metaanalyse eingeschlossenen Studien Transfereffekte mit mehreren Maßen für ein Konstrukt erfasst wurden (z.B. Colom et al., 2013; Redick et al., 2013), demonstrierten viele Studien Transfereffekte mit nur einem Maß für das jeweilige Konstrukt. In Übereinstimmung mit Shipstead et al. (2012), sollten mehrere Maße für die Erfassung des Transfers auf ein Konstrukt (z.B. Gf) verwendet werden, um reliablere statistische Werte zu bekommen.

Darüber hinaus untersuchten nur wenige Studien langfristige Transfereffekte. Mehr Studien sollten langfristige Effekte für Konstrukte betrachten, auf die es kurzfristige Transfer-effekte gab (z.B. verbale Fähigkeit). Da das AG für viele, höhere kognitive Leistungen wichtig ist und AG-Trainings anscheinend zu nahen Transfereffekten führen aber nur in begrenz-

tem Ausmaß weite Transfereffekte produzieren, sollten sich zukünftige Studien mit den Mechanismen befassen, die für dieses Ergebnismuster verantwortlich sind (z.B. eine erhöhte Menge an Informationen, die im PM gespeichert werden können).

Innerhalb dieses Ergebnismusters wurden Moderatoren gefunden, die teilweise die Variabilität in Transfereffekten erklären können. Einige der Moderatoren wurden sehr breit definiert, weshalb der Einfluss spezifischerer Kategorien auf Transfereffekte von AG-Trainings untersucht werden könnte. Insbesondere die Kategorien der instruktionalen Unterstützung, des Feedbacks und des Trainingsorts könnten elaboriert und experimentell manipuliert werden, um den Einfluss dieser und anderer Trainingsbedingungen auf Transfereffekte besser zu verstehen.

AG-Trainings zielen darauf ab, praktisch relevante Transfereffekte durch das Training einer grundlegenden kognitiven Funktion (AG) in einem eng definierten Kontext (z.B. Training mit einer dual n-back Aufgabe im Labor) zu erreichen. Im Hinblick auf die Transfereffekte von AG-Trainings in dieser Studie stellt sich die Frage, ob AG-Trainings andere kognitive Funktionen außer KZG und AG verbessern können. Es gibt Evidenz dafür, dass kognitive Funktionen in beachtlichem Ausmaß in komplexeren Kontexten verbessert werden können, die über längere Zeiträume andauern. Beispielsweise zeigte eine Studie, dass Kinder, die eine Montessori Schule besuchten, bessere exekutive Funktionen als gleichaltrige Kinder hatten, die auf andere Schulen gingen. Zudem waren die Schüler der Montessori Schule besser im Lesen und Mathematik und zeigten mehr Sorge für Gerechtigkeit und Fairness. In Montessori Klassenräumen gibt es nur ein Exemplar von jedem Material, weshalb Kinder lernen, abzuwarten. Zudem sind die Kinder an Instruktionsaktivitäten beteiligt (Diamond & Lee, 2011). Solche Aktivitäten beanspruchen kBf (z.B. ist das AG beansprucht, wenn Schüler Fragen stellen und gleichzeitig Informationen für instruktionale Aktivitäten gespeichert werden müssen) und könnten diese folglich verbessern. Ein weiteres Beispiel ist die schulische Ausbildung an sich. Es konnte gezeigt werden, dass die Schuldauer den IQ verbessert (Cahan & Cohen, 1989; Ceci, 1991). Kontexte, die komplexer als Settings in AG-Trainingsstudien sind (z.B. das Montessori Klassenzimmer), beanspruchen über einen längeren Zeitraum mehrere kognitive Funktionen und erfordern das Zusammenspiel kognitiver Funktionen. Insgesamt könnten nach aktueller Befundlage optimierte AG-Trainings größere Transfereffekte erzielen. Komplexe Kontexte stellen allerdings eine vielversprechende Alternative dar, kognitive Funktionen zu verbessern.

6.4.6 Schlussfolgerungen

Mehrere, kürzlich veröffentlichte Studien haben die Effekte von AG-Trainings in unterschiedlichen Stichproben untersucht. Auf der Grundlage der vorliegenden Metaanalyse wird geschlossen, dass AG-Trainings nahe Transfereffekte auf KZG- und AG-Komponenten erzielen. Diese nahen Transfereffekte werden über Monate aufrechterhalten. Allerdings unterstützt die derzeitige Befundlage zu den Effekten von AG-Trainings nicht die Behauptung, dass AG-Trainings zu bedeutsamen Transfereffekten auf andere kognitive Fähigkeiten und Fertigkeiten wie nonverbale Fähigkeit oder mathematische Fähigkeiten und Fertigkeiten führen. Die Bedingungen für diese weiten Transfereffekte sind noch schlecht verstanden und es existieren zu wenige Studien, die langfristige Transfereffekte untersuchen.

Die vorliegende Studie könnte durch die Analyse der moderierenden Rolle verschiedener Trainingsbedingungen zu einem besseren Verständnis der Bedingungen für Transfereffekte beitragen. Das Ergebnismuster zeigt deutlich, dass ein substantieller Anteil der Variabilität in den Transfereffekten dadurch bedingt ist, wie das Training gestaltet ist und in welchem Kontext es stattfindet. Mehr systematische Primärstudien sind notwendig, um spezifischere Trainingsbedingungen für Transfereffekte zu untersuchen.

7 Gesamtdiskussion

Im Rahmen der Gesamtdiskussion werden zunächst die Ergebnisse der beiden Studien der vorliegenden Arbeit zusammenfassend dargestellt. Im Anschluss daran werden der theoretische Erkenntnisfortschritt der Ergebnisse der beiden Studien dieser Arbeit diskutiert und Limitationen sowie Ansatzpunkte für weitere Forschung erläutert. Schließlich wird auf die praktische Bedeutung der Ergebnisse der beiden Studien dieser Arbeit eingegangen.

7.1.1 Zusammenfassung der Ergebnisse der Studien

Studie 1 widmete sich der moderierenden Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung. Die Teilnehmer lernten mit Problemen aus der Domäne Statistik entweder in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) oder einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele). In vorausgehenden Analysen zeigte sich, dass Teilnehmer in einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung und in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung konzeptuelles Wissen erwarben. Der Erwerb konzeptuellen Wissens war jedoch unabhängig vom Grad an instruktionaler Unterstützung. Teilnehmer in einer Lernumgebung mit einem hohen Grad an instruktionaler Unterstützung erwarben mehr anwendungsorientiertes Wissen als Teilnehmer in einer Lernumgebung mit einem niedrigen Grad an instruktionaler Unterstützung. Zwischen den Lernumgebungen gab es keinen signifikanten Unterschied im cognitive load (kognitive Belastung). Das anwendungsorientierte Vorwissen hatte keinen moderierenden Einfluss auf den Erwerb anwendungsorientierten Wissens.

Hinsichtlich der moderierenden Rolle kognitiver Basisfunktionen und fluider Intelligenz zeigte sich, dass der Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen von Shifting und fluider Intelligenz, jedoch nicht von der Arbeitsgedächtniskapazität abhängt. Der Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb war umso größer, je geringer die Fähigkeit zum Shifting und die fluide Intelligenz waren. Der Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb verringerte sich umso mehr, je höher die Fähigkeit zum Shifting und die fluide Intelligenz waren.

Mit der Studie 2 (Metaanalyse von Arbeitsgedächtnistrainings) wurden nahe und weite Transfereffekte von Arbeitsgedächtnistrainings sowie eine Vielzahl an Trainingsbedingungen

als potentielle Moderatoren von Transfereffekten untersucht. Die Ergebnisse der Studie 2 zeigten, dass Arbeitsgedächtnistrainings langfristige, nahe Transfereffekte auf Kurzzeitgedächtnis- und Arbeitsgedächtniskomponenten erzielen. Weite Transfereffekte waren auf die nonverbale und verbale Fähigkeit beschränkt. Die weiten Transfereffekte waren zudem kurzfristig und klein. Die Trainingsbedingungen Interventionstyp, Art der Kontrollgruppe, Trainingsdosis, Sessiondauer, Supervision und Trainingsort hatten einen moderierenden Einfluss auf Transfereffekte.

Zusammengenommen zeigen die Befunde der beiden Studien, dass kognitive Basisfunktionen und fluide Intelligenz eine Rolle für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung spielen können (Studie 1). Trainings zur Verbesserung der kognitiven Basisfunktion Arbeitsgedächtnis scheinen eher für die Verbesserung von Arbeitsgedächtnis- und Kurzzeitgedächtniskomponenten und weniger für die Erzielung von Transfereffekten auf andere kognitive Fähigkeiten und Fertigkeiten wie mathematische Fähigkeiten und Fertigkeiten geeignet zu sein. Trainingsbedingungen können eine moderierende Rolle für nahe und weite Transfereffekte spielen.

7.1.2 Theoretischer Erkenntnisfortschritt

Die Ergebnisse der vorliegenden Arbeit erweitern die Erkenntnisse bezüglich der Bedeutung von kognitiven Basisfunktionen und fluider Intelligenz für kognitive Leistungen, der Forschung zur Aptitude-Treatment-Interaktion und der Trainierbarkeit kognitiver Basisfunktionen.

Bedeutung kognitiver Basisfunktionen und fluider Intelligenz für kognitive Leistungen

Wie in Abschnitt 2.3 dargelegt, kommt kognitiven Basisfunktionen und fluider Intelligenz eine wichtige Rolle für kognitive Leistungen zu. Speziell dem Arbeitsgedächtnis als kognitive Basisfunktion wird im Rahmen der cognitive load theory (z.B. Sweller et al., 1998) ein zentraler Stellenwert zugeschrieben. Es wurde postuliert, dass instruktionale Maßnahmen die Kapazität des Arbeitsgedächtnisses berücksichtigen müssen (z.B. Sweller, 2011). Diese Annahme wurde jedoch kaum systematisch überprüft. Die Studie 1 dieser Arbeit liefert empirische Evidenz bezüglich der Prüfung dieser Annahme. Den Ergebnissen der Studie 1 zu Folge scheint die Arbeitsgedächtniskapazität global keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Wissenserwerb zu haben. Demnach kommt der Arbeitsgedächtniskapazität nicht die erwartete zentrale Rolle zu. Hintergrund der

Annahme einer moderierenden Rolle der Arbeitsgedächtniskapazität ist, dass das Arbeitsgedächtnis in Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung unterschiedlich beansprucht werden sollte. Die kognitive Belastung sollte in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung höher sein als in einer Lernumgebung mit hohem Grad an instruktionaler Unterstützung. Die kognitive Belastung war in der Studie 1 in der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) jedoch nur deskriptiv höher als in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele). Der Unterschied in der kognitiven Belastung war möglicherweise aufgrund einer zu geringen Teststärke nicht signifikant. Allerdings war dieser Unterschied gering und vermutlich ohne praktische Bedeutung. Trotz dem fehlenden Unterschied in der kognitiven Belastung zwischen den beiden Lernumgebungen erwarben Teilnehmer in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung mehr anwendungsorientiertes Wissen als in der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung. Nach Brünken et al. (2003) können Leistungsmaße zur Beurteilung von Effekten der kognitiven Belastung herangezogen werden. Im Kontext der Forschung zum multimedialen Lernen wird davon ausgegangen, dass eine erhöhte Wissensaneignung mit einer reduzierten kognitiven Belastung einhergeht (Brünken et al., 2003). Konform mit dieser Annahme gehen Paas und Van Merriënboër (1994b) davon aus, dass bei Aufgaben mit einer geringen kognitiven Belastung meist die Leistung besser ist. Hinsichtlich der Ergebnisse der Studie 1 dieser Arbeit legt der Unterschied zwischen den Lernumgebungen im Erwerb anwendungsorientierten Wissens bei fehlenden Unterschieden in der kognitiven Belastung somit nahe, dass die subjektiv eingeschätzte kognitive Belastung die tatsächliche kognitive Belastung nicht adäquat widerspiegelt. In Übereinstimmung damit ist nach Brünken et al. (2003) unklar, wie die, durch die Ratingskala erfasste mentale Anstrengung mit der tatsächlichen kognitiven Belastung in Verbindung steht. Des Weiteren gab es keinen Zusammenhang zwischen der kognitiven Belastung und der Arbeitsgedächtniskapazität. Dieser fehlende Zusammenhang könnte darauf hindeuten, dass die in der Studie 1 verwendete Ratingskala eine von der Arbeitsgedächtniskapazität weitgehend unabhängige kognitive Belastung erfasst. Die kognitive Belastung wurde mit der Arbeitsgedächtniskapazität in Verbindung gebracht (Sweller, 2011). Allerdings könnte die kognitive Belastung auch beispielsweise mit Anforderungen an schlussfolgerndes Denken zusammenhängen. Andere Aspekte der kognitiven Belastung als die Arbeitsgedächtniskapazität könnten womöglich durch die Ratingskala erfasst worden sein, zumal deren Validität bisher kaum in Frage gestellt wurde (de Jong, 2010). Dennoch scheinen Ratingskalen bei einem Großteil der aktuellen

Forschung die bevorzugte Methode zur Erfassung der kognitiven Belastung zu sein (Schmeck, Opfermann, Van Gog, Paas, & Leutner, 2014). Bezüglich der, die kognitive Belastung beeinflussenden Faktoren ist im Allgemeinen denkbar, dass in Abhängigkeit von Aufgabenanforderungen unterschiedliche kognitive Funktionen in unterschiedlichem Ausmaß beansprucht werden, was zur kognitiven Belastung beiträgt. Des Weiteren ist denkbar, dass die kognitive Belastung in Abhängigkeit der Ausprägung kognitiver Funktionen variiert. Konform hiermit wird die kognitive Belastung nach Paas und Van Merriënboër (1994a, 1994b) durch die kognitiven Fähigkeiten von Lernern beeinflusst. Zu diesen kognitiven Fähigkeiten könnten beispielsweise auch Aufmerksamkeitsprozesse von Lernern gehören (Brünken & Leutner, 2001). Eine weitere Variable, die einen Einfluss auf die kognitive Belastung haben kann, ist das Lernverhalten. Das Lernverhalten wird neben dem Instruktionsdesign und dem Vorwissen von individuellen Eigenschaften der Lerner wie beispielsweise Metakognition und Einstellungen beeinflusst (Opfermann, 2008). In der vorliegenden Studie hatte den Ergebnissen zu Folge die fluide Intelligenz einen, von der Arbeitsgedächtniskapazität unabhängigen Einfluss auf die kognitive Belastung. Dagegen beeinflusste die Shifting-Fähigkeit die eingeschätzte kognitive Belastung nicht. Die Shifting-Fähigkeit hatte jedoch einen moderierenden Einfluss auf den Wissenserwerb und wurde vermutlich in Abhängigkeit des Grads an instruktionaler Unterstützung unterschiedlich beansprucht. Die fehlende Korrelation zwischen Shifting und der kognitiven Belastung könnte darauf zurückzuführen sein, dass die Einschätzungen der Teilnehmer der kognitiven Belastung auf der Ratingskala nicht von den Shifting-Anforderungen beeinflusst waren. Die kognitive Belastung wurde neben der fluiden Intelligenz auch von anderen Variablen beeinflusst. So war die kognitive Belastung umso geringer, je höher das konzeptuelle und anwendungsorientierte Vorwissen war. Konform hiermit stellt das Vorwissen nach Paas und Van Merriënboër (1994b) eine, die kognitive Belastung beeinflussende Lernercharakteristik dar. Ein höheres Vorwissen reduziert nach Sweller et al. (1998) die Belastung im Arbeitsgedächtnis. Da die Arbeitsgedächtniskapazität jedoch nicht signifikant mit der kognitiven Belastung korrelierte, hängt ein höheres Vorwissen womöglich mit anderen Aspekten der kognitiven Belastung zusammen. Zu diesen Aspekten könnten die Anforderungen an die fluide Intelligenz gehören, da die fluide Intelligenz mit dem anwendungsorientierten Vorwissen und der kognitiven Belastung korreliert war. Weiterhin können den Ergebnissen der Studie 1 zu Folge auch die, mit dem Fragebogen zur aktuellen Motivation (Rheinberg et al., 2001) erfassten Variablen Misserfolgsbefürchtung, Erfolgswahrscheinlichkeit und Herausforderung zur kognitiven Belastung beitragen. Diese Skalen korrelierten mit der kognitiven Belastung. Beispielsweise legen die Ergebnisse nahe, dass erfolgs-

zuversichtlichere Personen eher eine geringere kognitive Belastung erfahren. Zu beachten ist bei den Skalen des Fragebogens zur aktuellen Motivation jedoch speziell die niedrige Reliabilität der Skala Herausforderung (siehe Abschnitt 5.2.5). Aufgrund der hohen Fehlervarianz wird eine Korrelation mit anderen Variablen unwahrscheinlicher. Dennoch war die Skala Herausforderung mit der kognitiven Belastung korreliert. Daher könnte der Einfluss der subjektiv wahrgenommenen Herausforderung durch die Aufgaben bzw. statistischen Probleme auf die kognitive Belastung bei einer höheren Reliabilität der Skala noch höher sein oder zumindest genauer geschätzt werden. Der mögliche Einfluss motivationaler Variablen ist konform mit der Sichtweise von Paas, Tuovinen, Van Merriënboër und Darabi (2005). Den Autoren zu Folge hat die mentale Anstrengung auch eine motivationale Komponente (Paas et al., 2005). Da sich die kognitive Belastung in der mentalen Anstrengung widerspiegelt (Paas & Van Merriënboër, 1994b), könnten motivationale Variablen die eingeschätzte kognitive Belastung beeinflussen. Insgesamt ist nach den Ergebnissen der Studie 1 fraglich, ob mit der Ratingskala von Paas (1992) die kognitive Belastung im Arbeitsgedächtnis erfasst werden kann. Es kann nicht ausgeschlossen werden, dass die Beanspruchung des Arbeitsgedächtnisses in beiden Lernumgebungen hoch war und die Arbeitsgedächtniskapazität daher global keinen moderierenden Einfluss auf den Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens hatte. Die Ergebnisse der Studie 1 liefern jedoch Hinweise auf eine mögliche, moderierende Rolle der Arbeitsgedächtniskapazität. In einem kleinen Wertebereich war der hohe Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit dem niedrigen Grad an instruktionaler Unterstützung (Problemlösen) hinsichtlich des Erwerbs von anwendungsorientiertem Wissen umso vorteilhafter, je geringer die Arbeitsgedächtniskapazität war.

Neben den Befunden zur moderierenden Rolle der Arbeitsgedächtniskapazität konnten in der ersten Studie dieser Arbeit neue Erkenntnisse zur Bedeutung der kognitiven Basisfunktion Shifting und der fluiden Intelligenz in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung gewonnen werden. Shifting ist für komplexe Aufgaben relevant, die einen Wechsel zwischen Aufgabenteilen erfordern (Best et al., 2009). Auch die fluide Intelligenz ist für komplexe Aufgaben relevant, die zudem neu sind (Primi et al., 2010). Vermutlich erforderte die Lösung statistischer Probleme in der Studie 1 dieser Arbeit die Fähigkeit zum Shifting sowie fluide Intelligenz. Lerner mit geringer Shifting-Fähigkeit und fluider Intelligenz hatten möglicherweise in der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) aufgrund hoher Anforderungen an diese kognitiven Fähigkeiten Schwierigkeiten. Wahrscheinlich aus diesen Gründen profitierten diese Teilnehmer im Hin-

blick auf den Erwerb anwendungsorientierten Wissens am meisten von der hohen instruktionalen Unterstützung (Lösungsbeispiele). Die Befunde geben erste Hinweise über die Relevanz von Shifting und fluider Intelligenz in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung. Neben Variablen wie Vorwissen und Arbeitsgedächtniskapazität könnten Shifting und fluide Intelligenz für die Anpassung des Grades an instruktionaler Unterstützung wichtig sein. Darauf wird nachfolgend in Bezug auf die Forschung zur Aptitude-Treatment-Interaktion näher eingegangen.

Aptitude-Treatment-Interaktion

Im Rahmen der Forschung zur Aptitude-Treatment-Interaktion wird davon ausgegangen, dass die Effektivität von Interventionsmaßnahmen bzw. treatments von Lernercharakteristiken bzw. aptitudes abhängt (z.B. Cronbach & Snow, 1977; Snow & Lohman, 1984). Wie in Kapitel 3 dargestellt, wurde in Bezug auf den Grad an instruktionaler Unterstützung auf Basis des *expertise reversal effect* gefolgert, dass Lösungsbeispiele (hoher Grad an instruktionaler Unterstützung) für Lerner mit niedrigem Vorwissen vorteilhaft sind. Dagegen sollte Problemlösen (niedriger Grad an instruktionaler Unterstützung) umso effektiver sein, je höher das Vorwissen von Lernern ist (für einen Überblick siehe Kalyuga, 2007). In der Studie 1 dieser Arbeit war das anwendungsorientierte Vorwissen der Teilnehmer im Mittel gering und in der Kovarianzanalyse für den Vergleich der Mittelwerte im Erwerb anwendungsorientierten Vorwissens zwischen den Lernumgebungen zeigte sich erwartungskonform ein Vorteil der hohen instruktionalen Unterstützung (Lösungsbeispiele) im Vergleich mit der niedrigen instruktionalen Unterstützung (Problemlösen). Allerdings konnte in der Moderatoranalyse für das anwendungsorientierte Vorwissen ein *expertise reversal effect* global nicht nachgewiesen werden. Grund hierfür ist wahrscheinlich eine starke Varianzeinschränkung, vor allem im unteren Bereich des anwendungsorientierten Vorwissens. Der *expertise reversal effect* kann nach Kalyuga, Rikers und Paas (2012) in Bezug zu Studien zur Aptitude-Treatment-Interaktion gesetzt werden. In der Moderatoranalyse für das anwendungsorientierte Vorwissen zeigte sich keine Interaktion zwischen dem anwendungsorientierten Vorwissen und dem Grad an instruktionaler Unterstützung. Es lag jedoch eine Interaktion zwischen der Shifting-Fähigkeit sowie fluider Intelligenz und dem Grad der instruktionalen Unterstützung vor. Der Vorteil des hohen Grads an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit dem niedrigen Grad an instruktionaler Unterstützung (Problemlösen) verringerte sich mit steigender Shifting-Fähigkeit und fluider Intelligenz. Somit lagen Interaktionen zwischen dem

Grad an instruktionaler Unterstützung und der Fähigkeit zum Shifting sowie dem Grad an instruktionaler Unterstützung und der fluiden Intelligenz vor. Diese Interaktionen legen nahe, in der Forschung zum Vergleich von Lösungsbeispielen (hoher Grad an instruktionaler Unterstützung) mit Problemlösen (niedriger Grad an instruktionaler Unterstützung) neben dem Vorwissen auch kognitive Fähigkeiten wie Shifting und fluide Intelligenz zu berücksichtigen.

Trainierbarkeit kognitiver Basisfunktionen

Insbesondere das Arbeitsgedächtnis erhielt in Studien zur Bedeutsamkeit kognitiver Basisfunktionen für kognitive Leistungen Beachtung (siehe Abschnitt 2.3.1). Vor dem Hintergrund der vielfältigen Zusammenhänge der Arbeitsgedächtniskapazität mit anderen kognitiven Fähigkeiten und Fertigkeiten ist die Idee von Arbeitsgedächtnistrainings, dass Verbesserungen im Arbeitsgedächtnis auf andere kognitive Fähigkeiten wie nonverbale Fähigkeit oder mathematische Fähigkeiten und Fertigkeiten generalisieren. Ausgangspunkt für die Studie 2 (Metaanalyse von Arbeitsgedächtnistrainings) dieser Arbeit war, dass mehrere Trainingsbedingungen in Studien zu Arbeitsgedächtnistrainings möglicherweise einen Einfluss auf Transfereffekte haben, aber bisher weitestgehend vernachlässigt wurden. Zudem wurde seit der Veröffentlichung einer Metaanalyse über Arbeitsgedächtnistrainings von Melby-Lervåg und Hulme (2013) eine Vielzahl weiterer Trainingsstudien veröffentlicht. Aufgrund der Vielzahl an neueren Trainingsstudien war die Befundlage zu nahen und weiten Transfereffekten von AG-Trainings unklar. Die Ergebnisse der Metaanalyse der vorliegenden Arbeit legen nahe, dass Arbeitsgedächtnistrainings nahe Transfereffekte auf Kurzzeitgedächtnis- und Arbeitsgedächtniskomponenten erzielen. Da in bisherigen Metaanalysen keine Trennung zwischen Kurzzeitgedächtnis und Arbeitsgedächtnis vorgenommen wurde, liefert die vorliegende Metaanalyse ein differenzierteres Bild von nahen Transfereffekten. Weite Transfereffekte auf andere kognitive Fähigkeiten und Fertigkeiten scheinen kaum möglich zu sein. Allerdings geben die Ergebnisse der Moderatoranalysen der Metaanalyse Hinweise darüber, welche Trainingsbedingungen einen Einfluss auf Transfereffekte von Arbeitsgedächtnistrainings haben könnten. In bisherigen Studien zu Arbeitsgedächtnistrainings waren die Trainingsbedingungen sehr variabel und wenig innovativ. Zum Beispiel gab es zwischen den, in die Metaanalyse dieser Arbeit eingeschlossenen Studien zu Arbeitsgedächtnistrainings eine hohe Variabilität in den Arten von elaboriertem Feedback, das über reines Ergebniswissen hinausgeht. Die Effekte unterschiedlicher Arten von Feedback auf Transfereffekte von Arbeitsgedächtnistrainings wurden in diesen Studien nicht systematisch untersucht. Unterschied-

liche Arten von Feedback können differenzierte Effekte auf den Lernerfolg haben (Hattie & Timperley, 2007). Es kann nicht ausgeschlossen werden, dass in die Metaanalyse der vorliegenden Arbeit Studien einfließen, in denen das Feedback über reines Ergebniswissen hinaus positive und negative Einflüsse auf Transfereffekte hatte. Die positiven und negativen Einflüsse könnten sich ausgeglichen und zu einem nichtsignifikanten Effekt des Feedbacks geführt haben. Möglicherweise haben verschiedene Arten von Feedback einen positiven Einfluss auf Transfereffekte von Arbeitsgedächtnistrainings. Insgesamt haben Arbeitsgedächtnistrainings mit optimierten Trainingsbedingungen das Potential, höhere Transfereffekte zu erzielen.

7.1.3 Limitationen und weitere Forschung

Nachfolgend werden die wichtigsten Limitationen der beiden Studien dieser Arbeit mit Empfehlungen für zukünftige Forschung dargestellt.

Im Kontext der ersten Studie muss beachtet werden, dass die Teilnehmer ein relativ hohes konzeptuelles Vorwissen und ein relativ geringes anwendungsorientiertes Vorwissen aufwiesen. Das hohe konzeptuelle Vorwissen könnte zum Teil für den fehlenden Vorteil der Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) verglichen mit der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) verantwortlich sein. Die Begründung ist, dass Lösungsbeispiele in erster Linie für vorwissenschwache Lerner vorteilhaft sind (Kalyuga, 2007). Ein Vorteil der Lernumgebung mit hohem Grad an instruktionaler Unterstützung verglichen mit der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung könnte bei Themengebieten nachgewiesen werden, zu denen Teilnehmer ein geringes konzeptuelles Vorwissen aufweisen. Zukünftige Studien könnten die moderierende Rolle von kognitiven Basisfunktionen und fluider Intelligenz für den Erwerb konzeptuellen Wissens untersuchen, wenn das konzeptuelle Vorwissen der Teilnehmer gering ist. Ob ein *expertise reversal effect* (Kalyuga, 2007) in dem Sinne vorlag, dass der Effekt der hohen instruktionalen Unterstützung (Lösungsbeispiele) verglichen mit der niedrigen instruktionalen Unterstützung (Problemlösen) auf den Erwerb konzeptuellen Wissens mit steigendem konzeptuellen Vorwissen abnimmt, konnte nicht untersucht bzw. nachgewiesen werden. Die Gründe hierfür waren fehlende Unterschiede im Erwerb konzeptuellen Wissens zwischen den Lernumgebungen mit niedrigem vs. hohem Grad an instruktionaler Unterstützung (Problemlösen vs. Lösungsbeispiele) und eine fehlende Varianz im Bereich niedrigen konzeptuellen Vorwissens (nur eine Person hatte ein niedriges konzeptuelles Vorwissen). Bei Verwendung von Fragen zur Erfassung des konzeptuellen Wissens, die eine hohe inhaltliche Überlappung zu den Lösungsbeispielen aufweisen, könnte ein Vorteil der Lernum-

gebung mit hohem Grad an instruktionaler Unterstützung gegenüber der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung eher nachgewiesen werden. Vor dem Hintergrund der niedrigen Reliabilitäten der Tests zur Erfassung des konzeptuellen Wissens, sollte eine größere Anzahl inhaltsähnlicher Fragen zur Erhöhung der Reliabilitäten dieser Tests verwendet werden. Die Fragen sollten zudem im unteren Bereich des konzeptuellen Wissens gut zwischen Teilnehmern differenzieren, um die moderierende Rolle des konzeptuellen Vorwissens für den Erwerb konzeptuellen Wissens nachweisen zu können. Des Weiteren könnte eine stark eingeschränkte Varianz des anwendungsorientierten Vorwissens im unteren Wertebereich die ausbleibende Moderatorenrolle dieses Wissens erklären. Ähnlich wie für die Fragen zur Erfassung konzeptuellen Vorwissens gefordert, sollten die Fragen zur Erfassung des anwendungsorientierten Vorwissens im unteren Wertebereich zwischen Teilnehmern gut differenzieren. Folglich könnte nachgewiesen werden, dass das anwendungsorientierte Vorwissen den Effekt des Grads an instruktionaler Unterstützung auf den Erwerb anwendungsorientierten Wissens moderiert. Da neben den Tests zur Erfassung des konzeptuellen Wissens auch die Tests zur Erfassung des anwendungsorientierten Wissens eine geringe Reliabilität aufwiesen und wenig Fragen beinhalteten, sollte eine größere Anzahl inhaltsähnlicher Fragen zur Erfassung des anwendungsorientierten Wissens für eine höhere Reliabilität verwendet werden. Reliablere Wissenstests ermöglichen genauere Schätzungen des konzeptuellen und anwendungsorientierten Wissens und könnten die Ergebnisse zukünftiger Studien untermauern.

Weitere wichtige Limitationen betreffen wiederum die verwendeten Wissenstests und statistischen Probleme in der Interventionsphase der Studie 1. Die Fragen der Wissenstests erfassten spezifische Wissensarten und die Inhalte sowie Aufgaben der statistischen Probleme waren hochspezifisch (komplexe statistische Probleme und zugehörige Informationen aus dem Themenbereich „Allgemeines lineares Modell“). Die Lösung der statistischen Probleme in der Bedingung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) stellte wahrscheinlich hohe Anforderungen an die Shifting-Fähigkeit und die fluide Intelligenz. Diese Anforderungen wurden in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) scheinbar signifikant reduziert, womit eine moderierende Rolle der Shifting-Fähigkeit und fluider Intelligenz begründet werden kann. Denkbar ist, dass Probleme in anderen Studien zur Effektivität von Lösungsbeispielen geringere Anforderungen an diese kognitiven Fähigkeiten stellen. Anhand eines Problems aus der Studie von Renkl et al. (2003) wurde in Abschnitt 5.4.3 erläutert, dass Shifting-Anforderungen in einer Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung auch deutlich geringer sein können als in der Studie 1 dieser Arbeit. Ähnlich wie in der Studie von Renkl et al. (2003) waren

die Probleme in anderen Studien weniger komplex (z.B. Leppink et al., 2014; Paas, 1992) und deren Lösung könnte geringere Anforderungen an kognitive Basisfunktionen und fluide Intelligenz stellen. Eine bedeutsame Reduzierung dieser Anforderungen durch einen hohen Grad an instruktionaler Unterstützung (Lösungsbeispiele) scheint kaum möglich zu sein. Folglich spielen interindividuelle Unterschiede in kognitiven Basisfunktionen und fluider Intelligenz wahrscheinlich eine untergeordnete Rolle und eine moderierende Rolle kognitiver Basisfunktionen und fluider Intelligenz könnte ausbleiben. Generell werden in Abhängigkeit von Aufgabenanforderungen kognitive Basisfunktionen und fluide Intelligenz möglicherweise in unterschiedlichem Ausmaß beansprucht. Ein hoher Grad an instruktionaler Unterstützung könnte in unterschiedlichem Ausmaß zur Reduzierung der Aufgabenanforderungen beitragen. Somit hängt die moderierende Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung vermutlich von der Beanspruchung dieser kognitiven Fähigkeiten bei der Lösung von Problemen und dem Ausmaß der Reduzierung der Beanspruchung durch instruktionale Unterstützungsmaßnahmen ab. Insgesamt ist mehr Forschung notwendig, um zu untersuchen, ob sich eine moderierende Rolle kognitiver Basisfunktionen und fluider Intelligenz für den Erwerb anderer Wissensarten wie beispielsweise prozedurales Wissen und/oder bei inhaltlich unterschiedlichen, statistischen Problemen und Problemen aus anderen Domänen nachweisen lässt.

Im Hinblick auf die ausbleibende, moderierende Rolle der Arbeitsgedächtniskapazität muss eingeräumt werden, dass ohne Bezugnahme auf die, mit der Ratingskala gemessene kognitive Belastung möglicherweise die Anforderungen an die Arbeitsgedächtniskapazität durch die Lösungsbeispiele nicht signifikant reduziert werden konnten. Falls die Anforderungen an die Arbeitsgedächtniskapazität in der Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) im Vergleich mit der Lernumgebung mit niedrigem Grad an instruktionaler Unterstützung (Problemlösen) signifikant reduziert sind, könnten Lerner mit geringer Arbeitsgedächtniskapazität besonders von der instruktionalen Unterstützung durch Lösungsbeispiele profitieren. Eine Reduzierung der Anforderungen an die Arbeitsgedächtniskapazität ist in Anlehnung an Ayres und Sweller (2005) durch ein integriertes Format wie beispielweise Grafiken möglich, die Lösungsbeispiele und die Informationen zur Begründung der Lösungsschritte abbilden. In der Lernumgebung mit hohem Grad an instruktionaler Unterstützung (Lösungsbeispiele) könnten durch diese Grafiken die Notwendigkeit der mentalen Integration getrennter Informationen (Lösungsschritte und zugehörige Informationen zu deren Begründung waren auf unterschiedlichen Folien) reduziert

und die Anforderungen an das Arbeitsgedächtnis verringert werden. Eine moderierende Rolle der Arbeitsgedächtniskapazität könnte folglich aufgedeckt werden.

Eine weitere Einschränkung betrifft das Ergebnismuster der Studie 1 dieser Arbeit. Diesem Ergebnismuster zu Folge ist die selbstberichtete kognitive Belastung ein Prädiktor mit geringerer Validität für Lernerfolgsmaße in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung als objektivere Maße kognitiver Basisfunktionen und fluider Intelligenz. Zudem ist die Konstruktvalidität der kognitiven Belastung aufgrund der fehlenden Korrelation zwischen kognitiver Belastung und Arbeitsgedächtniskapazität fragwürdig. Wie in Abschnitt 5.4.4 beschrieben, könnte dieser ausbleibende Effekt auf eine statistische Unsicherheit zurückgeführt werden. Für einen Ausschluss dieser Erklärung müssen die Ergebnisse der Studie 1 mit unterschiedlichen und eventuell größeren Stichproben repliziert werden. Unter der Annahme, dass der ausbleibende Effekt in dem Setting und für die Stichprobe der Studie 1 nicht existiert, könnten sich wichtige Konsequenzen ergeben. Ähnlich wie in der Studie 1 der vorliegenden Arbeit könnte sich herausstellen, dass die kognitive Belastung von mehreren Faktoren wie beispielsweise Vorwissen, motivationalen Variablen und fluider Intelligenz beeinflusst wird und somit in diesem Sinne ein multidimensionales Konstrukt darstellt. Wie in Abschnitt 5.4.4 näher beschrieben, könnten abhängig von den verwendeten Aufgaben und dem Grad an instruktionaler Unterstützung kognitive Basisfunktionen und die fluide Intelligenz unterschiedlich stark beansprucht werden. Ein moderierender Einfluss objektiv gemessener kognitiver Basisfunktionen und fluider Intelligenz für den Wissenserwerb in Lernumgebungen mit unterschiedlichen Graden an instruktionaler Unterstützung könnte zukünftig im Vergleich zur subjektiv eingeschätzten kognitiven Belastung besser Auskunft geben über die Belastung dieser kognitiven Fähigkeiten durch bestimmte Aufgaben bei unterschiedlichen Graden an instruktionaler Unterstützung. Ist beispielweise wie in der Studie 1 dieser Arbeit der Effekt eines hohen Grads an instruktionaler Unterstützung verglichen mit einem niedrigen Grad an instruktionaler Unterstützung umso größer, je niedriger die Shifting-Fähigkeit ist, waren die Shifting-Anforderungen bei einem niedrigem Grad an instruktionaler Unterstützung wahrscheinlich hoch und könnten für Lerner mit geringer Shifting-Fähigkeit überfordernd gewesen sein. Vor dem Hintergrund der Effektivität des hohen Grads an instruktionaler Unterstützung zur Reduzierung der Shifting-Anforderungen könnten diese Lerner identifiziert werden und der Grad an instruktionaler Unterstützung bei zukünftigen Aufgabenstellungen für diese Lerner erhöht werden.

Eine wichtige Limitation der Studie 2 dieser Arbeit betrifft die Analyse einiger langfristiger Transfereffekte von Arbeitsgedächtnistrainings und Moderatoren auf Basis einer ge-

ringen Anzahl an Studien. Insbesondere die Analyse langfristiger, weiter Transfereffekte (z.B. auf die verbale Fähigkeit) basierte auf wenigen Gruppenvergleichen. Bei einigen Moderatoren (z.B. die trainierte Modalität) waren einzelne Kategorien unterbesetzt und/oder es waren zu wenige Gruppenvergleiche vorhanden, um einen möglichen Effekt mit ausreichender Teststärke statistisch abzusichern. Deshalb sind die Ergebnisse mit Vorsicht zu interpretieren. Zukünftige Studien sollten langfristige Transfereffekte auf Konstrukte untersuchen, auf die es kurzfristige Transfereffekte gab (z.B. verbale Fähigkeit). Des Weiteren sollte der Einfluss einzelner Moderatoren systematisch in Primärstudien untersucht werden. Ein Grund hierfür ist, dass die (signifikanten und nicht signifikanten) Effektstärken bei einzelnen Moderatoren Hinweise auf einen möglichen Einfluss dieser Variablen auf Transfereffekte von Arbeitsgedächtnistrainings geben. Beispielsweise war die Effektstärke (η^2) des Einflusses von Supervision auf Transfereffekte auf das visuell-räumliche KZG groß (.17) aber womöglich wegen einer geringen Anzahl an Studien und einer, damit verbundenen geringen Teststärke in der Analyse nicht signifikant. Die möglicherweise mit einer geringen Anzahl an Studien in den Kategorien einiger Moderatoren verbundene geringe Teststärke könnte nichtsignifikante Effekte dieser Moderatoren erklären. Vor dem Hintergrund der theoretisch zu erwartenden Einflüsse der nichtsignifikanten Moderatoren, sollten die Signifikanzen für deren Effekte nicht überbewertet werden. Zudem waren die Kategorien mehrerer Moderatoren sehr breit definiert (z.B. instruktionale Unterstützung, Feedback), weshalb eine differenziertere Interpretation von Moderatoren nicht möglich war. Dies spricht ebenfalls dafür, den Einfluss einzelner Moderatoren systematisch in Primärstudien zu überprüfen.

Abgesehen von Arbeitsgedächtnistrainings könnte aufgrund der Bedeutsamkeit von Shifting und fluider Intelligenz für kognitive Leistungen und deren moderierenden Rolle für den Wissenserwerb (Studie 1) auch eine Förderung dieser Variablen von Interesse sein. Die Trainierbarkeit von Shifting konnte nachgewiesen werden, allerdings untersuchten wenige Trainingsstudien Transfereffekte auf andere kognitive Fähigkeiten und Fertigkeiten (siehe Abschnitt 2.4.2). Bei Trainings des induktiven Denkens als wesentliche Komponente fluider Intelligenz werden im Wesentlichen Vergleichsprozesse bzw. die Strategie des Vergleichens trainiert (Klauer, 2014). Dagegen wird im Rahmen von *Core Trainings* zur Verbesserung des Arbeitsgedächtnisses versucht den Einsatz domänenspezifischer Strategien zu minimieren (Morrison und Chein, 2010). In einer Metaanalyse von Trainings des induktiven Denkens konnten positive Effekte auf fluide Intelligenz, Lernen in Unterrichtslektionen und andere kognitive Variablen wie Konzentration (subsumiert in einer Kategorie) festgestellt werden. Lernen in Unterrichtslektionen meint in dieser Analyse, dass im Anschluss an das Training

eine Unterrichtsstunde zu einem Thema aus einem bestimmten Bereich (z.B. Deutsch) durchgeführt und im Anschluss daran der Lernerfolg mit einem lehrzielorientierten Test erfasst wurde (Klauer, 2014).

In Trainingsstudien werden zumeist spezifische Aufgaben isoliert trainiert, die bestimmte kognitive Funktionen erfassen. Lernumgebungen, in denen *mehrere* kognitive Funktionen angesprochen werden, haben das Potential, kognitive Funktionen zu fördern (für einen Überblick siehe Diamond, 2011). Auch die Lernumgebungen in Studie 1 stellten durch unterschiedliche Grade an instruktionaler Unterstützung scheinbar unterschiedliche Anforderungen an die Shifting-Fähigkeit und fluide Intelligenz. Denkbar ist, dass kognitive Basisfunktionen und fluide Intelligenz gefördert werden können, wenn die Anforderungen an diese kognitiven Fähigkeiten stets hochgehalten werden. Möglicherweise können die Anforderungen an kognitive Basisfunktionen und fluide Intelligenz hoch gehalten werden, indem der Grad an instruktionaler Unterstützung an das Niveau kognitiver Basisfunktionen und fluider Intelligenz adaptiert wird. Die Effektivität der Adaptation der Lösungsschritte von Lösungsbeispielen an das Wissen von Teilnehmern konnte bereits nachgewiesen werden (Salden et al., 2010). Es ist jedoch eine offene Frage, ob instruktionale Unterstützungsmaßnahmen auch zur Förderung von kognitiven Basisfunktionen und fluider Intelligenz angepasst werden können. Ein möglicher Vorteil dieser Anpassung von instruktionalen Unterstützungsmaßnahmen könnte darin bestehen, dass neben kognitiven Basisfunktionen und fluider Intelligenz vermutlich auch der Wissenserwerb gefördert wird.

7.1.4 Praktische Bedeutung

Statistik ist in vielen Berufen und Studiengängen ein wichtiges Fach (Leppink et al., 2013). Im universitären Kontext wurde für einzelne Teilgebiete des Fachs gezeigt, dass Lösungsbeispiele als eine Form von Lernumgebung mit hohem Grad an instruktionaler Unterstützung effektiv für den Wissenserwerb sind (Paas, 1992; Leppink et al., 2014). Nach den Ergebnissen der Studie 1 dieser Arbeit kann der Wissenserwerb auch im Themenbereich „Allgemeines lineares Modell“ bei vorwissenschwachen Studierenden durch Lösungsbeispiele gefördert werden.

Neben dem Vorwissen könnten auch kognitive Basisfunktionen wie die Fähigkeit zum Shifting und fluide Intelligenz eine praktische Bedeutung im Hinblick auf den Grad der instruktionalen Unterstützung haben. Werden Lerner in frühen Phasen des Fertigkeitserwerbs mit offenem Problemlösen konfrontiert, könnte es aufgrund der moderierenden Rolle der Shifting-Fähigkeit und fluider Intelligenz vorzugsweise für Lehrende hilfreich sein, zwischen

Lernern mit unterschiedlicher Shifting-Fähigkeit und fluider Intelligenz zu differenzieren. Ein hoher Grad an instruktionaler Unterstützung (Lösungsbeispiele) könnte im Vergleich mit einem niedrigen Grad an instruktionaler Unterstützung (Problemlösen) außer für vorwissenschwache Lerner auch für Lerner mit geringen Ausprägungen in der kognitiven Basisfunktion Shifting und fluider Intelligenz von Vorteil sein. Kognitive Basisfunktionen und fluide Intelligenz könnten wie in der Studie 1 dieser Arbeit vor den instruktionalen Unterstützungsmaßnahmen erhoben werden.

Einschränkend muss jedoch betont werden, dass die Erfassung kognitiver Basisfunktionen und fluider Intelligenz mit einem hohen Zeitaufwand verbunden ist (siehe Abschnitt 5.2.4). Die möglichen praktischen Konsequenzen der Studie 1 müssen insgesamt aufgrund mehrerer Limitationen vorsichtig interpretiert werden (siehe Abschnitte 5.4.3 und 7.1.3).

Den Ergebnissen der Studie 2 zu Folge kann konstatiert werden, dass die praktische Bedeutsamkeit von Arbeitsgedächtnistrainings eingeschränkt ist. Der Grund hierfür ist, dass weite Transfereffekte scheinbar kaum mit Arbeitsgedächtnistrainings erzielt werden können. Eine pessimistische Position bezüglich der Transfereffekte wäre, dass Arbeitsgedächtnistrainings in erster Linie die trainierten Funktionen verbessern. Allerdings ist aufgrund der Ergebnisse der Moderatoranalysen der Studie 2 nicht ausgeschlossen, dass Arbeitsgedächtnistrainings durch optimierte Trainingsbedingungen praktisch relevante Transfereffekte erzielen können.

Tabellenverzeichnis

Tabelle 1. Stichprobenszusammensetzung.....	33
Tabelle 2. Design von Studie 1.....	34
Tabelle 3. Mittelwerte und Standardabweichungen für Pretest- und Posttestwerte des konzeptuellen Wissens getrennt nach Experimentalbedingungen.....	57
Tabelle 4. Mittelwerte und Standardabweichungen für Pretest- und Posttestwerte des anwendungsorientierten Wissens getrennt nach Experimentalbedingungen.....	57
Tabelle 5. Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte des anwendungsorientierten Vorwissens.....	59
Tabelle 6. Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte der AG-Kapazität	61
Tabelle 7. Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte von Shifting.....	62
Tabelle 8. Bedingter Effekt des Grads an instruktionaler Unterstützung auf den Erwerb von anwendungsorientiertem Wissen für unterschiedliche Werte von Gf	64
Tabelle 9. Effektstärken η^2 und R^2 für Moderatoren von nahen und weiten Transfereffekten	108

Abbildungsverzeichnis

Abbildung 1. Erstes statistisches Problem mit Aufgabenstellung.....	36
Abbildung 2. Teil der Informationen für die Lösung des ersten statistischen Problems.....	37
Abbildung 3. Erster Lösungsschritt für das erste statistische Problem.....	38
Abbildung 4. Beispiele für Durchgänge bei der color-shape Aufgabe mit der Klassifizierung von Farbe (links) bzw. Form (rechts).....	42
Abbildung 5. Beispiele von Durchgängen bei der number-letter Aufgabe mit der Klassifikation einer Zahl (links) bzw. eines Buchstabens (rechts).....	43
Abbildung 6. Beispiele für Durchgänge bei der category-switch Aufgabe mit der Kategorisierung als „lebendig/nicht lebendig“ (links) bzw. „kleiner/größer als ein Fußball“ (rechts).....	45
Abbildung 7. Beispielitem aus dem Subtest „Numerisch-induktives Denken“ der INSBAT (unverändert übernommen aus Arendasy et al., 2012).....	47
Abbildung 8. Beispielitem aus dem Subtest „Figural-induktives Denken“ der INSBAT (unverändert übernommen aus Arendasy et al., 2012).....	47
Abbildung 9. Beispielitem aus dem Subtest „Verbal-deduktives Denken“ der INSBAT (unverändert übernommen aus Arendasy et al., 2012).....	48
Abbildung 10. Zweite MC-Frage aus dem Wissenstest zur Erfassung konzeptuellen Vorwissens. Richtige Antwortmöglichkeiten sind kursiv gedruckt.....	49
Abbildung 11. Erste offene Frage aus dem Wissenstest zur Erfassung anwendungsorientierten Vorwissens.....	50
Abbildung 12. Forest plot für kurzfristige Transfereffekte auf das verbale KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar.....	92

Abbildung 13. Forest plot für langfristige Transfereffekte auf das verbale KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar..... 93

Abbildung 14. Forest plot für kurzfristige Transfereffekte auf das visuell-räumliche KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 94

Abbildung 15. Forest plot für langfristige Transfereffekte auf das visuell-räumliche KZG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 95

Abbildung 16. Forest plot für kurzfristige Transfereffekte auf das verbale AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar..... 96

Abbildung 17. Forest plot für langfristige Transfereffekte auf das verbale AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar..... 97

Abbildung 18. Forest plot für kurzfristige Transfereffekte auf das visuell-räumliche AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 98

Abbildung 19. Forest plot für langfristige Transfereffekte auf das visuell-räumliche AG. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 99

Abbildung 20. Forest plot für kurzfristige Transfereffekte auf die nonverbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 100

Abbildung 21. Forest plot für langfristige Transfereffekte auf die nonverbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 101

Abbildung 22. Forest plot für kurzfristige Transfereffekte auf die verbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar... 102

Abbildung 23. Forest plot für langfristige Transfereffekte auf die verbale Fähigkeit. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar... 103

Abbildung 24. Forest plot für kurzfristige Transfereffekte auf die Fertigkeit zur Wortdekodierung. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 104

Abbildung 25. Forest plot für langfristige Transfereffekte auf die Fertigkeit zur Wortdekodierung. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 105

Abbildung 26. Forest plot für kurzfristige Transfereffekte auf mathematische Fähigkeiten und Fertigkeiten. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 106

Abbildung 27. Forest plot für langfristige Transfereffekte auf mathematische Fähigkeiten und Fertigkeiten. Punkte und horizontale Linien zeigen Effektstärken (Hedge's g) bzw. Konfidenzintervalle. Der Diamant stellt den Gesamteffekt und sein Konfidenzintervall (Weite des Diamanten) dar. 107

Literaturverzeichnis

Literaturangaben mit einem Sternchen kennzeichnen Studien, die in die Metaanalyse (Studie 2) der vorliegenden Arbeit eingeschlossen wurden.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities.

Journal of Experimental Psychology: General, 131(4), 567–589.

Ackerman, P. L., & Cianciolo, A. T. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology: Applied*, 8(3), 194–208.

doi:10.1037//1076-898X.8.3.194

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology*,

105(4), 286–305. doi:10.1016/j.jecp.2009.09.006

Aiello, J. R., & Douthitt, E. A. (2001). Social facilitation from triplatt to electronic performance monitoring. *Group Dynamics: Theory, Research, and Practice*, 5(3), 163–180.

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy:

learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive*

Science, 26(2), 147–179. doi:10.1207/s15516709cog2602_1

*Alloway, T. (2012). Can interactive working memory training improve learning? *Journal of Interactive Learning Research*, 23(3), 197–207.

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews*, 1(4), 134–139.

*Alloway, T. P., Bibile, V., & Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior*, 29(3),

632–638. doi:10.1016/j.chb.2012.10.023

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77(6),

1698–1716.

- Allport, A., & Wylie, G. (2000). Task switching, stimulus-response bindings, and negative priming. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (S. 35-70). Cambridge, MA: MIT Press.
- Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology, 102*(1), 115–134. doi:10.1037/a0016838
- Arendasy, M., Hornke, L. F., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., ... T. Körtner (2012). *Intelligenz-Struktur-Batterie (INSBAT)*. Manual. Mödling: Schuhfried GmbH.
- Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (Vol. 2, S. 135–147). Cambridge, U.K: Cambridge University Press.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839. doi:10.1038/nrn1201
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford, England: Oxford University Press.
- Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49*(6), 1393–1400. doi:10.1016/j.neuropsychologia.2010.12.042
- Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21*(8), 627–635.
- Banas, S., & Sanchez, C. A. (2012). Working memory capacity and learning underlying conceptual relationships across multiple documents. *Applied Cognitive Psychology, 26*(4),

594–600. doi:10.1002/acp.2834

Berryhill, M. E., & Hughes, H. C. (2009). On the minimization of task switch costs following long-term training. *Attention, Perception & Psychophysics*, *71*(3), 503–514.

doi:10.3758/APP.71.3.503

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*(3), 180–200. doi:10.1016/j.dr.2009.05.002

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample.

Learning and Individual Differences, *21*(4), 327–336. doi:10.1016/j.lindif.2011.01.007

Blair, C., Knipe, H., & Gamson, D. (2008). Is there a role for executive functions in the development of mathematics ability? *Mind, Brain, and Education*, *2*(2), 80–89.

doi:10.1111/j.1751-228X.2008.00036.x

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, *74*(4), 245–248.

*Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging*, *25*(4),

767–778. doi:10.1037/a0020683

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.).

Berlin, Heidelberg, New York: Springer.

*Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., & Bäckman, L.

(2011). Neural correlates of training-related working-memory gains in old age. *NeuroImage*, *58*(4), 1110–1120. doi:10.1016/j.neuroimage.2011.06.079

*Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuro-*

science, 6. doi:10.3389/fnhum.2012.00063

Broadbent, D. E. (1975). The magic number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (S. 3–18). New York: Wiley.

Brünken, R., & Leutner, D. (2001). Aufmerksamkeitsverteilung oder Aufmerksamkeitsfokussierung? Empirische Ergebnisse zur “Split-Attention-Hypothese” beim Lernen mit Multimedia. *Unterrichtswissenschaft*, 29, 357–366.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53–61.
doi:10.1207/S15326985EP3801_7

Bühner, M., König, C. J., Pick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance*, 19(3), 253–275.

Bühner, M., Kröner, S., & Ziegler, M. (2008). Working memory, visual–spatial-intelligence and their relationship to problem-solving. *Intelligence*, 36(6), 672–680.
doi:10.1016/j.intell.2008.03.008

Bühner, M., Krumm, S., Ziegler, M., & Plücken, T. (2006). Cognitive abilities and their interplay. *Journal of Individual Differences*, 27(2), 57–72. doi:10.1027/1614-0001.27.2.57

Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler* (1. Aufl.). München: Pearson.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 60, 1239–1249.

*Carretti, B., Borella, E., Fostinelli, S., & Zavagnin, M. (2013a). Benefits of training working memory in amnesic mild cognitive impairment: specific and transfer effects. *International Psychogeriatrics*, 25, 617–626. doi:10.1017/S1041610212002177

*Carretti, B., Borella, E., Zavagnin, M., & de Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults: Working memory

- training in older adults. *International Journal of Geriatric Psychiatry*, 28(5), 539–546.
doi:10.1002/gps.3859
- Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, 86(3), 360–367. doi:10.1037/0022-0663.86.3.360
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1–22. doi:10.1037/h0046743
- Cattell, R. B. (1971). *Abilities: their structure, growth, and action*. Boston: Houghton Mifflin.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27(5), 703–722. doi:10.1037/0012-1649.27.5.703
- Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, 37(5), 715–730. doi:10.1037//0012-1649.37.5.715
- *Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., ... Ramon, M. (2014). A randomized clinical trial of Cogmed Working Memory Training in school-age children with ADHD: a replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*, 55, 247–255.
doi:10.1111/jcpp.12146
- *Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17(2), 193–199. doi:10.3758/PBR.17.2.193
- Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition*, 28(1), 8–17.

- *Chooi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*, 531–542.
doi:10.1016/j.intell.2012.07.004
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. doi:10.1016/j.intell.2008.01.002
- Colom, R., Abad, F. J., Rebollo, I., & Chun Shih, P. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, *33*(6), 623–642.
doi:10.1016/j.intell.2005.05.006
- *Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., ... Jaeggi, S. M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence*, *41*(5), 712–727. doi:10.1016/j.intell.2013.09.002
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185.
doi:10.1017/S0140525X01003922
- Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

- *Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*, 1510–1512.
doi:10.1126/science.1155466
- *Dahlin, E., Nyberg, L., Bäckman, L., & Stigsdotter Neely, A. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, *23*, 720–730. doi:10.1037/a0014296
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18. doi:10.1016/0749-596X(86)90018-5
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.
- Danili, E., & Reid, N. (2004). Some strategies to improve performance in school chemistry, based on two cognitive factors. *Research in Science and Technological Education*, *22*, 203–226.
- De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105–134. doi:10.1007/s11251-009-9110-0
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*(2), 111–122. doi:10.1016/j.learninstruc.2009.02.010
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 Years Old. *Science*, *333*(6045), 959–964.
doi:10.1126/science.1204529
- *Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized

- controlled trial. *Developmental Science*, 16(6), 915–925. doi:10.1111/desc.12068
- Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89–98. doi:10.1080/01621459.2000.10473905
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.
- *Egeland, J., Aarli, A. K., & Saunes, B.-K. (2013). Few effects of far transfer of working memory training in ADHD: A randomized controlled trial. *PLoS ONE*, 8(10), e75660. doi:10.1371/journal.pone.0075660
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, 315(7109), 629–634.
- Engle, R. W. (2001). What is working memory capacity? In H.L. Roediger, J.S. Nairne, I. Neath, & A.M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (S. 297–314). Washington, DC.: American Psychological Association Press.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, S. 145–199). NY: Elsevier.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. doi:10.1037/0096-3445.133.1.101
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006).

- Not all executive functions are related to intelligence. *Psychological Science*, 17(2), 172–179. doi:10.1111/j.1467-9280.2006.01681.x
- Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, 47(5), 1410–1430. doi:10.1037/a0023750
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, 137(2), 201–225. doi:10.1037/0096-3445.137.2.201
- Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2014). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. Manuscript submitted for publication.
- Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, 17(5), 673–679. doi:10.3758/17.5.673
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18(1), 1–16. doi:10.1002/acp.934
- Gathercole, S. E., & Alloway, T. P. (2007). Understanding working memory. A classroom guide. London, UK: Harcourt Assessment. Retrieved from <http://www.york.ac.uk/res/wml/Classroom%20guide.pdf>
- Gimino, A. (2002). Students' investment of mental effort. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237–288.
- *Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., ... Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: a randomized controlled trial: Computerized working memory training in adolescents with severe LD/ADHD. *Journal of Child Psychology and Psychiatry*, 53, 1277–1284. doi:10.1111/j.1469-7610.2012.02592.x
- Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Experimental Social Psychology*, 22(1), 38–77. doi:10.1016/0022-1031(86)90040-5
- Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, 24(8), 1149–1167. doi:10.1002/acp.1624
- Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In Dempster, FN.; Brainerd, CJ. (Eds.), *Interference and inhibition in cognition*. (S. 175–204). San Diego, CA: Academic Press.
- *Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, 24(12), 2409–2419. doi:10.1177/0956797613492984
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. New York: Guilford Press.

- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from <http://www.afhayes.com/public/process2012.pdf>
- Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology*, *20*(5), 511–528. doi:10.1037/0894-4105.20.5.511
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, CA: Academic Press.
- *Heinzel, S., Schulte, S., Onken, J., Duong, Q.-L., Riemer, T. G., Heinz, A., ... Rapp, M. A. (2014). Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Aging, Neuropsychology, and Cognition*, *21*(2), 146–173. doi:10.1080/13825585.2013.790338
- Hiebert, J., & Lefevre, P. (1984). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (S. 1-29). Hillsdale, NJ: L. Erlbaum Associates.
- *Holmes, J., & Gathercole, S. E. (2013). Taking working memory training from the laboratory into schools. *Educational Psychology*, *34*(4), 440–450. doi:10.1080/01443410.2013.797338
- *Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15. doi:10.1111/j.1467-7687.2009.00848.x
- *Horowitz-Kraus, T., & Breznitz, Z. (2009). Can the error detection mechanism benefit from training the working memory? A comparison between dyslexics and controls — An ERP study. *PLoS ONE*, *4*(9), e7141. doi:10.1371/journal.pone.0007141

- *Hovik, K. T., Saunes, B.-K., Aarlien, A. K., & Egeland, J. (2013). RCT of working memory training in ADHD: Long-term near-transfer effects. *PLoS ONE*, *8*(12), e80561.
doi:10.1371/journal.pone.0080561
- *Hubacher, M., Weiland, M., Calabrese, P., Stoppe, G., Stöcklin, M., Fischer-Barnicol, D., ... Penner, I.-K. (2013). Working memory training in patients with chronic schizophrenia: A pilot study. *Psychiatry Journal*, *2013*, 1–8. doi:10.1155/2013/154867
- Huedo-Medina, T. B., Sanchez-Meca, J., Martin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychological Methods*, *11*(2), 193–206. doi:10.1037/1082-989X.11.2.193
- *Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.
- *Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*(25), 10081–10086. doi:10.1073/pnas.1103228108
- *Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*. doi:10.3758/s13421-013-0364-z
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412.
doi:10.1080/09658211003702171
- *Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*(6), 625–635.
doi:10.1016/j.intell.2010.09.001

- James, W. (1890). *Principles of psychology*. New York: Holt.
- *Jaušovec, N., & Jaušovec, K. (2012). Working memory training: Improving intelligence – Changing brain activity. *Brain and Cognition*, *79*(2), 96–106.
doi:10.1016/j.bandc.2012.02.007
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. doi:10.1007/s10648-007-9054-3
- Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. doi:10.1007/s10648-010-9150-7
- Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, *24*(2), 313–337. doi:10.1007/s10648-012-9195-x
- Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in Working Memory* (S. 21–48). Oxford University Press. Retrieved from
<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195168648.001.0001/acprof-9780195168648-chapter-2>
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. doi:10.1037/0096-3445.133.2.189
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978–990. doi:10.1111/j.1467-7687.2009.00846.x

- *Karch, J., Strobach, T., & Schubert, T. (2014). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. doi:10.1080/09297049.2014.899336
- Kessels, R. P. C., van den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment*, 15(4), 426–434. doi:10.1177/1073191108315611
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, 136(5), 849–874. doi:10.1037/a0019842
- Kimberg, D. Y., Aguirre, G. K., & D’Esposito, M. (2000). Modulation of task-related neural activity in task-switching: an fMRI study. *Cognitive Brain Research*, 10(1-2), 189–196. doi:10.1016/S0926-6410(00)00016-1
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661–667. doi:10.1111/j.0956-7976.2004.00737.x
- Klauer, K. J. (2014). Training des induktiven Denkens – Fortschreibung der Metaanalyse von 2008. *Zeitschrift Für Pädagogische Psychologie*, 28(1), 5–19. doi:10.1024/1010-0652/a000123
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317–324. doi:10.1016/j.tics.2010.05.002
- *Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... Westerberg, H. (2005). Computerized training of working memory in children with ADHD: A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(2), 177–186.

- *Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24(6), 781–791. doi:10.1076/jcen.24.6.781.8395
- Koch, I. (2001). Automatic and intentional activation of task sets. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 27(6), 1474–1486.
- Koch, I., Prinz, W., & Allport, A. (2005). Involuntary retrieval in alphabet-arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research*, 69(4), 252–261. doi:10.1007/s00426-004-0180-y
- König, C. J., Bühner, M., & Mürling, G. (2005). Working memory, fluid Intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, 18(3), 243–266. doi:10.1207/s15327043hup1803_3
- Kray, J., Karbach, J., Haenig, S., & Freitag, C. (2012). Can task-switching training enhance executive control functioning in children with attention deficit/-hyperactivity disorder? *Frontiers in Human Neuroscience*, 5. doi:10.3389/fnhum.2011.00180
- Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, 15(1), 126–147. doi:10.1037//0882-7974.15.1.126
- Krumm, S., Schmidt-Atzert, L., Bühner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, 37(4), 347–364. doi:10.1016/j.intell.2009.02.003
- Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence*, 36(5), 422–436. doi:10.1016/j.intell.2007.08.004
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load.

- Behavior Research Methods*, 45(4), 1058–1072. doi:10.3758/s13428-013-0334-1
- Leppink, J., Paas, F., Van Gog, T., Van der Vleuten, C. P. M., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32–42. doi:10.1016/j.learninstruc.2013.12.001
- Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 881–904. doi:10.1037/a0027298
- *Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, 20(1), 135–141. doi:10.3758/s13423-012-0335-6
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- *Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*, 18, 62–78. doi:10.1080/09297049.2011.575772
- Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology*, 21(6), 747–764. doi:10.1002/acp.1347
- Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, 75(2), 314–318. doi:10.1037/0022-0663.75.2.314
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1124–1140.

- Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41*(3), 211–253. doi:10.1006/cogp.2000.0736
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1423–1442. doi:10.1037/0278-7393.22.6.1423
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270–291. doi:10.1037/a0028228
- Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., ... Shoda, Y. (2011). “Willpower” over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience, 6*(2), 252–256. doi:10.1093/scan/nsq081
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. doi:10.1006/cogp.1999.0734
- Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. doi:10.1177/0963721411429458
- Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: the effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*(2-3), 123–142. doi:10.1016/j.actpsy.2003.12.004
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621–640. doi:10.1037//0096-3445.130.4.621

- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698.
doi:10.1073/pnas.1010076108
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.
doi:10.1016/S1364-6613(03)00028-7
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. doi:10.1177/1094428106291059
- Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60. doi:10.3758/s13423-010-0034-0
- Mosteller, F., & Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual Review of Public Health*, *17*(1), 1–23. doi:10.1146/annurev.pu.17.050196.000245
- Murphy, K. R., & Myers, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246.
- *Nutley, S. B., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. *Developmental Science*, *14*(3), 591–601.
doi:10.1111/j.1467-7687.2010.01022.x
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3),

411–421. doi:10.1037//0278-7393.28.3.411

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652.

doi:10.1016/j.intell.2008.01.007

Opfermann, M. (2008). *There's more to it than instructional design: The role of individual learner characteristics for hypermedia learning*. Berlin, Germany: Logos.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3*(3), 354–379. doi:10.1037/1082-989X.3.3.354

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434.

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. doi:10.1007/BF02504795

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

Paas, F., & Van Merriënboer, J. J. G. (1994b). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 51–71.

Paas, F., & Van Merriënboer, J. J. G. (1994a). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133. doi:10.1037/0022-0663.86.1.122

Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1 Pt 2), 419–430.

doi:10.2466/pms.1994.79.1.419

*Penner, I.-K., Vogt, A., Stöcklin, M., Gschwind, L., Opwis, K., & Calabrese, P. (2012).

Computerised working memory training in healthy adults: A comparison of two different training schedules. *Neuropsychological Rehabilitation*, 22(5), 716–733.

doi:10.1080/09602011.2012.686883

Primi, R. (2014). Developing a fluid intelligence scale through a combination of Rasch modeling and cognitive psychology. *Psychological Assessment*, 26(3), 774–788.

doi:10.1037/a0036712

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning:

A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, 20(5), 446–451. doi:10.1016/j.lindif.2010.05.001

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., &

Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28(3), 164–171.

doi:10.1027/1015-5759/a000123

*Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ...

Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359–379. doi:10.1037/a0029082

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working

memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology*, 24(7), 844–854. doi:10.1080/20445911.2012.704359

Reinmann, G., & Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten. In A. Krapp

& B. Weidenmann (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (S. 613–658).

Weinheim: Beltz.

- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. doi:10.1111/cogs.12086
- Renkl, A., Gruber, H., Weber, S., Lerche, T., & Schweizer, K. (2003). Cognitive Load beim Lernen aus Lösungsbeispielen. *Zeitschrift Für Pädagogische Psychologie*, 17(2), 93–101. doi:10.1024//1010-0652.17.2.93
- Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review*, 21(1), 67–78. doi:10.1007/s10648-008-9093-4
- Renkl, A., & Atkinson, R. K. (2007). An example order for cognitive skill acquisition. In F. E. Ritter, J. Nerb, E. Lehtinen, & T. M. O’Shea (Eds.), *In order to learn: how the sequence of topics influences learning* (S. 95–105). New York: Oxford University Press.
- Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. *Diagnostica*, 47(2), 57–66. doi:10.1026//0012-1924.47.2.57
- *Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, 26(4), 813–822. doi:10.1037/a0023631
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231. doi:10.1037/0096-3445.124.2.207
- Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology*, 4(1), 61–81. doi:10.1007/s11292-007-9046-9
- Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1477–1491. doi:10.1037/0278-7393.31.6.1477

- Salden, R. J. C. M., Alevén, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*(3), 289–307. doi:10.1007/s11251-009-9107-8
- *Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00166
- Schmeck, A., Opfermann, M., Van Gog, T., Paas, F., & Leutner, D. (2014). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science*. doi:10.1007/s11251-014-9328-3
- Schmidt, F.L., & Hunter, J.E. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods 1975–2001. In K.R. Murphy, *Validity generalization: A critical review* (S. 31–65). Mahwah, NJ: Erlbaum.
- *Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*. doi:10.3389/fnagi.2010.00027
- Schneider, W. J., & McGrew, K. (2012). The Cattell–Horn–Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (S. 99–144). New York, NY: Guilford Press.
- Schweizer, K., & Koch, W. (2002). Fluide Intelligenz und mittelfristig erworbenes Wissen in einer spezifischen Wissensdomäne. *Zeitschrift Für Psychologie*, *210*(3), 111–121. doi:10.1026//0044-3409.210.3.111
- Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Alevén, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*(2), 258–266. doi:10.1016/j.chb.2008.12.011
- Schwonke, R., Renkl, A., Salden, R., & Alevén, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning

outcomes. *Computers in Human Behavior*, 27(1), 58–62.

doi:10.1016/j.chb.2010.03.037

*Shavelson, R. J., Yuan, K., Alonzo, A. C., Klingberg, T., & Andersson, M. (2008). On the impact of computerized cognitive training on working memory and fluid intelligence. In D. C. Berliner & H. Kuppermintz (Eds.), *Contributions of educational psychology to changing institutions, environments, and people* (S. 1–11). New York, NY: Routledge.

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72, 116–141. doi:10.1016/j.jml.2014.01.004

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. doi:10.1037/a0027473

Shipstead, Z., Redick, T. S., & Engle, Randall W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50, 245–276.

*Shiran, A., & Breznitz, Z. (2011). The effect of cognitive training on recall range and speed of information processing in the working memory of dyslexic and skilled readers. *Journal of Neurolinguistics*, 24, 524–537. doi:10.1016/j.jneuroling.2010.12.001

Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research*, 7(1), 1–24. doi:10.2190/VQJD-T1YD-5WVB-RYPJ

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition*, 20(5), 539–548.

Snow, R. E., & Lohman, D. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology*, 76, 347–376.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelli-*

gence (Vol. 2, S. 47–103). Hillsdale, NJ: Lawrence Erlbaum Associates.

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, 22(2), 286–296.

doi:10.1080/09541440902771299

St Clair-Thompson, H. L., & Allen, R. J. (2013). Are forward and backward recall the same? A dual-task study of digit recall. *Memory & Cognition*, 41(4), 519–532.

doi:10.3758/s13421-012-0277-2

*St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, 30, 203–219.

doi:10.1080/01443410903509259

*Stepankova, H., Lukavsky, J., Buschkuehl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults. *Developmental Psychology*, 50, 1049–1059.

doi:10.1037/a0034913

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (S. 99–110). Chichester, UK: John Wiley & Sons, Ltd.

Strobach, T., Liepelt, R., Schubert, T., & Kiesel, A. (2012). Task switching: effects of practice on switch and mixing costs. *Psychological Research*, 76(1), 74–83.

doi:10.1007/s00426-011-0323-x

Strobach, T., Salminen, T., Karbach, J., & Schubert, T. (2014). Practice-related optimization and transfer of executive functions: a general review and a specific realization of their mechanisms in dual tasks. *Psychological Research*. doi:10.1007/s00426-014-0563-7

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive

- load. *Educational Psychology Review*, 22(2), 123–138. doi:10.1007/s10648-010-9128-5
- Sweller, J. (2011). Cognitive load theory. In J. Mestre, & B. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (Vol. 55, S. 37 – 76). Oxford: Academic Press.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296.
- Talebi, M. (2013). Study of publication bias in meta-analysis using trim and fill method. *International Research Journal of Applied and Basic Sciences*, 4(1), 31–36.
- *Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE*, 8(5), e63614.
doi:10.1371/journal.pone.0063614
- *Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, 12(1), 106–113. doi:10.1111/j.1467-7687.2008.00745.x
- Titz, C., & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. *Psychological Research*. doi:10.1007/s00426-013-0537-1
- Tsaparlis, G. (2005). Non-algorithmic quantitative problem solving in university physical chemistry: A correlation study of the role of selective cognitive factors. *Research in Science & Technological Education*, 23, 125–148.
- Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. doi:10.1037/0033-295X.114.1.104

- Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. doi:10.1037/0033-2909.133.6.1038
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.
- Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635–654. doi:10.1080/09658210902998047
- *Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities: Working memory training in M-BID. *Journal of Intellectual Disability Research*, *54*(5), 433–447. doi:10.1111/j.1365-2788.2010.01285.x
- Van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*(5), 427–449. doi:10.1016/j.intell.2006.09.001
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26. doi:10.1080/00461520701756248
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*(2), 154–164. doi:10.1016/j.learninstruc.2006.02.003
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155–174.

doi:10.1007/s10648-010-9134-7

- Von Bastian, C. C., & Oberauer, K. (2013). Effects and mechanisms of working memory training: a review. *Psychological Research*. doi:10.1007/s00426-013-0524-6
- Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine*, 75(6), 431–439.
- Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education*, 75, 218–228.
doi:10.1016/j.compedu.2014.02.016
- Wecker, C., Rachel, A., Heran-Dörr, E., Waltner, C., Wiesner, H., & Fischer, F. (2013). Presenting theoretical ideas prior to inquiry activities fosters theory-level knowledge. *Journal of Research in Science Teaching*, 50(10), 1180–1206. doi:10.1002/tea.21106
- *Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Östensson, M.-L., Bartfai, A., & Klingberg, T. (2007). Computerized working memory training after stroke—A pilot study. *Brain Injury*, 21, 21–29. doi:10.1080/02699050601148726
- Wilson, D. B. (2005). Meta-analysis macros for SAS, SPSS, and Stata. Retrieved from <http://mason.gmu.edu/~dwilsonb/ma.html>
- Woolfolk, A. (2008). *Pädagogische Psychologie* (10. Aufl.). München: Pearson Studium.
- Wouters, P., Paas, F., & Van Merriënboer, J. J. G. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology*, 34(1), 1–8. doi:10.1016/j.cedpsych.2008.03.001
- Yeniad, N., Malda, M., Mesman, J., van Ijzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, 23, 1–9. doi:10.1016/j.lindif.2012.10.004
- Yuan, K., Steedle, J., Shavelson, R., Alonzo, A., & Oppezzo, M. (2006). Working memory,

fluid intelligence, and science learning. *Educational Research Review*, 1(2), 83–98.

doi:10.1016/j.edurev.2006.08.005

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology (London, England: 1953)*, 80 (Pt. 3), 351–361.

Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality*, 46(2), 173–183. doi:10.1016/j.jrp.2012.01.002

Anhang A: Vorwissenstests mit Lösung

Code (erster Buchstabe des Vornamens der Mutter, erster Buchstaben des Vornamens des Vaters, Tag des Geburtsdatums (Zahl), letzte Buchstabe des eigenen Vornamens):

Arbeitsplatznummer (links oben auf PC; eine Zahl von 01 bis 06): _____

Alter (Jahre): _____

Geschlecht: _____

Studiengang: _____

Semesterzahl: _____

Test zur Erfassung konzeptuellen Vorwissens (richtige Antworten sind kursiv gedruckt; 1 Punkt pro richtiger Antwort; insgesamt 20 Punkte)

Bitte kreuzen Sie bei den folgenden Fragen die richtigen Antworten (Buchstaben) an. Bei jeder Frage können mehrere Antworten richtig sein. Für jede richtige Antwort bzw. nicht angekreuzte falsche Antwort gibt es einen Punkt.

1. Welche der folgenden Aussagen treffen im Hinblick auf den Begriff Skalenniveau zu?

- a) *Bildungsabschlüsse können als ordinalskalierte Daten betrachtet werden.*
- b) Der Median ist bei nominalskalierten Daten ein zulässiges Maß zur Ermittlung der zentralen Tendenz.
- c) *Bei intervallskalierten Daten kann der Mittelwert als Maß der zentralen Tendenz herangezogen werden.*
- d) Ein Item mit fünf geordneten Antwortkategorien (1=„Stimme nicht zu“ bis 5=„Stimme voll zu“) ist nominalskaliert.
- e) *Je höher das Skalenniveau der Variable ist, desto höher ist der Informationsgehalt ihrer Werte.*

2. Kreuzen Sie die Voraussetzungen an, die bei einer zweifaktoriellen Varianzanalyse ohne Messwiederholung gegeben sein müssen:

- a) Die abhängige Variable ist mindestens nominal skaliert.
- b) Die Interaktion zwischen den beiden Faktoren ist ordinal.

- c) *Die abhängige Variable ist in der Population in allen Zellen der Faktorstufenkombinationen normalverteilt.*
- d) *Die Varianzen müssen in jeder Teilpopulation gleich sein.*
- e) *Die unabhängige Variable muss mindestens ordinalskaliert sein.*

3. Bitte kreuzen Sie Zutreffendes zum Thema Kovarianzanalyse an!

- a) *Kovariaten können nie kausal interpretiert werden.*
- b) *Variablen die außer den unabhängigen Variablen einen Einfluss auf die abhängige Variable haben, nennt man Kovariaten.*
- c) *Mit einer Kovarianzanalyse kann der Einfluss einer unabhängigen Variablen auf eine abhängige Variable berechnet werden, wobei der Einfluss einer anderen unabhängigen Variablen auf die unabhängige Variable kontrolliert wird.*
- d) *Die Kovariate bei einer Kovarianzanalyse muss mindestens intervallskaliert sein.*
- e) *Die Voraussetzungen einer Kovarianzanalyse sind die gleichen wie bei einer Varianzanalyse.*

4. Welche der folgenden Aussagen in Bezug auf die Regression treffen zu?

- a) *Die lineare Regression dient der Merkmalsvorhersage, indem durch Kenntnis der Ausprägung eines Prädiktors die Ausprägung eines Kriteriums vorhergesagt wird.*
- b) *Im Gegensatz zur einfachen linearen Regression können bei der multiplen Regression mehrere Prädiktoren eingeschlossen werden.*
- c) *Mit einer linearen Regression können bei Querschnittsdaten keine kausalen Aussagen getroffen werden.*
- d) *Die abhängige Variable darf bei der multiplen linearen Regression dichotom sein.*
- e) *Um eine möglichst hohe Kollinearität zu erzielen, sollten Prädiktoren in die Regression eingeschlossen werden, die hoch miteinander korrelieren.*

Test zur Erfassung anwendungsorientierten Vorwissens (1 Punkt pro richtiger Antwort; insgesamt 6 Punkte)

1. *In einer Untersuchung soll eine neue Fruchtjoghurtsorte hinsichtlich der Beurteilung des Geschmacks (intervallskaliertes Merkmal) mit einer bereits auf dem Markt befindlichen Fruchtjoghurtsorte und einem Joghurt ohne Fruchtzubereitung verglichen wer-*

den. Zusätzlich sollen unterschiedliche Geschmackspräferenzen von Männern und Frauen berücksichtigt werden. Von Interesse ist darüber hinaus auch, ob die Präferenz einer Joghurtsorte vom Geschlecht abhängt. *Um welche Art von Design handelt es sich bei der vorliegenden Untersuchung? Wie können die Fragestellungen statistisch beantwortet werden? Bitte begründen Sie ihre Antworten kurz.*

Antwort:

Es handelt sich um ein zweifaktorielles (2x3 oder 3x2) Design (ohne Messwiederholung) mit den unabhängigen Variablen bzw. Faktoren Art des Joghurts (zweistufig) sowie Geschlecht (zweistufig). Als Auswertungsmethode ist eine zweifaktorielle Varianzanalyse (ohne Messwiederholung) / mehrfaktorielle Varianzanalyse mit einem drei- und einem zweistufigen Faktor geeignet. *Bewertung: 1 Punkt für Nennung des Designs (Teilfrage 1.a); 1 Punkt für Nennung der unabhängigen Variablen sowie abhängigen Variable (Teilfrage 1.b); 1 Punkt für Nennung der Auswertungsmethode (Teilfrage 1.c).*

2. Eine Studie hat untersucht, inwiefern die Anstrengungsbereitschaft von Schülern zum Lernen auf eine Mathematikschulaufgabe (intervallskaliertes Merkmal) vom Interesse für das Fach (intervallskaliertes Merkmal), dem Vorwissen (intervallskaliertes Merkmal) sowie dem Geschlecht (dichotomes Merkmal) abhängt. Zusätzlich ist von Interesse, welcher der Faktoren im Hinblick auf die Anstrengungsbereitschaft am bedeutendsten ist. *Was ist/sind die unabhängige/n Variable/n, was die abhängige/n Variable/n? Wie können die Fragestellungen statistisch beantwortet werden? Bitte begründen Sie ihre Antwort/en kurz.*

Antwort:

Für die Beantwortung der Fragestellungen ist eine multiple Regression mit dem Kriterium Anstrengungsbereitschaft der Schüler/innen sowie den Prädiktoren Interesse für das Fach, Vorwissen sowie Geschlecht geeignet. Die relative Bedeutung der Prädiktoren für die Vorhersage der Anstrengungsbereitschaft kann mittels der Beta-Gewichte bestimmt werden.

Bewertung: 1 Punkt für Nennung der multiplen Regression (Teilfrage 2.a), 1 Punkt für Nennung von Kriterium und Prädiktoren oder abhängigen Variable und unabhängigen Variablen (Teilfrage 2.b); 1 Punkt für Nennung der Beta-Gewichte (Teilfrage 2.c).

Anhang B: Wissenstests zum Posttest mit Lösung

Code (erster Buchstabe des Vornamens der Mutter, erster Buchstaben des Vornamens des Vaters, Tag des Geburtsdatums (Zahl), letzte Buchstabe des eigenen Vornamens):

Arbeitsplatznummer (links oben auf PC; eine Zahl von 01 bis 16): _____

Test zur Erfassung konzeptuellen Wissens zum Posttest (richtige Antworten sind kursiv gedruckt; 1 Punkt pro richtiger Antwort; insgesamt 20 Punkte)

Bitte kreuzen Sie bei den folgenden Fragen die richtigen Antworten (Buchstaben) an. Bei jeder Frage können mehrere Antworten richtig sein. Für jede richtige Antwort bzw. nicht angekreuzte falsche Antwort gibt es einen Punkt.

1. **Welche der folgenden Aussagen treffen im Hinblick auf den Begriff Skalenniveau zu?**
 - a) Platzierungen bei einer Olympiade liegen auf Nominalskalenniveau vor.
 - b) *Liegen Daten auf Ordinalskalenniveau vor, kann der Median als Maß der zentralen Tendenz dienen.*
 - c) Bei nominalskalierten Daten beinhaltet der Mittelwert am meisten Informationen.
 - d) *IQ-Werte können als Beispiel für intervallskalierte Daten angesehen werden.*
 - e) *Daten auf Intervallskalenniveau haben einen höheren Informationsgehalt als Daten auf Ordinalskalenniveau.*

2. **Bitte kreuzen Sie die Voraussetzungen an, die bei einer Varianzanalyse ohne Messwiederholung mit zwei unabhängigen Variablen (Faktoren) gegeben sein müssen:**
 - a) *Die abhängige Variable ist mindestens intervallskaliert.*
 - b) Die Interaktion zwischen den beiden Faktoren sollte hybrid sein.
 - c) *Auf allen Stufen der Faktoren ist die abhängige Variable normalverteilt.*
 - d) Die abhängige Variable ist über alle Stufen nur eines Faktors hinweg normalverteilt.
 - e) *Die unabhängige Variable muss nominal oder ordinalskaliert, also kategorial sein.*

3. Bitte kreuzen Sie Zutreffendes zum Thema Kovarianzanalyse an!

- a) Kovariaten können nur unter bestimmten Umständen kausal interpretiert werden.
- b) *Kontrollvariablen üben neben den unabhängigen Variablen einen Einfluss auf die abhängige Variable aus.*
- c) *Bei einer Kovarianzanalyse wird der Einfluss einer Kovariaten aus dem Einfluss einer oder mehrerer unabhängiger Variablen auf eine abhängige Variable herausgerechnet.*
- d) Die Kovariate darf bei einer Kovarianzanalyse auf Nominalskalenniveau vorliegen.
- e) *Bei einer Kovarianzanalyse muss außer den Voraussetzungen für eine Varianzanalyse berücksichtigt werden, dass der Einfluss der Kovariaten in den Stufen der unabhängigen Variablen homogen sein muss.*

4. Welche der folgenden Aussagen in Bezug auf die Regression treffen zu?

- a) *Mit Hilfe einer Regressionsgleichung können bei einer linearen Regression die Werte in einem Kriterium durch die Werte eines Prädiktors vorhergesagt werden.*
- b) Bei einer multiplen Regression können im Gegensatz zur einfachen Regression mehrere Kriterien einbezogen werden.
- c) *Bei Längsschnittdaten ist es möglich, dass bei einer linearen Regression Kausalaussagen getroffen werden.*
- d) *Bei einer einfachen linearen Regression muss das Kriterium mindestens intervallskaliert sein.*
- e) *Um Kollinearität zu vermeiden, sollten keine Prädiktoren in die Regression eingeschlossen werden, die hoch miteinander korrelieren.*

Test zur Erfassung anwendungsorientierten Wissens zum Posttest (1 Punkt pro richtiger Antwort; insgesamt 6 Punkte)

1. Die Effektivität zweier Diätmaßnahmen soll evaluiert werden. Hierzu wird das Gewicht der Teilnehmer/innen in drei Gruppen (zwei Diätmaßnahmen, eine Kontrollgruppe) jeweils vor und nach der Diätmaßnahme gemessen. Neben einem Vergleich der Effektivität der unterschiedlichen Diätmaßnahmen soll bestimmt werden, ob alle

Gruppen im Mittel über die Zeit eine Gewichtsabnahme zeigen und die Gewichtsabnahme in den drei Gruppen von der Zeit abhängt. *Um welche Art von Design handelt es sich bei der vorliegenden Untersuchung? Wie kann die Fragestellung statistisch beantwortet werden? Bitte begründen Sie ihre Antworten kurz.*

Antwort:

Es handelt sich um ein zweifaktorielles (2x3 oder 3x2) Design (mit Messwiederholung) oder Prä-Post-Kontrollgruppenplan mit den unabhängigen Variablen oder Faktoren Zeitpunkt (zweistufig) sowie Art der Diätmaßnahme (dreistufig). Als Auswertungsmethode ist eine zweifaktorielle Varianzanalyse (mit Messwiederholung) / mehrfaktorielle Varianzanalyse mit einem drei- und einem zweistufigen Faktor oder *gemischtes Design* geeignet. Es können Haupteffekte für den Messzeitpunkt sowie die Art der Diätmaßnahme als auch ein Interaktionseffekt zwischen diesen beiden Faktoren bestimmt werden.

Bewertung: 1 Punkt für Nennung des Designs (Teilfrage 1.a); 1 Punkt für Nennung der unabhängigen Variablen sowie abhängigen Variable oder Faktoren, auch wenn diese dargestellt wurden (Teilfrage 1.b); 1 Punkt für Nennung der Auswertungsmethode (Teilfrage 1.c).

2. Ein Personalleiter möchte mit einer Untersuchung herausfinden, von welchen Faktoren die Hilfsbereitschaft der Arbeiter/innen einer Abteilung abhängt. Hierzu erhebt er das Vertrauen der Mitarbeiter untereinander, das Commitment der Mitarbeiter zum Unternehmen sowie das von den Mitarbeitern eingeschätzte Arbeitsklima. Der Personalleiter vermutet, dass einige der Variablen leicht miteinander korrelieren und möchte herausfinden, von welcher Variable die Hilfsbereitschaft am meisten abhängt. *Was ist/sind die unabhängige/n Variable/n, was die abhängige/n Variable/n? Wie können die Fragestellungen statistisch beantwortet werden? Bitte begründen Sie ihre Antwort/en kurz.*

Antwort:

Für die Beantwortung der Fragestellungen ist eine multiple Regression mit dem Kriterium Hilfsbereitschaft der Arbeiter/innen sowie den Prädiktoren Vertrauen der Mitarbeiter untereinander, das Commitment der Mitarbeiter zum Unternehmen sowie das von den Mitarbeitern eingeschätzte Arbeitsklima geeignet. Die relative Bedeutung der Prädiktoren für die Vorhersage der Anstrengungsbereitschaft kann mittels der Beta-Gewichte bestimmt werden. Der Prädiktor mit dem höchsten Beta-Gewicht ist für die Vorhersage der Anstrengungsbereitschaft am bedeutsamsten.

Bewertung: 1 Punkt für Nennung der multiplen Regression (Teilfrage 2.a); 1 Punkt für Nennung von Kriterium und Prädiktoren oder abhängige Variable und unabhängige Variablen (Teilfrage 2.b); 1 Punkt für Nennung der Beta-Gewichte (Teilfrage 2.c).

Anhang C: Fragebogen zur aktuellen Motivation (FAM; Rheinberg et al., 2001)

Code (erster Buchstabe des Vornamens der Mutter, erster Buchstaben des Vornamens des Vaters, Tag des Geburtsdatums (Zahl), letzte Buchstabe des eigenen Vornamens):

Arbeitsplatznummer (links oben auf PC; eine Zahl von 01 bis 16): _____

Nun wollen wir wissen, wie deine **momentane Einstellung** zu der beschriebenen Aufgabe ist. Dazu findest du auf dieser Seite Aussagen. Kreuze bitte jene Zahl an, die auf dich am Besten passt.

	trifft				trifft		
	nicht zu				zu		
1. Ich mag solche Rätsel und Knobeleyen. (I)	1	2	3	4	5	6	7
2. Ich glaube, der Schwierigkeit dieser Aufgabe gewachsen zu sein. (E)	1	2	3	4	5	6	7
3. Wahrscheinlich werde ich die Aufgabe nicht schaffen. (E)	1	2	3	4	5	6	7
4. Bei der Aufgabe mag ich die Rolle des Wissenschaftlers, der Zusammenhänge entdeckt. (I)	1	2	3	4	5	6	7
5. Ich fühle mich unter Druck, bei der Aufgabe gut abschneiden zu müssen. (M)	1	2	3	4	5	6	7
6. Die Aufgabe ist eine richtige Herausforderung für mich. (H)	1	2	3	4	5	6	7
7. Nach dem Lesen der Instruktion erscheint mir die Aufgabe sehr interessant. (I)	1	2	3	4	5	6	7
8. Ich bin sehr gespannt darauf, wie gut ich hier abschneiden werde. (H)	1	2	3	4	5	6	7
9. Ich fürchte mich ein wenig davor, dass ich mich hier blamieren könnte. (M)	1	2	3	4	5	6	7
10. Ich bin fest entschlossen, mich bei dieser Aufgabe voll anzustrengen. (H)	1	2	3	4	5	6	7
11. Bei Aufgaben wie dieser brauche ich keine Belohnung, sie machen mir auch so viel Spaß. (I)	1	2	3	4	5	6	7
12. Es ist mir etwas peinlich, hier zu versagen. (M)	1	2	3	4	5	6	7
13. Ich glaube, dass kann jeder schaffen. (E)	1	2	3	4	5	6	7
14. Ich glaube, ich schaffe diese Aufgabe nicht. (E)	1	2	3	4	5	6	7
15. Wenn ich die Aufgabe schaffe, werde ich schon ein wenig stolz auf meine Tüchtigkeit sein. (H)	1	2	3	4	5	6	7

Anhang C: Fragebogen zur aktuellen Motivation

16. Wenn ich an die Aufgabe denke, bin ich etwas beunruhigt. **(M)** 1 2 3 4 5 6 7
17. Eine solche Aufgabe würde ich auch in meiner Freizeit bearbeiten. **(I)** 1 2 3 4 5 6 7
18. Die konkreten Leistungsanforderungen hier lähmen mich. **(M)** 1 2 3 4 5 6 7

(I): Skala Interesse

(M): Skala Misserfolgsbefürchtung

(E): Skala Erfolgswahrscheinlichkeit

(H): Skala Herausforderung

Anhang D: Statistische Probleme und zugehörige Informationen

Herzlich willkommen zu den Aufgaben mit statistischen Problemen!

Auf den folgenden Folien bekommen Sie drei statistische Probleme präsentiert.

Zu jeder Problemstellung finden Sie auf den darauf folgenden Folien die Lösungsschritte und im Anschluss daran Informationen zur Lösung der Probleme.

Bitte begründen Sie die Lösungsschritte auf dem Antwortbogen.

Problem 1

Der Kognitionsforscher Max Musterhirn möchte herausfinden, ob die Anwesenheit einer Person einen Einfluss auf die Effektivität eines kognitiven Trainings hat. Als Probanden hat er Studierende der Pädagogik und Psychologie ($N = 120$) akquiriert, die er zufällig zwei Gruppen zuwies. Eine Gruppe absolvierte ein kognitives Training über einen Zeitraum von 4 Wochen allein in einem Labor, die andere Gruppe absolvierte dasselbe kognitive Training über den gleichen Zeitraum, allerdings unter der Aufsicht eines Experimentators. Die Trainingsleistung (intervallskaliertes Merkmal) hat Max Musterhirn zu Beginn, nach zwei Wochen Training und am Ende des Trainings erhoben. Da er noch keine Untersuchung mit dem kognitiven Training an einer studentischen Stichprobe durchgeführt hat, ist Max Musterhirn daran interessiert, ob sich die Trainingsleistung der Versuchspersonen zwischen den drei Messzeitpunkten signifikant unterscheidet. Er vermutet zudem, dass der Effekt der Anwesenheit eines Experimentators nicht sofort erkennbar ist, sondern von der Dauer des Trainings abhängt. Um in Hinblick auf die statistische Auswertung korrekt vorzugehen, möchte Max Musterhirn auch die Voraussetzungen für seine Analyse berücksichtigen. *Wie kann die Fragestellung von Max Musterhirn statistisch beantwortet werden? Bitte begründen Sie ihre Antworten wenn möglich.*

Informationen zu Problem 1

Für den Vergleich von Mittelwerten aus mehr als zwei Stichproben gibt es verschiedene statistische Verfahren. Von Bedeutung sind die Anzahl zu vergleichender Stichproben und ob diese unabhängig voneinander sind. Für den Vergleich mehrerer unabhängiger Stichproben, die Realisationen eines Faktors bzw. einer unabhängigen Variablen sind, ist eine einfaktorielles Varianzanalyse ohne Messwiederholung geeignet. Dabei liegt ein einfaktorielles Design vor, da nur Mittelwerte aus den Stufen eines Faktors bzw. einer unabhängigen Variable betrachtet werden. Wird eine Stichprobe zu mehreren Zeitpunkten untersucht, sind die Werte in der abhängigen Variablen zu den verschiedenen Zeitpunkten abhängig voneinander. In diesem Fall liegt ebenfalls ein einfaktorielles Design mit einem Messwiederholungsfaktor (unabhängige Variable) vor. Vergleicht man dagegen Mittelwerte aus Kombinationen von Faktorstufen, liegt ein mehrfaktorielles Design vor. Hat man bspw. drei unabhängige Variablen bzw. Faktoren mit jeweils 3 Stufen, liegt ein $3 \times 3 \times 3$ -faktorielles Design vor. Die Auswertung erfolgt dann mit einer dreifaktoriellen Varianzanalyse ohne Messwiederholung. Einen Spezialfall stellt die Kombination eines Messwiederholungsfaktors und einem oder mehreren anderen Faktoren dar. In diesem Fall kann man auch von einem gemischten Design sprechen. Eine Auswertung kann mit einer mehrfaktorielles Varianzanalyse mit Messwiederholung erfolgen. Wäre in dem Beispiel mit drei Faktoren ein Faktor ein Messwiederholungsfaktor (mit drei Stufen), könnte man Unterschiede zwischen den Mittelwerten aus den Faktorstufenkombinationen mit einer dreifaktoriellen Varianzanalyse mit Messwiederholung feststellen.

Informationen zu Problem 1

Die Durchführung von Varianzanalysen ist an bestimmte Voraussetzungen gebunden. Zum einen muss die abhängige Variable intervallskaliert sein. Ordinale Daten wie Rangplätze können folglich nicht mit einer Varianzanalyse ausgewertet werden. Hierfür sind nonparametrische Verfahren wie die Rangvarianzanalyse nach Kruskal und Wallis geeignet. Eine weitere Voraussetzung stellt die Normalverteilung der Werte in der abhängigen Variablen über alle Stufen(kombinationen) der unabhängigen Variablen dar. Bei dem oben genannten 3x3x3-Design müssen folglich die Werte aus allen Faktorstufenkombinationen (insgesamt 27) aus normalverteilten Grundgesamtheiten stammen. Die Stichprobenwerte in den einzelnen Faktorstufenkombinationen müssen also normalverteilt sein. Ein Test zur Überprüfung dieser Annahme ist der Kolmogorov-Smirnov-Test. Die Werte in allen Faktorstufenkombinationen müssen zudem aus Populationen mit gleichen Varianzen stammen. Die Varianzen der Stichprobenwerte in den einzelnen Faktorstufenkombinationen müssen also homogen sein. Ein Test zur Überprüfung dieser Annahme ist der Levene-Test. Bei Designs mit Messwiederholung ist zudem die sog. Sphäritätsannahme zu berücksichtigen. Diese besagt, dass die Varianzen der Differenzen zwischen verschiedenen Messzeitpunkten gleich sein müssen. Hat man beispielsweise einen Messwiederholungsfaktor mit vier Stufen müssen die Varianzen der Differenzen zwischen den vier Messzeitpunkten gleich sein. Diese Annahme kann mit Hilfe des Mauchly-Test auf Sphärität überprüft werden. Bei einer Verletzung der Sphäritätsannahme empfiehlt sich eine Korrektur der Freiheitsgrade der Prüfgröße.

Problem 2

Prof. Dr. Marc Löffler möchte in diesem Semester weniger Zeit für die Korrektur von Hausarbeiten aufwenden. Deshalb sollen seine Studenten die Arbeiten, anstatt alleine, in Vierergruppen erstellen. Damit alle Studierenden fair und möglichst gut bewertet werden können, überlegt er die Gruppen systematisch zu bilden. Er möchte wissen, ob sich die Gruppennote (intervallskaliert) aus der mittleren Problemlösekompetenz (intervallskaliert) und der mittleren Studienmotivation (intervallskaliert) der Mitglieder vorhersagen lässt und würde die Gruppen hinsichtlich dieser Kriterien homogen bilden. Damit will er verhindern, dass einige Gruppen nur aus sehr demotivierten Studenten mit wenig Problemlösefähigkeit bestehen, während andere Gruppen hinsichtlich dieser beiden Kennwerte sehr hohe Ausprägungen aufweisen. Zudem ist Prof. Dr. Löffler daran interessiert, ob die Problemlösekompetenz oder die Studienmotivation besser zur Vorhersage der Gruppennote geeignet ist. Um vorher entscheiden zu können, ob dieser Aufwand der systematischen Gruppenbildung überhaupt sinnvoll ist, will er Daten von seinem Kollegen Prof. Dr. Immertreu untersuchen. Dieser hatte im vorigen Semester die Gruppeneinteilungen dem Zufall überlassen und die mittlere Gruppen-Problemlösekompetenz, Motivation und die Gruppennote zu einem Zeitpunkt (Querschnittsdesign) erhoben. Die Stichprobe von Prof. Dr. Immertreu umfasst 180 Versuchspersonen. Der Anteil an Männern und Frauen ist in etwa ausgeglichen. Aus den Daten von Prof. Dr. Immertreu weiß Prof. Dr. Löffler, dass die Problemlösekompetenz mit der Studienmotivation nicht all zu hoch ($r = .03$) korreliert ist. Falls Prof. Dr. Löffler feststellen sollte, dass sich die vergebenen Noten signifikant ($p = 0.05$) durch die Problemlösekompetenz und Motivation der Gruppenmitglieder vorhersagen lassen, will er seine Gruppen in dieser Hinsicht homogen bilden. Um in Hinblick auf die statistische Auswertung korrekt vorzugehen, möchte Prof. Dr. Marc Löffler auch die Voraussetzungen für seine Analyse und Kausalitätsschlüsse berücksichtigen. *Wie kann die Fragestellung von Prof. Dr. Marc Löffler statistisch beantwortet werden? Bitte begründen Sie ihre Antworten wenn möglich.*

Informationen zu Problem 2

Der Zusammenhang zwischen einer Variablen und einer anderen Variablen kann durch eine Korrelation quantifiziert werden. Eine Korrelation zwischen zwei Variablen bedeutet, dass bestimmte Werte auf einer Variablen mit bestimmten Werten auf einer anderen Variablen einhergehen. Im Falle einer positiven linearen Korrelation sind hohe Werte auf der einen Variablen mit hohen Werten auf der anderen Variablen verbunden. Im Falle einer negativen Korrelation treten bei hohen Werten auf der einen Variablen eher niedrige Werte auf der anderen Variablen auf. Da eine Korrelation einen ungerichteten Zusammenhang quantifiziert, sind keine Kausalaussagen möglich.

Prinzipiell gilt das gleiche für eine einfache lineare Regression. Bei einer einfachen linearen Regression werden die Werte in einer abhängigen Variablen, auch Kriterium genannt, durch die Werte einer unabhängigen Variablen, auch Prädiktor genannt, vorhergesagt. Werden Kriterium und Prädiktor zu einem Messzeitpunkt erhoben, können keine Kausalaussagen getroffen werden. Kriterium und Prädiktor könnten in der Regression nämlich auch vertauscht werden, wodurch das ehemalige Kriterium (jetzt Prädiktor) zur Vorhersage des ehemaligen Prädiktors (jetzt Kriterium) dient. Nur bei Längsschnittdaten sind unter bestimmten Bedingungen Kausalaussagen möglich. Dies gilt auch für multiple Regressionen.

Informationen zu Problem 2

Erweiterungen der einfachen linearen Regression stellen die multiple logistische Regression sowie die multiple lineare Regression dar. Sie erlauben den Einschluss mehrerer Prädiktoren, die kategorial (nominal- oder ordinalskaliert) oder intervallskaliert (bzw. höheres Skalenniveaus) sein können. Die multiple logistische Regression wird bei einer dichotomen abhängigen Variablen bzw. Kriterium verwendet, die multiple lineare Regression bei einer mindestens intervallskalierten abhängigen Variablen. Bei beiden Arten von Regressionen kann die relative Vorhersagekraft eines Prädiktors im Vergleich zu einem anderen Prädiktor durch das sog. Beta-Gewicht bestimmt werden. Bei einem Beta-Gewicht für einen Prädiktor wird die Stärke der Vorhersage eines Kriteriums durch diesen Prädiktor bestimmt, wobei der gemeinsame Anteil des Prädiktors mit anderen Prädiktoren herausgerechnet wird.

Sowohl für die multiple logistische als auch die multiple lineare Regression müssen bestimmte Voraussetzungen gelten. Unter anderem sollten bei beiden Arten von Regressionen keine Kollinearität vorliegen, d.h., die Prädiktoren sollten nicht zu hoch miteinander korrelieren. Dies kann z.B. mit Konditionsindizes überprüft werden. Zudem sollte der Zusammenhang zwischen jedem Prädiktor und dem Kriterium sowie den Prädiktoren untereinander bei einer multiplen linearen Regression linear sein, was mit Hilfe von Streudiagrammen überprüft werden kann. Neben linearen Zusammenhängen sind bspw. auch hyperbolische Zusammenhänge denkbar. Sind die Zusammenhänge zwischen Prädiktoren und Kriterium nicht linear, gibt es verschiedene Methoden mit denen nicht-lineare Regressionen berechnet werden können.

Problem 3

Der junge Forscher Harald möchte der Frage nachgehen, inwiefern Schulleistungsunterschiede in Mathematik zwischen drei Klassen (Klassengrößen: $N_1=N_2=N_3=30$) durch Unterschiede in der Motivation bedingt sind. Hierzu erhebt er neben der Mathematikleistung (intervallskaliert) die Motivation (intervallskaliert) sowie die Variablen Geschlecht (nominalskaliert), Alter (rationalskaliert) und IQ (intervallskaliert) der Schüler/innen. Harald geht aufgrund von Ergebnissen aus früheren Studien davon aus, dass die Schulleistungswerte der Schüler neben der Motivation auch noch von der Intelligenz abhängen. Dies bestätigte sich auch in ersten Analysen, die zeigten, dass die Intelligenz einen positiven Einfluss auf die Schulleistungen in den Klassen hatte. Dieser Einfluss war jedoch von Klasse zu Klasse etwas verschieden. Harald möchte nun möglicherweise vorhandene Unterschiede in den Schulleistungen nur auf die Motivation zurückführen. Um in Hinblick auf die statistische Auswertung korrekt vorzugehen, möchte Harald auch die Voraussetzungen für seine Analyse berücksichtigen. *Wie kann die Fragestellung von Harald statistisch beantwortet werden? Bitte begründen Sie ihre Antworten wenn möglich.*

Informationen zu Problem 3

Möchte man Mittelwerte zwischen verschiedenen unabhängigen Stichproben statistisch vergleichen, hat man oft das Problem, dass andere Variablen (sogenannte Kovariaten oder Kontrollvariablen) als die Gruppierungsvariable einen Einfluss auf die abhängige Variable haben. D.h., die Korrelation zwischen der Kovariaten und der abhängigen Variable ist signifikant. Möchte man den Einfluss von Kovariaten auf die abhängige Variable herausrechnen, gibt es mehrere Möglichkeiten. Ein Weg besteht in der Berechnung von Semipartialkorrelationen, bei denen der Einfluss von Kovariaten in einer der beiden Variablen, für die ein Zusammenhang berechnet werden soll, herauspartialisiert wird. Im Gegensatz dazu wird bei Partialkorrelationen der Einfluss von Kovariaten bei beiden Variablen, zwischen denen ein Zusammenhang berechnet werden soll, kontrolliert. Zu beachten ist bei allen Korrelationen, dass diese keine kausalen Schlüsse zulassen. Korrelationen stellen somit ungerichtete Zusammenhänge dar. Quadriert man Korrelationen, kann der gemeinsame Varianzanteil zwischen zwei Variablen bestimmt werden. Ein anderer Weg als die Berechnung von Semipartialkorrelationen besteht darin, aus den Quadratsummen der normalen Varianzanalyse zum Vergleich von mehreren Stichproben mittels Regressionen neue Quadratsummen zu berechnen, welche um den Einfluss der Kovariaten bereinigt sind.

Informationen zu Problem 3

Für die Quadratsumme innerhalb der Gruppen wird dabei ein gemeinsamer Steigungskoeffizient (Regressionskoeffizient) verwendet. Folglich wird eine einfaktorielle Varianzanalyse mit den neuen Quadratsummen berechnet. Diese Art der Varianzanalyse wird auch Kovarianzanalyse genannt. Die Voraussetzungen der Kovarianzanalyse sind bis auf einige Besonderheiten die gleichen wie bei einer normalen Varianzanalyse. So muss die Kovariate mindestens intervallskaliert sein, was nicht immer gegeben ist. Der Einfluss der Kovariate auf die abhängige Variable muss über die Stufen der unabhängigen Variablen hinweg homogen sein. D.h., die Regressionskoeffizienten bzw. Steigungen dürfen sich nicht über die Gruppen hinweg unterscheiden. Die Verletzung spielt allerdings dann eine geringe Rolle im Hinblick auf den Alpha-Fehler und die Teststärke, wenn die Gruppengrößen gleich sind.

Anhang E: Lösungsbeispiele

Lösung Problem 1

1. Schritt: Bestimmung des Designs und der unabhängigen Variablen sowie der abhängigen Variable

Es handelt sich um ein zweifaktorielles (2x3 bzw. 3x2-faktorielles) Design mit den unabhängigen Variablen (UVs) bzw. den Faktoren „Anwesenheit einer Person während des Trainings“ (zweistufig, Ja vs. Nein) sowie Messzeitpunkt (dreistufig). Die abhängige Variable (AV) ist die Trainingsleistung.

Lösung Problem 1

2. Schritt: Auswahl der statistischen Auswertungsmethode

Für die statistische Auswertung ist eine zweifaktorielle Varianzanalyse mit Messwiederholung (gemischtes Design) geeignet.

Lösung Problem 1

3. Schritt: Berücksichtigung der Voraussetzungen

Es muss die Annahme erfüllt sein, dass die abhängige Variable intervallskaliert ist und die Werte der abhängigen Variablen in allen Faktorstufen(kombinationen) aus normalverteilten Grundgesamtheiten stammt, also die Stichprobendaten in den einzelnen Faktorstufen(kombinationen) jeweils normalverteilt sind. Des Weiteren sollte die Voraussetzung der Sphärizität (Homogenität der Varianzen der Differenzen zwischen den einzelnen Messzeitpunkten) erfüllt sein.

Lösung Problem 2

1. Schritt: Bestimmen der unabhängigen Variablen (UVs) und abhängigen Variable (AV)

AV: Gruppennote

UV 1: durchschnittliche Problemlösekompetenz

UV 2: durchschnittliche Motivation

Lösung Problem 2

2. Schritt: Auswahl des geeigneten Testverfahrens

Für die statistische Auswertung eignet sich eine multiple Regression mit den Prädiktoren (UVs) Problemlösekompetenz und Motivation und dem Kriterium (AV) Gruppennote.

Lösung Problem 2

3. Schritt: Berücksichtigung der Voraussetzungen und der Kausalität der Ergebnisse

Es sollte keine Kollinearität vorliegen, d.h., die Prädiktoren sollten nicht zu hoch miteinander korrelieren. Zudem sollte der Zusammenhang zwischen jedem Prädiktor und dem Kriterium sowie den Prädiktoren untereinander linear sein. Es dürfen keine Kausalaussagen getroffen werden.

Lösung Problem 3

1. Schritt: Bestimmung des Designs und der unabhängigen Variable sowie der abhängigen Variable

Es handelt sich um ein einfaktorielles Design mit der unabhängigen Variable (UV) bzw. dem dreistufigen Faktor Klasse sowie der abhängigen Variable (AV) Motivation. Der IQ stellt eine Kovariate dar.

Lösung Problem 3

2. Schritt: Auswahl der statistischen Auswertungsmethode

Für die statistische Auswertung ist eine Kovarianzanalyse mit der Kovariate IQ geeignet.

Lösung Problem 3

3. Schritt: Berücksichtigung der Voraussetzungen

Berücksichtigt werden muss der in Abhängigkeit der unterschiedlichen Klassen variierende Einfluss der Intelligenz auf die Mathematikleistungen. Dieser Einfluss darf nicht signifikant sein.

Anhang F: Antwortbogen zu den statistischen Problemen

Code (erster Buchstabe des Vornamens der Mutter, erster Buchstaben des Vornamens des Vaters, Tag des Geburtsdatums (Zahl), letzte Buchstabe des eigenen Vornamens):

Arbeitsplatznummer (links oben auf PC; eine Zahl von 01 bis 16): _____

Antwortbogen zu den Aufgaben mit statistischen Problemen

Problem 1

Problem 2

Problem 3

Anhang G: Cognitive load Skala (Paas, 1992; ins Deutsche übersetzt)

Code (erster Buchstabe des Vornamens der Mutter, erster Buchstaben des Vornamens des Vaters, Tag des Geburtsdatums (Zahl), letzte Buchstabe des eigenen Vornamens):

Bitte kreuzen Sie an, wie hoch die kognitive Belastung bei der Bearbeitung der statistischen Probleme war.

1. sehr, sehr niedrige kognitive Belastung
2. sehr niedrige kognitive Belastung
3. niedrige kognitive Belastung
4. eher niedrige kognitive Belastung
5. weder niedrige noch hohe kognitive Belastung
6. eher hohe kognitive Belastung
7. hohe kognitive Belastung
8. sehr hohe kognitive Belastung
9. sehr, sehr hohe kognitive Belastung

Anhang H: Korrelationen innerhalb der Aufgaben zur Erfassung der Arbeitsgedächtniskapazität und der Shifting-Fähigkeit

Korrelationen zwischen Aufgaben zur Erfassung der AG-Kapazität

		aospan	asymspan	arspan
aospan	<i>r</i>	1	.498**	.657**
	<i>N</i>	76	76	76
asymspan	<i>r</i>	.498**	1	.452**
	<i>N</i>	76	76	76
arspan	<i>r</i>	.657**	.452**	1
	<i>N</i>	76	76	76

Anmerkung. *N* = Stichprobengröße.

** $p < .01$.

Korrelationen zwischen Aufgaben zur Erfassung der Shifting-Fähigkeit

		category-switch	color-shape	number-letter
category-switch	<i>r</i>	1	.431**	.565**
	<i>N</i>	76	76	76
color-shape	<i>r</i>	.431**	1	.436**
	<i>N</i>	76	76	76
number-letter	<i>r</i>	.565**	.436**	1
	<i>N</i>	76	76	76

Anmerkung. *N* = Stichprobengröße.

** $p < .01$.