



Universität Trier
Fachbereich I – Psychologie

Dissertation zur Erlangung des akademischen Doktorgrades der
Naturwissenschaften

Alterskorrelierte Veränderungen von falschen Erinnerungen?

Eine experimentelle Studie an Kindern und jungen Erwachsenen

vorgelegt von

Anette Jammers

Begutachtung von

Prof. Dr. Karl Friedrich Wender
Dr. Silvia Mecklenbräuer

Trier, im Oktober 2007

Inhaltsverzeichnis

Einleitung	1
I. Theoretischer Hintergrund	5
1 Das Forschungsfeld der Gedächtnisillusionen.....	7
1.1 Begriffsbestimmung zu den Gedächtnisillusionen.....	7
1.2 Begriffsbestimmung zu falschen Erinnerungen	9
1.3 Historischer Überblick	12
2 Das Deese/Roediger-McDermott (DRM)-Paradigma	19
2.1 Entwicklung und Bewertung des DRM-Paradigmas	19
2.1.1 Forschungsarbeiten von Deese (1959a, 1959b): Die Basis des DRM-Paradigmas.....	19
2.1.2 Forschungsarbeiten von Roediger und McDermott (1995).....	21
2.1.3 Bewertung des DRM-Paradigmas	23
2.2 Theoretische Erklärungsansätze.....	26
2.2.1 Diverse Ansätze.....	26
2.2.2 Der Aktivierungs-Monitoring Ansatz (AMA)	30
2.2.3 Vergleich des Aktivierungs-Monitoring Ansatzes (AMA) mit anderen Theorien im DRM-Paradigma	37
2.3 Einflussfaktoren auf falsche Erinnerungen	43
2.3.1 Rückwärtsassoziationsstärke (RAS)	46
2.3.2 Länge der Wortlisten.....	46
2.3.3 Instruktion	47
2.3.4 Güte des Listenlernens	48
2.3.5 Ausmaß der konzeptuellen Verarbeitung.....	49
2.3.6 Testart.....	51
2.3.7 Beeinflussung durch den vorherigen Test.....	52

3	Gedächtnisentwicklung im Kindesalter	53
3.1	Gedächtniskapazität	54
3.2	Gedächtnisstrategien	56
3.2.1	Enkodierstrategien.....	56
3.2.2	Abrufstrategien.....	57
3.3	Metagedächtnis.....	58
3.3.1	Deklaratives Metagedächtnis	58
3.3.2	Prozedurales Metagedächtnis.....	59
3.4	Bereichsspezifisches Wissen.....	61
4	Alterskorrelierte Veränderungen falscher Erinnerungen.....	64
4.1	Vorhersagen des Aktivierungs-Monitoring Ansatzes	64
4.2	Studien und deren Probleme	67
5	Prozeduren zur getrennten Erfassung von Aktivierung und Quellendiskrimination	74
5.1	Zeitdruck: Ausschluss von Monitoringprozessen	74
5.2	Implizite Gedächtnistests	75
5.3	Förderung von Aktivierung durch Lerninstruktionen.....	77
5.4	Inklusionsinstruktionen	78
6	Ziele der eigenen Studie	82
II.	Eigene Untersuchungen	84
7	Alterskorrelierte Differenzen in impliziten Tests?.....	85
7.1	Experiment 1	85
7.1.1	Vorversuch	87
7.1.2	Methode.....	89
7.1.3	Ergebnisse	92
7.1.3.1	Implizite Tests.....	93
7.1.3.2	Explizite Tests: Wiedererkennen	95
7.1.4	Diskussion	101

8	Alterskorrelierte Differenzen im falschen Reproduzieren und im Wiedererkennen?	109
8.1	Experiment 2: Falsches Reproduzieren: Einflüsse von Aktivierung und Quellendiskrimination	109
8.1.1	Vorversuch	111
8.1.2	Methode	112
8.1.3	Ergebnisse	114
8.1.3.1	Listenitems	114
8.1.3.2	Kritische Items	116
8.1.4	Diskussion	121
8.2	Experiment 3A: Falsches Wiedererkennen: Einflüsse von Aktivierung und Quellendiskrimination	128
8.2.1	Vorversuch	129
8.2.2	Methode	130
8.2.3	Ergebnisse	132
8.2.3.1	Einflüsse der Testabfolge	133
8.2.3.2	Ergebnisse der separaten Analysen mit den zuerst erteilten Tests	137
8.2.4	Diskussion	143
8.3	Experiment 3B: Falsches Wiedererkennen: Effekte von Aktivierung fördernden Lerninstruktionen	148
8.3.1	Methode	151
8.3.2	Ergebnisse	151
8.3.2.1	Einflüsse der Testabfolge	152
8.3.2.2	Ergebnisse der separaten Analysen mit den zuerst erteilten Tests	155
8.3.3	Vergleich der Experimente 3A und 3B	164
8.3.3.1	Listenitems	164
8.3.3.2	Kritische Items	165
8.3.4	Diskussion	168
9	Alterskorrelierte Differenzen im Quellendiskriminationstest?	178
9.1	Experiment 4	178
9.1.1	Methode	181
9.1.2	Ergebnisse	183
9.1.2.1	Listenitems	183
9.1.2.2	Kritische Items	186
9.1.3	Diskussion	191

10	Gesamtdiskussion	199
10.1	Statistisch-methodische Probleme.....	199
10.2	Theoretische Einordnung der Befunde.....	200
10.2.1	Aktivierungs-Monitoring Ansatz (AMA)	200
10.2.2	Fuzzy-Trace Theorie (FTT)	207
10.3	Ausblick	209
11	Zusammenfassung	213
12	Literaturverzeichnis.....	216
Anhang		230
	Anhang A: Experiment 1	231
	Anhang B: Experiment 2.....	239
	Anhang C: Experimente 3A und 3B.....	244
	Anhang D: Experiment 4.....	250
	Anhang E: Eltern- und Schülerinformationsbriefe.....	257

Danksagung

An dieser Stelle möchte ich mich bei Frau Dr. Silvia Mecklenbräuer ganz herzlich für die hervorragende Betreuung bedanken, die mir mit engagierter Unterstützung in den verschiedenen Phasen der Dissertation hilfreich zur Seite stand.

Auch Herrn Prof. Dr. Karl Friedrich Wender danke ich für die Betreuung der Arbeit.

Mein Dank gilt ebenfalls den Direktoren, Schulleitern und Lehrern/innen, Leiterinnen und Erzieherinnen zahlreicher Gymnasien, Grundschulen und Kindergärten, die damit einverstanden waren, dass Untersuchungen an ihren Schulen und Kindergärten durchgeführt wurden und Räume zur Verfügung stellten. Vielen Dank nicht zuletzt auch an die Kinder, Jugendlichen und Freunde/innen, die interessiert und geduldig an den Experimenten teilnahmen.

Einen großen Dank möchte ich Frau Clementine Hüsken und Herrn Hans-Joachim Hüsken aussprechen, die mir die Planung der Untersuchung in vielen Schulen und Kindergärten erleichterten.

Ein herzlicher Dank meinem Mann Robert, der durch liebevolle Unterstützung, Interesse und Aufmunterungen die Arbeit begleitete.

Einleitung

Ein Mensch, vertrauend auf sein klares
Gedächtnis sagt getrost „So war es!“
Er ist ja selbst dabei gewesen –
Doch bald schon muss er's anders lesen.
Es wandeln sich, ihm unter Händen,
Wahrheiten langsam zu Legenden.
Des eigenen Glaubens nicht mehr froh
Fragt er sich zweifelnd: „War es so?“
Bis schließlich überzeugt er spricht:
„Ich war dabei – so war es nicht!“
Legendenbildung von Eugen Roth

Das Gedächtnis hilft uns das alltägliche Leben zu meistern. Ohne unser Gedächtnis müssten wir jeden Tag neu herausfinden, wer die Personen sind, denen wir begegnen, wie ein Auto oder ein Computer funktioniert. Die Fähigkeit, Wissen über uns und unsere Umwelt zu erwerben und zu behalten, stellt die Basis für Entscheiden und Handeln dar und ermöglicht eine effektive Anpassung an die Umwelt.

Jedoch können im Gedächtnis repräsentierte Erinnerungen fehlerhaft sein. Im Alltag zeigen sie sich in Erinnerungen an nie geschehene Ereignisse oder in verzerrten Erinnerungen von subjektiven Erfahrungen (Roediger & McDermott, 1995). Solche so genannten „falschen Erinnerungen“ stellen eine Gedächtnisillusion dar, welche durch eine starke Abweichung der Erinnerung von einem tatsächlichen Ereignis, resultierend aus Hinzufügungen, Auslassungen oder Substitutionen von Informationen, gekennzeichnet sind (Roediger, 1996).

Einhergehend mit dem Aufkommen zahlreicher Vorwürfe sexuellen und körperlichen Missbrauchs an Kindern wurden Reformen im Rechtssystem der USA in den 80er Jahren des vergangenen Jahrhunderts vorgenommen, wodurch sehr junge Kinder als Zeugen vor Gericht vernommen werden konnten. Jedoch wurde an der Glaubwürdigkeit von kindlichen Zeugenaussagen gezweifelt und Beeinflussungsmöglichkeiten wurden wahrscheinlich gehalten, weshalb sich eine systematische Untersuchung falscher Erinnerungen als nötig erwies, welche die gravierenden Folgen von Falschaussagen verhindern sollte. Während viele Studien sich auf die generelle Suggestibilität von Zeugenaussagen konzentrierten (Ceci & Bruck, 1993), zogen vermehrt die zugrunde liegenden Faktoren von Gedächtnisillusionen Interesse an sich.

Das vor über zehn Jahren entwickelte Paradigma zur experimentellen Erforschung falscher Erinnerungen, das **Deese/Roediger-McDermott-** (DRM) Paradigma, stellt ein zuverlässiges und schon vielfach angewandtes Vorgehen dar, um falsche Erinnerungen zu induzieren. Anhand von Wortlisten (z.B. „fahren, Straße, Reifen, Lenkrad, Bus, Führerschein, Benzin, Porsche, Motor, Unfall“), bei denen alle Wörter semantisch mit einem nicht präsentierten so genannten kritischen Wort („Auto“) assoziiert sind, werden falsche Erinnerungen an dieses Wort im Test der Freien Wiedergabe und des Wiedererkennens hervorgerufen (Roediger & McDermott, 1995).

Einer der derzeit am häufigsten diskutierten theoretischen Ansätze zur Erklärung falscher Erinnerungen im DRM-Paradigma ist der Aktivierungs-Monitoring Ansatz (AMA), welcher von zwei entgegengesetzt wirkenden Prozessen ausgeht: Lexikalische/semantische Aktivierung der kritischen Items (primär in der Lernphase) und Monitoring, genauer Quellendiskrimination (primär in der Testphase), d.h. Entscheidung, ob ein Item tatsächlich präsentiert (externe Quelle) oder selbst generiert wurde (interne Quelle). Während der Präsentation der Lernliste wird das kritische Item als Resultat einer Aktivierungsausbreitung im semantischen Netzwerk (Collins & Loftus, 1975) aktiviert und ist entsprechend leichter zugänglich. Nach dem AMA sollte die Aktivierung kritischer Items zu falschen Erinnerungen an diese führen, aber nur wenn es zu Fehlern bei der Quellendiskrimination kommt, d.h. wenn die Probanden die eigenständig generierten kritischen Items (interne Quelle) fälschlicherweise für Listenitems halten (externe Quelle).

Nach dem AMA lassen sich Vorhersagen zu alterskorrelierten Veränderungen falscher Erinnerungen treffen: Sowohl die Aktivierung als auch die Quellendiskrimination sollten sich mit zunehmendem Alter verbessern. Jüngere Kinder verfügen über geringere Fähigkeiten in der Quellendiskrimination und neigen folglich zu einer höheren Rate falscher Erinnerungen, allerdings *nur* dann, wenn zuvor die kritischen Items aktiviert worden sind. Wurden die kritischen Items nicht aktiviert, ergibt sich auch kein Quellendiskriminationsproblem in der Testphase. Demzufolge ist die Aktivierungskomponente für Vorhersagen zu alterskorrelierten Differenzen falscher Erinnerungen entscheidend.

In diesem Kontext lässt sich prüfen, ob jüngere Kinder tatsächlich, wie Befunde der Zeugenaussageforschung zeigen, empfänglicher für falsche Erinnerungen sind als ältere und Erwachsene. Bis zum Zeitpunkt der Planung der eigenen Untersuchungen liegen leider nur wenige publizierte Studien zu alterskorrelierten Veränderungen falscher Erinnerungen im Rahmen des DRM-Paradigmas vor, die zu konträren Ergebnissen gelangen (Brainerd, Reyna & For-

rest, 2002a; Ghetti, Qin & Goodman, 2002). Angesichts der spärlichen Befundlage alterskorrelierter Veränderungen falscher Erinnerungen sind weitere Untersuchungen erforderlich. Ausgehend vom AMA soll in der vorliegenden Arbeit geprüft werden inwieweit mögliche alterskorrelierte Differenzen falscher Erinnerungen auf entsprechende Unterschiede in den Fähigkeiten zur Quellendiskrimination und/oder auf Unterschiede in der Wahrscheinlichkeit der Aktivierung der kritischen Wörter zurückgehen.

Im ersten Teil der vorliegenden Arbeit wird der theoretische Hintergrund dargestellt mit dem Forschungsfeld der Gedächtnisillusionen (Kapitel 1), dem DRM-Paradigma (Kapitel 2) und dessen Entwicklung (2.1), theoretischen Erklärungsansätzen (2.2) sowie den Einflussfaktoren falscher Erinnerungen (2.3).

Die eigene Untersuchung basiert auf dem AMA und entsprechend seinen theoretischen Annahmen werden die Hypothesen formuliert. Da der AMA aber keine entwicklungspsychologische Theorie ist, sondern eine aus der Kognitionspsychologie stammende, müssen die Grundlagen der Gedächtnisentwicklung im Kindesalter hinzugezogen werden (Kapitel 3), um alterskorrelierte Annahmen zu formulieren und zu begründen. Daran anschließend werden aus theoretischer Sicht alterskorrelierte Veränderungen falscher Erinnerungen erläutert und die wenigen dazugehörigen Studien mit Kindern und Erwachsenen im DRM-Paradigma berichtet und problematisiert (Kapitel 4).

Ziel der eigenen Studie ist es zu untersuchen, inwieweit mögliche alterskorrelierte Differenzen falscher Erinnerungen auf entsprechende Unterschiede in der Quellendiskrimination und/oder auf Unterschiede in der Wahrscheinlichkeit der Aktivierung der kritischen Wörter zurückgehen. Nur durch die separate Erfassung der beiden Prozesse lassen sich differenzierte Aussagen zu einem möglichen Alterstrend falscher Erinnerungen treffen. Daher werden im fünften Kapitel Prozeduren zur getrennten Erfassung von Aktivierung und Quellendiskrimination vorgestellt.

Nachdem das theoretische Rahmengerüst gebildet ist und die dazugehörigen Forschungsbefunde erläutert wurden, gilt es die Ziele der eigenen Studie vorzustellen (Kapitel 6), die den Übergang zum zweiten Teil der Arbeit mit den eigenen Untersuchungen bilden. Anhand unterschiedlicher Prozeduren, die in verschiedenen Experimenten realisiert werden sollen, wird das Ziel verfolgt alterskorrelierte Differenzen in Aktivierung und Quellendiskrimination von der frühen Kindheit bis zum jungen Erwachsenenalter zu untersuchen. Es werden stets drei Altersgruppen untersucht: Vorschüler bzw. Erstklässler, Dritt- bzw. Viertklässler und junge Erwachsene.

Das erste Experiment (Kapitel 7) soll prüfen, ob sich alterskorrelierte Differenzen falscher Erinnerungen in impliziten Tests zeigen. Da diese Tests nur Aktivierung und keine Quellediskrimination erfassen, lassen sich gezielt Aussagen zum Aktivierungsprozess und dessen alterskorrelierten Veränderungen treffen.

Im achten Kapitel sollen alterskorrelierte Differenzen falscher Erinnerungen in der Freien Reproduktion und im Wiedererkennen geprüft werden, um auch Aussagen zur Quellediskrimination treffen zu können. Als Erstes soll die Freie Reproduktion (8.1) als Testverfahren eingesetzt und eine spezielle Instruktion erteilt werden anhand derer nur die Aktivierung getestet werden soll. In einem Nachtest soll zudem separat die Quellediskrimination geprüft werden. Dieses Experiment soll mit einem Wiedererkennenstest repliziert werden (8.2). Um zu prüfen, ob alterskorrelierte Unterschiede in der Aktivierung auf Verarbeitungsstrategien zurückgehen, die ältere Probanden spontan einsetzen können, soll im darauf folgenden Experiment (8.3) eine Instruktion erteilt werden, die den Einsatz von Verarbeitungsstrategien fördern, damit bei den jüngeren Probanden kritische Items besser aktiviert werden können. Dieses Experiment mit dem Wiedererkennenstest ist abgesehen von der Instruktion identisch mit dem vorherigen, so dass beide Experimente miteinander verglichen werden können, was Aufschlüsse über die Aktivierungsförderung gibt und deren Einfluss auf falsche Erinnerungen.

Das letzte Experiment (Kapitel 9) hat ebenfalls die Zielsetzung die Aktivierung kritischer Items anhand von Lerninstruktionen zu fördern, aber auf eine andere Weise als im vorherigen Experiment, denn hier sollen zusätzlich die Listenitems tiefer gehend verarbeitet werden. Im Gegensatz zu den bisherigen Experimenten soll ein differenzierter Quellediskriminationstest eingesetzt werden, der die Quellenzuordnung zur internen (selbst gesagt bzw. gedacht) und externen Quelle (vorgelesen) separat prüfen und damit aufschlussreicher sein soll als ein Wiedererkennenstest, der nur ein „alt/neu“-Urteil verlangt.

Einen Gesamtüberblick über alle Ergebnisse der vorliegenden Arbeit soll der Leser im Kapitel der Gesamtdiskussion (10) erhalten. Hier sollen zunächst problematische Aspekte der eigenen Untersuchung erläutert werden (10.1). Darauf folgend sollen die Ergebnisse in Bezug zueinander gesetzt und Interpretationsmöglichkeiten nach den beiden populären Theorien (AMA und Fuzzy Trace Theorie, FTT) vorgestellt werden (10.2). Abschließend soll ein Ausblick auf Fragestellungen möglicher zukünftiger Untersuchungen gegeben werden (10.3), die aus den eigenen Befunden resultieren und die Praxisrelevanz verdeutlicht werden.

Zusammenfassend werden die wesentlichen Befunde der eigenen Arbeit berichtet (11).

I. Theoretischer Hintergrund

Das Forschungsfeld der falschen Erinnerungen stellt ein noch sehr junges Gebiet der kognitiven Psychologie dar, jedoch sind im Laufe der letzten Jahre zahlreiche Studien durchgeführt worden, die sich dem allgemeinen Phänomen, den Beeinflussungsmöglichkeiten falscher Erinnerungen und praktischen Implikationen widmen.

In den folgenden Abschnitten geht es um die Darstellung des Forschungsfeldes der Gedächtnisillusionen und wie sich dieses im Laufe der Zeit weiterentwickelt hat (1). In einem weiteren Abschnitt wird auf das DRM-Paradigma ausführlich eingegangen (2), weil es die Basis für die eigene Studie ist. Neben der Entwicklung des Paradigmas werden theoretische Ansätze erläutert, die falsche Erinnerungen erklären sollen. Um mögliche alterskorrelierte Unterschiede in falschen Erinnerungen zu verstehen, wird die Gedächtnisentwicklung bei Kindern dargestellt, wobei Schwerpunkte auf die relevanten Anteile der eigenen Fragestellung gelegt werden (3). Darüber hinaus werden in einem anderen Abschnitt (4) bisherige Studien zu alterskorrelierten Veränderungen falscher Erinnerungen aufgeführt und deren Schwierigkeiten in der Formulierung klarer Prognosen herausgestellt. In einem separaten Abschnitt (5) werden Lösungsansätze erarbeitet, um mit besseren Prozeduren falsche Erinnerungen in verschiedenen Altersgruppen zu untersuchen. Abschließend (6) werden die Ziele der eigenen Studie vorgestellt.

1 Das Forschungsfeld der Gedächtnisillusionen

Die systematische Erforschung des Gedächtnisphänomens falscher Erinnerungen kann erst auf knapp ein Jahrzehnt zurückschauen (Roediger & McDermott, 1995), jedoch liegt diesem Gebiet eine lange Forschungsgeschichte der Wahrnehmungs- und Gedächtnisillusionen zugrunde, welches bis auf die Anfänge der experimentellen Psychologie zurückverfolgt werden kann. Damit stellen die Untersuchungen zu Wahrnehmungs- und Gedächtnisillusionen den Ausgangspunkt der Erforschung falscher Erinnerungen dar.

Bevor ein historischer Überblick über das Forschungsfeld der Gedächtnisillusion gegeben wird, soll die etymologische Bedeutung der Illusion und die inhaltliche Bedeutung der "false memories" geklärt und die verschiedenen Zugänge zu diesem Forschungsgebiet erläutert werden, um zu einem besseren Verständnis zu gelangen und um eine Basis für weitere Erläuterungen zu schaffen.

1.1 Begriffsbestimmung zu den Gedächtnisillusionen

Der Begriff „Illusion“ kann etymologisch auf das aus dem lateinischen stammende Verb „illudere“ zurückverfolgt werden, was „vortäuschen“ oder „verspotten“ bedeutet. Roediger (1996, p. 76) erklärt dies wie folgt:

Illusions mock our belief that what we perceive, remember, and know is in perfect accord with the state of the external world.

Illusionen können sich auf die Wahrnehmung oder auf das Gedächtnis beziehen. Drever (1964, zit. n. Roediger, 1996, S.76) kennzeichnet diese unterschiedlichen Gebiete auf folgende Weise:

In the case of sense perception, a subjective perversion of the objective content, or actual sense data; in the case of memory a subjective falsification by addition, omission, or substitution, in the recall of a past event.

Helmholtz (1881, zit. n. Roediger, 1996) führte zahlreiche Studien zu Wahrnehmungstäuschungen durch, wobei er von einem konstruktiven Prozess unbewusster Inferenzen ausging, die zu Illusionen führen sollten. Beispielhaft wird hier die dreidimensionale Raumwahrnehmung beschrieben, welche trotz zweidimensionaler Eigenschaften der Retina zustande kommt. Auch Opper (1854-1855, zit. n. Roediger, 1996) konnte Wahrnehmungstäuschungen

durch das Zeichnen zweier senkrechter Linien mit gleicher Länge hervorrufen. Jene Linie wirkte länger, die in Segmente unterteilt war, als jene, die nur durch Anfangs- und Endstriche gekennzeichnet war.

Im Unterschied zu den Wahrnehmungsillusionen, die in einer auslösenden Situation unmittelbar entstehen und direkt widerlegt werden können, treten Erinnerungsillusionen häufiger im Alltag auf, oft auch unbemerkt, und sind schwerer falsifizierbar.

Titchener (1928, zit. n. Roediger, 1996) widmete sich der Untersuchung von Gedächtnisillusionen, wobei er illusionäre Erinnerungen und illusionäres Wiedererkennen unterschied. Er ging davon aus, dass Menschen sich an Ereignisse erinnern könnten, die ihnen nie passiert sind, aber ebenso könnten Menschen wirklich erlebte Erfahrungen vergessen, als wären sie nie geschehen. Titchener beschrieb diese zwei Illusionen mit „déjà vu“ und „jamais vu“.

Nach Erdfelder (2001) können Gedächtnistäuschungen durch drei Effekte ausgelöst werden: durch (1) Kontexteffekte, (2) Interferenzeffekte und (3) Quellenverwechslungen. Zur ersten Gruppe gehören Erinnerungen, die in Bezug auf ein bestimmtes Schema verzerrt sind und so reproduziert werden wie sie typischerweise ablaufen. Die interferenzbedingten Gedächtnistäuschungen bezeichnen Beeinflussungen von Gedächtnisurteilen durch Informationen, die eine Person vor oder nach dem Ereignis aufgenommen hat. Quellenverwechslungen sollen nach Erdfelder (2001) die häufigsten Auslöser für alltägliche Gedächtnistäuschungen sein. Es können externe Quellen verwechselt werden (i.S. von „source monitoring“) oder auch vorgestellte und tatsächlich eingetretene Ereignisse (i.S. von „reality monitoring“).

Roediger und McDermott (2000) differenzieren Gedächtnisillusionen dahingehend, dass sie zwei Gedächtnisfehler benennen: „omission“ und „comission“. Während es im Fall der „omission“ um das alltägliche Vergessen von Ereignissen geht, wird bei der „comission“ das Phänomen dargestellt, dass Menschen Ereignisse anders erinnern als sie sie objektiv erlebt haben oder auch Erfahrungen berichten, die ihnen nie geschehen sind. Die Autoren bezeichnen diese Art des Gedächtnisfehlers auch als Erinnerungsverzerrung („memory distortion“), denn damit wird eine qualitative Abweichung der Erinnerung von einem Referenzereignis verdeutlicht. Drever (1964, zit. n. Roediger, 1996) betrachtet verschiedene Mechanismen für das Entstehen von Gedächtnisillusionen, die er Hinzufügen, Auslassen oder Substitution im Abruf eines vergangenen Ereignisses nennt.

1.2 Begriffsbestimmung zu falschen Erinnerungen

Eine inhaltliche Untersuchung des aus jüngerer Zeit stammenden Begriffs “false memories“ zeigt, dass dieser eine verfälschte Erinnerung beschreibt, welche nicht auf Vergessensprozesse zurückzuführen ist, sondern eine Verzerrung der Erinnerung darstellt (Echterhoff, 2001). “False memories“ sind den Gedächtnisillusionen subsumiert, da sie entweder auf eine Diskrepanz zwischen den Erfahrungen (Input) und den sie betreffenden Erinnerungen (Output) zurückgehen oder ein Erinnern eines nie geschehenen Ereignisses beschreiben. Das Phänomen der Veränderung von Erinnerungen zog früh das Interesse verschiedener Wissenschaftler auf sich, die die Erinnerungsverzerrung zunächst unter dem Aspekt von Vergessenstheorien betrachteten (Balwin & Shaw, 1895; Wulf, 1922). Heute widmen sich Forscher verschiedener theoretischer und praktischer Herkunft dem Verstehen von falschen Erinnerungen. Sowohl Gedächtnistheoretiker als auch Neurowissenschaftler zeigen großes Interesse an dieser Erinnerungssillusion. Auch Wissenschaftler, die den Zusammenhang von Gedächtnis und Altern erforschen, sowie jene, die die Effekte der Hypnose erkunden, wollen dem Gedächtnisphänomen näher kommen. In der Praxis sind es Psychotherapeuten, Gerichtsgutachter und auch Richter, die Interesse an diesem Gebiet zeigen (Benjamin, 2001).

Eine große Schwierigkeit bei der Erforschung falscher Erinnerungen ergibt sich daraus, dass Input und Output in Feldstudien und damit im Alltag kaum zuverlässig beobachtbar sind (Echterhoff, 2001). Dieses Problem wird mit kontrollierten Laborstudien gelöst, welche aber eine geringere externe Validität und praktische Relevanz haben. Das heute am häufigsten angewendete Paradigma zur Erforschung falscher Erinnerungen ist jenes, das von Deese, Roediger und McDermott (DRM) entwickelt wurde und auch in der eigenen Untersuchung verwendet wird. Durch die Präsentation von semantisch assoziierten Wortlisten (z.B. “thread, pin, eye, sewing, sharp, point, pricked, thimble, haystack, pain, hurt, injection“) zeigten sich falsche Erinnerungen an ein nicht präsentiertes kritisches Wort (in diesem Fall “needle“), das mit der Liste assoziiert war (Roediger & McDermott, 1995). Die Forscher konnten falsche Erinnerungen durch zwei Tests demonstrieren: In der Freien Reproduktion (“free recall“) und im Wiedererkennenstest (“recognition“) traten diese Gedächtnisillusionen auf.

Falsches Reproduzieren zeigt sich darin, dass ein Item, welches in der Lernliste nicht vorkam, dennoch in der Freien Reproduktion genannt wird. Nach Brainerd und Mitarbeitern (2002a) spricht man hier von Intrusionen, von denen es drei Arten gibt. Intrusionen können von kritischen Items entstehen, die selbst nicht in der Lernliste erscheinen, aber den Bedeutungsgehalt der Liste wiedergeben. Auch ergeben sich Intrusionen, wenn andere Items als die Listenitems

oder kritischen Items genannt werden, die aber in Beziehung zu den Listenitems stehen. Die letzte Art der Intrusionen bezieht sich auf die Reproduktion von Items der vorangehenden Liste (Brainerd et al., 2002a). Neben diesen drei gedächtnisbegründeten Intrusionen können durch Antworttendenzen oder Rateeffekten nichtpräsentierte Items falsch reproduziert werden.

Anhand von Wiedererkennenstests werden durch „alt/neu“-Urteile bezüglich präsentierter oder nicht präsentierter Items falsche Erinnerungen erhoben (Lockhart, 2000; Multhaup & Conner, 2002). Die Probanden werden gebeten die Items danach zu beurteilen, ob sie bereits in der Lernphase vorkamen. Auch hier kann es zu Rateeffekten kommen (Roediger, McDermott & Robinson, 1998), jedoch ist nach Brainerd, Reyna und Kneer (1995b) diese Testform sensitiver für Gedächtnisinhalte und wird bei entwicklungspsychologischen Studien häufiger verwendet. Es wird in der Auswertung zwischen falschen Alarmraten und falschen Zurückweisungen unterschieden (Lockhart, 2000). Bei der falschen Alarmrate wird ein nichtpräsentiertes, neues Item (Distraktor) als bereits gelernt beurteilt. Brainerd und Reyna (1998a) unterscheiden bedeutungskonsistente von bedeutungsinkonsistenten Distraktoren, wobei sich diese Differenzierung aus der Beziehung des Distraktors zum Zielitem ergibt. Erscheint das Zielitem „Collie“ und wird der Distraktor „Hund“ dargeboten, so liegt eine bedeutungskonsistente Beziehung vor. Demgegenüber handelt es sich um eine Bedeutungsinkonsistenz, wenn die Items nicht aus der gleichen Kategorie entstammen. In diesem Fall beruht der falsche Alarm auf Rateeffekten oder anderen Antworttendenzen (Brainerd & Reyna, 1998a). In Untersuchungen zum DRM-Paradigma werden bedeutungskonsistente Distraktoren eingesetzt, die das kritische Item repräsentieren, da sich auf diese Weise induzierte Erinnerungen messen lassen (Deese, 1959b).

Im Fall des falschen Zurückweisens wird ein bereits gelerntes Item fälschlicherweise als neu bezeichnet, was die „Trefferrate“ reduziert. Diese stellt die korrekte Wiedererkennensleistung der Probanden dar und wird in Bezug zu den falschen Erinnerungen gesetzt. Die korrigierte Wiedererkennensleistung (Pr) bezeichnet die Anzahl der Treffer, welche sich dadurch ergibt, dass alte Items als solche erkannt werden, wobei falsche „alt“-Urteile für nicht präsentierte Items von der Anzahl der Treffer abgezogen werden (Snodgrass & Corwin, 1988).

Neben diesen expliziten Tests gibt es auch implizite Tests, die im Rahmen der Erforschung von „false memories“ eingesetzt werden (McDermott, 1997). Bei dieser Testform erhalten die Probanden keine Instruktion sich zu erinnern. Es kann zu einer Nutzung des kritischen Items kommen, ohne dass sich der Proband dessen bewusst ist (Lockhart, 2000). Im Abschnitt 5.2

wird auf implizite Tests und ihre Relevanz in der Erforschung von falschen Erinnerungen im DRM-Paradigma ausführlich eingegangen.

Da falsche Erinnerungen eine subjektive mentale Repräsentation sind und diese ausschließlich über sprachliche Äußerungen erhoben werden können, ergibt sich ein weiteres Problem in deren Erforschung. Die verbale Darstellung muss nicht Notgedrungen der mentalen Repräsentation entsprechen, was das grundlegende Problem in der Überprüfung der Wahrheit oder Falschheit von Erinnerungen darstellt (Echterhoff, 2001). Auch hier erweisen sich Laboruntersuchung nach dem DRM-Paradigma als hilfreich, um anhand kontrollierter Studien dem Phänomen näher zu kommen.

Im Verlauf der Forschungsgeschichte stellte sich die Frage, wie und zu welchem Zeitpunkt falsche Erinnerungen entstehen. Während Underwood (1965) Enkodierungsprozesse untersuchte, konzentrierte sich Deese (1959b) auf Abrufprozesse und wiederum andere Wissenschaftler (Anisfeld & Knapp, 1968) sahen die Interaktion beider Informationsverarbeitungsphasen als maßgeblich zur Entstehung von "false memories" an.

Aufgrund dessen, dass einige Menschen von Ereignissen berichten, die ihnen nie widerfahren sind, ohne dass sie im Voraus Fehlinformationen erhalten hatten (spontanes falsches Erinnern), stellt sich die Frage, inwiefern falschen Erinnerungen ein automatischer Prozess zugrunde liegt (Brainerd & Mojardin, 1998; Gallo & Seamon, 2004; Seamon, Luo & Gallo, 1998). Loftus und Palmer (1974, zit. n. Roediger, 1996) untersuchten unter dem "misleading information effect" Geschwindigkeitseinschätzungen von Probanden, die einen Autounfall auf Video beobachtet hatten. In einem Fragebogen wurden die Probanden um die Einschätzung der Geschwindigkeit gebeten mit der ein Auto in das andere geriet. Als unabhängige Variable wurde das auf die Geschwindigkeit bezogene Verb ("smashed into" vs. "hit the car") variiert, was die Probanden tatsächlich in ihrer Geschwindigkeitseinschätzung beeinflusste. Bei den Verben, die eine höhere Geschwindigkeit implizierten ("smashed"), gaben die Personen auch signifikant höhere Geschwindigkeiten an. Nicht nur Kinder, sondern auch Erwachsene, unterlagen diesem Effekt. Seamon und Mitarbeiter (1998) sowie Gallo und Seamon (2004) kommen zu dem Schluss, dass eine bewusste Aktivierung in der Lern- bzw. Enkodierphase keine notwendige Voraussetzung für falsches Erinnern ist. Dies bedeutet, dass diese Erinnerungsverzerrungen der Person nicht bewusst sind, was weitreichende Konsequenzen in der Praxis haben kann. Zeugenaussagen könnten damit ohne wissentliche, bewusste Täuschung nicht der Wahrheit entsprechen (Roediger, 1996). Ist der Mensch Fehlinformationen ausgesetzt, so kann sich seine Erinnerung unbewusst verändern (Ceci & Bruck, 1993). Damit

stellt sich auch die Frage, inwiefern falsche Erinnerungen durch Suggestion induziert werden können. Zahlreiche Studien widmen sich aktuell vermeintlich „wiederbelebten Erinnerungen“ aus früher Kindheit, die sich auf vergessene sexuelle oder körperliche Missbrauchserfahrungen beziehen (Andrews, 1997; Gudjonsson, 1997; Loftus, 1993; Roediger, McDermott & Goff, 1997). Damit richtet sich immer mehr das Interesse auf das Therapeutenverhalten, das z.B. durch Hypnose und suggestive Fragen falsche Erinnerungen hervorrufen kann (Roediger, 1996).

Darüber hinaus werden “false memories“ dahingehend untersucht wie dauerhaft diese sind (McDermott, 1996; Roediger, et al., 1998). Es ist festgestellt worden, dass falsche Erinnerungen ein robustes Phänomen sind: Während richtige Erinnerungen mit der Zeit abnehmen, bleiben falsche Erinnerungen stabil oder steigen sogar an (Roediger et al., 1998).

Sehr viele Forscher haben großes Interesse daran zu überprüfen, ob es Unterschiede der “false memories“ in verschiedenen Altersgruppen gibt und inwiefern sich diese Gedächtnisillusion mit dem Alter verändert (Brainerd & Mojardin, 1998; Brainerd & Reyna, 1998a; Brainerd & Reyna, 2004; Brainerd et al., 2002a; Ghetti et al., 2002). Diesem Forschungszweig ist ein eigener Abschnitt (3) gewidmet, da sich die eigene Untersuchung diesem Aspekt widmet. Falschen Erinnerungen wird aus entwicklungspsychologischer Sicht besondere Aufmerksamkeit gewidmet, da Kinder gegenüber Suggestivfragen besonders beeinflussbar zu sein scheinen und es fraglich ist, ob den Aussagen ein Wahrheitsgehalt zugrunde liegt (Brainerd & Reyna, 1998a).

In der kognitiven Psychologie besteht aktuell die Auffassung, dass “false memories“ insofern Folge einer Fehlattribution sind, dass eine erinnerungsrelevante Information einer falschen Quelle zugeordnet wird (Echterhoff, 2001; Marsh & Bower, 2004; Multhaup & Conner, 2002; Roediger, 1996). Aus dieser Sicht werden selbständig generierte Informationen als tatsächliche Erfahrungen im Nachhinein erlebt. Den theoretischen Erklärungsansätzen ist ein eigener Unterpunkt gewidmet (2.2), in welchem die entsprechenden und zurzeit populärsten Theorien ausführlich dargestellt werden.

1.3 Historischer Überblick

Die Anfänge der Erforschung von Gedächtnisillusionen lassen sich außerhalb der akademischen Psychologie auf die Antike zurückführen. Schon Aristoteles (1966, zit. n. Roediger et

al., 1998) beschrieb die Prinzipien des Assoziationismus, welche sich auf die Verbindung mentaler Repräsentationen beziehen.

Mit dem Beginn der experimentellen Psychologie war es Ebbinghaus (1885/1964, zit. n. Roediger, 1996), der Gedächtnisillusionen in seinen gedächtnispsychologischen Studien beobachtete, sie jedoch als Störfaktoren nicht weiter untersuchte. Sein revolutionäres Konzept, welches er in der Monographie „Über das Gedächtnis“ beschrieb, gründete auf der Induktion von Erinnerungen in Laboruntersuchungen (Lockhart, 2000). Er erforschte Gedächtnisprozesse und Erinnerungsleistungen beim Lernen sinnloser Silben und war der Auffassung, dass das Gedächtnis über assoziative Prozesse funktioniert. Die Assoziationen waren entweder direkt, in dem Sinne, dass Items miteinander verbunden waren oder indirekt, insofern dass eine Assoziation zweier Items durch ein Item oder mehrere Items getrennt war (Roediger et al., 1998). Ebbinghaus zeigte ein innovatives Vorgehen, indem er die assoziativen Prozesse nicht nur genau beschrieb, sondern auch methodisch präzise untersuchte. Durch Laboruntersuchungen konnte er über die Kontrolle des Inputs feststellen, ob eine Erinnerung richtig oder falsch war (Roediger, 1996). Vorwiegend interessierte sich Ebbinghaus für Gedächtnisstrategien (Wiedererlern- und Ersparnismethoden) und erhob darüber hinaus auch Reaktionszeiten. Traten in seinen Untersuchungen fehlerhafte Antworten auf, so wurden sie als Störfaktoren betrachtet, die durch Rateeffekte oder Kriterienverschiebungen verursacht worden waren (Roediger, 1996). Aus diesem Grund wurde ihnen im weiteren Verlauf wenig Beachtung geschenkt.

Zur gleichen Zeit in der Ebbinghaus seine Gedächtnisstudien durchführte, widmete sich auch Kirkpatrick (1894, zit. n. Roediger et al., 2001) der experimentellen Erforschung des Gedächtnisses. Kirkpatrick stellte fest, dass bildhafte Items besser erinnert wurden als verbales Material und dass bildhafte Vorstellungen das Erinnern verbaler Items erleichterten. Anhand von Untersuchungen konnte gezeigt werden, dass durch die Präsentation bestimmter Wörter (z.B. „Messer“) an andere Wörter (z.B. „Gabel“) der gleichen Kategorie automatisch gedacht wurde. Die Probanden der Studie dachten, dass diese zugehörigen Wörter auch präsentiert worden waren. Kirkpatrick bezeichnete diesen Befund als Suggestionseffekt.

Andere Untersuchungen an Kindern bestätigten diesen Suggestionseffekt. Binet (1900, zit. n. Roediger & McDermott, 2000) konnte zeigen, dass durch Suggestivfragen falsche Antworten von den Kindern gegeben wurden, wohingegen es bei neutralen Fragen und freiem Abruf zu weniger fehlerhaften Antworten kam.

Stern (1910, zit. n. Roediger, 1996) führte ebenfalls Studien an Kindern durch und konnte durch fehlgeleitete Fragen Erinnerungsverzerrungen feststellen.

Von einem anderen, nicht experimentellen, sondern kasuistischen Standpunkt aus, untersuchte Freud (1895, zit. n. Roediger, 1996) Erinnerungsverzerrungen bei Patienten mit psychopathologischen Auffälligkeiten. Das Phänomen veränderter Erinnerungen beschrieb er als Repression, welches das Verbannen schmerzhafter früher Erinnerungen in einen unbewussten Zustand sein sollte. Jene unbewussten Erinnerungen sollten sich der Person nur im Traum oder in psychopathologischen Symptomen zeigen. Auch stellte sich für Freud die noch heute aktuelle Frage, inwiefern falsche Kindheitserinnerungen durch eine Therapie induziert werden können und wie Erinnerungen, die in einer Therapie aufgedeckt werden, zu bewerten sind. Er wies darauf hin, dass Kindheitserinnerungen im Erwachsenenalter nicht dem wirklichen Erleben in der Kindheit entsprechen und kaum von Phantasien zu unterscheiden sind. Früheste Erinnerungen sind durch vielfältige psychische Faktoren geformt, so dass sie zwar für die Person als der Realität entsprechend erlebt werden, aber objektiv eine Erinnerungsverzerrung darstellen.

In der akademischen Psychologie spielte die systematische Erforschung von Gedächtnistäuschungen bis zur zweiten Hälfte des letzten Jahrhunderts kaum eine Rolle. Es gab zwei Ausnahmen von denen nun berichtet wird. Zum einen waren es die Gestalttheoretiker, die davon ausgingen, dass Erinnerungen sich an den Gestaltgesetzen der Organisation orientieren und sich im Zeitverlauf an diese annähern. Menschen würden sich also an Ereignisse in einer besser organisierten Form erinnern (Roediger, 1996). Wulf (1922) zeigte dies beispielhaft an visuellen Formen, die im Zeitverlauf symmetrischer und geregelter erinnert wurden. Zudem wurde im Rahmen der Gestalttheorie auf die Bedeutung der Verbalisierung von Erinnerung hingewiesen, denn Wörter bestimmten die Art und Weise mit der Ereignisse später (verändert) erinnert wurden.

Zum anderen lenkte Bartlett (1932) die Aufmerksamkeit auf die Erforschung von Gedächtnisillusionen. In seiner 1932 veröffentlichten Monographie "Remembering: A Study in Experimental and Social Psychology" ging Bartlett davon aus, dass Erinnerungen nicht perfekt sind und Menschen Ereignisse nicht wortgetreu erinnern könnten. In seiner weltberühmten Studie sollten Studierende eine indianische Volkssage "The War of the Ghosts" lesen und diese später in eigenen Worten wiedergeben. Diese Volkssage war dadurch gekennzeichnet, dass sie übersinnliche und irrationale Elemente enthielt. Nach unterschiedlich langen Behaltensintervallen erzählten die Studenten die Volkssage nach und es konnte festgestellt werden, dass die

reproduzierte Geschichte deutlich von dem Original differierte. Neben zahlreichen Auslassungen konnten auch systematische inhaltliche Veränderungen dahingehend festgestellt werden, dass zusätzliche Elemente eingefügt wurden, um die Geschichte verständlicher zu machen. Die übersinnlichen Anteile wurden häufig ausgelassen und stattdessen wurde die Sage durch neue Elemente den englischen Volkssagen angepasst. In einigen Fällen wurde die indische Geschichte sogar so sehr verändert, dass eine Moral eingefügt wurde. Bartlett erklärte die Befunde damit, dass sowohl die Wahrnehmung als auch die Erinnerung ein konstruktiver Prozess seien, die sich an vorhandenen kognitiven Schemata der Person orientieren und dass gemäß diesen Ereignisse interpretiert werden würden. Durch die Veränderung der indischen Volkssage wurde diese den eigenen Schemata angepasst, die den Studierenden vertrauter waren. Ein Schema ist nach Bartlett eine organisierte Struktur, das Wissen und Erwartungen zusammenfasst. Schemata entstehen durch spezifische Erfahrungen und sind somit kulturabhängig. Werden vergangene Ereignisse erinnert, so sind es nicht die spezifischen Details, die die Erinnerung bilden, sondern vielmehr die übergeordneten Themen. Tritt nun eine Leerstelle in der Erinnerung auf, so wird diese durch schemakonsistente Information gefüllt, wodurch falsche Erinnerungen entstehen können.

Zwar stellen Bartletts Untersuchungen einen Meilenstein in der Erforschung von Gedächtnisillusionen dar, jedoch wurden seinen Befunden in den anschließenden Jahrzehnten der experimentellen Gedächtnisforschung keine Aufmerksamkeit gewidmet. Spätere Untersuchungen konnten methodische Mängel aufdecken und die Ergebnisse nicht replizieren (Roediger, 1996), aber zweifelsohne hat Bartlett durch seine innovative Studie der Erforschung falscher Erinnerung einen Anstoß gegeben.

Eine neue Phase in der Entwicklung des Forschungsfeldes der Gedächtnisillusionen brach mit den Informationsverarbeitungsansätzen in den 1950er Jahren an. Miller (1956) widmete sich auf revolutionäre Weise der Erforschung des menschlichen Gedächtnisses, indem er das Konzept der Rekodierung entwarf. In der Phase der Enkodierung wird Information von der Außenwelt aufgenommen und in der Phase der Dekodierung wird diese abgerufen. Zwischen diesen Phasen wird die Information auf veränderte Weise internal repräsentiert, was den Vorgang der Rekodierung beschreibt. Nach Miller (1956) werden die Input-Ereignisse zunächst gruppiert, dann neu benannt und auf die Weise erinnert, dass die neue Benennung und nicht mehr der eigentliche Input relevant ist. Der Vorgang der Rekodierung beschreibt somit einen aktiven Transformationsprozess, der eine gewisse Fehlerwahrscheinlichkeit im Rekodieren und anschließenden Dekodieren beinhaltet, da das externale Ereignis nicht perfekt mit der internalen Repräsentation übereinstimmt. Ebenso wie Bartlett ging auch Miller von kogniti-

ven Schemata aus, die die Erinnerung formen. Hört ein Mensch eine Geschichte, so wird er sie in seine eigenen Worte übersetzen und sie dementsprechend erinnern. Nach Miller (1956) beinhaltet der Abruf gleichzeitig eine Dekodierung und ist auf diese Weise empfänglich für Fehler. Anschließende Studien des Informationsverarbeitungsansatzes erforschten den Aspekt des Einflusses der Materialorganisation auf die Erinnerung und weniger die Gedächtnisfehler (Roediger, 1996).

Das Forschungsgebiet der Gedächtnisillusionen wurde zunehmend konkreter, was auf Underwood (1965) zurückgeführt werden kann, der den Einfluss von assoziativen Verbindungen zwischen einzelnen Wörtern auf die Erinnerung untersuchte und herausfand, dass diese im Wiedererkennenstest zu typischen Fehlern führen. Der Präsentation semantisch assoziierter Wörter folgte falsches Wiedererkennen. Wurde den Probanden in der Lernphase ein Wort dargeboten (z.B. "table"), das semantisch hoch assoziiert mit einem Wort aus der Testphase (z.B. "chair") war, dann resultierte für jene semantisch stark assoziierten Wörter eine hohe falsche Alarmrate. Diese falsche Alarmrate lag signifikant über der für nicht assoziierte, nicht präsentierte Wörter. In weiteren Studien versuchte Underwood die zugrunde liegenden Prozesse genauer zu erkunden. Während bei der Präsentation von konvergierenden (Lernphase: "bed, dream"; Testphase: "sleep") und semantisch übergeordneten Wörtern (Lernphase: "oak, elm, birch"; Testphase: "tree") falsches Wiedererkennen beobachtet werden konnte, war dies bei eindrucksbildenden Wörtern (Lernphase: "chalk, milk, rice, snow"; Testphase: "white") nicht der Fall. Auch Antonyme (Lernphase: "bottom"; Testphase: "top") führten nur durch wiederholte, randomisierte Präsentation zu falschem Wiedererkennen. Andere Autoren konnten diesen Effekt durch die Präsentation semantisch assoziierter Wörter replizieren, wenn auch in geringerem Ausmaß (Anisfeld & Knapp, 1968). Es konnten weitere Einflussfaktoren ausfindig gemacht werden, die das Phänomen des falschen Wiedererkennens beeinflussten: Mit der Anzahl der präsentierten semantisch assoziierten Wörter in der Lernphase nahmen falsche Erinnerungen in der Testphase zu. Underwood (1965) führte diesen Effekt auf eine "implicit associative response" (IAR) zurück. In der Enkodierphase soll es zu einer bewussten Aktivierung des Wortes "table" dadurch kommen, dass die Person sofort an ein, dem präsentierten Item nahe stehendes Wort denkt, was zu einem späteren falschen Wiedererkennen im Abruf führt. Gemäß diesen Ausführungen würde man den beschriebenen Effekt heute auf Fehler im "reality monitoring" (Johnson & Raye, 1981) zurückführen, was bedeutet, dass Probanden nicht fähig sind zwischen den Wörtern zu unterscheiden, welche sie in der Lernphase gehört und an welche sie nur gedacht haben. Underwoods Befunde werden heute unter dem Begriff "relatedness effect" subsumiert und diskutiert (Roediger & McDermott, 2000).

Hier geht es darum, dass Ereignisse, die im hohen Maße miteinander in Beziehung stehen, dazu führen, dass andere, nicht erfahrene Ereignisse erinnert werden, die mit jenen präsentierten zusammenhängen. Die Beziehung der Ereignisse kann aber nicht nur assoziativ, sondern auch kategorial und thematisch sein. Tulving (1972, zit. n. Roediger & McDermott, 2000) erklärt dies damit, dass das Vorwissen einer Person das Erinnern von Ereignissen dadurch beeinflusst, dass es hilft episodische Erinnerungen zu ergänzen.

Im Rahmen der kognitiven Wende gab es weitere Fortschritte in der Erforschung von Gedächtnisillusionen. Neisser (1967) knüpfte an Bartletts Theorie an, die besagte, dass Erinnerung ein konstruktiver Prozess ist. In diesem Rahmen wies er auf die Einheit kognitiver Prozesse hin, was bedeuten sollte, dass Wahrnehmung, Erinnerung und Denken als aktive und voneinander abhängige Prozesse zu konzeptualisieren sind. Hinsichtlich Wahrnehmungs- und Gedächtnistäuschungen verwies Neisser auf die konstruktive Aktivität, die für diese Phänomene verantwortlich sein sollte. Gespeicherte Information wird nicht wiederbelebt, sondern in der konstruktiven Aktivität des Abrufs verändert. Jene konstruktive Aktivität bearbeitet die Information und die wahrnehmende oder erinnernde Person ist sich nur dem Produkt der Aktivität bewusst. Weder in der Wahrnehmung noch in der Erinnerung gelangt die Bearbeitung der Information ins Bewusstsein bzw. bildet sich im Verhalten ab. Somit sind die Gedächtnistäuschungen unbewusst und außerhalb jeder Kontrolle. Durch Neissers Arbeiten erlebten die Studien zur Erforschung von Gedächtnisillusionen in Bartletts Tradition eine Renaissance und waren eine Art Impulsgeber für zahlreiche Untersuchungen von Gedächtnisillusionen. Durch die komplexe Konzeptualisierung der Gedächtnisphänomene als interaktive Prozesse konnte ein besseres Verständnis normaler Erinnerungsvorgänge erlangt werden, welche heutigen Theorien zur Erklärung der "false memories" zugute kommen.

In den letzten 15 bis 20 Jahren erlangte das Forschungsfeld der Gedächtnisillusionen erneut Relevanz im Hinblick auf gesellschaftliche Probleme. Im Rahmen von Gerichtsprozessen zweifelte man an der Genauigkeit von Zeugenaussagen, besonders jener von Kindern. Aus diesem Grund sollte überprüft werden inwiefern Kinder wahrheitsgemäße Aussagen machen können und inwiefern sie durch suggestive oder manipulative Fragen beeinflussbar sind (Roediger, 1996). Darüber hinaus stellte sich die Frage wie wiedererlebte Erinnerungen, die im psychotherapeutischen Setting auftreten, zu bewerten sind. Häufig berichten Klienten von Erinnerungen des Missbrauchs aus ihrer frühen Kindheit und es ist unklar, ob jene, häufig unspezifischen, Erinnerungen der tatsächlichen Erfahrung entsprechen (Andrews, 1997; Gudjonsson, 1997; Loftus, 1993; Roediger et al., 1997). Gudjonsson (1997) stellt die Frage, inwiefern solche wiedererlebten Erinnerungen durch die Therapie hervorgerufen sind. Es wird

davon ausgegangen, dass ein falscher Glaube von einer falschen Erinnerung unterschieden werden muss. Der Glaube geht der Erinnerung zeitlich voraus und wird der Erstere positiv bestärkt, so entsteht eine falsche Erinnerung. Aus diesen strafrechtlich und gesellschaftlich relevanten Gründen zog das Forschungsfeld der Gedächtnisillusionen erneut Interesse auf sich.

Rückblickend auf die lange Forschungsgeschichte kann festgehalten werden, dass zunächst Gedächtnisfehler als Störfaktor in Gedächtnisstudien auftraten und wenig Beachtung fanden. Im weiteren Verlauf widmete man sich der Erforschung der Faktoren, welche diese Gedächtnisverzerrungen bzw. -illusionen hervorriefen. Es stellte sich die Frage, warum bei jedem Menschen solche Erinnerungsverzerrungen auftreten und diese nicht bemerkt werden. Aus verschiedenen Forschungsrichtungen wurde versucht herauszufinden, was Menschen in ihren korrekten Erinnerungen beeinträchtigt, und im Laufe der Zeit entwickelte sich ein zunehmend differenziertes Verständnis dieser Gedächtnisprozesse.

2 Das Deese/Roediger-McDermott (DRM)-Paradigma

Das von Deese, Roediger und McDermott (DRM) entwickelte Paradigma zur Erforschung falscher Erinnerungen wurde erst vor gut 10 Jahren von Roediger und McDermott (1995) in das Forschungsgebiet der Gedächtnisillusionen eingeführt; dennoch hat es eine Fülle von Untersuchungen und theoretischen Überlegungen in dieser Zeit hervorgerufen. Erst seit 2001, nachdem Tulving einen eigenen Namen für das Paradigma vorschlug, sind Untersuchungen unter dem Begriff des DRM-Paradigmas in der Literatur zu finden (Roediger, Watson, McDermott & Gallo, 2001).

In diesem Unterpunkt steht das DRM-Paradigma im Mittelpunkt, wobei zunächst auf seine Entwicklungsgeschichte eingegangen werden soll (2.1). Um die Effekte von falschen Erinnerungen, hervorgerufen und abgebildet durch das DRM-Paradigma, zu verdeutlichen, werden die theoretischen Ansätze erläutert (2.2). Zur Klärung der Einflussfaktoren auf falsche Erinnerungen wird in Abschnitt 2.3 ein Überblick über diese gegeben.

2.1 Entwicklung und Bewertung des DRM-Paradigmas

Die Erforschung falscher Erinnerungen anhand des Deese/Roediger-McDermott (DRM)-Paradigmas hat in der kognitiven Psychologie einen besonderen Stellenwert erlangt. Zwar handelt es sich um ein noch sehr junges Paradigma (erste Untersuchungen von Roediger & McDermott, 1995), aber es zeigt auf vorbildliche Weise, wie ein innovatives Vorgehen zum Erkenntnisgewinn führen und eine Schrittmacherfunktion in der Forschung übernehmen kann. Die Besonderheit des DRM-Paradigmas besteht darin, dass Roediger und McDermott (1995) bereits bestehende Forschungsbefunde aufgegriffen haben, um sie in ihre Vorgehensweise zu integrieren.

2.1.1 Forschungsarbeiten von Deese (1959a, 1959b): Die Basis des DRM-Paradigmas

Die Anfänge dieses Paradigmas gründen auf den Forschungsarbeiten von Deese (1959a, 1959b), der sich in den 1950er Jahren der systematischen Untersuchung von Gedächtnisillusionen widmete. Dabei konzentrierte er sich auf die Erforschung assoziativer Prozesse und deren Einfluss auf die Erinnerung.

Deese (1959a, p.305-306) war anfänglich der Annahme, dass eine hohe Assoziationsstärke zwischen den Items auf einer stärkeren kognitiven Organisation beruhte und damit zu einer besseren Wiedergabeleistung führen sollte:

...lists with high inter-item associative strength are more highly organized than those with low inter-item strength. High organization means that the number of items recalled is increased and that resistance to outside intrusion is relatively high.

Demzufolge sollte die Assoziationsstärke ein Schutzfaktor gegen Intrusionen sein, welche das falsche Erinnern an nicht präsentierte Wörter darstellten. Er definierte die Assoziationsstärke als durchschnittliche Häufigkeit, mit der die Listenitems die anderen zugehörigen Items in der Freien Assoziation hervorriefen.

Zum einen stellte er fest, dass zufällig präsentierte Wörter, die sich in der Freien Assoziation als zusammenhängend erwiesen hatten, in der Testphase zusammen abgerufen wurden und die Reihenfolge sich nach deren Assoziationsstärke richtete (Deese, 1959a). Anhand von Wortlisten konnte er, im Einklang mit seiner ersten Annahme, eine Korrelation der Assoziationsstärke zwischen den Wörtern und deren korrekter Wiedergabe nachweisen. Zum anderen fiel ihm, entgegen seiner Annahme, auf, dass hoch assoziierte Wörter aus der Lernphase mit Intrusionen in der Testphase einhergingen. Diese Ergebnisse erklärte er folgendermaßen (Deese, 1959a, S.311):

...recall consists of a small core of words directly available through immediate memory and of strong free association of these. Recall is good or poor depending, then, upon the tendency of free associations from items within the list to converge upon other items within the list.

Somit nahm Deese an, dass die Intrusionshäufigkeit von der Assoziationsstärke der Kategorienitems abhängt. Wenn nicht alle zusammengehörigen Kategorienexemplare präsentiert wurden, kamen falsche Erinnerungen durch die Assoziation der Kategorie zustande.

In seiner nachfolgenden Studie griff Deese (1959b) seine Befunde auf und konzentrierte sich auf die Untersuchung von Intrusionsprädiktoren in der Freien Reproduktion. Die assoziative Struktur der Wortliste sah er als ausschlaggebend für das Auftreten dieser Intrusionen an.

Um dies empirisch zu belegen wollte er Intrusionen auf der Basis von Wortassoziationsnormen hervorrufen. Er konstruierte 36 Wortlisten, von denen jede Liste aus 12 Items bestand, die am stärksten mit einem nicht präsentierten Wort assoziiert waren. Beispielsweise wurde das kritische Item "sleep" am häufigsten durch die Liste der Wörter "bed, rest, awake, tired, dream, wake, night, blanked, doze, slumber, snore, pillow, peace, yawn, drowsy" genannt

(Roediger & McDermott, 1995). Den Versuchspersonen wurden jene Listen vorgelesen, die sie anschließend frei reproduzieren sollten. Hier interessierte sich Deese für die Wahrscheinlichkeit mit der das kritische, nicht präsentierte Item fälschlicherweise erinnert wurde.

Als Basis dieser Wortlisten verwendete Deese (1959b) die Minnesota Normen für die Kent-Rosanoff Items von Russell und Jenkins (1954), bei denen bereits die Vorwärtsassoziationsstärke (“forward associative strength“) der Listenitems zu den späteren kritischen Items vorlag. Die Vorwärtsassoziationsstärke repräsentiert die ersten Assoziationen von Personen zu dem späteren kritischen Item, die sich in einer Wortliste nach ihrer prozentualen assoziierten Häufigkeit anordnen lassen. Alle Listen wurden auf diese Weise konstruiert, jedoch ergab sich ein Unterschied in der Rückwärtsassoziationsstärke (“backward associative strength“). Die Rückwärtsassoziationsstärke stellt die durchschnittliche Wahrscheinlichkeit dar, mit der die Listenitems das kritische, nicht präsentierte Item im Freien Assoziationstest hervorrufen.

Anhand seiner Ergebnisse konnte Deese (1959b) feststellen, dass die Wahrscheinlichkeit der freien Reproduktion des kritischen Items in Bezug zur Rückwärtsassoziationsstärke stand. Damit konnte er seine anfängliche Annahme untermauern und seine bisherigen Befunde verständlich machen (Deese, 1959a), die auf die Bedeutung der Listenzusammensetzung für das Entstehen von Intrusionen hinwiesen. Richtige und falsche freie Reproduktionen schienen also entscheidend von der assoziativen Struktur der Wortliste abzuhängen.

2.1.2 Forschungsarbeiten von Roediger und McDermott (1995)

Aufgrund des steigenden Interesses an der Erforschung von “false memories” in den 1990er Jahren, ausgelöst durch die Debatte um die Glaubwürdigkeit von Zeugenaussagen und wiederbelebten Erinnerungen (Roediger, 1996), wollten Roediger und McDermott (1995) sich diesem Phänomen über kontrollierte Laboruntersuchungen nähern und falsche Erinnerungen reliabel erheben. Durch kritische Analysen der Ergebnisse aus Deeses (1959b) Studie stellten sich die Autoren die Frage, weshalb die Raten der falschen Reproduktion so hoch ausfielen (Roediger et al., 1998). Bisherige Studien wiesen im Gegensatz zu den hohen Intrusionsraten darauf hin, dass Probanden gewöhnlich wenig fehlerhaft in der Freien Reproduktion antworteten (Cofer, 1973; zit. n. Roediger & McDermott, 1995). Es kam der Gedanke auf, dass die Probanden nicht explizit darauf hingewiesen wurden auf das Erraten von Wörtern zu verzichten. Den Autoren war unklar, ob die Probanden den Test zur Freien Reproduktion nicht mit einem Test zur Freien Assoziation verwechselt hatten, wodurch eine solch hohe Intrusionsrate verständlich werden konnte. Mit anderen Worten ausgedrückt, zweifelten Roediger und Mc-

Dermott an der internen Validität der Studie, welche eine Voraussetzung für die zuverlässige Interpretation und Verallgemeinerung der Ergebnisse ist. Die Autoren waren daran interessiert die Ergebnisse aus Deeses Studie (1959b) zu replizieren (Roediger et al., 1998). Darüber hinaus wollten sie zur Erhebung falscher Erinnerungen zusätzlich einen Wiedererkennenstest verwenden und prüfen, ob sich Unterschiede richtiger und falscher Erinnerungen auf ein Konfidenzrating auswirken (Roediger & McDermott, 1995).

In ihrem ersten Experiment griffen Roediger und McDermott (1995) auf sechs der semantisch assoziierten Listen von Deese (1959b) zurück, die zu den höchsten Raten der falschen Reproduktion geführt hatten. Die aus 12 Items bestehenden Wortlisten waren mit einem kritischen, nicht präsentierten Item hoch assoziiert. Die Wörter waren innerhalb der Liste gemäß ihrer Assoziationsstärke absteigend angeordnet (Roediger & McDermott, 1995). Nach der Präsentation der Items sollten die Probanden zunächst alle Wörter, an die sie sich erinnerten ohne zu raten, reproduzieren, wobei sie mit dem zuletzt gehörten Item beginnen sollten. Die Listen wurden den Probanden so vorgelesen, dass der Abstand zwischen jedem Wort 1,5 Sekunden betrug. Im Anschluss an die Freie Reproduktion folgte ein Wiedererkennenstest, der aus 42 Items bestand, von denen 12 in der Lernphase präsentierte und 30 nichtpräsentierte Items waren. Auf einer vierstufigen Ratingskala wurden die Probanden gebeten die Sicherheit zu ihrem Wiedererkennensurteil anzugeben [(4)= „sicher alt“, (3)= „wahrscheinlich alt“, (2) = „wahrscheinlich neu“, (1)= „sicher neu“].

In der Freien Reproduktion wurden die in der Lernphase präsentierten Wörter mit einer Wahrscheinlichkeit von .65 reproduziert. Die kritischen nicht präsentierten Wörter wurden mit einer Wahrscheinlichkeit von .40 fälschlicherweise reproduziert, was der Reproduktionswahrscheinlichkeit der präsentierten Listenitems aus mittlerer Position entsprach. Andere Wörter wurden kaum genannt (.14), daher konnten Rateeffekte ausgeschlossen werden.

Die Ergebnisse des Wiedererkennenstests zeigten, dass nicht präsentierte kritische Items mit einer Wahrscheinlichkeit von .84 annähernd genauso häufig wie die präsentierten Listenitems als „alt“ (.86) bezeichnet wurden. Darüber hinaus fiel dieses Urteil in 58% der Fälle mit hoher Sicherheit aus, womit Listen- und kritische Items ein ähnlich hohes Konfidenzrating erhielten. Im Gegensatz dazu wurden die nicht assoziierten Items selten (.02) und leicht assoziierte Items etwas häufiger (.21) fälschlicherweise wiedererkannt. Auch erhielten diese Items ein geringeres Sicherheitsurteil.

Das zweite Experiment wurde unter den Zielsetzungen durchgeführt, die Ergebnisse des ersten Experiments mit einer größeren Bandbreite von Wortlisten zu replizieren, zu überprüfen,

ob ein Zusammenhang zwischen den reproduzierten Items mit den später wiedererkannten Items bestand und wie sich die Rate der falschen Alarme für kritische Items im Wiedererkennenstest veränderte, wenn die entsprechenden Wortlisten vorher nicht präsentiert wurden. Anstelle eines Sicherheitsurteils wurde zur Erfassung der phänomenologischen Erfahrung ein "remember/know"-Urteil erhoben (Tulving, 1985). Insgesamt wurden 24 Listen verwendet, die jeweils 15 assoziierte Items enthielten. Von diesen 24 Listen wurden den Probanden jedoch nur 16 präsentiert. Nach der Präsentation von acht Wortlisten sollten die Probanden die Items frei reproduzieren und nach den anderen acht Listen lösten die Personen Mathematikaufgaben. Die verbliebenen acht Listen dienten im Wiedererkennenstest als Distraktoren. In diesem Experiment wurden die Listen durch ein Tonbandgerät mit einem Abstand von 1,5 Sekunden zwischen den Listenwörtern vorgegeben. In der folgenden Testphase wurde der Wiedererkennenstest bearbeitet. Zu „alt“ bezeichneten Items sollte ein "remember/know"-Urteil abgegeben werden. Im Fall einer lebhaften Erinnerung an ein Item sollten die Probanden ein "remember"-Urteil angeben, das somit eine Erinnerung an spezifische Aspekte beinhaltete. Demgegenüber sollte im Fall einer unspezifischen Erinnerung ein "know"-Urteil abgegeben werden. Insgesamt bestand der Wiedererkennenstest aus 96 Items, von denen 48 Listenitems aus der Lernphase und 48 neue Items waren. Bei den 48 neuen Items handelte es sich um die 24 kritischen Items aller 24 Listen und um je drei Items aus den acht nicht präsentierten Listen.

Die Ergebnisse konnten auch in diesem Experiment eine hohe falsche Reproduktion demonstrieren. Mit einer Wahrscheinlichkeit von .55 wurden die kritischen Items häufiger reproduziert als Listenitems aus mittlerer Position (.47).

Im Wiedererkennenstest zeigte sich im Fall der vorherigen Freien Reproduktion eine höhere falsche Wiedererkennensleistung kritischer Items (.81) verglichen mit der Bearbeitung von Rechenaufgaben (.72). Dem Reproduzieren folgte ein höherer Anteil von falschen "remember"-Urteilen (.58) als dem Lösen von Rechenaufgaben (.38). Die Ergebnisse zu den "remember/know"-Urteilen überraschten, da falsche Erinnerungen an nicht präsentierte Items bisher durch "know"-Urteile gekennzeichnet waren (Rajaram, 1993), jedoch in dieser Studie mit einem hohen "remember"-Urteil einhergingen.

2.1.3 Bewertung des DRM-Paradigmas

Rückblickend kann zum DRM-Paradigma festgehalten werden, dass Deeses (1959a, 1959b) Arbeiten den Anstoß für die systematische Erforschung von falschen Erinnerungen für Roe-

diger und McDermott (1995) gaben. Das grundlegende Paradigma von Deese (1959a, 1959b) zur Induktion und Erhebung von Gedächtnisillusionen griffen sie auf und weiteten es auf den Wiedererkennstest aus (Roediger et al., 1998). Man spricht von einem DRM-Paradigma, wenn durch semantisch assoziierte Wortlisten falsche Reproduktionen und falsches Wiedererkennen hervorgerufen werden.

Das DRM-Paradigma hat entscheidend zum Erkenntnisgewinn falscher Erinnerungen beigetragen, da es zuverlässig durch die semantisch assoziierten Wortlisten diese Erinnerungseffekte hervorrufen kann (Blair, Lenton & Hastie, 2002; Roediger & McDermott, 1995). Anhand dieses Paradigmas lässt sich nicht nur die Anzahl, sondern auch die Genauigkeit der Erinnerungen erheben. Da das DRM-Paradigma durch ein standardisiertes Vorgehen im „Labor“ gekennzeichnet ist, erreicht der Untersucher ein hohes Maß an experimenteller Kontrolle einhergehend mit einer sehr zufrieden stellenden Objektivität und internen Validität (Payne, Elie, Blackwell & Neuschatz, 1996). Auch Ghetti et al. (2002) verweisen auf diesen Aspekt, indem sie das DRM-Paradigma mit klassischen Interviews zu tatsächlich erlebten Erinnerungen vergleichen. Während in den Interviews die Antworten durch soziale Erwünschtheit verfälscht sein können, wird dies im DRM-Paradigma dadurch umgangen, dass die Probanden instruiert sind, möglichst viele Listenitems und nur diese zu erinnern. Um die Validität des Ergebnisses im DRM-Paradigma zu sichern, werden Antwortkriterien und Sensitivitätsmaße bestimmt (Gallo, Roediger & McDermott, 2001b; Miller & Wolford, 1999; Wickens & Hirshman, 2000; Wixted & Stetch, 2000), so dass geprüft werden kann, ob die Personen sich an die Instruktion gehalten und gemäß dieser geantwortet haben.

Blair et al. (2002) widmeten sich der Untersuchung der Reliabilität des DRM-Paradigmas und stellten fest, dass sowohl falsche als auch richtige Erinnerungen als reliabel bezeichnet werden können. Retestkoeffizienten des zweiwöchigen Zeitabstandes der Untersuchungen weisen auf eine bedeutende Stabilität der Erinnerungen hin. Zu beiden Messzeitpunkten produzierten die Probanden ein ähnliches Maß richtiger und falscher Erinnerungen. Obwohl es sich nicht um identische Listen bei beiden Messzeitpunkten handelte, erzielten jene Versuchspersonen, die im ersten Test einen hohen Anteil falscher Erinnerungen hatten auch im zweiten Test ein vergleichbares Ausmaß, womit die falschen Erinnerungen des ersten Messzeitpunktes zuverlässige Prädiktoren für die falschen Erinnerungen des zweiten Messzeitpunktes waren. Die Autoren schlussfolgern, dass die Reliabilität falscher Erinnerungen im DRM-Paradigma nicht durch allgemeine Antworttendenzen erklärt werden kann.

Wortintrusionen der kritischen Items können in der Freien Reproduktion reliabel erzielt werden (Deese, 1959b); zudem lassen sich im Wiedererkennenstest zuverlässig falsche Erinnerungen testen (Roediger & McDermott, 1995). Der Vorteil des DRM-Paradigmas zeigt sich darin, dass es nicht an eine bestimmte Testart gebunden ist: Nicht nur verschiedene explizite Tests (Reproduktion und Wiedererkennen), sondern auch verschiedene implizite Tests (konzeptuelle und perzeptuelle) bilden das Gedächtnisphänomen falscher Erinnerungen ab (McDermott, 1997). Zwar sind die Raten falscher Erinnerungen dahingehend unterschiedlich, dass sie im Wiedererkennenstest höher ausfallen als in der Freien Reproduktion (Roediger & McDermott, 1995) und dass sie deutlicher in konzeptuellen als in perzeptuellen Tests vorzufinden sind (McDermott, 1997), aber diese Befunde liefern damit ein differenzierteres Verständnis falscher Erinnerungen.

Ein Kritikpunkt der Untersuchungen zum DRM-Paradigma richtet sich auf die geringe externe Validität, da viele Untersuchungen im Labor stattfanden und Studenten als Stichprobe nutzten (Roediger & McDermott, 1995). Auch ist es schwierig diese Untersuchungen auf Fragestellungen zur Missbrauchsdebatte zu übertragen (s.o.). Demnach wären die Ergebnisse nur bedingt interpretierbar und generalisierbar.

Das DRM-Paradigma hat in der Forschung trotz solcher ungeklärter Probleme einen hohen Stellenwert erlangt, da es zum einen eine zuverlässig hohe Anzahl falscher Erinnerungen hervorrufen kann und zum anderen einen phänomenologischen Vergleich richtiger und falscher Erinnerungen ermöglicht, was sich im "remember"-Urteil der Probanden widerspiegelt (Miller & Wolford, 1999; Payne et al., 1996; Roediger & McDermott, 1995).

Auch liegt ein großer Vorteil des Paradigmas darin, dass es verschiedenen Stichproben vorgegeben werden kann. Demnach lassen sich falsche Erinnerungen älterer und jüngerer Erwachsener miteinander vergleichen (Norman & Schacter, 1997; Schacter, Israel & Racine, 1999), ebenso ist ein Kontrast dieser Gruppen mit älteren und jüngeren Kindern, aber auch eine alleinige Untersuchung der falschen Erinnerungen im Kindesalter möglich (Brainerd et al., 2002a). Auf diese Weise lassen sich Aussagen über Entwicklungstrends falscher Erinnerungen treffen. Darüber hinaus kann durch das DRM-Paradigma das Verständnis unterschiedlicher Krankheitsbilder erweitert werden: Einige Untersuchungen nutzten es, um die kognitiven Beeinträchtigungen Amnesie- und Demenzerkrankter genauer zu erfassen und um uneingeschränkte kognitive Funktionen ausfindig zu machen (Norman & Schacter, 1997; Schacter, Verfaillie & Pradere, 1996).

Zudem erweist sich das DRM-Paradigma nicht nur in der Grundlagenforschung als nützlich, es hat auch zu einem Erkenntnisgewinn in angewandten Disziplinen geführt. Nicht nur Kinder, auch Erwachsene unterliegen dem Gedächtnisphänomen falscher Erinnerungen, welches die Genauigkeit von Zeugenaussagen bei Gericht beeinträchtigen kann (Miguel & Garcia-Bajos, 1999). Auch die Erkenntnis, dass falsche Erinnerungen durch ein Gefühl der lebhaften Erinnerung gekennzeichnet sein können, relativiert die Glaubwürdigkeit wiedererlebter Kindheitserinnerungen, besonders im Rahmen von (Psycho-)Therapien (Andrews, 1997; Gudjonsson, 1997; Loftus, 1993; Roediger, et al., 1997).

2.2 Theoretische Erklärungsansätze

Ausgelöst durch zahlreiche Befunde falscher Erinnerungen im Rahmen des DRM-Paradigmas entwickelten verschiedene Forschungsgruppen Erklärungen dieser Gedächtnisillusion. Im Folgenden werden zunächst verschiedene Theorien zusammenfassend vorgestellt; dem folgt die ausführliche Darstellung des Aktivierungs-Monitoring Ansatzes (AMA), welches die Grundlage der eigenen Studie bildet. Überblicksartig sollen die Theorien miteinander verglichen werden, so dass sich Gemeinsamkeiten und auch Unterschiede abbilden lassen.

2.2.1 Diverse Ansätze

Um das Phänomen falscher Erinnerungen im DRM-Paradigma zu erklären wurden verschiedene Theorien postuliert, von denen jedoch nur zwei, der AMA und die Fuzzy Trace Theorie (FTT), einen besonderen Stellenwert in der Forschung erlangen konnten. Während im nächsten Abschnitt auf den AMA separat eingegangen wird, sollen zunächst die Kriteriumsverschiebungstheorie ("Criterion-Shift"; Miller & Wolford, 1999) und die Distinktivitätsheuristik (Dodson & Schacter, 2001, 2002; Israel & Schacter, 1997; Schacter et al., 1999) erörtert werden, denen sich eine ausführliche Darstellung der FTT anschließt.

Die Vertreter der Kriteriumsverschiebungstheorie (Miller & Wolford, 1999) sind der Auffassung, dass falsche Erinnerungen im DRM-Paradigma das Resultat eines liberalen Antwortkriteriums sind. Diese Theorie wurde besonders heftig angegriffen und als methodisches Artefakt dargestellt, denn eine gezielte Anweisung zur Veränderung des Antwortkriteriums durch spezielle (Vor-)Warnungsinstruktionen konnte kein konservativeres Antwortverhalten hervorrufen und damit zu einer willkürlichen Reduktion falscher Erinnerungen führen (Gallo et al., 2001b; Roediger & McDermott, 1999). Gegen die Kriteriumsverschiebungstheorie ist einzu-

wenden, dass Probanden im Metagedächtnisurteil kritischen Items eine “remember“-Antwort zuteilen, die durch klare Details gekennzeichnet ist. Demnach erinnern sie tatsächlich die vermeintliche Präsentation des kritischen Items. Zudem stellten Wixted und Stretch (2000) korrespondierende mathematische Modelle sowohl für Roediger und McDermotts (1995) AMA und für Miller und Wolfords (1999) Kriteriumsverschiebungstheorie auf und konnten anhand von vergleichenden Analysen zeigen, dass beide Modelle zu ähnlichen Vorhersagen kamen, da ihre Parameter der Signalentdeckungstheorie identisch waren. Somit trägt die Kriteriumsverschiebungstheorie von Miller und Wolford zu keinem Erkenntnisgewinn bei.

Nach der Distinktivitätsheuristik können Entscheidungsprozesse falsches Wiedererkennen im Rahmen des DRM-Paradigmas beeinflussen (Dodson & Schacter, 2001; Dodson & Schacter, 2002; Israel & Schacter, 1997; Schacter et al., 1999). Diese Heuristik widmet sich der Erklärung durch welche Faktoren sich falsche Erinnerungen *reduzieren* lassen. Richtige Erinnerungen sind durch distinktive Details gekennzeichnet. Stimmen Reiz (d.h. das kritische Item) und die Details des Gedächtnisinhalts überein, so wird der Reiz als „alt“ wiedererkannt. Je deutlicher der Reiz von den Details der Gedächtnisinhalte differiert desto eher wird er als „neu“ erkannt, d.h. falsche Erinnerungen werden reduziert. Es wird beispielsweise davon ausgegangen, dass Bildmaterial mehr Details enthält und damit distinktiver ist als Wortmaterial. Erhalten Versuchspersonen in der Lernphase Bildmaterial, so wird davon ausgegangen, dass beim Abruf für gelernte Listenitems bildhafte Vorstellungen zugänglich sind, welche dazu führen, dass diese als „alt“ bezeichnet werden. Demgegenüber sollen diese Bildinformationen bei kritischen Items fehlen und damit die Entscheidung zum „neu“-Urteil vereinfachen. Verglichen mit Wortmaterial zeigten Schacter et al. (1999) zudem, dass die Enkodierbedingung Bildinformation zu weniger falschen Erinnerungen führt. Auch beinhaltet die Listenpräsentation „sagen vs. hören“ unterschiedlich distinktive Informationen. Dodson und Schacter (2001) überprüften dies und konnten feststellen, dass die von den Probanden selbst vorgelesenen Listen spezifischere Informationen beinhalteten und demzufolge die Anzahl falscher Erinnerungen reduzierten.

Einige Befunde, wie zum Beispiel die zum Verlauf korrekter und falscher Erinnerungen in Abhängigkeit des Listenlernens (Seamon, Goodkind, Dumey, Dick, Aufsesser, Strickland, Woulfin & Fung, 2003; s. Abschnitt 2.2.2 und 2.3.4), lassen sich nicht mit Ein-Prozess-Theorien erklären. Hier geht es darum, dass ein verbessertes Lernen der Wortliste einerseits einen Anstieg korrekter Erinnerungen, andererseits aber einen umgekehrt U-förmigen Verlauf falscher Erinnerungen bedeutet. Anhand der FTT, ebenso wie anhand des AMA, lassen sich diese Befunde durch zwei Prozesse erklären (im Gegensatz zu der Kriteriumsverschiebungs-

theorie und der Distinktivitätsheuristik, die einen Prozess annehmen). Während der AMA einen Aktivierungs- und einen Quellendiskriminationsprozess postuliert, bezeichnet die FTT einen ersten Prozess als “gist“ und einen anderen, Kontrollprozess, als “memory editing“. Falsche Erinnerungen entstehen durch Generierungsprozesse, die “gist“-Spuren in der Lernphase erzeugen, wobei das “memory editing“ falschen Erinnerungen in der Testphase durch Prüfung der Vertrautheit entgegenwirkt. Die Grundannahmen der FTT beruhen auf Forschungsarbeiten von Reyna und Kiernan (1994), die im falschen Wiedererkennen den Einfluss höherer kognitiver Prozesse vermuteten. Während Vertreter des Konstruktivismus und der Schematheorie altersbedingte Veränderungen im falschen Wiedererkennen als konstruktive Inferenz betrachteten und eine Steigerung der Gedächtnisleistungen auf eine zunehmende Entwicklung des logischen Denkens zurückführten (Piaget & Inhelder, 1973), hatten Reyna und Kiernan (1994) aufgrund bisheriger Studien zu Gedächtnisleistungen bei Kindern eine andere Auffassung. Erlebte Erfahrungen entstehen durch zwei getrennte Gedächtnisspuren, die als “verbatim“ und “gist“ bezeichnet werden (Reyna & Kiernan, 1994). Während die verbatim-Gedächtnisspuren die Oberflächenmerkmale einer Erfahrung mit ihren spezifischen Details (z.B. Stimme oder Lautstärke mit der Listenitems präsentiert werden) beinhalten, wird in den gist-Gedächtnisspuren der wesentliche Bedeutungsgehalt erfasst (Brainerd, Holliday & Reyna, 2004; Brainerd & Reyna, 1998a, 1998b; Brainerd & Reyna, 2002b; Brainerd & Reyna, 2004; Brainerd, Reyna & Brandse, 1995a; Brainerd et al., 1995b; Reyna, 1992). Die verbatim- und gist-Gedächtnisspuren beruhen beide auf Inputreizen, jedoch werden sie getrennt voneinander gespeichert. Schacter (1999) beschreibt die beiden Gedächtnispräsentationen als ereignisspezifisches Wissen (“verbatim“) und allgemeines Ereigniswissen (“gist“) und stellt die Beziehung so dar, dass die Enkodierung spezifischer Merkmale eines Ereignisses zu einer Speicherung von ereignisspezifischem Wissen führt und gleichzeitig allgemeines Ereigniswissen gespeichert wird. Erfahrungen werden somit in Bezug zum Vorwissen gesetzt und interpretiert.

Während die Enkodierung von “verbatim“- und “gist“-Gedächtnisspuren parallel verläuft, erfolgt der Abruf sequentiell. Hier sind Abrufhilfen, Zeitintervall zwischen Darbietung und Testung der Items und Verfügbarkeit der Gedächtnisinhalte zu nennen. Ein Beispiel für Abrufhilfen der “verbatim“-Gedächtnisspuren (z.B. für eine präsentierte Liste wie “hound, poodle, retriever, spaniel“) ist die Darbietung eines bereits präsentierten Items (z.B. “spaniel“). Für “gist“-Gedächtnisspuren sind nie präsentierte Items, die mit der Bedeutung der präsentierten Ereignisse übereinstimmen, bessere Abrufhilfen (z.B. “collie“ für “dog“) (Reyna & Brainerd, 1995).

Zu den Kontrollprozessen, die von der FTT als “memory editing“ bezeichnet werden, gehört die so genannte “recollection rejection“ (Brainerd & Reyna, 2002b; Odegard & Lampinen, 2005). Durch die Testitems werden “verbatim“- oder “gist“-Spuren hervorgerufen. Im Fall der “verbatim“-Spuren findet ein Vergleich mit den “verbatim“-Merkmalen der Testitems statt und es wird ein Identitätsurteil gefällt, wenn das kritische Item der “verbatim“-Spur vollständig entspricht (Odegard & Lampinen, 2005). Ansonsten wird es als „neu“ erkannt, auch wenn es mit der Bedeutung der Listenitems übereinstimmt (Brainerd et. al., 1995b). Analog dazu werden “gist“-Spuren mit dem Bedeutungsgehalt der Testitems verglichen und es werden Ähnlichkeitsurteile durch Merkmalsüberschneidungen hervorgerufen.

Für richtiges und falsches Wiedererkennen spielen unterschiedliche Gedächtnisspuren eine Rolle. Anhand verschiedener Studien war es möglich beide Gedächtnisspuren getrennt zu untersuchen, so dass richtige Erinnerungen auf “verbatim“- und “gist“-Spuren zurückzuführen waren (Brainerd & Reyna, 2004; Brainerd et al., 2004; Reyna, 1995; Reyna & Kiernan, 1994). Waren die spezifischen “verbatim“-Merkmale mit den präsentierten Items identisch, so bezeichneten die Probanden das Item im Wiedererkennenstest korrekt als gelernt, gaben also ein Identitätsurteil; bei mangelnder Übereinstimmung erkannten sie es als nicht präsentiert (Brainerd et al., 1995b; Brainerd et al., 2004; Brainerd & Reyna, 2002a; Brainerd & Reyna, 2004). Im Fall der Freien Reproduktion ermöglichten die “verbatim“-Gedächtnisspuren einen Zugriff auf die gelernten Items, so dass deren Abruf möglich war (Brainerd, Wright, Reyna & Payne, 2002b). Die “gist“-Gedächtnisspuren wirkten sich insofern auf richtige Erinnerungen aus, dass die Bedeutung eines Items vertraut war (Brainerd & Reyna, 2002a). Erinnerungen lassen sich nach phänomenologischen Merkmalen unterteilen. In der FTT liegt dem Abruf von “verbatim“-Gedächtnisspuren eine lebhaftere Form der Erinnerung mit bewussten Erinnerungen an das Item und den Kontext zugrunde (“recollection“), demgegenüber zeichnet sich der Abruf von “gist“-gestützten Gedächtnisspuren durch eine unspezifische Vertrautheit (“familiarity“) aus. Durch experimentelle Manipulation der Lernbedingungen kann “phantom recollection“ entstehen, wenn die Erinnerung an nie dargebotene Items jener der präsentierten phänomenologisch ähnlich ist (Brainerd & Reyna, 1998; Brainerd & Reyna, 2002a). Sind die “gist“-Spuren stark ausgeprägt, weil z.B. das Listenthema während der Enkodierphase mehrfach präsentiert wurde, kann ein Distraktor, der dieses Thema repräsentiert, eine falsche lebhaftere Erinnerung an das nicht gelernte Item hervorrufen.

2.2.2 Der Aktivierungs-Monitoring Ansatz (AMA)

Als Roediger und McDermott (1995) anhand ihres Vorgehens zeigten, dass falsche Erinnerungen durch ein neues experimentelles Design zuverlässig hervorgerufen werden konnten, wollten sie dieses Phänomen auch erklären. Von ihnen wurde der AMA vorgestellt, der mit dem Beginn des DRM-Paradigmas entwickelt und in der darauf folgenden Zeit durch weitere Befunde differenziert wurde (Gallo & Roediger, 2002; McDermott & Watson, 2001; Roediger & McDermott, 2000; Roediger et al., 2001).

Der AMA ist auch eine Zwei-Prozess-Theorie (s.o.), nach welcher die Prozesse Aktivierung und Monitoring zusammen zum Entstehen falscher Erinnerungen im DRM-Paradigma beitragen. Durch die Listenitems wird das zugehörige nicht präsentierte kritische Item aktiviert. Das Monitoring wirkt falschen Erinnerungen durch Quellendiskriminationsprozesse entgegen, wobei geprüft wird, ob Items tatsächlich präsentiert wurden. Während die Aktivierung sich vorrangig auf die Enkodierphase bezieht, ist das Monitoring hauptsächlich der Abrufphase zugeordnet. Jedoch können diese Prozesse auch in beiden Phasen wirksam werden (Roediger et al., 2001). Für die Bedeutung der Aktivierungsprozesse spricht, dass sich in einer multiplen Regressionsanalyse (Roediger et al., 2001) die Stärke der Rückwärtsassoziation ("backward associative strength"), d.h. der assoziativen Verbindungen der Listenitems, als bester Prädiktor für falsches Reproduzieren und Wiedererkennen erwies. Je stärker die Listenitems mit dem kritischen Item assoziiert sind, umso mehr Aktivierung breitet sich zum kritischen Item aus.

In diesem Zusammenhang ist die von Collins und Loftus (1975) entwickelte Theorie der Aktivierungsausbreitung im semantischen Netzwerk relevant. Es wird davon ausgegangen, dass die Netzwerke hierarchisch aufgebaut sind und aus Inhalten oder Knoten bestehen, die über verschiedene Assoziationen (z.B. Kategorienzugehörigkeit, Merkmale, assoziierte visuelle Vorstellungen) unterschiedlich stark miteinander verknüpft sind. Die Stärke einer Verbindung von Knoten A zu Knoten B stellt die Wahrscheinlichkeit der Aktivierung des Knotens B als Folge der Aktivierung des Knotens A dar. Diese Zugänglichkeit resultiert durch die Häufigkeit mit der die Verbindung in der Vergangenheit aktiviert wurde. Die Inhalte oder Knoten repräsentieren Objekte oder Konzepte (z.B. Gattung Tier) und ähnliche Inhalte sind miteinander assoziiert.

Die Organisation semantischer Netzwerke resultiert aus dem Prinzip der semantischen Ähnlichkeit, d.h. je mehr Eigenschaften zwei Begriffe miteinander verbinden desto vielfältiger

und enger sind die beiden Begriffe miteinander vernetzt. Die Verbindungsstärke zwischen zwei Knoten wird als semantische Ähnlichkeit bezeichnet.

Bei der Verarbeitung eines Begriffs wird der dazugehörige Knoten im semantischen Netzwerk aktiviert und die Aktivierung breitet sich über die Verbindungen im Netzwerk auf andere Knoten aus, die wiederum die erhaltene Aktivierung an andere verbundene Knoten weitergeben. Mit zunehmender Weiterleitung der Aktivierung nimmt diese über die Knoten ab, wird schwächer.

Die Aktivierung einzelner Verbindungen ist abhängig von der Verbindungsstärke. Es wird angenommen, dass nicht nur Knoten, sondern auch Verbindungen aktiviert werden können.

Anhand einer Computersimulation des Konzepts der Aktivierungsausbreitung wurde das Modell von Collins und Loftus von Rabinowitz und Chi (1978) spezifiziert. Die Aktivierung eines Knotens wird nur dann im Netzwerk weitergeleitet, wenn ein Schwellenwert überschritten wird. Somit hängt die Aktivierungsausbreitung nicht nur von der Verbindungsstärke zwischen zwei Knoten ab, sondern auch vom Aktivierungsniveau des Ursprungsknotens. Dies lässt sich an einem Beispiel verdeutlichen: Wenn der Knoten „bellen“ aktiviert wird, breitet sich die Aktivierung von „bellen“ zu „Hund“ dann mit hoher Wahrscheinlichkeit aus, wenn zwischen den beiden Knoten eine starke Verbindung besteht oder wenn bei mittlerer Verbindungsstärke die Aktivierungsstärke des Knotens „bellen“ durch itemspezifische Verarbeitungsprozesse erhöht wird.

Diese grundlegende Theorie der Aktivierungsausbreitung wurde von Hutchison und Balota (2005) aufgegriffen und aktualisiert. Es wird auf die Bedeutung assoziativer Netzwerke im Gegensatz zu semantischen hingewiesen. Bei der assoziativen Verbindung von Wörtern muss nicht zwingend eine semantische Ähnlichkeit bestehen. Untersuchungen zur lexikalischen vs. semantischen Aktivierung zeigten (Balota & Paul, 1996), dass Priming für homographische (unterschiedliche Bedeutung gleich geschriebener kritischer Items wie z.B. “kidney-piano-ORGAN“) und nicht homographische kritische Items (gleiche Bedeutung unterschiedlich geschriebener kritischer Items wie z.B. “lion-stripes-TIGER“) vergleichbar war, was verdeutlicht, dass die semantische Aktivierung nicht überschätzt werden sollte.

Ein Netz kann Konzepte enthalten, die gleichzeitig Mitglieder anderer Netze sind. Demnach besteht die Möglichkeit, dass verschiedene Netze über gemeinsame Konzepte miteinander in Verbindung stehen und damit ein größeres Netzwerk entsteht. Die von einer Wortliste ausgehende Aktivierung kann sich im assoziativen Netzwerk ausbreiten und auch Konzepte ansprechen, die zwar nicht präsentiert wurden, aber mit den Listenitems verbunden sind.

Diese Annahmen zu den Netzwerkmodellen bilden die Grundlage zum Verständnis des AMA, genauer gesagt, des Aktivierungsprozesses.

Die Aktivierung des kritischen Items findet nach Roediger und McDermott (1995) durch die semantisch assoziierten Listenwörter statt. Nach dem AMA kann die Aktivierung das Resultat automatischer oder kontrollierter, semantisch-elaborierter Verarbeitung sein. Underwood (1965) war der Auffassung, dass im Sinne der "implicit associative response" (IAR), eine bewusste Aktivierung des kritischen Items durch das Lernen semantisch assoziierter Wortlisten zustande kam. Nach Hutchison und Balota (2005) könnte aber auch eine automatische Aktivierungsausbreitung über assoziative Verbindungen Netzwerk erfolgen. Das unterschiedliche Verständnis über den Aktivierungsprozess löste weitere Untersuchungen aus, welche zu der Annahme führten, dass sowohl die bewusste als auch die automatische Aktivierung zusammen eine Rolle spielen könnten (Roediger et al., 1998). McDermott (1997) konnte in ihrer Studie Hinweise auf bewusste und automatische Prozesse erbringen. Durch die Präsentation von Wortlisten nach dem DRM-Paradigma zeigte sich in konzeptuellen und perzeptuellen impliziten Tests Priming des nicht präsentierten kritischen Items. Aufgrund dessen, dass eine automatische Aktivierung nicht zu perzeptuellem Priming führen kann, schlussfolgerte sie, dass die Versuchspersonen während der Präsentation der Listen bewusst an das kritische Item gedacht hatten. Zeelenberg, Plomp und Raaijmakers (2003) unterstützen die Annahme bewusster Prozesse, indem sie anhand einer Studie mit dem DRM-Paradigma zeigen konnten, dass im Fall einer sehr kurzen (20 ms) Darbietung der Items richtige und falsche Erinnerungen beeinträchtigt waren. Zu ähnlichen Schlussfolgerungen gelangten Untersuchungen, die den Effekt der Vorwarnung prüften (Gallo, Roberts & Seamon, 1997; McDermott & Roediger, 1998). Durch die Vorwarnung ließen sich falsche Erinnerungen reduzieren.

Demgegenüber sprechen andere Befunde für eine automatische Aktivierung: Seamon et al. (1998) entdeckten Effekte falschen Erinnerns, wenn Listenitems in der Lernphase kurz dargeboten wurden (ebenso 20 ms wie bei Zeelenberg et al., 2003) und die Möglichkeit der korrekten Erinnerung sehr eingeschränkt war.

Anhand einer Untersuchung zum gerichteten Vergessen konnten Kimball und Bjork (2002) zeigen, dass sich kein Effekt für kritische Items zeigte. Unabhängig von der Lernbedingung (Erinnerungs- oder Vergessensinstruktion) wurden kritische Items genannt, womit die Bedeutung automatischer Prozesse herausgestellt wurde. Nur Listenitems wurden in der Erinnerungsinstruktion besser wiedergegeben als in der Vergessensinstruktion. Die Anweisung, ge-

lernte Wörter zu vergessen, scheint demnach einen Einfluss auf den Abruf zu haben, nicht aber auf die Aktivierung.

Auch Seamon, Lee, Toner, Wheeler, Goodkind und Birch (2002a) erbrachten Befunde, die für automatische Aktivierungsprozesse sprechen. Während eine Hälfte der Probanden die Aufforderung erhielt laut zu memorieren, wurde der anderen Hälfte der Probanden eine normale Lerninstruktion erteilt. Durch das Memorieren eines kritischen Items in der Lernphase wurde dieses in der Freien Reproduktion häufiger falsch erinnert. Aber die Wahrscheinlichkeit des falschen Wiedererkennens eines kritischen Items wurde von der "rehearsal"-Bedingung nicht beeinflusst [was im Gegensatz zu den Befunden einer ähnlichen Studie von Marsh und Bower (2004) steht, die in Abschnitt 5.3 erläutert wird]. Vorher memorierte und nicht memorierte kritische Items wurden mit einer ähnlichen Wahrscheinlichkeit falsch wiedererkannt, so dass davon ausgegangen werden kann, dass keine bewusste Aktivierung in der Enkodierphase notwendig war. Im Fall geteilter Aufmerksamkeit konnten Seamon und Mitarbeiter (2003) eine Verschlechterung der korrekten Erinnerung feststellen, jedoch keine Veränderung der falschen Erinnerung. Demnach schienen zwar bewusste Prozesse auf das richtige Erinnern, nicht aber auf falsches einzuwirken. Gallo und Seamon (2004) führten eine weitere Studie durch die Klarheit schaffen sollte. In der Lernphase erhielten die Probanden auf einem Computerbildschirm kurz präsentierte Wortlisten (20 ms pro Listenitem). Vor und nach jedem Listenitem folgte ein Maskierungsreiz. Nach jeder Liste, in Form einer Freien Reproduktion, sollten die Probanden so viele Wörter aufschreiben wie sie erinnern konnten. Im Anschluss an die Wortlisten wurde, ohne es den Probanden vorher mitgeteilt zu haben, ein "forced-choice" Wiedererkennenstest durchgeführt, der in aufeinander folgenden Wortpaaren auf dem Bildschirm präsentiert wurde. Die Probanden wurden angewiesen, dass immer ein Item des Wortpaares gelernt und das andere neu war, wobei sie das gelernte Item herausfinden sollten. Die Wiedererkennensleistung dieses Tests konnte in Bezug zu der Freien Reproduktion gesetzt werden und die Ergebnisse zeigten, dass eine bewusste Generierung der kritischen Items nicht notwendig für das Wiedererkennen war, denn falsches Wiedererkennen trat auch ohne vorherige (falsche) Reproduktion auf. Jedoch war das Wiedererkennen kritischer Items von der Anzahl der Listenitems in der Freien Reproduktion abhängig. Wurden keine Listenitems genannt oder wiedererkannt, dann lag die Rate falschen Wiedererkennens des kritischen Items im Zufallsbereich (.49). Wurde hingegen nur ein Listenitem erkannt, so erhöhte sich das Wiedererkennen des kritischen Items über den Zufall (.66). Dies deutet darauf hin, dass bei falschem Wiedererkennen des kritischen Items eine minimale bewusste Verarbeitung der Listenitems stattfand. Es erscheint offensichtlich, dass die Aktivierungsprozesse bewusst und auto-

matisch ablaufen, daher fordern die Autoren eine Untersuchung der Wirkweise der beiden Prozesse, was auch in der eigenen Studie von Relevanz ist.

Der Aktivierungsprozess führt nicht alleine zu falschen Erinnerungen, sondern es tritt ein weiterer Prozess hinzu: Das Monitoring, genauer Quellendiskrimination, bezieht sich primär auf die Testphase und beschreibt einen Kontrollprozess. Hier geht es darum zu entscheiden, ob ein Item in der Lernphase tatsächlich präsentiert (externe Quelle) oder selbst generiert wurde (interne Quelle). Roediger et al. (2001) übernehmen den aus dem Forschungsfeld der Quellenattribution stammenden "Monitoring Ansatz" von Johnson und Raye (1981). Das "reality monitoring" bezeichnet einen Entscheidungsprozess nach dem eine Erinnerung auf eine interne oder externe Quelle zurückgeführt wird. Die Quellendiskrimination ("source monitoring") tritt sowohl in der Enkodier- als auch in der Abrufphase auf und wird bewusst eingesetzt, um Erinnerungen einer Quelle zuzuschreiben. In der Enkodierphase müssen Ereignisse danach eingeteilt werden, ob sie tatsächlich geschehen sind oder sich nur gedanklich abspielten. Beim Abruf geht es darum kritischen Items eine interne Quelle zuzuschreiben und Listenitems eine externe Quelle. Wurden kritische Items in der Lernphase aktiviert, dann treten nur falsche Erinnerungen auf, wenn es in der Testphase zu Fehlern bei der Quellendiskrimination oder genauer im "reality monitoring" kommt, d.h. wenn die Probanden die selbst generierten kritischen Items fälschlicherweise für kritische Items halten (Seamon, Luo, Schwartz, Jones, Lee & Jones, 2002c).

Gemäß des Quellendiskriminationsansatzes (Johnson et al., 1993; Johnson & Raye, 1981) wird die Quelle einer Information nicht zusammen mit dieser abgespeichert, sondern in der Testphase aus enkodierten Attributen rekonstruiert. Informationen aus einer externen Quelle sind durch mehr perzeptuelle (Stimme, Schreibweise) und kontextuelle Attribute gekennzeichnet als Informationen aus einer internen Quelle, welche demgegenüber mehr Attribute über kognitive Operationen aufweisen. Die Informationen aus externen und internen Quellen werden von Johnson und Raye wie folgt voneinander abgegrenzt (1981, p. 71):

(...) externally generated memories in general may have more spatial and temporal contextual attributes coded in the representation of the event than internally generated memories do. (...) they should also have more sensory attributes (...) we propose that externally generated representations are more semantically detailed - that is contain more information or more specific information - than internally generated representations.

(...) internally generated memories may typically have more operational attributes associated with them.

Diese Unterschiede in der Information nach externen und internen Merkmalen sollen für die Entscheidung ausschlaggebend sein, ob Ereignisse tatsächlich aufgetreten sind. Selbst generierte Ereignisse können extern präsentierten sehr ähnlich sein, so dass das "reality monitoring" erschwert ist. Kritische Items können während der Enkodierung der Wortliste Merkmale der Listenitems annehmen, was sich z.B. in den "remember"-Urteilen widerspiegelt (Tulving, 1985; s. Abschnitt 2.1.2). Die Aufgabe zur Quellendiskrimination im DRM-Paradigma ist demzufolge eine schwierige.

Johnson, Hashtroudi und Lindsay (1993) unterscheiden zwei Arten von Entscheidungsprozessen bei der Quellendiskrimination: Zum einen einen heuristischen Prozess, der einer geringen Kontrolle bedarf und bei dem zur Entscheidung, ob ein Ereignis tatsächlich stattfand oder nicht einfache Regeln angewendet werden. Zum anderen wird ein systematischer Prozess postuliert, der strategischer abläuft und einer größeren Kontrolle bedarf. Hier spielt das Metagedächtnis eine Rolle, genauer das metamnestische Wissen. Welcher dieser Prozesse beim Abruf genutzt wird ist von einer Vielzahl von Faktoren abhängig (z.B. Qualität und Quantität an Informationen; McCabe & Smith, 2002). Johnson und Mitarbeiter (1993, p.5) stellen beide Prozesse wie folgt gegenüber und betonen die Abhängigkeit voneinander:

Source judgements are typically made heuristically; systematic processes are engaged less often and tend to be slower and more subject to disruption. Both can provide important checks on each other. For example, systematic processes can, on the basis of implausibility, provide a challenge to memories that, say, passed a heuristic check for reality monitoring because of high vividness. Conversely, heuristic processes can challenge recollections (on basis such as lack of sensory detail) that would otherwise be accepted because they fitted with one's general knowledge and beliefs.

Um Enkodierprozesse und ihren Bezug zu Abrufprozessen besser zu verstehen unterscheiden Hunt und McDaniel (1993) itemspezifische und semantisch-relationale Verarbeitung. Bei der itemspezifischen Verarbeitung werden Merkmale eines Items enkodiert, die es von anderen Items der Liste unterscheiden. Richtige Erinnerungen werden gefördert und falsche reduziert (Roediger et al., 2001). Demgegenüber bezieht sich die semantisch-relationale Verarbeitung auf die Ähnlichkeitsbeziehung zwischen den Listenitems, wodurch die Unterscheidung richtiger und falscher Erinnerungen erleichtert wird. Während der Enkodierung können durch Assoziation oder Inferenzen gemeinsame Merkmale der Items zu späteren falschen Erinnerungen führen, da in Beziehung stehende Informationen aktiviert werden. Konkret bedeutet dies für falsche Erinnerungen im DRM-Paradigma, dass durch die Präsentation von Wortlisten, bei denen semantisch verbundene Listenitems in absteigender Assoziationsstärke zum nicht prä-

sentierten kritischen Item dargeboten werden, die entsprechenden kritischen Items aktivieren. Während itemspezifische und semantisch-relationale Verarbeitungsprozesse der Erinnerung an Listenitems zugrunde liegen, basiert die Erinnerung an kritische Items hauptsächlich auf semantisch-relationalen Verarbeitungsprozessen (Norman & Schacter, 1997; Roediger et al., 2001). Wird das kritische Item während der Lernphase stark aktiviert, so kann dieses zusätzlich sensorisch-kontextuelle Merkmale der Listenitems, ähnlich der itemspezifischen Verarbeitung, annehmen (Roediger et al., 2001). Negative Korrelationen von korrekter Erinnerung an Listenitems und falscher Erinnerung des kritischen Items (Roediger et al., 2001) lassen sich dadurch erklären, dass Probanden itemspezifische Informationen nutzen, um zuvor präsentierte Listenitems von kritischen Items zu unterscheiden. Das kritische Item wird nicht reproduziert, da es weniger itemspezifische Informationen enthält. Durch eine Verbesserung der itemspezifischen Verarbeitung (Präsentation von bildhaftem Material: Schacter et al., 1999; langsame Präsentationsrate: Gallo & Roediger, 2002), lassen sich korrekte Erinnerungen fördern und falsche reduzieren. Nach dem so genannten “more is less“ Effekt (s. Abschnitt 2.3.6) unterstützt die semantisch-relationale Verarbeitung gleichermaßen richtige *und* falsche Erinnerungen (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001).

Aufschluss über das Zusammenspiel von Aktivierung und Quellendiskrimination gibt eine Studie von McDermott und Watson (2001) über die Effekte der Darbietungsdauer auf falsche Erinnerungen. Nach ihren Befunden scheinen Aktivierung und Quellendiskrimination antagonistisch zu wirken. Die Wahrscheinlichkeit ein kritisches Item fälschlicherweise wiederzugeben nimmt in Abhängigkeit der Darbietungsdauer einen umgekehrt U-förmigen Verlauf an. Diese Befunde werden in der Abbildung zur Güte des Listenlernens (“level of list learning“) von Seamon et al. (2003) zusammengefasst und im Abschnitt 2.3.4 aufgegriffen. Richtige und falsche Erinnerungen werden in Bezug zur Güte des Listenlernens gesetzt, die der Präsentationsdauer oder der wiederholten Darbietung der Wortlisten entsprechen. Wurde bei sehr kurzen Darbietungsraten die Darbietungsdauer von 20 auf 250 ms pro Wort verlängert, so stieg der Anteil falscher Erinnerungen in der Freien Reproduktion von .14 auf .31 (McDermott & Watson, 2001). Bei längerer Darbietungsdauer (1000, 3000 oder 5000 ms pro Wort) reduzierte sich der Anteil falscher Erinnerungen von .22 auf .14. Demgegenüber stieg der Anteil richtiger Erinnerungen mit zunehmender Darbietungsdauer. Die vorgefundenen Ergebnisse bestätigten die Hypothese der Autoren, dass Aktivierung und Quellendiskrimination antagonistische Prozesse sind. Im Fall zunehmender Darbietungsdauer der Wortlisten bei kurzen Darbietungszeiten wurde der Anstieg falscher Erinnerungen darauf zurückgeführt, dass die Aktivierung des kritischen Items im semantisch-lexikalischen Netzwerk zunahm.

Umgekehrt kam es bei einer zunehmenden Darbietungsdauer bei längeren Darbietungszeiten zu einer Reduktion falscher Erinnerungen, was dadurch erklärt wird, dass Listenitems besser itemspezifisch enkodiert werden konnten und sich die Quellendiskrimination verbesserte. Nach Meinung der Autoren wirkten kontrollierende Quellendiskriminationsprozesse der Aktivierung entgegen. Zusammenfassend weisen McDermott und Watson (2001) auf die Notwendigkeit einer Zweiprozesstheorie hin, welche die kurvilineare Beziehung erklären kann.

Anhand des AMA lassen sich vielfältige Befunde zu Gedächtnistäuschungen im Rahmen des DRM-Paradigmas erklären, aber dennoch gibt es in diesem Ansatz keine konkreten Annahmen und Studien zum Auftreten falscher Erinnerungen bei Kindern. Die Vertreter dieses Ansatzes sind lediglich der Meinung, dass besonders Kinder und ältere Erwachsene vermehrt falsche Erinnerungen aufgrund veränderter Quellendiskriminationsfähigkeiten aufweisen (Roediger & McDermott, 2000), jedoch gibt es von Verfechtern dieses Ansatzes keine empirischen Studien mit Kindern. Untersuchungen zu alterskorrelierten Veränderungen wurden durchgeführt, wobei es sich bei den Probanden um junge und ältere Erwachsene handelt (Benjamin, 2001; Dehon & Brédart, 2004; Gallo & Roediger, 2003; McCabe & Smith, 2002).

2.2.3 Vergleich des Aktivierungs-Monitoring Ansatzes (AMA) mit anderen Theorien im DRM-Paradigma

Eine vergleichende Betrachtung erfolgt in diesem Abschnitt hauptsächlich zwischen dem AMA und der FTT, denn beide Theorien lassen sich dadurch gut gegenüberstellen, dass zwei Prozesse postuliert werden, die jeweils entgegengesetzt auf diese Gedächtnisillusion einwirken. Zudem sind sie die populärsten und am häufigsten geprüften Ansätze, die eine differenzierte Erklärung falscher Erinnerungen ermöglichen.

Eine kurze und übersichtliche Zusammenschau aller vier Theorien ermöglicht die folgende Tabelle 2.1, welche die wichtigsten Aspekte gegenüberstellt.

Tabelle 2.1:

Zusammenfassung der vier Theorien zur Erklärung falscher Erinnerungen im DRM-Paradigma nach den wichtigsten Aspekten.

Vergleichsebene	Kriteriumsverschiebungstheorie	Distinktivitätsheuristik	FTT	AMA
1. Begründer	Miller & Wolford (1999)	Israel & Schacter (1997)	Reyna & Kiernan (1994); Brainerd et al. (1995a)	Roediger & McDermott (1995)
2. Basis	Signalentdeckungstheorie (Green & Swets, 1974)	Heuristiken, Informationsverarbeitung (Chai-ken, Lieberman & Eagly, 1989)	Konstruktivismus, Schematheorie (Piaget & Inhelder, 1973)	“implicit associative response“ (Underwood, 1965); Aktivierungsausbreitung (Collins & Loftus, 1975); “reality monitoring“ (Johnson & Raye, 1981)
3. Art des Prozesses	Verschiebung des Antwortkriteriums	Kognitive Heuristiken	“gist“-Spuren und “memory editing“	Aktivierung und Monitoring
4. (Primäre) Entstehungsphase falscher Erinnerungen	Abruf	Enkodierung und Abruf	Enkodierung und Abruf	Enkodierung und Abruf
5. Prinzip und Wirkmechanismen	Erhöhung falscher Erinnerungen durch liberales Antwortkriterium	Reduktion falscher Erinnerungen durch item-spezifische Verarbeitung	Erhöhung oder Reduktion falscher Erinnerungen durch “gist“-Spuren und “memory editing“	Erhöhung oder Reduktion falscher Erinnerungen durch Aktivierung und Quellendiskrimination
6. Zielsetzung der Studien	Bewusste Entscheidung für liberales Antwortkriterium	Variation der Enkodierbedingungen; Distinktivität fördern	Interaktion und Wirkweise “verbatim“- und “gist“-Spuren	Interaktion und Wirkweise Aktivierung und Quellendiskrimination
7. Studien mit Altersvergleich	Nein	Ja: Vergleich jüngerer mit älteren Erwachsenen	Ja: Vergleich aller Altersgruppen	Ja: aber nur Vergleich jüngerer mit älteren Erwachsenen

Wie unschwer zu erkennen ist, unterscheidet sich der AMA auf den ersten Blick kaum von der FTT. Daher soll nun eine differenzierte Gegenüberstellung dieser Theorien erfolgen.

Während bei dem AMA die Aktivierung des nicht präsentierten Items falsches Erinnern begünstigt, wirkt die Quellendiskrimination dieser entgegen (McDermott & Watson, 2001).

Demgegenüber schreibt die FTT den “verbatim“- und “gist“-Repräsentationen die Verantwortung für falsche Erinnerungen zu. Auch diese wirken gegeneinander, dadurch, dass die “ver-

batim“-Spuren die Oberflächenmerkmale mit den Details beinhalten, die “gist“-Spuren aber den zugrunde liegenden Bedeutungsgehalt (Brainerd & Reyna, 1998a; Brainerd & Reyna, 2002b; Brainerd & Reyna, 2004; Brainerd et al., 2004). In Gedächtnistests können falsche Erinnerungen demnach vorwiegend durch “gist“-Repräsentationen zustande kommen, richtige Erinnerungen jedoch durch ein Zusammenspiel beider Spuren (Brainerd & Reyna, 2002b; s. Abschnitt 2.2.2).

Beide Theorien lassen sich dahingehend verbinden, dass Probanden zwei Arten von Informationen speichern, wenn sie DRM-Wortlisten lernen (Brainerd et al., 1995b; Gallo & Roediger, 2002). Nach Roediger et al. (2001) entspricht “gist“ der FTT der Aktivierung des kritischen Items im AMA. Die itemspezifische Verarbeitung des AMA ist in der FTT mit der “verbatim“-Spur zu vergleichen und die semantisch-relationale Enkodierung (AMA) mit der “gist“-Spur (FTT). Jedoch lassen sich diese Komponenten nicht jeweils gleichsetzen: Die itemspezifische Verarbeitung geht über die “verbatim“-Spur hinaus, da erstere zusätzlich z.B. auch die Verarbeitung von Einzelitems in Form einer bildhaften Vorstellung beinhalten kann (Roediger et al., 2001).

Beide Theorien ähneln sich darin, dass sie Kontrollprozesse annehmen. Im AMA wird dieser Prozess durch die Quellendiskrimination abgebildet, die sowohl während der Enkodierung als auch des Abrufs wirksam wird (Roediger et al., 2001). Wie bereits erwähnt, dient dieser im Sinne des “reality monitoring“ dazu, das kritische Item einer internen (selbst generierten) oder externen (in der Liste präsentierten) Quelle zuzuordnen. Im Fall der Attribution auf eine externe Quelle resultiert eine falsche Erinnerung.

Auch die FTT postuliert Kontrollprozesse, die als “memory editing“ bezeichnet werden. Eine spezielle Form des “memory editing“ ist die “recollection rejection“ (Brainerd & Reyna, 2002b). Die durch die Testitems hervorgerufenen “verbatim“- oder “gist“-Spuren werden mit den entsprechenden “verbatim“- oder “gist“-Merkmalen der Testitems verglichen und es werden Ähnlichkeitsurteile durch Merkmalsüberschneidungen hervorgerufen.

Beide Theorien führen demnach falsche Erinnerungen darauf zurück, dass das Thema einer Liste erfasst bzw. im semantischen Netzwerk aktiviert wird und die anschließenden Kontrollprozesse fehlerhaft ablaufen.

Grundlegend unterscheiden sich beide Theorien darin, dass der AMA die Aktivierung des kritischen Items als Voraussetzung für falsche Erinnerungen ansieht, wobei die Repräsentation des kritischen Items sowohl semantische als auch Oberflächenmerkmale beinhalten muss. Demgegenüber setzt die FTT nur eine Erfassung des Listenthemas mit semantischen Informa-

tionen voraus. Demnach reicht es für das falsche Erinnern aus, wenn das kritische Item ausreichend mit dem “gist“ einer Liste übereinstimmt.

Die Befundlage ist gemischt. Beispielsweise konnten Cabeza und Lenartson (2005) zeigen, dass ein Wechsel der Sprachen in der Lern- und Testphase nichtsdestotrotz zu falschen Erinnerungen führte, obwohl offensichtlich die Oberflächenmerkmale verschieden waren. Wurden die DRM-Wortlisten in englischer Sprache dargeboten, die Items des Wiedererkennentests aber in französischer Sprache, ließ sich ein Effekt falscher Erinnerungen an nicht präsentierte kritische Items feststellen. Dieser Befund lässt sich besser mit den Annahmen der FTT in Einklang bringen, die die Extraktion des Listenthemas (also der Bedeutungsmerkmale) als ausschlaggebend ansehen.

Anhand der Ergebnisse einer Untersuchung von Hutchison und Balota (2005) wird andererseits an der FTT gezweifelt, ob diese falsche Erinnerungen mit den “gist“-Repräsentationen hinreichend erklären kann. Sie konnten zeigen, dass die assoziative Aktivierung von entscheidender Bedeutung für falsche Erinnerungen ist und weniger die semantisch-relationale Verarbeitung. Sie konstruierten zwei verschiedene Listenbedingungen, die alle aus 12 Items bestanden: in den nicht-homographischen (DRM-) Listen waren zwei Sets von jeweils sechs Items mit der gleichen Bedeutung des kritischen Items assoziiert (z.B. “slumber, lay, motel, trance, lazy, nightmare“ und “snooze, wake, bedroom, unconscious, deep, blanket“ für das kritische Item “sleep“). Die andere Hälfte der Listen bestanden aus zwei Sets von Items, die zwei unterschiedlichen Bedeutungen desselben kritischen Items zugehörig waren (z.B. “stumble, slip, rise, trip, faint, clumsy“ und “autumn, season, spring, leaves, brisk, harvest“ für das kritische Item “fall“). In allen Experimenten zeigte sich, dass falsche Erinnerungen unbeeinflusst von der Variation der Listenkonstruktion waren, d.h. der Anteil falscher Erinnerungen war nicht von der Anordnung der Items nach einem gemeinsamen Thema abhängig, sondern von der Assoziationsstärke der Listen. Dies impliziert, dass die FTT an Bedeutung verliert, denn nach diesen Befunden ist es nicht wichtig, das gemeinsame Thema, also den “gist“ der Liste zu erfassen und zu extrahieren.

Neben der unterschiedlichen Erklärung der Entstehung falscher Erinnerungen lassen sich die FTT und der AMA darin unterscheiden, dass sie verschiedene Annahmen über die Stabilität falscher Erinnerungen treffen. Während die FTT davon ausgeht, dass durch die “gist“-Spuren falsche Erinnerungen gegenüber zeitlichen Einflüssen resistent sind, ist dies laut dem AMA nicht der Fall. Nach dem AMA ist die Aktivierung kritischer Items, bewusst oder automatisch, eine notwendige Voraussetzung damit falsche Erinnerungen entstehen können. Erfolgt

die Aktivierung automatisch, dann ist nicht von einer zeitlichen Stabilität falscher Erinnerungen auszugehen. Eine Übereinstimmung beider Theorien zeigt sich nur dann, wenn die Aktivierung kontrolliert unter kognitiver Anstrengung erfolgt und somit zu langlebigen falschen Erinnerungen führen kann. Sowohl die FTT als auch der AMA gehen davon aus, dass sich richtige Erinnerungen wegen ihrer "verbatim"-Spuren oder itemspezifischen Merkmale im Zeitverlauf reduzieren (Gallo & Roediger, 2002).

Im Hinblick auf die Erklärung alterskorrelierter Veränderungen falscher Erinnerungen unterscheiden sich die Theorien voneinander. Während sich die FTT diesem Aspekt in verschiedenen Studien intensiv widmet (Brainerd & Reyna, 1996; Brainerd et al., 1995b; Brainerd et al., 2004; Brainerd & Mojardin, 1998; Reyna & Kiernan, 1994), zeigen die Vertreter des AMA ein geringeres Interesse. Dies ist nicht verwunderlich, denn beide Theorien entstammen aus unterschiedlichen Gebieten der Psychologie. Während der AMA aus der kognitiven Psychologie entstammt, wurde die FTT zur Erklärung entwicklungspsychologischer Befunde konzipiert. Wie im vierten Kapitel zu alterskorrelierten Veränderungen falscher Erinnerungen ausführlich gezeigt wird, gibt es bedeutsame Unterschiede der Theorien. Die FTT geht davon aus, dass sich "verbatim"- und "gist"-Fähigkeiten mit dem Alter verbessern und der Anteil richtiger und falscher Erinnerungen mit dem Alter zu-, abnehmen aber auch stabil bleiben kann, je nachdem welche Spur durch den Test angesprochen wird (Brainerd & Mojardin, 1998; Brainerd & Reyna, 1998a; Brainerd & Reyna, 2004; Brainerd et al., 2002a; Brainerd et al., 2004).

Hingegen vertritt der AMA nur die allgemeine Auffassung, dass Kinder und Senioren gegenüber falschen Erinnerungen besonders empfänglich sind, da die Quellendiskrimination schlecht der Aktivierung der kritischen Items entgegenwirken kann (Roediger & McDermott, 2000). Es konnte gezeigt werden, dass ältere Personen jüngeren in der Aktivierung gleichgestellt sind, jedoch scheinen jüngere über bessere Quellendiskriminationsfähigkeiten zu verfügen (Balota, Cortese, Duchek, Adams, Roediger, McDermott & Yerys, 1999; Dehon & Brédart, 2004). Da ältere Menschen schlecht zwischen Listen- und kritischen Items unterscheiden können, erkennen sie kritische Items falsch wieder. Verglichen mit jüngeren Erwachsenen werden die Listenitems von ihnen ähnlich semantisch-relational verarbeitet, aber Details werden aufgrund der schlechteren itemspezifischen Verarbeitung weniger beachtet. Ob dieser Effekt auch bei Kindern auftritt, bleibt wegen mangelnder Befunde unklar.

Ein weiterer bedeutsamer Unterschied zeigt sich darin, dass die FTT vorrangig in ihren Studien zum DRM-Paradigma Wiedererkennenstests einsetzt, wohingegen der AMA sowohl

Tests zur Freien Reproduktion als auch zum Wiedererkennen verwendet (Roediger et al., 2001).

Neben diesen Unterschieden konnte aber anhand verschiedener Studien gezeigt werden, dass der AMA und die FTT zumeist gleichermaßen geeignet sind die Befunde falscher Erinnerungen zu erklären, womit sich ihre Verschiedenheit in Frage stellen lässt. In einer Untersuchung zur Darbietungsdauer und Wiederholung der Listenitems konnten Seamon, Luo, Kopecky, Price, Rothschild, Fung und Schwartz (2002b) die Befunde mit beiden Theorien verständlich machen. Die FTT geht davon aus, dass bei einer kurzen Darbietungsdauer richtiges und falsches Wiedererkennen gering ist und beides durch Wiederholung gesteigert wird, da "verbatim"- und "gist"-Spuren sich noch ausformen. Bei einer langen Darbietungsdauer sind richtige und falsche Erinnerungen hoch. Hier nimmt das richtige Wiedererkennen mit der Wiederholung zu und das falsche Erinnern reduziert sich aufgrund einer vollständigen "verbatim"-Repräsentation und einer verbesserten Unterscheidung von "verbatim"- und "gist"-Spuren. Auch der AMA trifft ähnliche Aussagen: Die Darbietungsdauer beeinflusst deutlich die Quellendiskrimination. Wird die Liste schnell dargeboten, so kann der Quellendiskriminationsprozess, auch bei wiederholter Darbietung der Liste, falschen Erinnerungen nicht entgegenwirken, weil die Listenitems unzureichend itemspezifisch verarbeitet werden können. Die Wiederholung bewirkt eine verstärkte Aktivierung ohne entgegenwirkende Quellendiskrimination, was falsche Erinnerungen begünstigt. Im Fall einer langsamen Darbietung führt die Wiederholung zu einem Anstieg richtiger und einer Reduktion falscher Erinnerungen. Hier wird auf die Bedeutung der Quellendiskrimination hingewiesen, die bei einer langsamen Darbietung, durch die Ermöglichung einer guten itemspezifischen Verarbeitung, richtige Erinnerungen unterstützen und falsche unterbinden kann (Gallo & Roediger, 2002).

Gemäß diesem Befund sind beide Theorien nicht nur gleichermaßen geeignet die Ergebnisse zu erklären, darüber hinaus deuten sie auch auf die antagonistische Beziehung zweier Prozesse im Entstehen richtiger und falscher Erinnerungen hin. Die Grundannahmen und Vorhersagen ähneln sich.

Ausschlaggebend für die Wahl des AMA als Basis für die eigene Untersuchung, anhand dessen die Hypothesen formuliert werden, ist, dass sich aus diesem Vorhersagen für alterskorrelierte Differenzen ableiten und prüfen lassen, was bisher in der Forschung noch nicht geschah. Damit wird der ursprünglich aus der Kognitionspsychologie stammende AMA um entwicklungspsychologische Aspekte erweitert.

2.3 Einflussfaktoren auf falsche Erinnerungen

In diesem Abschnitt werden Faktoren berichtet, welche neben der Altersvariablen bedeutende Auswirkungen auf das Entstehen falscher Erinnerungen im DRM-Paradigma haben können.

Die folgende Tabelle (2.2) ermöglicht einen Überblick über die Vielfalt erforschter Variablen, von denen in diesem Abschnitt nur jene erläutert werden, welche für die eigene Untersuchung von Bedeutung sind. In der Tabelle werden einige potenzielle Einflussvariablen und eine Auswahl zugehöriger Studien aufgeführt. Die Befundlage wird zudem dahingehend zusammengefasst, in welche Richtung die Einflussvariablen sich auf falsche Erinnerungen auswirken (erhöhen, reduzieren oder gemischte Befunde).

Tabelle 2.2:

Potenzielle Einflussfaktoren auf falsche Erinnerungen und die Richtung des Einflusses nach den Ergebnissen zugehöriger Studien.

Bezugspunkte	Beispiele
Personenmerkmale	<p>1) Alter</p> <p>1.1 Kinder vs. Erwachsene: gemischte Befunde (s. Kapitel 4 ; Brainerd et al., 2002a; Brainerd et al., 2004; Ghetti et al., 2002)</p> <p>1.2 Junge Erwachsene vs. ältere Erwachsene: Erhöhung mit Alter (Dehon & Brédart, 2004; Gallo & Roediger, 2003; McCabe & Smith, 2002; Norman & Schacter, 1997; Schacter et al., 1999)</p> <p>2) Arbeitsgedächtniskapazität: Reduktion (Watson, Bunting, Poole & Conway, 2005)</p>
Testmerkmale	<p>1) Rückwärtsassoziationsstärke (RAS): Erhöhung (Gallo & Roediger, 2002; Hicks & Hancock, 2002; McEvoy, Nelson & Komatsu, 1999; Roediger et al., 2001)</p> <p>2) Länge der Wortlisten (Anzahl der Items): Erhöhung (Gallo & Roediger, 2003; Ghetti et al., 2002; Hancock, Hicks, Marsh & Ritschel, 2003; Marsh & Bower, 2004; Robinson & Roediger, 1997)</p> <p>3) Itemmerkmale</p> <p>3.1 Wortlänge: Reduktion (Roediger et al., 1998; Roediger et al., 2001)</p> <p>3.2 Konkrete und übliche Items: Erhöhung (Roediger et al., 1998; Roediger et al., 2001)</p> <p>4) Zusammensetzung der Wortlisten (im Block präsentiert): Erhöhung (McDermott, 1996)</p>
<i>Listenkonstruktion</i>	
<i>Lernbedingung (Enkodierphase)</i>	<p>5) Instruktion</p> <p>5.1 Vorwarnung: gemischt (Dehon & Brédart, 2004; Gallo et al., 1997; Gallo et al., 2001b; Heit, Brockdorff & Lamberts., 2004; McCabe & Smith, 2002; Multhaup & Conner, 2002; Neuschatz, Benoit & Payne, 2003)</p> <p>5.2 Inklusion: Erhöhung (Brainerd & Reyna, 1998b; Hege & Dodson, 2004; Heit et al., 2004)</p>

6) Güte des Listenlernens

6.1 Wiederholte Darbietung der Liste: gemischt

(Benjamin, 2001; McKone & Murphy, 2000; Seamon et al., 2003; Seamon et al., 2002a)

6.2 Darbietungsdauer: gemischt

(Gallo & Roediger, 2002; McDermott & Watson, 2001; Thapar & McDermott, 2001)

7) Verarbeitungstiefe: gemischt

(Chan, McDermott, Watson & Gallo, 2005; Gallo & Seamon, 2004; Marsh & Bower, 2004; McDermott & Watson, 2001; Read, 1996; Soraci, Carlin, Togli, Chechile & Neuschatz, 2003)

8) Distinktivität: Reduktion

(Dodson & Schacter, 2001; Dodson & Schacter, 2002; Gallo, McDermott, Percer & Roediger, 2001a; Ghetti et al., 2002; Hege & Dodson, 2004; Israel & Schacter, 1997; Schacter et al., 1999)

Testbedingung (Abrufphase)

9) Behaltensintervall: gemischt

(Blair et al., 2002; McDermott, 1996; Payne et al., 1996; Seamon et al., 2002b; Thapar & McDermott, 2001)

10) Testart

10.1 Freie Reproduktion vs. Wiedererkennen: Wiedererkennen mehr falsche Erinnerungen

(Hege & Dodson, 2004; Lampinen, Neuschatz & Payne, 1999; Roediger & McDermott, 1995; Seamon et al., 2003)

10.2 Explizite Tests vs. Implizite Tests: Explizite Tests mehr falsche Erinnerungen

(Deliberto-Macaluso, 2005; McDermott, 1997; McKone & Murphy, 2000)

10.3 Implizit perzeptuelle Tests vs. Implizit konzeptuelle Tests: Implizit konzeptuelle Tests mehr falsche Erinnerungen

(McDermott, 1997)

11) Testabfolge (Beeinflussung durch vorherigen Test): Erhöhung

(Lampinen et al., 1999; McEvoy et al., 1999; Payne et al., 1996; Robinson & Roediger, 1997; Roediger & McDermott, 1995)

2.3.1 Rückwärtsassoziationsstärke (RAS)

Von großem Interesse sind die manipulierbaren Variablen, die sich auf die Listenkonstruktion, Lern- und Testbedingung beziehen. In der eigenen Studie wurde der Listenkonstruktion besondere Aufmerksamkeit gewidmet. Schon Deese (1959a, 1959b) sah in der Zusammenstellung der Listen zwei ursächliche Faktoren für falsche oder richtige Erinnerungen, die er Assoziationsstärke zwischen den Items (ASZI; "interitem associative strength") und Rückwärtsassoziationsstärke (RAS; "backward associative strength") nannte. Die ASZI ist die mittlere, relative Häufigkeit, mit der die Listenitems sich gegenseitig in der Freien Assoziation hervorrufen. Zum einen korreliert dieses Maß positiv mit der Anzahl richtig erinnerter Wörter in der Freien Reproduktion und ähnlicher Intrusionen der Probanden, zum anderen weist die ASZI einen negativen Zusammenhang zur Anzahl der nicht mit der Liste assoziierten Intrusionen auf. Eine starke assoziative Verbindung innerhalb einer Wortliste führte zu korrekten Erinnerungen vieler Listenitems, aber auch häufig zu Intrusionen des assoziierten, nicht präsentierten Items. Die mittlere RAS stellt die durchschnittliche Wahrscheinlichkeit dar mit der die Listenitems das kritische Item in der Freien Assoziation hervorrufen. Dieses Maß korreliert stark mit der Wahrscheinlichkeit das kritische Item in der Freien Reproduktion und im Wiedererkennen falsch zu erinnern (Hicks & Hancock, 2002). Um multiple Einflussfaktoren auf falsche Erinnerungen herauszufinden, griffen Roediger und Mitarbeiter (2001) auf die mittlere RAS und ASZI zurück, erweiterten ihre Analyse aber um die Variablen Vorwärtsassoziationsstärke (VAS; "forward associative strength": Häufigkeit, mit welcher die Listenitems zu dem kritischen Item assoziiert werden) und korrekte Reproduktion ("veridical recall": Maß für die mittlere Wahrscheinlichkeit, die Listenitems in einer Wortliste frei zu reproduzieren). Es konnte gezeigt werden, dass die RAS der beste Prädiktor in der multiplen Regressionsanalyse für falsche Erinnerungen und die korrekte Reproduktion mit negativem Vorzeichen den zweitbesten Prädiktor darstellt. Die Untersuchungen von McEvoy et al. (1999) bestätigen dies. Da im Deutschen keine Normen für die RAS vorliegen, wird stattdessen die Gesamt-Rückwärtsassoziationsstärke (Gesamt-RAS) durch Ergebnisse von Vorläuferstudien beachtet. Diese berechnet sich durch die Häufigkeit mit der das kritische Item durch die gesamte DRM-Liste und nicht durch einzelne Listenitems assoziiert wird.

2.3.2 Länge der Wortlisten

Die Listenlänge hat einen bedeutenden Einfluss auf falsche Erinnerungen, da eine hohe Anzahl von assoziierten Items sich laut der FTT auf die gist-Gedächtnisspur auswirkt und damit

falsche Erinnerungen wahrscheinlicher hervorruft als weniger Listenitems (Ghetti et al., 2002). Auch der AMA vertritt diese Auffassung, da mit der Anzahl der Listenitems (Gesamt-RAS) die Aktivierung des kritischen Items, und damit falsches Erinnern, steigen soll (Robinson & Roediger, 1997). Hancock et al. (2003) untersuchten anhand des DRM-Paradigmas Einflussfaktoren falscher Erinnerungen und stellten fest, dass Wortlisten, die aus 14 anstatt 10 semantisch assoziierten Items bestanden, eine hohe semantische Aktivierung mit einer größeren Anzahl von falschen Erinnerungen zur Folge hatten. Um den Einfluss der Wortlistenlänge zu prüfen und um zu testen ob und wie sich diese auf die Gedächtnisleistung von Kindern auswirkt, wird die Länge der Wortlisten in einem Experiment der eigenen Untersuchung manipuliert.

2.3.3 Instruktion

Neben der Listenkonstruktion werden falsche Erinnerungen auch durch Merkmale der Enkodierphase beeinflusst.

Als Erstes erweist sich die Instruktion als wichtig, durch welche die Probanden bereits auf mögliche falsche Erinnerungen hingewiesen werden können. Im Abschnitt 2.2.2 wurde bereits auf einige der vielen Untersuchungen mit Variation der Instruktion eingegangen. Gallo und Mitarbeiter (1997) untersuchten den Einfluss einer solchen Vorwarnung auf die Erinnerungsqualität und unterteilten die Probanden in eine nicht-vorgewarnte, eine eindeutig vorgewarnte Gruppe und in eine Gruppe, die vage Hinweise erhielt. Die Probanden der vagen Warnungsbedingung erzielten weniger Treffer, aber nicht weniger falsche Alarme. Hingegen reduzierte sich für die vorgewarnten Probanden falsches Wiedererkennen deutlich, wenn auch nicht vollständig. Auch andere Untersuchungen zeigten ein ähnliches Ergebnis (Heit et al., 2004; McDermott & Roediger, 1998; Multhaup & Conner, 2002). Andere Forscher vergleichen eine nichtinformierte Gruppe mit zwei Gruppen, die eine Vorwarnung zu unterschiedlichen Zeitpunkten der Untersuchung erhalten, entweder vor der Lernphase oder vor der Testphase (McCabe & Smith, 2002). Wurde die Vorwarnung vor der Lernphase ausgesprochen, so reduzierte sich falsches Wiedererkennen, während die Warnung nach der Lern- und vor der Testphase falsche Erinnerungen nur dann reduzierte, wenn die Darbietungsgeschwindigkeit langsam war. (Gallo et al. 2001b; McDermott & Roediger, 1998). Neuschatz et al. (2003) fanden in ihrer Studie zum Effekt der Vorwarnung eine Mediatorvariable, die falsche Erinnerungen beeinflusst. Es scheint eine Rolle zu spielen, ob das Item leicht identifizierbar ist, denn

eine Vorwarnung reduzierte deutlich falsche Erinnerungen bei leicht identifizierbaren Listen und erhöhte damit die Effektivität der Warnung.

In der eigenen Untersuchung wird der Instruktion besondere Aufmerksamkeit gewidmet. Anhand spezieller Inklusionsinstruktionen (Listenitems erinnern und Items, an die der Proband selbst stark gedacht hat) sollen Erinnerungen an nicht präsentierte kritische Items experimentell *erhöht* werden. Im fünften Abschnitt („Prozeduren zur getrennten Erfassung von Aktivierung und Quellendiskrimination“) wird diese Einflussvariable ausführlich dargestellt.

2.3.4 Güte des Listenlernens

In der Lernphase können verschiedene Variablen auf falsche Erinnerungen einwirken. Seamon und Mitarbeiter (2003) schlugen eine dem LOP-Ansatz verwandte aber umfassendere Sichtweise vor, die auf den Ergebnissen von McDermott und Watson (2001) zur Auswirkung der Präsentationsdauer sowie den Ergebnissen von Seamon et al. (2002c) zum Effekt der Listenwiederholung basiert. Zusammenfassend lassen sich einige in der Abbildung von Seamon und Mitarbeiter (2003) erläutern. Anhand einer idealisierten Grafik (s. Abbildung 2.3.1) werden die Verlaufskurven korrekter und falscher Erinnerungen in Abhängigkeit der Güte des Listenlernens („level of list learning“) dargestellt.

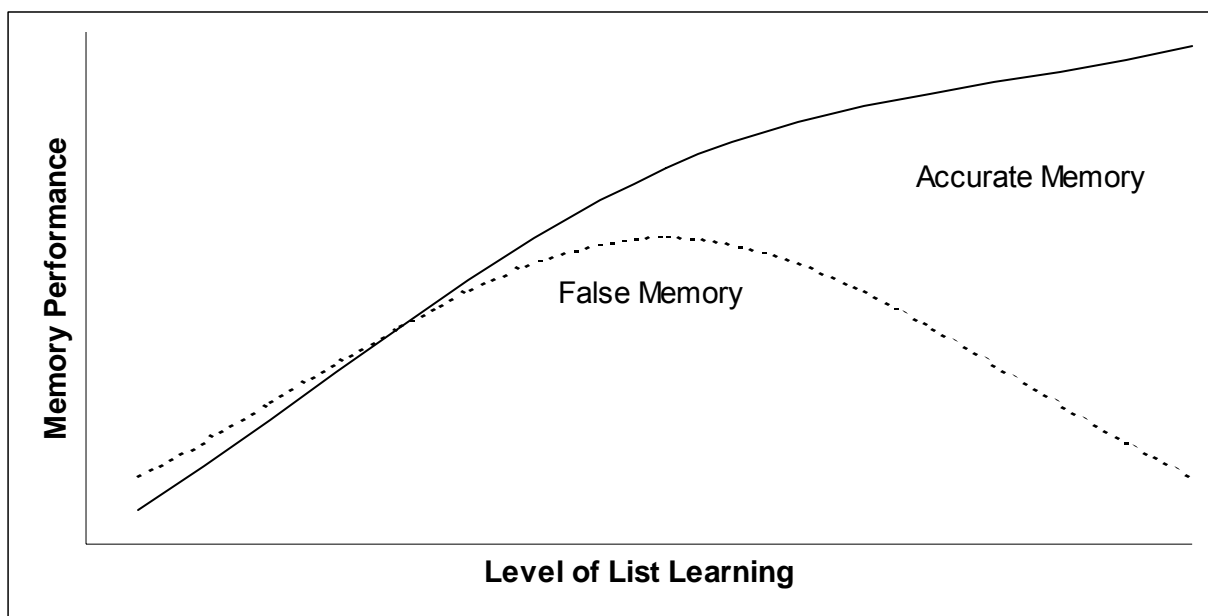


Abbildung 2.3.1: Verlauf korrekter und falscher Erinnerungen im DRM-Paradigma in Abhängigkeit der Güte des Listenlernens nach Seamon et al. (2003, S.454).

Demnach sei das Ausmaß korrekter aber auch falscher Erinnerungen davon abhängig wie gut die Liste gelernt wurde. Während korrektes Erinnern monoton ansteigt, zeigen falsche Erinne-

rungen einen umgekehrt U-förmigen Verlauf. Können die Listen kaum gelernt werden, so resultieren wenige falsche aber auch wenige korrekte Erinnerungen. In dieser Bedingung können sogar mehr falsche als korrekte Erinnerungen auftreten. Mit zunehmendem Listenlernen erhöhen sich korrekte und falsche Erinnerungen, wobei die falschen Erinnerungen bei einem mittleren Ausmaß der Güte des Listenlernens gipfeln und anschließend stetig abnehmen, während die korrekten Erinnerungen weitergehend mit besserem Listenlernen ansteigen. Die dem hypothetischen Verlauf zugrunde liegenden Studien von McDermott und Watson (2001) sowie von Seamon et al. (2002a) wurden bereits im Abschnitt 2.2.2 erläutert. Diese Befunde lassen sich nicht mit Ein-Prozess-Theorien erklären, sondern verdeutlichen die Notwendigkeit einer Zwei-Prozess-Theorie. Beispielsweise erklärt der AMA diesen Verlauf korrekter und falscher Erinnerungen damit, dass die Aktivierung mit zunehmendem Listenlernen zunimmt, ebenso aber auch die der Aktivierung entgegenwirkende Quellendiskrimination. Daher reduziert sich der Anteil falscher Erinnerungen ab einer bestimmten Güte des Listenlernens.

2.3.5 Ausmaß der konzeptuellen Verarbeitung

Dem Effekt der konzeptuellen Verarbeitung kommt in der eigenen Untersuchung eine besondere Rolle zu, denn jüngere Kinder neigen dazu, nach einer Lerninstruktion weniger konzeptuell zu verarbeiten als ältere (vgl. Kapitel 3), was Einfluss auf die Erinnerungsleistung hat. In Abhängigkeit von der Verarbeitungstiefe (gemäß des Verarbeitungsprozessansatzes “levels of processing“ = LOP) werden Wortlisten besser oder schlechter erinnert. Die Verarbeitungstiefe lässt sich auf einem Kontinuum zunehmender Tiefe von oberflächlich über phonologisch bis semantisch unterteilen, wobei die oberflächliche Verarbeitung sich auf physikalische Merkmale (z.B. Farbe oder Größe eines Objekts) bezieht. Auf einer tieferen Ebene werden Informationen phonologisch (z.B. nach Klang) verarbeitet und die tiefste Verarbeitung findet semantisch statt, in welcher die Bedeutung herausgearbeitet wird (Craik & Lockhart, 1972). Wenn die Verarbeitung eine vermehrte Analyse und Interpretation bedeutet, dann sollte eine bessere Erinnerung resultieren (Craik & Tulving, 1975). Viele Studien zeigten, dass eine semantische Verarbeitung der Wortlisten (z.B. bildhafte Vorstellung, Gruppierung der Items in Kategorien) gegenüber einer oberflächlichen Verarbeitung der orthographischen oder phonologischen Merkmale der Listenitems richtige Erinnerungen unterstützt (Brainerd et al., 1995b; Chan et al., 2005; Gallo et al., 2001b; Israel & Schacter, 1997; McDermott, 1996; Soraci et al., 2003; Smith & Hunt, 1998; Thapar & McDermott, 2001). Vertreter des AMA (McDer-

mott & Watson, 2001; Payne et al., 1996; Thapar & McDermott, 2001) waren der Auffassung, dass eine tiefere Verarbeitung eine verstärkte semantische Aktivierung der kritischen Items bewirkt, die damit zu vermehrten richtigen aber auch falschen Erinnerungen führt. Nach dem von Toggia, Neuschatz und Goodwin (1999) benannten "more is less" Effekt wird der AMA unterstützt.

Aus der Modifikation des LOP-Ansatzes resultiert das Konzept der „Elaboriertheit“, welches „...das Ausmaß der kognitiven Analyse auf der jeweiligen Stufe der Verarbeitung bezeichnet“ (Hupbach, 2000, S.49). Kann der Proband auf semantisch elaborierte Verarbeitungsstrategien zurückgreifen, dann wird tatsächlich präsentiertes Material genauer und dauerhafter erinnert (Craik & Tulving, 1975; Tulving, 1985).

Soraci und Mitarbeiter (2003) untersuchten den Einfluss der Elaboriertheit auf Erinnerungen und variierten die Elaboriertheit dadurch, dass die Probanden entweder Wörter lesen (wenig elaboriert) oder Anagramme lösen (elaborierte Verarbeitung) sollten. Zwar erhöhte sich mit der semantischen Elaboriertheit der Anteil richtiger aber nicht falscher Erinnerungen.

Thapar und McDermott (2001) testeten den Einfluss der Verarbeitungstiefe auf die Rate falscher Erinnerungen bei Erwachsenen. Die Verarbeitungstiefe wurde variiert indem eine Gruppe der Versuchspersonen die dargebotenen Items auf einer Ratingskala von angenehm bis unangenehm beurteilen sollte, während eine zweite Gruppe die Farben der Wörter benennen und eine dritte Gruppe die Vokale zählen sollte. Somit wurde über diese drei Bedingungen sukzessive die Verarbeitungstiefe von tief bis oberflächlich verändert. Die Ergebnisse weisen darauf hin, dass falsche Erinnerungen sich eindeutig mit der Verarbeitungstiefe erhöhten.

Andere Studien demonstrierten ebenfalls den "more is less" Effekt und unterstützten damit den AMA. Rhodes und Anastasi (2000) instruierten eine Gruppe von Probanden, die nach dem DRM-Paradigma präsentierten Items dahingehend zu verarbeiten, dass sie die Anzahl der Vokale zählen und aufschreiben sollten (geringe Verarbeitungstiefe). In der Bedingung der tiefen Verarbeitung sollten die Probanden sich jedes präsentierte Item bildhaft vorstellen und es in Kategorien einordnen, je nachdem ob es sich um einen konkreten („Baum“) oder abstrakten („Gerechtigkeit“) Begriff handelte. Die Ergebnisse zeigen im Einklang mit den AMA, dass eine tiefe Verarbeitung zu vermehrt richtigen aber auch falschen Erinnerungen führt.

Eine Weiterentwicklung des Prozessansatzes der Informationsverarbeitung (LOP) führte zu dem Konzept der transferangemessenen Verarbeitung ("transfer appropriate processing" = TAP; Bransford, Franks, Morris & Stein, 1979). Je nach Testart werden unterschiedliche Pro-

zesse gefordert, semantische oder phonemische. Eine semantische Verarbeitung während der Lernphase führt zu besten Erinnerungen, wenn auch in der Testphase semantische Prozesse angesprochen werden. Umgekehrt trifft dies auf phonemische Verarbeitung zu. Stimmen die in der Lern- und Testphase geforderten Prozesse überein, zeigt sich optimale Erinnerungsleistung.

Anhand einer neuen Studie untersuchten Chan et al. (2005) den Einfluss der Verarbeitungstiefe und der transferangemessenen Verarbeitung auf falsche Erinnerungen. Wurden semantisch assoziierte [Listenitems sind mit dem kritischen Item semantisch verbunden; im Gegensatz zu phonologisch (d.h. nach Wortklang, Reimen) assoziierten] DRM-Wortlisten vorgegeben, so konnte der Effekt der Verarbeitungstiefe (Ähnlichkeit des Klangs vs. Ähnlichkeit der Bedeutung der Listenitems) bestätigt werden. Durch die Verwendung phonologisch assoziierter Listen ("sweep, steep, sleet" mit dem kritischen Item "sleep") kehrte sich der Effekt bei Übereinstimmung von Listen- und Verarbeitungsart um: Eine phonologische Verarbeitung (Ähnlichkeit des Klangs) führte gegenüber der tieferen semantischen Verarbeitung (Ähnlichkeit der Bedeutung) zu vermehrt richtigen *und* auch falschen Erinnerungen. Demzufolge wurde der "more is less effect" bestätigt, jedoch dahingehend differenziert, dass nicht generell eine tiefere Verarbeitung diese bewirkt, sondern dass auf die Übereinstimmung von Listen- und Verarbeitungsart geachtet werden muss.

2.3.6 Testart

Auch in der Abrufphase gibt es verschiedene Einflussfaktoren auf falsche Erinnerungen. Bisherige Studien zum DRM-Paradigma prüften falsche Erinnerungen explizit mit einem Freien Reproduktions- und einem Wiedererkennenstest (Ghetti et al., 2002; Lampinen et al., 1999; Roediger & McDermott, 1995; Stadler, Roediger & McDermott, 1999). Dadurch zeigte sich, dass die Rate falscher Erinnerungen je nach Testart unterschiedlich ausfiel. In der Freien Reproduktion war der Anteil falscher Erinnerungen deutlich geringer als im Wiedererkennenstest (Roediger & McDermott, 1995). Zudem zeigten sich stärkere Unterschiede zwischen dem Anteil richtiger und falscher Erinnerungen im Freien Reproduktionstest als im Wiedererkennenstest. Dies weist darauf hin, dass richtige und falsche Erinnerungen entscheidend von der Sensitivität des Tests beeinflusst werden.

Neben expliziten Gedächtnistests können im DRM-Paradigma auch implizite (perzeptuelle und konzeptuelle) Tests eingesetzt werden in denen keine Erinnerungsinstruktion gegeben wird, sondern die Probanden sollen eine Aufgabe so schnell oder so gut wie möglich erledigen.

gen. Aufgrund dessen besteht keine Notwendigkeit zwischen präsentierten und selbst generierten Items zu unterscheiden. Die impliziten Tests erfassen die Aktivierung kritischer Items ohne entgegenwirkende Quellendiskrimination. Sowohl perzeptuelle als auch konzeptuelle Tests sollen eher eine bewusste Aktivierung testen: In perzeptuellen Tests ist Priming nur dann zu beobachten, wenn der Proband gezielt an das Item denkt, da eine Übereinstimmung der perzeptuellen Prozesse zwischen Lern- und Testphase gewährleistet sein muss. Einige Befunde sprechen dafür, dass konzeptuelle Tests auch eine bewusste Aktivierung testen, denn das konzeptuelle Priming profitiert von konzeptueller Verarbeitung wie die Ergebnisse zu den LOP-Effekten zeigen (z.B. Challis & Sidhu, 1993; Hamann, 1990). Zudem führt eine Aufmerksamkeitsablenkung in der Lernphase im Allgemeinen zu negativen Effekten.

McDermott (1997) und McKone und Murphy (2000) führten entsprechende Studien mit impliziten Tests durch, die im Abschnitt 5.2 ausführlich erläutert werden. Eine Betrachtung des Einflusses der Testart spielt besonders im Hinblick auf die eigene Untersuchung eine Rolle, da in dieser verschiedene explizite und implizite Tests eingesetzt werden.

2.3.7 Beeinflussung durch den vorherigen Test

Falsche Erinnerungen können auch von einem vorherigen Test beeinflusst sein. Roediger und McDermott (1995) prüften diese Einflussvariable in ihrer zweiten Studie, indem nur in der Hälfte der Listen dem Wiedererkennenstest eine Freie Reproduktion vorausging. Fand eine Freie Reproduktion statt, so fiel im Wiedererkennenstest die Rate richtiger aber auch falscher Erinnerungen höher aus (s. auch Gallo & Roediger, 2002; Lampinen et al., 1999). Auch Robinson und Roediger (1997) sowie McEvoy et al. (1999) verweisen auf die Reihenfolge der Tests und trennen in der Darstellung ihrer Ergebnisse des Wiedererkennenstests falsche Erinnerungen danach, ob sie im vorhergehenden Reproduktionstest genannt wurden. Falsches Wiedererkennen war enorm hoch, wenn das kritische Item bereits genannt wurde. Diese Ergebnisse werden als "carry over" Effekte bezeichnet, bei denen der vorherige Test die nachfolgende Testleistung beeinflusst. In einem Experiment der eigenen Untersuchung wird die Testabfolge als Kontrollvariable überprüft, um eventuelle "carry over" Effekte bei der Interpretation der Ergebnisse ausschließen zu können.

3 Gedächtnisentwicklung im Kindesalter

Im Zusammenhang mit der Diskussion von Untersuchungen falscher Erinnerungen bei Kindern wurde auf die Relevanz der Berücksichtigung von Altersunterschieden in den Gedächtnisleistungen hingewiesen. Kognitionspsychologische Studien zu alterskorrelierten Unterschieden falscher Erinnerungen bedienen sich entwicklungspsychologischer Theorien, welche die zugrunde liegenden Einflussvariablen benennen. In diesem Zusammenhang ist die historische Untersuchung von Brunswick, Goldscheider und Pilek (1932, zit. nach Schneider & Büttner, 1998) relevant, die bereits früh die Einzelleistungen von Kindern und Jugendlichen im Alter von sechs bis 18 Jahren in verschiedenen Gedächtnisaufgaben zusammenfassten. Auf diese Weise konnte für jede Person ein Gesamtmaß ihrer Gedächtnisstärke ermittelt werden. Abbildung 3.1 verdeutlicht das Entwicklungsprofil, welches auch durch aktuelle Studien bestätigt wird. Die Gedächtnisleistung verbessert sich mit zunehmendem Alter, zeigt aber zwischen sechs und 10 Jahren, was der Kindergarten- und Grundschulzeit entspricht, den stärksten Anstieg.

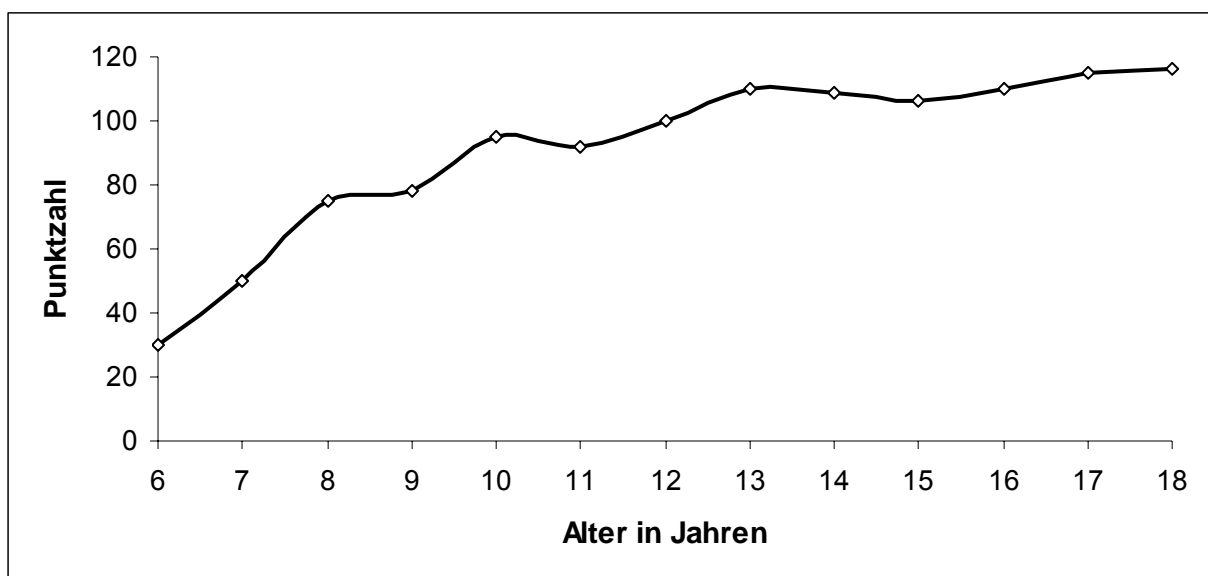


Abbildung 3.1: Gesamtentwicklung des Gedächtnisses bei Kindern und Jugendlichen im Alter von sechs bis 18 Jahren (nach Brunswick et al., 1932, aus Schneider & Büttner, 1998, S.668).

In der neueren entwicklungspsychologischen Forschung geht es weniger um die Beschreibung der Veränderung der Gedächtnisleistung, sondern um die Entwicklung der zugrunde liegenden Mechanismen. Schneider und Büttner (2002) fassen vier Determinanten zusammen, wel-

che in den folgenden Abschnitten mit Bezug zur eigenen Arbeit erläutert werden: Entwicklung der Gedächtniskapazität, der Gedächtnisstrategien, des bereichsspezifischen Wissens und des Metagedächtnisses.

3.1 Gedächtniskapazität

Die Gedächtniskapazität ist die Menge an Information, die in verschiedenen Gedächtnisspeichern gehalten werden kann (Schneider & Bjorklund, 1998). Atkinson und Shiffrin (1968) entwickelten ein Modell über Gedächtnisspeicher, das als Grundlage der Mehrspeichermodelle¹ diente. Diese Gedächtnisspeicher werden Kurz- und Langzeitspeicher nach der Dauer genannt mit welcher Informationen gehalten werden. Indem der Informationsfluss als linearer Übertragungsprozess konzipiert wird (Wippich, 1984), gelangt die Information zunächst über einen sensorischen Register zu dem Kurzzeitspeicher, der eine begrenzte Kapazität und verschiedene Kontrollprozesse (Wiederholung oder Gruppierung der Information) besitzt. Kann die Information aufrechterhalten werden, dann kann sie den Langzeitspeicher erreichen, in welchem sie langfristig gespeichert wird.

Die Kapazität des Kurzzeitspeichers kann über die Gedächtnisspanne erfasst werden, indem Probanden Material (Zahlen, Buchstaben, Wörter) in kurzen Abständen präsentiert wird. Die Gedächtnisspanne ist die Anzahl des in richtiger Reihenfolge erinnerten Materials und beläuft sich bei Erwachsenen auf fünf bis neun Items. Die Gedächtnisspanne zeigt einen alterskorrelierten Zuwachs, was auf vermehrte Anwendung von Strategien (Wiederholung und Gruppierung), Wissenszuwachs, verbesserte metakognitive Kompetenzen und eine Zunahme der Informationsverarbeitungsgeschwindigkeit zurückgeführt wird (Dempster, 1981).

In dem Arbeitsgedächtnismodell von Baddeley und Hitch (Baddeley, 1990) wird die alterskorrelierte Erhöhung der Gedächtnisspanne mit Unterschieden der Verarbeitungsgeschwindigkeit erklärt. Das Arbeitsgedächtnis besteht aus drei Subsystemen: einem visuell-räumlichen Skizzenblock, einer phonologischen Schleife und einer zentralen Exekutive. Letztere koordiniert die Gedächtnisprozesse und teilt Ressourcen zu. Visuelle, räumliche Reize und verbale Informationen, die bildhaft kodiert sind, werden vom visuell-räumlichen Skizzenblock verarbeitet. Sprachliche Informationen werden der phonologischen Schleife zugeordnet, die sich aus einer Speicherkomponente und einer artikulatorischen Kontrollkompo-

¹ Mehrspeichermodelle stehen im Gegensatz zu den Modellen der Analysestufen von Craik und Lockhart (1972, LOP; Tiefe der Verarbeitung, s. oben). Während Mehrspeichermodelle Speicherstrukturen in den Vordergrund stellen, fokussieren Modelle der Analysestufen Verarbeitungsprozesse.

nente zusammensetzt. Werden Informationen von der artikulatorischen Komponente eingelesen und wiederholt, so wird ein Verfall der Informationen verhindert. Jedoch ist die Kapazität der phonologischen Schleife begrenzt und vor allem von der Artikulationsgeschwindigkeit abhängig. Wird das Arbeitsgedächtnis aus entwicklungspsychologischer Perspektive untersucht, so steht zumeist die phonologische Schleife im Zentrum des Interesses. Durch die alterskorrelierte verstärkte Nutzung der Wiederholungsstrategie und die Verbesserung der Artikulationsgeschwindigkeit erhöht sich die Gedächtnisspanne, denn mehr Informationen können im Arbeitsgedächtnis aufrechterhalten werden.

Die Verbesserung der Qualität der Wiederholungsprozesse zeigt sich darin, dass jüngere Kinder (sieben/acht Jahre) oft ein einzelnes Item, das zumeist auch gerade präsentiert wurde, wiederholen („passive“ Wiederholungsstrategie). Jedoch neigen ältere Kinder und Erwachsene dazu mehrere Items zu wiederholen („kumulative“ Wiederholungsstrategie). Damit gelingt es ihnen eine höhere Anzahl von Items im Arbeitsgedächtnis zu halten (Chi, 1978; Ornstein, Baker-Ward & Naus, 1988).

Ein Modell von Case (1985) ist in diesem Zusammenhang interessant, da es entwicklungspsychologische Befunde zur Gedächtniskapazität integrieren kann. Nach Case (1985) wird zwischen einer Speicherkomponente („storage space“) und einer Verarbeitungskomponente („operating space“) unterschieden. Die Gesamtkapazität des Kurzzeitspeichers ist altersinvariant, allerdings verschiebt sich das Verhältnis von Verarbeitungs- und Speicherkapazität mit zunehmendem Alter dahingehend, dass die Speicherkapazität überwiegt. Der alterskorrelierte Zuwachs in der Gedächtnisspanne ist nach Case (1985) darauf zurückzuführen, dass im Verlauf des Vor- und Grundschulalters zur Bewältigung der gleichen Aufgaben wegen steigender Informationsgeschwindigkeit weniger Platz für die erforderlichen mentalen Operationen (die Speicherkomponente) benötigt wird. Kinder und Erwachsene verfügen demnach über eine ähnliche Verarbeitungskapazität, aber sie können diese noch nicht so effizient und automatisch wie Erwachsene nutzen, was sich infolgedessen auch in der geringeren Gedächtnisspanne widerspiegelt.

In der eigenen Untersuchung spiegelt sich die Rolle der Gedächtnisspanne in der ausgewählten Anzahl der Listenitems in den DRM-Listen wieder, so dass jüngere Kinder nicht überfordert sind aber für Erwachsene die Aufgabe auch nicht zu einfach ist und sie die Listen perfekt lernen können.

3.2 Gedächtnisstrategien

Gedächtnisstrategien sind „...potentiell bewusste, intentionale kognitive Aktivitäten (...), die dabei helfen sollen, eine Gedächtnisaufgabe besser zu bewältigen“ (Schneider & Büttner, 2002, S. 501).

Auf die Entwicklung der Gedächtnisstrategien sind viele Forscher eingegangen, um differenziert alterskorrelierte Unterschiede im Einsatz von Strategien und deren Einflussfaktoren herauszuarbeiten. Vor dem Eintritt in die Grundschule und auch in der frühen Grundschulzeit verwenden Kinder Gedächtnisstrategien häufig nicht spontan, sind aber nach gezielter Anweisung fähig, diese effizient zu nutzen (so genanntes Produktionsdefizit; Bjorklund & Coyle, 1995) und damit ihre Gedächtnisleistungen zu verbessern. Jüngere Kinder setzen Strategien auch ein, jedoch ohne ihre Gedächtnisleistung zu verbessern. Dieses so genannte Nutzungsdefizit wird im Sinne einer Übergangsphase erklärt, in welcher Strategien bereits angewendet werden, aber ein hohes Maß an mentalen Ressourcen beanspruchen, welche im Folgenden die Kapazitäten zur Aufgabenausführung reduzieren.

3.2.1 Enkodierstrategien

In der Enkodierphase lassen sich vier Strategien voneinander unterscheiden: Wiederholen („rehearsal“), Organisieren, Kategorisieren und Elaborieren. Das Wiederholen wird vorwiegend zwischen fünf und 10 Jahren genutzt (Flavell, Beach & Chinsky, 1966), jedoch liegt der bereits erwähnte Vorteil älterer Kinder darin, dass sie mehrere Items auf einmal wiederholen (s. Abschnitt 3.1) und nicht wie jüngere jedes Item einzeln (Ornstein, Naus & Liberty, 1975).

Die Strategie des Organisierens mit der Unterform des Kategorisierens entwickelt sich auch mit zunehmendem Alter, wobei komplexe Organisationsstrategien meist erst ab dem 13. Lebensjahr gezielt eingesetzt werden (Bjorklund, 1985). Jüngere Kinder weisen Defizite in diesen Verarbeitungsformen auf, was sich in Freien Reproduktionstests und Wiedererkennentests manifestiert. Zwar können jüngere Kinder von semantischen Assoziationen im Lernmaterial durch eine automatische, unbewusste Aktivierung profitieren, aber sie sind nicht fähig, semantisch-relationale Verarbeitungsstrategien einzusetzen, um so das Lernmaterial zu organisieren (Bjorklund, 1985; Bjorklund & Jacobs, 1985). Nach Hasselhorn (1996) können Kinder schon ab dem vierten Schuljahr Organisationsstrategien spontan und flexibel nutzen. Hasselhorn (1995, 1996) fasst in seiner Strategie-Emergenz Theorie die Bedingungen zum Einsatz der Organisationsstrategie zusammen. Während beim automatischen kategorialen Orga-

nisieren, (bei Achtjährigen) eine ausreichende Informationsgeschwindigkeit und eine hinreichend starke kategoriale Inter-Item Relation vonnöten sind, wird für die strategische Nutzung des kategorialen Organisierens (bei 10-Jährigen) ein angemessen entwickeltes aufgabenspezifisches Metagedächtnis vorausgesetzt. Bei Achtjährigen wird das automatische kategoriale Organisieren maßgeblich von der Wissensbasis bestimmt, hingegen resultiert bei 10-Jährigen das strategische kategoriale Organisieren durch das Erkennen der Listenstruktur, welches vom Metagedächtnis beeinflusst wird.

Das Elaborieren ist eine komplexe Enkodierstrategie, die erst relativ spät, d.h. in der frühen Adoleszenz, angewendet wird (Pressley, 1982). Durch verbessertes Wissen und Strategien können bildliche oder sprachliche Assoziationen zwischen Wortpaaren aufgebaut werden, wodurch eine integrierte Repräsentation resultiert (Bjorklund, 1989). Da das Elaborieren das höchste Niveau der Enkodierstrategien darstellt, wird es selbst von vielen Erwachsenen nicht spontan eingesetzt. Werden aber die Probanden gezielt darauf hingewiesen, diese Strategie zu nutzen, dann verbessert sich die Gedächtnisleistung in allen Altersgruppen. Jedoch können jüngere Kinder unter diesen Umständen weniger effektiv elaborieren als ältere und Erwachsene.

Werden diese Annahmen und Befunde auf die eigene Untersuchung übertragen, dann könnte davon ausgegangen werden, dass jüngere Kinder von semantischen Assoziationen im Lernmaterial hinsichtlich ihrer Gedächtnisleistung profitieren könnten. Allerdings ist zu erwarten, dass sie im Gegensatz zu älteren Kindern keine semantisch-relationalen Verarbeitungsstrategien einsetzen (Bjorklund, 1985). Vielmehr könnten jüngere Kinder die Aufgabe durch eine automatische Aktivierungsausbreitung im semantisch-lexikalischen Netzwerk bewältigen.

3.2.2 Abrufstrategien

Alterskorrelierte Unterschiede zeigen sich auch im Einsatz von Abrufstrategien. Viele Studien weisen darauf hin, dass jüngere Kinder aufgrund ihrer wenig elaborierten Verarbeitungsstrategien nicht spontan fähig sind, Inferenzen in der Lernphase zu ziehen und als Erinnerungshilfen in der Testphase zu nutzen (Paris, 1978). Paris und Lindauer (1976) präsentierten drei verschiedenen Altersgruppen (fünf-, sieben- und neunjährigen Kindern) Sätze, von denen jeder eine Handlung beschrieb, aber das notwendige Handlungsinstrument fehlte, das von den Kindern selbstständig generiert werden musste (Beispiel: "Our neighbor unlocked the door."; Fehlendes Handlungsobjekt: "key"). Im folgenden "cued recall" Test erhielten die Kinder ein aus der Lernphase bereits präsentiertes Subjekt, Verb oder Objekt (explizite "Cues") oder

ihnen wurde das implizierte Handlungsinstrument (implizite “Cues“) vorgestellt. Während Fünfjährige explizite Cues besser nutzen konnten als implizite, gab es diesen Unterschied bei Neunjährigen nicht mehr. Sie konnten sowohl explizite als auch implizite Cues nutzen. Um den Alterseffekt mit älteren Kindern zu überprüfen, wurde die Studie mit sieben-, neun- und 11-jährigen Kindern repliziert (Paris & Lindauer, 1976). Jüngere Kinder profitierten erneut eher von expliziten als impliziten Cues, aber nun konnten nur die 11-Jährigen beide Cues gleichermaßen nutzen.

Keniston und Flavell (1979) untersuchten Schüler der ersten, dritten und siebten Klasse sowie Collegestudierende, um den Einsatz von Abrufstrategien zu prüfen. Nur die beiden älteren Versuchsgruppen konnten eine effektive Abrufstrategie spontan einsetzen, wohingegen die jüngeren Probanden die Strategie nur nach gezielter Instruktion anwendeten. Ceci und Howe (1978) bestätigten dies in einer Studie mit Kindern im Alter von vier, sieben und 10 Jahren. Alle Altersgruppen verwendeten gleichermaßen taxonomische und thematische Hinweisreize (“cued recall“), jedoch generierten nur die älteren Kinder diese im Suchprozess auch spontan (“free recall“) und wechselten häufiger zwischen ihnen.

Nach Schneider und Büttner (2002) sollten die vorgefundenen Altersunterschiede weniger auf eine verminderte Verfügbarkeit, vielmehr auf eine verminderte Zugänglichkeit der enkodierten Informationen zurückgeführt werden.

Zusammenfassend spiegelt sich die Überlegenheit älterer Probanden gegenüber jüngeren in den Abrufstrategien darin wider, dass sie eher auf interne Erinnerungshilfen zurückgreifen, über geeignete Suchstrategien verfügen und damit gezielter und flexibler Informationen im eigenen Gedächtnis finden.

3.3 Metagedächtnis

Das Konzept des Metagedächtnisses wurde von Flavell (1971) eingeführt und bezeichnet das bewusste Wissen über das Gedächtnis und dessen Inhalte. Es findet eine weitergehende Differenzierung zwischen deklarativem und prozeduralem Metagedächtnis statt.

3.3.1 Deklaratives Metagedächtnis

Deklaratives Metagedächtnis bezieht sich auf das verfügbare und verbalisierbare Wissen über das Gedächtnis, darüber hinaus auf das Wissen über Person-, Aufgaben- und Strategiemerkmale (Flavell & Wellman, 1977). Konkret handelt es sich um das Wissen über das eigene Ge-

dächtnis und das Gedächtnis anderer Personen, über Merkmale von Gedächtnisaufgaben, die diese erschweren oder erleichtern und um das Wissen, wann einzelne Enkodier- und Abrufstrategien effektiv eingesetzt werden können.

Interviewstudien eignen sich, um das deklarative Metagedächtnis zu erfassen. Bisherige Untersuchungen zeigten, dass das Metagedächtnis zu Person-, Aufgaben- und Strategiemerkmale bei Kindergartenkindern bereits vorhanden, wenn auch rudimentär ausgebildet ist und dass es sich von der Grundschulzeit bis in die Adoleszenz verbessert (Hasselhorn, 1995). Kindergartenkinder haben Kenntnisse über Funktionen und Leistungen des Gedächtnisses, sind mit Erinnerungshilfen vertraut und wissen, dass Gedächtnisleistungen von der Lernzeit, der Menge und der Art des Lernmaterials abhängig sind (z.B. Kreuzer, Leonard & Flavell, 1975). Dritt- und Fünftklässler können in Lern- und Testbedingungen planvoller reagieren und auf unterschiedliche Strategien zurückgreifen.

3.3.2 Prozedurales Metagedächtnis

Prozedurales Metagedächtnis bezeichnet die Fähigkeit eigene Gedächtnisaktivitäten zu überwachen und zu regulieren. Parallel zu gerade ablaufenden Gedächtnisaktivitäten schalten sich diese metakognitiven Prozesse ein und erfüllen ihre Kontroll- und Steuerungsfunktion.

Wie das deklarative Metagedächtnis zeigt die Entwicklung des prozeduralen Metagedächtnisses einen ähnlichen Verlauf: Kindergartenkinder können ihre Gedächtnisaktivitäten nur rudimentär kontrollieren und regulieren, was sich aber mit zunehmendem Alter verbessert. Im Grundschulalter entwickelt sich eine höhere Sensibilität für internale Gedächtniserfahrungen.

Methoden zur Erfassung des prozeduralen Metagedächtnisses sind beispielsweise die Einschätzung von Lernzeit. Dufresne und Kobasigawa (1989) erteilten verschiedenen Altersgruppen (Erst-, Dritt-, Fünft- und Siebtklässler) die Instruktion, Wortpaare so lange zu lernen bis sie sicher waren, dass sie die Items korrekt erinnern würden. Sie fanden heraus, dass bei selbstbestimmter Lernzeit die älteren Probanden (Fünft- und Siebtklässler) für schwieriger zu lernende (d.h. niedrig assoziierte) Wortpaare mehr Zeit investierten als für einfach zu lernende Paare (d.h. hoch assoziierte). Bei Drittklässlern war diese Differenz geringer und Erstklässler zeigten keinen Unterschied im Zeitaufwand des Lernens schwieriger und leichter Wortpaare. Dementsprechend differierten die Altersgruppen in ihrer Wiedergabeleistung, denn die älteren Probanden konnten mehr Items erinnern als die jüngeren Probanden.

In Studien zur Einschätzung der eigenen Gedächtnisleistung überschätzten sich Kindergartenkinder. Im Laufe der Grundschulzeit schätzen sich die Kinder zunehmend realistischer ein und unterscheiden sich Ende der Grundschulzeit nicht mehr von der Selbsteinschätzung Erwachsener (Schneider, 1989).

Anhand einer umfassenden Metaanalyse konnte ein Zusammenhang mittlerer Größe ($r = .41$) zwischen Metagedächtnis und Gedächtnisleistung festgestellt werden (Schneider, 1985). Mit zunehmendem Alter stieg der Zusammenhang und hing von anderen Faktoren wie Aufgabenart und -schwierigkeit ab. Beispielsweise zeigte sich für Vorschüler nur ein schwacher Zusammenhang, wenn es sich um lange Listen handelte und komplexe Gedächtnisstrategien genutzt werden sollten. Inwiefern die Gedächtnisleistung nun vom Metagedächtnis abhängt, ist insbesondere für altersvergleichende Studien nur zu beantworten, wenn die Testanforderungen beachtet werden.

In der eigenen Arbeit werden die Metagedächtnisfähigkeiten in Form der strategischen Quellendiskriminationsprozesse, nicht aber der heuristischen untersucht, da letztere nicht ausreichen, um die geforderten Aufgaben im DRM-Paradigma zu bewältigen. Jüngere Kinder können die strategische Quellendiskrimination wahrscheinlich weniger effizient nutzen verglichen mit älteren Kindern und Erwachsenen. Dies kann sich wiederum im höheren Anteil falscher Erinnerungen widerspiegeln. Bisherige Befunde zum "source monitoring" (s. Abschnitt 1.1 und 1.3) verweisen auf die Schwierigkeiten jüngerer Kinder in der Unterscheidung zweier sehr ähnlicher Quellen (Ceci & Bruck, 1993; Foley & Johnson, 1985). Für Sechsjährige ist es eine schwierige Aufgabe zu beurteilen, ob sie eine Handlung tatsächlich ausgeführt oder ob sie sich diese nur vorgestellt haben ("internal source monitoring"; Foley & Johnson, 1985). Jedoch waren sie nicht schlechter als erwachsene Probanden in der Unterscheidung selbst vorgestellter Handlungen und durch andere ausgeführte Handlungen ("reality monitoring"; s. Abschnitt 1.1 und 2.2.2), darüber hinaus waren sie Erwachsenen nicht unterlegen, wenn die Aufgabe lautete zu unterscheiden, wer von zwei Personen bestimmte Handlungen ausgeführt hatte ("external source monitoring"). Ähnliche Ergebnisse erbrachten Foley, Johnson und Raye (1983) in einer Studie zur Unterscheidung vorgestellter von gesprochenen Wörtern. Demnach haben Kinder besondere Schwierigkeiten in der Unterscheidung zweier interner Quellen. Jedoch weisen Lindsay, Johnson und Kwon (1991) darauf hin, dass die schlechteren Leistungen in der Quellendiskrimination weniger darauf zurückgehen, dass die Kinder selbst die Quelle der Information sind, sondern vielmehr auf die Aufgabenschwierigkeit. Sind zwei Quellen einander sehr ähnlich, dann haben v.a. jüngere Kinder Probleme in der Quellendiskrimination. In Untersuchungen zur Suggestibilität von Kindern zeigte sich, dass jüngere

Kinder tatsächliche Ereignisse und perzeptuell oder semantisch ähnliche suggerierte kaum voneinander unterscheiden können. Diese Befunde beziehen sich auf Untersuchungen außerhalb des DRM-Paradigmas.

Im Rahmen des Falschinformationsparadigmas konnten auch Ackil und Zaragoza (1995, 1998) zeigen, dass jüngere Kinder stärkere Suggestibilitätseffekte hatten als ältere und Erwachsene. Hierbei wurde von den Probanden verlangt, Informationen zwei verschiedenen externen Quellen zuzuweisen (1995) bzw. einer externen und einer internen Quelle (1998). Diese Studien werden im Rahmen der im Abschnitt der eigenen Untersuchungen (vgl. Abschnitt 9.1) differenzierter dargestellt, da sie ein Anhaltspunkt für die Konzeptualisierung eines eigenen Experiments sind.

3.4 Bereichsspezifisches Wissen

Nach den Ergebnissen verschiedener Studien kann das Vorwissen einer Person entscheidend zu ihrer Gedächtnisleistung beitragen (Bjorklund, 1987). Es wird zwischen dem Inhaltswissen, welches sich auf die spezifischen inhaltlichen Kenntnisse eines Gegenstandsbereiches bezieht, und dem strukturellen Vorwissen, welches das Wissen um allgemeine Organisationsformen (beispielsweise von Texten) umfasst, unterschieden (Liben, 1982). Beide Wissensformen sind dem semantischen Langzeitgedächtnis zugehörig und es wird davon ausgegangen, dass sie die Gedächtnisleistung unbewusst und automatisch beeinflussen können (Schneider & Büttner, 2002).

Anhand von Netzwerkmodellen (vgl. Abschnitt 2.2.2) lässt sich die Entwicklung bereichsspezifischen Vorwissens verdeutlichen. Das Wissen von Kindern über beispielsweise Tiere könnte im Netzwerk so organisiert sein, dass verschiedene Exemplare (z.B. Hund) auf verschiedenen hierarchischen Ebenen Verknüpfungen mit bestimmten Merkmalen des Exemplars (z.B. kann bellen, hat einen Schwanz) aufweisen. Im Laufe der Entwicklung wird das Wissen durch verschiedene Erfahrungen umfassender. Das Netzwerk verändert sich dahingehend dass es komplexer wird: Nicht nur die Anzahl der Knoten, sondern auch die Anzahl und damit die Stärke der Verknüpfungen zwischen ihnen steigen. Je stärker die Knoten miteinander vernetzt sind desto größer ist die Wahrscheinlichkeit, dass die Aktivierung eines Knotens sich auch auf benachbarte Bereiche auswirkt. Diese erhöhte Verfügbarkeit und Zugänglichkeit des Vorwissens führen zu schnelleren und effizienteren Verarbeitungsprozessen (Schneider & Büttner, 2002).

Durch das vielseitig vorhandene, wenn auch rudimentäre, Wissen von Kindern werden diese als „universelle Novizen“ bezeichnet. Wird versucht den Einfluss der Wissensbasis auf die Gedächtnisleistung zu untersuchen, so stellt sich ein Problem, denn Wissensstand und chronologisches Alter sind konfundiert. In der eigenen Studie muss der Wissensstand bei der Erstellung des Lern- und Testmaterials berücksichtigt werden, um eine solche Konfundierung zu vermeiden. Keinesfalls sollte der Fehler begangen werden, von einer generellen Bekanntheit auf eine altersinvariante Vertrautheit des Materials zu schließen. Darüber hinaus sollte beachtet werden, dass das gleiche Material in verschiedenen Altersgruppen unterschiedliche Bedeutungen haben kann, resultierend aus der unterschiedlichen Struktur der Netzwerke (Schneider & Büttner, 2002).

Eine Möglichkeit diese Konfundierung zu umgehen bietet das Experten-Novizen Paradigma, bei dem Vorwissen und Alter nicht korreliert sind. Chi (1978) verglich die Rekonstruktionsleistung von 10-jährigen Schachexperten mit erwachsenen Schachnovizen. Beide Gruppen sollten verschiedene kurzzeitig präsentierte Positionen von Schachfiguren auf einem leeren Schachbrett nachstellen und es zeigte sich, dass die Kinder den Erwachsenen überlegen waren. Demzufolge kann das bereichsspezifische Vorwissen erwartete Altersunterschiede in der Gedächtnisleistung ausgleichen oder sogar ins Gegenteilige wenden. In einem allgemeinen Test zur Erhebung der Gedächtnisspanne erbrachten die Kinder erwartungskonform geringere Leistungen als die Erwachsenen. In einer weiterführenden Studie von Opwis, Gold, Gruber und Schneider (1990) zeigte sich, dass kindliche Schachexperten erwachsenen Novizen am deutlichsten überlegen waren, wenn es sinnvolle Schachpositionen zu lernen galt. Bei sinnlosen Schachpositionen zeigte sich immerhin noch ein geringfügiger Altersunterschied zugunsten der kindlichen Experten. Kinder können dann bessere Leistungen erbringen, wenn Informationen aus bestimmten Gebieten erfragt werden, über die sie sehr gut Bescheid wissen. Durch ihr Vorwissen können Kinder Informationen anders enkodieren (die Verbindung zwischen den Items wird gestärkt und zugehörige Konzepte werden aktiviert) und auch schon elaborierte Gedächtnisstrategien spontan und gezielt einsetzen, um Informationen effizient zu erinnern. Die Bedeutung des Strategieeinsatzes bei Kindern wird durch die Analyse mehrerer Studien dahingehend relativiert, dass kindliche Experten hauptsächlich durch ihr Vorwissen zu überlegenen Leistungen gelangen können (Bjorklund & Douglas, 1997).

Nach Schneider und Büttner (2002) führen das Inhaltswissen und das strukturelle Vorwissen zu konstruktiver Verarbeitung, durch welche lückenhafte Erinnerungen mit Inhalten gefüllt werden, die nicht explizit genannt wurden (d.h. Inferenzen) oder nicht vorkamen (d.h. falschen Erinnerungen). Bereits Kleinkinder (und sogar Säuglinge) können auf strukturelles

Vorwissen in Form von Skripten (Wissensstrukturen über den typischen Ablauf von Ereignissen) zurückgreifen (Fivush, 1997). Das strukturelle Vorwissen hat einen bedeutenden Einfluss auf die Gedächtnisleistung und zeigt einen alterskorrelierten Anstieg: Im Abschnitt 3.2.2 wurde bereits die Untersuchung von Paris und Lindauer (1976) aufgeführt, die demonstrierte, dass ältere Kinder Abrufhilfen besser nutzen konnten als jüngere Kinder. Abrufhilfen waren für jüngere Kinder nur dann hilfreich, wenn sie in der Lernphase dargeboten worden waren. Somit können fehlende Informationen mit zunehmendem Alter besser spontan erschlossen werden. Einhergehend mit der besseren Strukturiertheit des semantischen Wissens können Informationen, die im Abrufprozess nicht mehr verfügbar sind, dennoch bedeutungskonsistent richtig erinnert werden. Allerdings können fehlende Informationen auch falsch rekonstruiert werden.

In der eigenen Studie würde dies bedeuten, dass jüngere Kinder weniger falsche Erinnerungen produzieren aufgrund ihrer verminderten Fähigkeiten fehlende Informationen zu erschließen.

In diesem Kapitel sollten die Grundlagen zum Verständnis der Gedächtnisentwicklung gegeben werden, welche für das Verständnis der eigenen Untersuchung relevant sind. Ausgehend von diesen theoretischen Grundlagen können im folgenden Kapitel Befunde zu alterskorrelierten Veränderungen falscher Erinnerungen dargestellt und erklärt werden.

4 Alterskorrelierte Veränderungen falscher Erinnerungen

Studien mit dem Ziel der Erforschung alterskorrelierter Veränderungen falscher Erinnerungen im DRM-Paradigma konzentrieren sich hauptsächlich auf den Vergleich jüngerer mit älteren Erwachsenen. Die eigene Untersuchung hingegen prüft die Entwicklung falscher Erinnerungen vom frühen Kindes- bis zum frühen Erwachsenenalter. Zu diesen Altersgruppen liegen bislang wenige veröffentlichte Studien (Brainerd et al., 2002a; Ghetti et al., 2002) mit widersprüchlichen Befunden vor, so dass eine umfassende Untersuchung besonders notwendig erscheint.

Zunächst werden aber erst die Vorhersagen des AMA erläutert, welcher die Basis der eigenen Untersuchung bildet. Dem folgt eine kritische Darstellung der Befunde im Rahmen des DRM-Paradigmas zu alterskorrelierten Veränderungen falscher Erinnerungen.

4.1 Vorhersagen des Aktivierungs-Monitoring Ansatzes

Theoretisch wird sowohl von dem AMA (Ghetti et al., 2002; Norman & Schacter, 1997; Schacter et al., 1999) als auch von der FTT (Brainerd & Mojardin, 1998; Brainerd & Reyna, 1998a; Brainerd & Reyna, 2004; Brainerd et al., 2002a; Brainerd et al., 2004) davon ausgegangen, dass sich verschiedene Aussagen zu alterskorrelierten Veränderungen falscher Erinnerungen treffen lassen.

Die Vertreter des AMA setzen sich nicht zum Hauptziel alterskorrelierte Veränderungen falscher Erinnerungen zu untersuchen, da sie dem kognitiven Forschungszweig zugehören und somit andere Schwerpunkte setzen (vgl. Abschnitt 2.2.3). Allgemein wird davon ausgegangen, dass die Aktivierung sich mit zunehmendem Alter erhöht und die Quellendiskrimination verbessert. Nach Roediger und Mitarbeitern (2001) kann die Aktivierung sowohl automatisch als auch bewusst sein. Die semantisch-relationale Verarbeitung erfolgt überwiegend bewusst und ist bei älteren Kindern stärker ausgeprägt. Diese Auffassung wird durch Befunde zur LOP-Manipulation gestützt, die darauf hinweisen, dass die Aktivierung beeinflusst werden kann und nicht nur unwillkürlich abläuft (vgl. Abschnitt 2.3.5). Die bewusste Aktivierung sollte bei Erwachsenen höher sein als bei Kindern, da sie über Verarbeitungsstrategien verfügen, besser und effektiver organisieren können (vgl. Abschnitt 3.2).

Mit zunehmendem Alter sollte sich auch die automatische Aktivierung erhöhen, da Erwachsene über mehr allgemeines und bereichsspezifisches Wissen verfügen als Kinder (vgl. Abschnitt 3.4).

Die Aktivierung ist die Grundvoraussetzung für falsche Erinnerungen, wird aber von der Quellediskrimination beeinflusst, die durch Kontrollprozesse falsche Erinnerungen verhindern soll. Aufgrund unzureichender Quellediskriminationsfähigkeiten sollten besonders Kinder und ältere Erwachsene falsche Erinnerungen aufweisen (Roediger & McDermott, 2000). Empirisch wurde diese Hypothese mit Kindern nicht untermauert. In einer Studie verglichen Gallo und Roediger (2003) jüngere mit älteren Erwachsenen und zeigten, dass ältere Erwachsene vermehrt falsche Erinnerungen produzierten, was die Autoren auf eine nachlassende Frontallappenfunktion zurückführten, die das Monitoring beeinträchtigte (vgl. auch Norman & Schacter, 1997; Schacter et al., 1999; Schacter et al., 1996).

Ältere Personen sollen Probleme in der Quellediskrimination haben, wenn die Quellen einander sehr ähnlich sind (Gallo & Roediger, 2003; Norman & Schacter, 1997). Nichtpräsen- tierte kritische Items werden fälschlicherweise als „alt“ erinnert, wenn sie nur gehört werden, da in diesem Fall wenige Details vermittelt werden und in Folge kann eine Zuordnung zu einer externen (Sprecher) oder internen Quelle (selbst vorgestellt) erschwert sein (Norman & Schacter, 1997). Zudem fanden die Autoren heraus, dass ältere Erwachsene Informationen schlechter itemspezifisch verarbeiten können, wodurch die Anzahl richtiger Erinnerungen in der Gruppe der älteren geringer ausfiel als in jener der jüngeren. Hinsichtlich der Anzahl richtiger Erinnerungen erweist es sich ebenfalls als sinnvoll auf die Unterteilung des Enkodierprozesses in itemspezifische und semantisch-relationale Verarbeitung von Hunt und McDaniel (1993) zurückzugreifen (s. Abschnitt 2.2.1). Jüngere Erwachsene können itemspezifische Informationen (Listenposition und Klang) zur Unterscheidung der Listenitems und kritischen Items besser nutzen und auf diese Weise den Anteil falscher Erinnerungen reduzieren und den der korrekten Erinnerungen erhöhen (Norman & Schacter, 1997).

In Abschnitt 3.3.2 wurde im Zusammenhang mit der Gedächtnisentwicklung bereits auf die alterskorrelierten Veränderungen der Quellediskrimination eingegangen. Die Probleme jüngerer Kinder in der Quellediskrimination können durch ihre geringe Arbeitsgedächtniskapazität erklärt werden, die sich auf Prozesse in der Lern- und Testphase auswirken. Während sie in der Lernphase aufgrund ihrer schlechteren itemspezifischen und semantisch-relationalen Verarbeitung weniger quellenrelevante Informationen verarbeiten können, ist in der Testphase die Nutzung quellenstrategischer Prozesse erschwert. Bei jüngeren Kindern werden Probleme

in der Quellediskrimination erwartet, wenn es um die Unterscheidung zweier interner oder sehr ähnlicher Quellen geht (Foley et al., 1983; Foley & Johnson, 1985). Dies lässt sich auf die eher schwierige intern-extern Entscheidung im DRM-Paradigma übertragen und zu der Vermutung führen, dass Kinder hier schlechter abschneiden als Erwachsene.

Bisher liegen keine Studien zur differenzierten Untersuchung der relevanten intern-externen Quellediskrimination im DRM-Paradigma an Kindern vor, nur eine Untersuchung an Erwachsenen: In einer Studie von Multhaup und Conner (2002) wurden den Probanden im Wiedererkennenstest zusätzlich zu den „alt/neu“ Antwortalternativen auch eine korrekte Antwortalternative für die nicht präsentierten kritischen Items in Form von den Kategorien „neu, aber fasst die Bedeutung der Listenitems zusammen“ bzw. „neu, ich habe selbst an das Item gedacht“ vorgegeben. Da keine vergleichbaren Untersuchungen an Kindern existieren, beruhen Vorhersagen zur Quellediskrimination bei Kindern auf Befunden zur Gedächtnisentwicklung (vgl. Kapitel 3).

Die geringere Arbeitsgedächtniskapazität, die im Allgemeinen wenig elaborierten Strategien und die im Speziellen schlechtere itemspezifische Verarbeitung führen zu der Annahme, dass jüngere Kinder älteren und Erwachsenen in Quellediskriminationsleistungen unterlegen sind. Darüber hinaus sollte die Gehirnentwicklung beachtet werden, da sich die zur Quellediskrimination nötigen Hirnregionen (Frontallappen) bis zur Adoleszenz erst vollständig ausbilden (zur Übersicht s. Zelazo, Carter, Resnick & Frye, 1996).

Aus den dargestellten Grundannahmen zur Aktivierung und Quellediskrimination lassen sich verschiedene Vorhersagen treffen, in Abhängigkeit davon, welcher Prozess überwiegt, d.h. ob die Aktivierung oder die Quellediskrimination ausschlaggebend ist.

Hat die Aktivierung einen größeren Einfluss auf falsche Erinnerungen als die Quellediskrimination, dann sollte sich ein alterskorrelierter Anstieg falscher Erinnerungen zeigen, da sowohl die automatische als auch die bewusste Aktivierung mit zunehmendem Alter zu steigen scheint.

Wenn die Quellediskrimination, die sich mit zunehmendem Alter verbessert, einen stärkeren Einfluss hat, dann sollten falsche Erinnerungen alterskorreliert abnehmen.

Wirken Aktivierung und Quellediskrimination gleichermaßen auf falsche Erinnerungen, so sollte sich kein Unterschied zwischen den Altersgruppen im Ausmaß falscher Erinnerungen zeigen. Zwar ist die Quellediskrimination bei Kindern schlechter als bei Erwachsenen, aber kritische Items werden bei ihnen auch weniger stark aktiviert.

4.2 Studien und deren Probleme

In den wenigen vor der Durchführung der eigenen Untersuchung veröffentlichten alterskorrelierten Studien zu falschen Erinnerungen im DRM-Paradigma wurden verschiedene Altersgruppen miteinander verglichen: Kindergarten- und Schulkinder, jüngere Erwachsene (Studenten) und Erwachsene im fortgeschrittenen Alter, wobei für die eigene Untersuchung nur die ersten drei Altersgruppen relevant sind. Je nach experimentellem Design und manipulierten Variablen ergaben sich unterschiedliche Resultate in den alterskorrelierten Veränderungen. In diesem Abschnitt werden die zwei bis zum aktuellen Zeitpunkt publizierten Studien im Rahmen des DRM-Paradigmas berichtet, die zu unterschiedlichen Aussagen gelangen.

Brainerd und Mitarbeiter (2002a) prüften in ihrer Studie, ob jüngere Kinder im DRM-Paradigma weniger zu falschen Erinnerungen neigen als ältere. Es wurde nach der FTT davon ausgegangen, dass mit dem Alter die den "verbatim"- und "gist"-Repräsentationen zugrunde liegenden Fähigkeiten zunehmen. Im DRM-Paradigma sollten die "verbatim"-Fähigkeiten nicht ausreichend sein um falsche Erinnerungen zu verhindern, da durch die Wortlisten insbesondere die "gist"-Repräsentationen angesprochen werden. Aus Sicht der FTT wurde somit ein alterskorrelierter Zuwachs falscher Erinnerungen erwartet (Brainerd & Reyna, 2002a). In ihrer ersten Studie verglichen Brainerd et al. (2002a) anhand von DRM-Wortlisten Kinder mit Erwachsenen und stellten fest, dass Kinder nur 6% der kritischen Items in der Freien Reproduktion falsch nannten, hingegen belief sich der Wert auf 50% in der Gruppe Erwachsener. Bei Kindergartenkindern kam es besonders häufig zu Intrusionen aus vorangehenden Listen, so dass davon ausgegangen wurde, dass sie den "gist" der Liste nicht erfassen konnten. In der zweiten Studie wurde die Erinnerungsleistung von Kindergartenkindern und Zweitklässlern in der Freien Reproduktion geprüft, wobei es auch darum ging das Ergebnis der ersten Studie mit DRM-Wortlisten zu replizieren, die bei Erwachsenen einen hohen Anteil falscher Reproduktionen hervorgerufen hatten. Insgesamt wurden ausgehend von Stadler et al. (1999) 16 DRM-Wortlisten mit je 15 Items verwendet, die anhand vorheriger Studien mit Erwachsenen zeigten, dass acht Listen hohe (im Folgenden als „hoch“ bezeichnet) und acht niedrige (im Folgenden als „niedrig“ bezeichnet) falsche Reproduktionsraten erzeugten. Die Ergebnisse zeigten, dass es mit zunehmendem Alter zu einem Anstieg in der richtigen Reproduktion kam. Die falsche Reproduktion fiel in beiden Altersgruppen (Kindergartenkinder und Zweitklässler) gering aus. Die Listen, die bei Erwachsenen hohe oder geringe falsche Reproduktionsraten erzeugt hatten, stimmten nicht mit denen überein, die bei Kindern zur falschen Reproduktion führten. Auch hier zeigten sich Intrusionen aus vorangegangenen Listen. Aufgrund dessen, dass kein Altersunterschied in den falschen Erinnerungen festgestellt werden konnte,

wurde im dritten Experiment ein größerer Abstand der Altersgruppen gewählt. Als Versuchspersonen wurden Kindergartenkinder, Sechstklässler und Studenten untersucht. Zusätzlich zur Freien Reproduktion wurde ein Wiedererkennenstest verwendet, wodurch das Vorgehen dem von Roediger und McDermott (1995; Experiment 2) ähnelte. In der Gruppe der Sechstklässler wurde aufgrund gesteigerter Elaborations- und Organisationsstrategien für die Freie Reproduktion ein Anstieg falscher Erinnerungen vermutet. Durch die bessere “gist“-Repräsentation würden 11-jährige eher als jüngere Kinder dazu neigen, die Items zu gruppieren und den gemeinsamen Bedeutungsgehalt zu erfassen. Durch den Aufbau der Untersuchung war ein direkter Vergleich der Ergebnisse des Reproduktions- und des Wiedererkennenstests möglich. Im Wiedererkennenstest wurden aufgrund der erhöhten Sensitivität vermehrt falsche Erinnerungen für Vorschüler erwartet als in der Freien Reproduktion. Darüber hinaus fand ein direkter Vergleich der Erinnerungen von Kindergartenkindern und Erwachsenen unter gleichen Bedingungen statt. Es wurden erneut die 16 Wortlisten verwendet, von denen acht direkt reproduziert wurden. Im Anschluss an die acht anderen Listen erfolgte keine Reproduktion; stattdessen wurden Buchstaben ausgemalt. Nach einer Distraktoraufgabe schloss sich ein Wiedererkennenstest für alle Probanden an. In der Freien Reproduktion zeigte sich ein alterskorrelierter Anstieg falscher Erinnerungen.

Tabelle 4.1:

Ergebnisse der Untersuchung von Brainerd et al. (2002a) für die kritischen und Listenitems in der Freien Reproduktion und im Wiedererkennen unterteilt nach Altersgruppen (in Klammern Werte der Wiedererkennensleistung ohne vorangegangene Reproduktion; alle Werte in relativen Häufigkeiten).

			Kindergarten	Sechstklässler	Erwachsene
Reproduktion	kritische Items	hoch	.10	.27	.53
		niedrig	.12	.18	.21
	Listenitems	hoch	.19	.39	.60
		niedrig	.22	.42	.66
Wiedererkennen	kritische Items	hoch	.77 (.77)	.83 (.91)	.92 (.91)
		niedrig	.76 (.73)	.81 (.79)	.82 (.79)
	Listenitems	hoch	.72 (.83)	.85 (.90)	.92 (.90)
		niedrig	.78 (.76)	.87 (.88)	.90 (.90)

Anmerkung: hoch/niedrig bezeichnet die Unterteilung in hohe oder geringe falsche Reproduktionsraten der verwendeten Listen.

Wie aus Tabelle 4.1 hervorgeht zeigten Kindergartenkinder weniger falsche Reproduktionen der kritischen nicht präsentierten Items als Sechstklässler, deren Rate falscher Erinnerungen wiederum geringer war als die der Studenten. Es konnte auch beobachtet werden, dass der Anteil der richtigen Reproduktion der Listenitems mit dem Alter kontinuierlich über die drei

Gruppen zunahm. Anhand der Ergebnisse des Wiedererkennenstests konnten für falsche Erinnerungen die Hauptbefunde des Reproduktionstests teilweise bestätigt werden: Kindergartenkinder hatten weniger falsche Erinnerungen als Sechstklässler, jedoch war die Differenz zwischen Sechstklässlern und Erwachsenen nicht signifikant. Richtige Erinnerungen im Wiedererkennenstest erhöhten sich mit zunehmendem Alter. Zudem wurden aber relativierte Werte berechnet (A' -Werte²; s. Howe, Cicchetti, Toth & Cerrito, 2004; Snodgrass & Corwin, 1988), die zeigten, dass es keinen signifikanten Unterschied der Altersgruppen gab: Zwar erzielten Erwachsene mehr Treffer, aber sie produzierten auch mehr falsche Alarme. Die Hauptbefunde des Reproduktions- und Wiedererkennenstests ähneln sich, jedoch weisen Brainerd et al. (2002a) auf einen bedeutenden Unterschied hin: In der Freien Reproduktion nannten Kindergartenkinder und Sechstklässler mehr richtige Erinnerungen als falsche, unabhängig ob die Liste bei Erwachsenen einen hohen oder geringen Anteil falscher Reproduktionen bewirkte. Im Wiedererkennenstest hatten Kindergartenkinder einen gleichen Anteil richtiger und falscher Erinnerungen, was verdeutlicht, dass sie hier eher der Gedächtnisillusion unterlagen.

Gemäß der FTT lassen sich die Befunde dahingehend interpretieren, dass der alterskorrelierte Anstieg falscher Erinnerungen auf bessere "gist"-Gedächtnisspuren älterer Probanden zurückgeführt werden kann. Durch das DRM-Paradigma kann auch die verbesserte "verbatim"-Spur erwachsener Probanden nicht falschen Erinnerungen der "gist"-konsistenten Distraktoren entgegenwirken. Howe et al. (2004) replizierten die Untersuchung von Brainerd et al. (2002a) und kamen zu ähnlichen Ergebnissen in der Freien Reproduktion, welche einen alterskorrelierten Anstieg falscher Erinnerungen zeigten. Auch sie führten diese Befunde auf die schlechtere Nutzung der "gist"-Spur jüngerer Kinder zurück.

Eine andere Studie zu alterskorrelierten Veränderungen im DRM-Paradigma wurde von Ghetti et al. (2002) durchgeführt, die drei Altersgruppen (Kindergartenkinder, Zweitklässler und Erwachsene) in ihren Anteilen falscher Erinnerungen verglichen. Ghetti et al. (2002) machten keine eindeutigen Vorhersagen zum Entwicklungstrend falscher Erinnerungen, vielmehr gingen sie davon aus, dass gemäß der FTT sowohl eine alterskorrelierte Zunahme (durch die bessere Extraktion des "gist" der Wortliste mit zunehmendem Alter) als auch eine alterskorrelierte Abnahme falscher Erinnerungen (mit zunehmendem Alter können durch bessere "verba-

² Für Listenitems: $A' = .5 + [(H - FA_L)(1 + H - FA_L) \div [4H(1 - FA_L)]]$
Für kritische Items: $A' = .5 + [(FA_{KIP} - FA_{KINP})(1 + (FA_{KIP} - FA_{KINP})) \div [4FA_{KIP}(1 - FA_{KINP})]]$.

H = Treffer richtig wiedererkannter Listenitems

FA_L = Falsch „alt“ bezeichnete Listenitems

FA_{KIP} = Falsch „alt“ bezeichnete kritische Items präsentierter Listen

FA_{KINP} = Falsch „alt“ bezeichnete kritische Items nicht präsentierter Listen

tim“-Fähigkeiten eher Nicht-Identitätsurteile der nicht präsentierten kritischen Items getroffen werden, d.h. das kritische Item kann als nicht gelernt identifiziert werden) erwartet werden könnte. Um die Rolle distinktiver Information zu überprüfen, verglichen die Autoren semantisch assoziierte Wortlisten mit Bildmaterial. Somit erweiterten sie das Vorgehen von Seamon und Mitarbeitern (2000), indem ein direkter Vergleich des Einflusses von verbalem und bildlichem Material auf die Gedächtnisleistung unterschiedlicher Altersgruppen möglich war. Ghetti und Mitarbeiter (2002) wollten überprüfen, ob durch die Präsentation von Bildern falsche Erinnerungen bei Kindern und Erwachsenen gleichermaßen reduziert werden können. Bei Kindern sollte nur dann eine Verringerung falscher Erinnerungen eintreten, wenn die Bilder itemspezifische Informationen liefern, die der Gedächtnisillusion entgegenwirken. Nach der Distinktivitätsheuristik (Schacter et al., 1999) sollte dieser Effekt bei Kindern aber nicht eintreten, weil ihre itemspezifische Verarbeitung durch die geringere Arbeitsgedächtniskapazität schlechter sein sollte. Darüber hinaus wollten die Autoren alterskorrelierte Unterschiede in der externen Quellediskrimination (zwei Sprecher unterschiedlichen Geschlechts) untersuchen.

Die Probanden erhielten 10 Wortlisten mit sieben Items in der Lernphase. Durch die Begrenzung auf sieben Items sollte eine Überforderung jüngerer Kinder vermieden werden. Die Präsentation der Listen erfolgte zum einen durch einen männlichen und zum anderen durch eine weibliche Versuchsleiter/in. Eine Variation der symbolischen Darbietungsform wurde dadurch erreicht, dass einer Hälfte der Versuchspersonen semantisch assoziierte Wortlisten und der anderen entsprechendes Bildmaterial der Listenitems präsentiert wurde. Es erfolgte nach jeder Liste eine Freie Reproduktion und im Anschluss an alle Listen ein Wiedererkennenstest.

Tabelle 4.2:

Ergebnisse der Untersuchung von Ghetti et al. (2002) für die kritischen und Listenitems in der Freien Reproduktion und im Wiedererkennen unterteilt nach Altersgruppen (Werte in relativen Häufigkeiten).

			Fünfjährige	Siebenjährige	Erwachsene
Reproduktion	kritische Items	ohne	.15	.27	.53
		mit	.15	.18	.21
	Listenitems	ohne	.88	.39	.60
		mit	.90	.42	.66
Wiedererkennen	kritische Items	ohne	.36	.83 (.91)	.92 (.91)
		mit	.28	.81 (.79)	.82 (.79)
	Listenitems	ohne	.88	.85 (.90)	.92 (.90)
		mit	.91	.87 (.88)	.90 (.90)

Anmerkung: mit/ohne bezieht sich auf die Lernbedingung mit oder ohne Bildmaterial.

Tabelle 4.2 fasst die Ergebnisse zusammen. In der Freien Reproduktion wurden insgesamt mehr Listenitems als kritische Items genannt. Zudem zeigten ältere Probanden mehr richtige Erinnerungen an Listenitems als jüngere, während in der falschen Reproduktion kein Altersunterschied festgestellt wurde. Eine relativierte Betrachtung der Werte der falschen Reproduktion (die reproduzierten kritischen Items wurden zu der Anzahl der insgesamt reproduzierten Items in Beziehung gesetzt), führte jedoch zu einem anderen Ergebnis. Kindergartenkinder hatten im Vergleich zu den anderen Altersgruppen eine höhere Rate falscher Erinnerungen. Die bildliche Darbietung erbrachte keine Verbesserung der korrekten Erinnerung an Listenitems, aber sie konnte falsches Erinnern an kritische Items reduzieren. Im Wiedererkennenstest war der Anteil erinnerter Listenitems auch größer als jener der kritischen Items. Erwachsene erkannten mehr Listenitems wieder als die jüngeren Altersgruppen. Bei falschen Erinnerungen konnte kein Altersunterschied gefunden werden, denn alle Altersgruppen beurteilten im gleichen Maße kritische Items als „alt“. Das Bildmaterial erhöhte im Wiedererkennenstest die richtigen und reduzierte die falschen Erinnerungen in allen Altersgruppen. Für die Quellenattribution zeigte sich nur in der Bildbedingung ein Altersunterschied. Kindergartenkindern fiel es schwer Listenitems der richtigen Quelle zuzuordnen und kritische Items keinem der Sprecher zuzuweisen.

Die dargestellten Befunde verweisen darauf, dass in der Freien Reproduktion und im Wiedererkennenstest kein signifikanter Altersunterschied im Ausmaß falscher Erinnerungen zwischen den Altersgruppen festgestellt werden konnte. Dieser Befund führt nicht zu eindeutigen Schlussfolgerungen: Einerseits könnten Erwachsene die kritischen Items stärker aktiviert haben, aber aufgrund ihrer verbesserten Quellendiskriminationsfähigkeiten wurde der Aktivierung entgegengewirkt, so dass sie nicht vermehrt falsche Erinnerungen produzierten. Ande-

rerseits könnte dieses Ergebnis darauf verweisen, dass die Aktivierung und die Quellediskrimination durch die Testkonstruktion in allen Altersgruppen vergleichbar waren.

Obwohl beide Studien dem DRM-Paradigma folgten, führten sie doch zu unterschiedlichen Ergebnissen. Die konträre Befundlage über die alterskorrelierte Veränderung falscher Erinnerungen trägt, ohne eine differenzierte Betrachtung des Designs, wenig zum Erkenntnisgewinn bei. Während Brainerd et al. (2002a) zu dem Schluss kommen, dass ältere Kinder und Erwachsene eher zu falschen Erinnerungen neigen als jüngere Kinder, treffen Ghetti et al. (2002) eine andere Vorhersage: Sie konnten keine Altersunterschiede feststellen; durch das Hinzuziehen relativierter Werte konnten sie sogar eine gegenteilige Prognose formulieren. Bei Ghetti et al. (2002) wurden die Ergebnisse zur Bild-/Nichtbildbedingung zusammen anstatt separat angegeben, was einen Vergleich mit den Ergebnissen von Brainerd et al. (2002a) erschwert.

Wird die Anzahl der Listenitems betrachtet, so zeigt sich ein Grund für die unterschiedlichen Resultate: Während Brainerd et al. (2002a) 12 oder 15 Items pro Liste präsentierten, wurden bei Ghetti et al. (2002) nur sieben Listenitems verwendet, um eine Überforderung der jüngeren Kinder zu vermeiden. Bei Ghetti et al. (2002) lag in der Gruppe der Vorschüler die falsche Reproduktionsrate mit .25 deutlich höher als bei Brainerd et al. (2002a) mit .06. Zudem wurde bei Ghetti et al. (2002) auf die Arbeitsgedächtniskapazität und den Wortschatz (konkrete Wörter) jüngerer Kinder geachtet, so dass die für falsche Erinnerungen verantwortlichen Prozesse/Repräsentationen anscheinend wirksam werden konnten. Es konnte gezeigt werden, dass Kinder im Alter von fünf Jahren bereits falschen Erinnerungen gegenüber empfänglich sein können.

Allgemein scheinen kürzere Wortlisten zu einer geringeren Rate falscher Erinnerungen zu führen, hingegen lange Wortlisten vermehrt zu falschen Reproduktionen oder Wiedererkennen (Robinson & Roediger, 1997). Die totale und nicht die mittlere Assoziationsstärke sollten sich auf falsche Erinnerungen auswirken, was bedeutet, dass mit jedem zusätzlichen Item in der Liste das kritische Wort aktiviert werden soll. Dies könnte aber nur für Erwachsene gelten, denn in einer Studie mit zunehmender Listenlänge reproduzieren Kinder weniger falsche Erinnerungen, da die Aktivierung des kritischen Items bei ihnen abzunehmen scheint (Price, Metzger, Phelps & Phelps, 2001; zit. n. Ghetti et al., 2002). Dies lässt sich anhand des AMA erklären. Da die Listenitems absteigend nach ihrer mittleren RAS in der Liste angeordnet sind, haben die letzten Items die geringste RAS. Daraus könnten Interferenzen (Aktivierung eines anderen Items) resultieren anstatt der Aktivierung des kritischen Items.

Aus der FTT ließe sich ableiten, dass ältere Kinder kurze Listen durch ihren verstärkten Zugriff auf die “verbatim“-Spur besser erinnern können und damit falsche Reproduktionen der kritischen Items unterbinden können. In der Untersuchung von Ghetti et al. (2002) lagen die Raten falscher Erinnerungen für ältere Kinder niedriger als bei Brainerd et al. (2002a). Es wäre denkbar, dass durch die kurzen, kindgerechten Wortlisten mit konkreten Items älteren Kindern die itemspezifische Verarbeitung erleichtert wurde, so dass die Listen von ihnen nahezu perfekt gelernt werden konnten. Zudem könnte es sein, dass ihnen durch die kurzen Listen und die bessere Verarbeitung die Quellediskrimination vereinfacht wurde. Jüngere Kinder hatten eine geringere Rate falscher Erinnerungen bei langen Listen, da sie diese itemspezifisch und nicht semantisch-relational verarbeiteten.

Zusammengefasst kann festgehalten werden, dass die Konstruktion der Wortlisten eine entscheidende Auswirkung auf die DRM-Illusion zu haben scheint. Während Ghetti et al. (2002) kindgerechte kurze Wortlisten verwendeten, griffen Brainerd et al. (2002a) auf Wortlisten zurück, die an Erwachsenen normiert und sehr lang waren. Dieses Problem der Konfundierung von Wortlisten mit dem Alter ergäbe sich nicht, wenn unterschiedliche Listenlängen für die verschiedenen Altersgruppen verwendet werden würden, die sich der Arbeitsgedächtniskapazität der jeweiligen Altersgruppe anpassen würden. Aufgrund der unterschiedlichen Befundlage lassen sich keine eindeutigen Aussagen bezüglich des Entwicklungstrends falscher Erinnerungen treffen.

Die Ergebnisse dieses Abschnitts verweisen auf eine Veränderung des Anteils falscher Erinnerungen mit dem Alter, wobei Uneinigkeit über die Richtung und die zugrunde liegenden Prozesse aufgrund der gegensätzlichen Effekte von Aktivierung und Quellediskrimination auf falsche Erinnerungen besteht. Von oberster Priorität erscheint es, dass Aktivierung und Quellediskrimination anhand spezieller Prozeduren separat erfasst werden sollten, was das Hauptziel der vorliegenden Untersuchung ist.

Im folgenden Abschnitt werden aktuelle Studien vorgestellt, die sich dieser Problematik widmen und auf verschiedene Weise versuchten, Aktivierung und/oder Quellediskrimination getrennt zu erfassen.

5 Prozeduren zur getrennten Erfassung von Aktivierung und Quellendiskrimination

Das Problem Aktivierung ohne den Einfluss von der entgegenwirkenden Quellendiskrimination zu erfassen, ist im Forschungsgebiet des AMA bekannt. Bisherige Studien versuchten auf unterschiedliche Weise Aufschlüsse über Aktivierung und Quellendiskrimination zu erlangen. In den folgenden Abschnitten werden Untersuchungen aufgeführt, welche durch verschiedene Testphasen versuchten, die Prozesse zu trennen oder anhand impliziter Gedächtnistests zu differenzierteren Aussagen zu gelangen. Andere Forscher entwickelten Instruktionen zur Förderung von Aktivierungsprozessen.

5.1 Zeitdruck: Ausschluss von Monitoringprozessen

Eine Möglichkeit die Quellendiskrimination in der Abrufphase auszuschließen könnte dadurch realisiert werden, dass die Probanden unter Zeitdruck gesetzt werden. Diese Variable soll aber nicht in der eigenen Untersuchung manipuliert werden, da in einer altersvergleichenden Studie mit Vorschülern dies nicht umsetzbar erscheint.

Um Aufschlüsse getrennt über Aktivierung und Quellendiskrimination in Wiedererkennentests zu erhalten, versuchte Benjamin (2001) in einem Experiment seiner Studie den Einfluss der strategischen Quellendiskrimination durch Zeitdruck zu minimieren. In dieser Untersuchung mit jüngeren und älteren Erwachsenen wurden zwei Variablen manipuliert: Präsentationshäufigkeit und Zeitdruck. Erwachsenen wurden die Wortlisten entweder ein- oder dreimal aufeinanderfolgend präsentiert und mit oder ohne Zeitdruck sollten sie die Wörter des Wiedererkennentests als „alt“ oder „neu“ beurteilen.

Für kritische Items zeigte sich eine Interaktion der beiden Variablen. In der Bedingung mit Zeitdruck führte die mehrfache Präsentation vermehrt zu falschen Erinnerungen verglichen mit der einmaligen Präsentation, was dadurch erklärt werden kann, dass die Aktivierung bei mehrfacher Präsentation stärker, aber strategische Quellendiskrimination kaum möglich ist. Zwar wird bei einmaliger Präsentation mit Zeitdruck die Quellendiskrimination minimiert, jedoch fällt die Aktivierung in diesem Fall auch geringer aus, so dass weniger falsche Erinnerungen resultieren.

Demgegenüber zeigte sich für die Bedingung ohne Zeitdruck ein gegenteiliger Effekt: Bei mehrfacher Präsentation der Listen war der Anteil falscher Erinnerungen geringer verglichen mit der einmaligen Präsentation. Hier führte die mehrfache Präsentation zu einer stärkeren Aktivierung als die einmalige Präsentation, aber durch die unbegrenzte Zeitvorgabe konnten die enkodierten Merkmale strategisch genutzt werden, was einen positiven Effekt auf die Quellendiskriminationsleistung und damit auf einen geringen Anteil falscher Erinnerungen hat.

Der Altersvergleich zeigte, dass ältere Probanden vermehrt zu falschen Erinnerungen neigten. Sie hatten auch ohne Zeitdruck beim Test einen ähnlichen Anteil falscher Erinnerungen wie jüngere Probanden, die unter Zeitdruck den Test bearbeitet hatten. In Einklang mit den Annahmen von Roediger und McDermott (2000) spricht dieser Befund dafür, dass die Quellendiskriminationsfähigkeiten älterer Probanden reduziert sind.

5.2 Implizite Gedächtnistests

Eine der effektivsten Möglichkeiten Aktivierung ohne entgegenwirkende Quellendiskrimination zu erfassen, die in einem der eigenen Experimente umgesetzt werden soll, stellen implizite Gedächtnistests dar in denen die Probanden die Aufforderung erhalten so schnell oder gut wie möglich eine bestimmte Aufgabe zu erledigen. Da es keine Erinnerungsinstruktion gibt, besteht keine Notwendigkeit zwischen tatsächlich präsentierten und selbst generierten Items zu unterscheiden (Roediger & McDermott, 1993). Ein impliziter oder Priming-Effekt ist dann demonstriert, wenn alte Informationen schneller/besser genutzt werden als vergleichbare neue Informationen (Graf & Schacter, 1985).

Einen Hinweis dafür, dass in impliziten Tests Aktivierung ohne Quellendiskrimination erhoben werden kann, erbrachten McKone und Murphy (2000). Sowohl in einem impliziten (Wortstammerngänzung) als auch in einem expliziten (wortstammbezogener Abruf) Test zeigten sich Erinnerungen an nicht präsentierte kritische Items. Indem Wortlisten mehrfach präsentiert wurden, erhöhte sich im expliziten Test der Anteil richtig erinnerter Items, während sich falsche Erinnerungen reduzierten. Im impliziten Test hingegen veränderte sich weder der Anteil der Erinnerung an präsentierte noch an nicht präsentierte Items, die zudem ähnlich hoch waren. Während sich die Erinnerungen in expliziten Tests von der wiederholten Lernbedingung beeinflussen ließen, blieb dieser Effekt in impliziten Tests aus, was darauf hinweist, dass strategische Quellendiskrimination in impliziten Tests nicht wirksam wird.

In einer ähnlichen Studie mit einem ebenfalls perzeptuell impliziten Test (in diesem Fall Wortfragmentergänzung) überprüfte Diliberto-Macaluso (2005) den Effekt impliziter und expliziter Testinstruktionen auf den Anteil falscher Erinnerungen bei 10-jährigen Kindern. Es resultierte signifikantes Priming sowohl für präsentierte als auch nicht präsentierte kritische Items nach beiden Instruktionen. Zudem war das Priming unabhängig vom Präsentationsstatus des kritischen Items in der Wortliste, demzufolge war die Aktivierung für präsentierte und nicht präsentierte kritische Items gleich.

In ihrer Studie stellte McDermott (1997) die Frage, ob sich Priming präsentierte und nicht präsentierte kritischer Items in impliziten Tests demonstrieren lässt. Sie überprüfte die assoziative Aktivierung anhand konzeptueller und perzeptueller impliziter Tests. McDermott (1997) ging davon aus, dass perzeptuelles Priming eine bewusste Aktivierung der nicht präsentierte kritischen Items bedeuten würde, da dieses Priming sich nur dann zeigen kann, wenn die Person sich die perzeptuellen Merkmale vorstellt. Konzeptuelles Priming demgegenüber sollte aus bewusster und unbewusster Aktivierung resultieren. Zeigen sich falsche Erinnerungen in einem konzeptuellen expliziten Test (Freie Reproduktion) auf, so sollte sich Priming dieser kritischen Items auch in konzeptuellen impliziten Tests zeigen. Das Vorgehen in ihren Experimenten richtete sich nach dem DRM-Paradigma, wobei aber etwas kürzere Listen (18 Wortlisten mit 12 Items) wegen des Wortergänzungstests visuell, statt wie meist üblich auditiv, dargeboten wurden. Es gab drei Listenbedingungen: In der ersten Bedingung kam das kritische Item in der Liste nicht vor, in der zweiten wurde es in der vierten Listenposition präsentiert und die dritte Bedingung diente der Erhebung der Basisrate im impliziten Test und daher wurde die Liste nicht präsentiert. In allen Experimenten wurde ein ähnliches Vorgehen in der Lernphase gewählt, welches nur durch den Einsatz unterschiedlicher Gedächtnistests in der Prüfphase verändert wurde.

In einem Experiment verwendete McDermott (1997) die Wortstammergänzung und die Wortfragmentergänzung als perzeptuelle implizite Tests anhand derer perzeptuelles Priming geprüft wurde. In der Wortstammergänzung zeigten sich Primingeffekte für präsentierte und nicht präsentierte kritische Items, die sich als gering erwiesen und sich nicht signifikant voneinander unterschieden. Im Test der Wortfragmentergänzung resultierte höheres Priming ($M = .12$) als in der Wortstammergänzung ($M = .08$). In einem weiteren Experiment wurde der Wortassoziationstest als konzeptuelles implizites Prüfverfahren eingesetzt. Bei der Konstruktion des Wortassoziationstests wurden Cues für die kritischen Items aller Listen bestimmt zu denen die Probanden ihre ersten sechs Assoziationen benennen sollten. Diese Cues (z.B. Knochen) waren Wörter, die nicht in der Liste vorkamen, aber schwach mit dem kritischen Item

(z.B. Hund) assoziiert waren. In diesem konzeptuell impliziten Wortassoziationstest zeigte sich, dass signifikantes Priming sowohl für präsentierte (8%) als auch nicht präsentierte kritische Items (8%) zustande kamen und dass das Priming über der Basisrate von 10% in der nicht gelernten Listenbedingung lag. Hinsichtlich des Priming unterschieden sich die Listenbedingungen präsentiert und nicht präsentiert also nicht voneinander, woraus geschlossen werden kann, dass die Aktivierung der nicht präsentierten kritischen Items anscheinend genauso hoch war als wären sie tatsächlich präsentiert worden.

Die Ergebnisse zeigen, dass implizite Verfahren geeignet sind Priming nicht präsentierter kritischer Items zu erheben. Aber nur im Fall des konzeptuell impliziten Tests zeigte sich signifikantes Priming, da die assoziative Aktivierung der DRM-Wortlisten sich am besten in einem konzeptuellen Test abbilden lässt. Somit ist eine Übereinstimmung der geforderten konzeptuellen Prozesse der Lern- und Testphase gesichert. Zur ausschließlichen Erfassung der Aktivierung sollen daher in einem der eigenen Experimente konzeptuell implizite Tests eingesetzt werden.

5.3 Förderung von Aktivierung durch Lerninstruktionen

Eine andere Möglichkeit zu separaten Aussagen über die Aktivierung zu gelangen besteht in derer gezielten experimentellen Manipulation. Indem die Aktivierung gefördert wird, sind Rückschlüsse auf die strategische Aktivierung und indirekt auch auf die automatische Aktivierung möglich. Marsh und Bower (2004) untersuchten die Auswirkung der Listenlänge [0 (nicht präsentiert), 3, 6, 9 und 12 Items] und des Generierens auf falsche Erinnerungen. Die Hälfte der Probanden wurde der Generierungsbedingung zugeteilt und sollte im ersten Experiment nach der Darbietung einer jeden Liste zwei Wörter generieren und aufschreiben, die ihnen als Erstes zur Liste einfielen. Die andere Hälfte stellte die Standardbedingung dar und bearbeitete nicht diese Aufgabe. Abschließend sollten im Wiedererkennenstest kritische Items, Listenitems, verbundene und unverbundene Distraktoren beurteilt werden. Die Generierungsaufforderung führte erstens zu einer häufigen Nennung der zugehörigen nicht präsentierten kritischen Items, was als Hinweis auf eine stärkere Aktivierung gewertet wurde, und zweitens wurden im anschließenden Wiedererkennenstest die selbständig generierten Items häufiger als „alt“ bezeichnet. Davon ausgehend schlussfolgerten die Autoren, dass die Anzahl generierter kritischer Items als ein Prädiktor für falsche Erinnerungen gesehen werden kann.

Um die Aktivierung zu fördern, scheint diese Vorgehensweise besonders für Untersuchungen mit jüngeren Probanden geeignet zu sein, denn Kinder im Vorschulalter setzen auf eine Lern-

instruktion hin keine effizienten Verarbeitungsstrategien ein (z.B. Bjorklund, 1985; Bjorklund, 1989; Paris, 1978). Wird ihnen aber eine geeignete Strategie durch die Instruktion oder die Art der Präsentation des Materials nahegelegt, so können sie in der Regel von der Strategie profitieren (z.B. Schneider & Büttner, 2002). Es erscheint sinnvoll diese Instruktion in der eigenen Untersuchung zu erteilen, um zu prüfen, welche Art von Aktivierung der kritischen Items bei Erwachsenen stärker sein könnte als bei Kindern und ob sich die Aktivierung bei Kindern fördern lässt. Wenn Unterschiede in der Aktivierung auf fehlende elaborative Verarbeitungsstrategien der Kinder zurückgehen, dann sollten Kinder besonders von der Aufgabe profitieren und Altersunterschiede geringer ausfallen. Falls die Altersunterschiede auf automatischen Aktivierungsprozessen gründen, dann sollten sie trotz fördernder Lerninstruktion bestehen bleiben.

5.4 Inklusionsinstruktionen

Eine Möglichkeit gezielt die Aktivierung kritischer Items ohne den Einfluss von Quellendiskrimination zu erfassen, zeigten sowohl Dehon und Brédart (2004), Hege und Dodson (2004) als auch Heit und Mitarbeiter (2004) anhand spezieller Instruktionen. Neben einer Standardinstruktion wurde eine so genannte „Inklusionsinstruktion“ in verschiedenen Varianten eingesetzt, in welcher die Probanden in der Freien Reproduktion sowohl die zuvor präsentierten Items nennen sollten als auch jene Items, an die sie während der Lernphase selbst sehr stark gedacht hatten (z.B. die entsprechenden kritischen Wörter). Da eine Inklusionsinstruktion keine Quellendiskrimination verlangt, konnte die Anzahl genannter kritischer Items als Indikator ihrer Aktivierungsstärke gelten. Aufschluss über die Güte der Quellendiskrimination erbrachte der Vergleich von Standard- und Inklusionsinstruktion bzw. einer zweiten Testphase.

Dehon und Brédart (2004) zeigten eine entsprechende Prozedur zur Erfassung der Aktivierung der kritischen Items: Anhand des DRM-Paradigmas untersuchten sie das Ausmaß falscher Erinnerungen junger und älterer Erwachsener. Die Probanden erhielten im ersten Experiment in der Lernphase sechs DRM-Wortlisten mit jeweils 15 Items und nach jeder Liste erfolgte ein Test der Freien Reproduktion, in welchem die Versuchspersonen ohne zu raten so viele Items wie möglich aufschreiben sollten (Phase 1). In der zweiten Phase sollten sie in einem Nachtest weitere Wörter angeben, an die sie entweder während der Lern- oder Abrufphase gedacht, sie aber nicht aufgeschrieben hatten, weil sie der Meinung waren, dass diese

Items nicht präsentiert worden waren. Dazu wurden ihnen ihre reproduzierten Items Liste für Liste nacheinander vorgezeigt und denen sollten sie weitere gedachte Items hinzufügen.

Es lässt sich festhalten, dass anhand von Phase 1 Aktivierung und Quellendiskrimination im Gegeneinanderspiel erhoben werden konnte, während die Summe der ersten und zweiten Phase Auskunft über die Aktivierung ohne Quellendiskrimination gab. In diesem Sinn handelt es sich hierbei um eine Inklusionsinstruktion mittels derer die Aktivierung der kritischen Items ohne entgegenwirkende Quellendiskrimination abgebildet werden soll. Die Summe der aus beiden Reproduktionstests genannten Items repräsentierte die gesamte Aktivierung, die keinen Altersunterschied zeigte. Indem die Erinnerungsleistungen der Freien Reproduktion zu jener des Nachttests in Relation gesetzt wurden, konnten Aussagen zu der Quellendiskrimination getroffen werden, welche auf alterskorrelierte Differenzen hinwiesen: Jüngere Erwachsene nannten zwar weniger kritische Items, gaben diese jedoch im Nachttest häufiger an als ältere Erwachsene, was ihre bessere Quellendiskriminationsfähigkeiten verdeutlicht.

Im Rahmen der Distinktivitätsheuristik (die Abrufstrategie soll falsche Erinnerungen reduzieren, indem die Abwesenheit von Gedächtnisinhalten als Hinweis für das Nichtauftreten eines Ereignisses gewertet wird; z.B. Dodson & Schacter, 2001; 2002; Ghetti et al., 2002; Hancock et al, 2003; Johnson et al., 1993; s. Abschnitt 2.2) untersuchten Hege und Dodson (2004) die Auswirkung einer Wort- und Bildenkodierung und einer Standard- und Inklusionsinstruktion auf falsche Erinnerungen. Sie wollten untersuchen, ob die geringen falschen Erinnerungen bei Bildmaterial darauf zurückgehen, dass dieses nur besser itemspezifisch oder auch noch weniger semantisch-relational verarbeitet wird.

Es wurde angenommen, dass die Inklusionsinstruktion Aktivierung ohne entgegenwirkende Quellendiskrimination erfasst. Die Auswahl der Instruktion richtete sich nach Brainerd und Reyna (1998b), die eine Standardinstruktion mit einer bedeutungsbasierten ("meaning recognition") Instruktion verglichen. Gemäß dieser Instruktion sollten alle Items reproduziert werden, die zuvor gelernt wurden und in Bezug zu den gelernten Listenitems standen ("...to report as many studied items as they were able but also to report any items related to the studied items that came to mind." Hege & Dodson, 2004, p. 789). Im ersten Experiment wurde in der ersten Testphase Aktivierung ohne entgegenwirkende Quellendiskrimination erhoben, denn es sollten möglichst viele der gelernten Items frei reproduziert werden und zusätzlich Items, die mit dem Inhalt der gelernten Liste in Bezug standen und an die die Probanden während der Lernphase gedacht hatten. In der nächsten Phase wurde die Quellendiskrimination getestet und die Probanden sollten ihre genannten Items dahingehend selektieren, welche

wirklich in der Lernphase vorkamen. Im zweiten Experiment, welches eine Standard- mit einer Inklusionsinstruktion verglich, wurden die Probanden entweder der Wort- oder Bildenkodierung zugewiesen und in der Testphase bearbeiteten sie entweder nach einer Standard- oder Inklusionsinstruktion eine Freie Reproduktionsaufgabe mit anschließendem Wiedererkennenstest (d.h. Beurteilung der eigenen reproduzierten Items hinsichtlich der tatsächlichen Präsentation in der Lernphase). Es zeigte sich, dass in der Bedingung der Wortenkodierung nach der Standard- und der Inklusionsinstruktion insgesamt mehr kritische Items genannt wurden als in der Bedingung der Bildenkodierung und dass auch im Nachtest die kritischen Items in der Wortbedingung eher als „alt“ bezeichnet wurden. Aus diesen Ergebnissen schlussfolgern die Autoren, dass Bildmaterial besser itemspezifisch und weniger semantisch-relational verarbeitet zu werden scheint als Wortmaterial. Da auch nach der Standardinstruktion in der Bedingung der Bildenkodierung weniger kritische Items genannt wurden, ist anzumerken, dass neben der geringeren Aktivierung auch eine bessere Quellendiskrimination zu dem Ergebnis geführt haben könnte und nicht eindeutige Schlussfolgerungen möglich sind.

Für die eigene Untersuchung ist relevant, dass die Inklusionsinstruktion sowohl nach Wort- als auch Bildenkodierung zu vermehrter Reproduktion kritischer Items führte verglichen mit der Standardinstruktion und daher geeignet erscheint, Aktivierung ohne Quellendiskrimination zu erfassen. Anhand des Nachtests konnte zudem die Quellendiskrimination separat erfasst werden. Hier wurden die genannten kritischen Items hinsichtlich des Präsentationsstatus beurteilt und es zeigte sich, dass in Folge weniger kritische Items als „präsentiert“ bezeichnet wurden.

In einer anderen Studie zu Inklusionsinstruktionen untersuchten Heit et al. (2004) die Bedeutung strategischer Prozesse im Entstehen falscher Erinnerungen in Wiedererkennenstests. Ausgehend von dem Vorgehen der „response signal technique“ (Reed, 1973), in der die Darbietungszeit von Items zu einer automatischen (unter 500ms) oder bewussten Verarbeitung führen soll, setzten Heit et al. (2004) diese Technik im DRM-Paradigma ein. Sie gingen davon aus, dass hauptsächlich automatische Prozesse zu falschen Erinnerungen führen und um dies zu prüfen, variierten sie die Darbietungsdauer der Items (kurze Zeit: vorwiegend automatische Prozesse, längere Zeit: auch strategische Prozesse) und versuchten durch verschiedene Instruktionen entweder die strategischen Prozesse zu stärken oder diese zu minimieren. Durch den Vergleich des Anteils falscher Erinnerungen könnten somit Rückschlüsse auf automatische und strategische Prozesse ermöglicht werden. Zunächst verglichen die Autoren den Anteil falscher Erinnerungen nach unterschiedlicher Darbietungsdauer der Items auf einem Monitor in der Lernphase (200, 400, 600 oder 1,100 ms) und gaben entweder eine Standard-,

Vorwarnungs- oder Inklusionsinstruktion. Nach der Inklusionsinstruktion sollten die Probanden alle Items, die entweder in der Lernphase präsentiert worden waren oder die mit den in der Lernphase präsentierten Items verbunden waren (also die kritischen Items) als „alt“ bezeichnen. Auffallend war, dass nach der Inklusionsinstruktion erwartungskonform mehr kritische Items wiedererkannt wurden, so dass geschlussfolgert wurde, dass die Inklusionsinstruktion weniger strategische (Quellendiskriminations-) Prozesse involviert. Demgegenüber wurde zwischen der Standard- und Vorwarnungsinstruktion kein Unterschied hinsichtlich der „alt“-Urteile kritischer Items gefunden. Mit zunehmender Präsentationsdauer reduzierten sich die falschen Erinnerungen nach der Standard- und Vorwarnungsinstruktion gleichermaßen, jedoch blieben sie nach der Inklusionsinstruktion konstant, was abermals die Annahme des Fehlens strategischer (Quellendiskriminations-) Prozesse stützt.

6 Ziele der eigenen Studie

In den vorangegangenen Abschnitten wurden die Grundlagen des DRM-Forschungsparadigmas gelegt, die Einflussfaktoren falscher Erinnerungen erläutert, und es wurde auf die Gedächtnisentwicklung eingegangen. Ausgehend von den theoretischen Annahmen des AMA, den Befunden zur Gedächtnisentwicklung und den bisherigen Studien zu alterskorrelierten Veränderungen falscher Erinnerungen, wurden Vorhersagen zu einem möglichen Alterstrend formuliert. Da die eigene Arbeit sich zum Ziel setzt zu erforschen ob mögliche alterskorrelierte Differenzen falscher Erinnerungen auf Unterschieden in der Quellediskrimination und/oder in Aktivierungsprozessen beruhen, interessieren Verfahren zur separaten Erfassung dieser beiden Prozesse. Einige der bisherigen Studien wurden in diesem Zusammenhang dargestellt.

Durch die eigene Studie wird erhofft neue Erkenntnisse zur Gedächtnisentwicklung zu gewinnen, aber auch zur allgemeinen Kognitionspsychologie. Der aus dieser Forschungsrichtung stammende AMA soll dahingehend geprüft werden, ob er geeignet ist entwicklungspsychologische Befunde zu erklären. Die FTT ist eine entwicklungspsychologische Theorie, aber hier von geringerer Bedeutung, da die untersuchten Variablen, Aktivierung und Quellediskrimination dem AMA entlehnt sind.

Im nun folgenden zweiten Teil der vorliegenden Arbeit werden die fünf eigenen Experimente dargestellt. Auf verschiedene Weisen sollen Aktivierung und Quellediskrimination separat getestet werden, um differenzierte Aussagen zu einem möglichen Alterstrend falscher Erinnerungen treffen zu können.

An allen Experimenten nehmen drei Altersgruppen teil: Vorschüler/Erstklässler, Dritt-/ Viertklässler und junge Erwachsene. Die Wahl dieser Altersgruppen gründet auf Befunden der Gedächtnisentwicklung (s. Kapitel 3), die eine Zunahme der Gedächtniskapazität, der -strategien und des Wissens in dieser Altersspanne annehmen. Während Vorschüler kaum über Gedächtnisstrategien verfügen sollten und ihr Wissen noch sehr gering sein sollte, wird bei Drittklässlern schon eine Verbesserung erwartet, während bei jungen Erwachsenen dieses vollständig entwickelt sein sollte.

Im ersten Experiment soll die Aktivierung der kritischen Items anhand impliziter Testverfahren geprüft werden. Diese Tests erfassen ausschließlich die Aktivierung und können auch von jüngeren Kindern problemlos bearbeitet werden.

Im darauf folgenden Experiment (2) soll die Aktivierung der kritischen Items anders getestet werden, indem in der Freien Reproduktion eine Inklusionsinstruktion erteilt wird, die den bisher eingesetzten Inklusionsinstruktionen (vgl. Abschnitt 5.4) darin ähnelt, dass alle präsentierten Items und jene Items an die während der Lernphase selbst gedacht wurden, reproduziert werden sollen. Im zweiten Experiment soll die Quellendiskrimination anhand einer Nachbefragung erhoben werden. Es lässt sich prüfen, ob Unterschiede falscher Erinnerungen auf Unterschieden in der Aktivierung der kritischen Items basieren und/oder auf Unterschieden in der Quellendiskrimination.

Das dritte Experiment gliedert sich in zwei Teilexperimente: Das Experiment 3A unterscheidet sich vorrangig vom vorherigen, indem ein anderer expliziter Test eingesetzt wird, das Wiedererkennen. In Experiment 3B wird zusätzlich eine Lerninstruktion erteilt, die die Aktivierung der kritischen Items fördern soll. Anhand eines Vergleichs der beiden Experimente lassen sich gezielt Aussagen zur möglichen Förderung der Aktivierung treffen und welche Auswirkungen diese auf falsche Erinnerungen hat. Die Quellendiskrimination wird auch hier in der Nachbefragung separat erfasst.

Im letzten Experiment (4) wird eine andere Lerninstruktion zur Förderung der Aktivierung eingesetzt und darüber hinaus soll die Quellendiskrimination in einem entsprechenden Quellendiskriminationstest differenzierter erhoben werden als in den bisherigen Nachbefragungen.

II. Eigene Untersuchungen

7 Alterskorrelierte Differenzen in impliziten Tests?

7.1 Experiment 1

Viele der vorliegenden Studien zu falschen Erinnerungen setzen explizite Testverfahren ein, also einen Freien Reproduktionstest oder einen Wiedererkennenstest. Jedoch sind die Ergebnisse ein Resultat der gegeneinander wirkenden Prozesse Aktivierung und Quellendiskrimination. Gezielte Aussagen zu jeweils einem dieser Prozesse sind anhand expliziter Verfahren nicht möglich. Im vorliegenden Experiment erhielten die Probanden zwei konzeptuelle implizite Gedächtnistests: zuerst einen bislang noch nicht eingesetzten „Ersparnistest“ und als Zweites einen Wortassoziationstest. Implizite Tests verlangen keine Quellendiskrimination und eignen sich daher zur ausschließlichen Erfassung der Aktivierung. Auch können sie problemlos schon von jüngeren Kindern bearbeitet werden. In impliziten Tests wird im Gegensatz zu den expliziten Tests keine Erinnerungsinstruktion gegeben, sondern die Probanden sollen eine Aufgabe so schnell oder so gut wie möglich erledigen (Roediger & McDermott, 1993). Aufgrund dessen besteht keine Notwendigkeit zwischen präsentierten und selbst generierten Items zu unterscheiden (Roediger & McDermott, 1993) und eine explizite Kontamination, die bewusste Abrufbemühungen einschließt, erscheint unwahrscheinlich. Es wurden konzeptuelle Tests gewählt, da die Aktivierung der Bedeutung der kritischen Items im Mittelpunkt stand.

Auf die Variation der Abfolge wurde verzichtet, da der neue, erstmals eingesetzte Ersparnistest von größerem Interesse war und deshalb immer zuerst bearbeitet wurde.

Nach dem Lernen eines Sets von DRM-Listen wurden die kritischen Items dieses Sets und die eines nicht präsentierten Sets (Lernset für die andere Hälfte der Versuchspersonen) mit einer Lerninstruktion präsentiert, wobei sich diese Wiederlernphase an der Ersparnisemethode von Ebbinghaus (1885/1964, zit. n. Roediger, 1996) orientierte. Priming wurde in der Freien Reproduktion getestet, welches dann demonstriert war, wenn kritische Items aus präsentierten Listen häufiger reproduziert wurden als kritische Items aus neuen, nicht präsentierten Listen. Damit wurde ein impliziter Effekt (Werden die „alten“ bereits gelernten kritischen Items häufiger genannt als die „neuen“ nicht gelernten?) anhand eines expliziten Tests (Freie Reproduktion) geprüft. Dieser Test dürfte sich auch der Kategorie konzeptueller Verfahren zuordnen lassen, denn es werden keine (physikalischen) Oberflächenmerkmale bearbeitet, sondern die präsentierten Wörter und deren zugrunde liegender Bedeutungsgehalt werden semantisch verarbeitet.

In Anlehnung an McDermott (1997, Experiment 2; s. Abschnitt 5.2) wurde als konzeptuell impliziter Test ein Wortassoziationstest ausgewählt, da sich in diesem signifikantes Priming zeigte. Hier wurden in der Testphase mit den kritischen Items semantisch assoziierte Cues vorgegeben, mit der Aufforderung spontan das erste Wort zu assoziieren. In der Hälfte der DRM-Listen wurden kritische Items mit präsentiert, in der anderen Hälfte nicht. Zudem gab es Kontrolllisten, die zur Erhebung der Basisrate dienten, die nicht in der Lernphase präsentiert wurden. Priming war dann demonstriert, wenn zu den mit den alten kritischen Items verbundenen Cues mehr kritische Items assoziiert wurden als zu jenen mit den neu verbundenen. Nach dem Prinzip des aufgabenangemessenen Transfers (Priming ist abhängig vom Ausmaß der Überlappung der in Lern- und Testphase geforderten Prozesse) sollte sich im Wortassoziationstest deutliches Priming zeigen, da es bei semantisch assoziierten DRM-Wortlisten um eine primär semantische Aktivierung [nach den Befunden von Hutchison und Balota (2005) ist jedoch die assoziative Aktivierung ausschlaggebend; s. Abschnitt 2.2.3] der kritischen Items geht.

Aufgrund der widersprüchlichen Befundlage eines ähnlichen Experiments mit dem Wortassoziationstest (Hüsken, 2004), war diese Replikation mit dem zusätzlichen Ersparnistest besonders interessant. Unklarheiten könnten so aufgeklärt werden und die Ergebnisse des Priming beider konzeptueller Verfahren verglichen werden.

Da es sich bei den verwendeten impliziten Tests um konzeptuelle Tests handelte, für deren Bearbeitung konzeptuelle Prozesse relevant sind, können alterskorrelierte Verbesserungen im Priming angenommen werden. Daher ist es wichtig zu prüfen, ob in allen Altersgruppen implizite Effekte unabhängig von der Listenbedingung sind (mit vs. ohne Präsentation der kritischen Items). Ein solches Ergebnis würde für eine vergleichbare Aktivierung der kritischen Items in allen Altersgruppen sprechen.

Anhand von Wiedererkennenstests nach beiden impliziten Verfahren sollten falsche Erinnerungen erfasst werden, welche durch die gegeneinander wirkenden Prozesse Aktivierung und Quellendiskrimination resultierten. Die Ergebnisse, die mit den impliziten Effekten verglichen werden können, lassen Aussagen zu eventuellen alterskorrelierten Veränderungen der Aktivierung zu.

Für das erste Experiment wurden ausgehend von theoretischen Überlegungen und empirischen Befunden folgende Hypothesen formuliert:

- (1) Priming wird sowohl für präsentierte als auch nicht präsentierte kritische Items erwartet und zwar
 - 1.1 im Ersparnistest und
 - 1.2 im Wortassoziationstest.
- (2) Da davon ausgegangen wird, dass mit zunehmendem Alter die automatische (umfangreicheres und stärker vernetztes Wissen) als auch bewusste Aktivierung (bessere semantisch-relationale Verarbeitungsstrategien) steigen, sollte sich eine alterskorrelierte Zunahme im Priming für präsentierte und nicht präsentierte kritische Items zeigen.
 - 2.1 Dies wird für den Ersparnistest erwartet und
 - 2.2 für den Wortassoziationstest.
- (3) Für das falsche Wiedererkennen nicht präsentierte kritischer Items wird keine gerichtete Annahme formuliert. Sowohl Aktivierung als auch Quellendiskrimination wirken zusammen auf das Wiedererkennen ein und es ist ungeklärt, welcher der beiden Prozesse stärker ist. Demnach ist es
 - 3.1 für das falsche Wiedererkennen nach dem Ersparnistest unklar, ob und in welche Richtung sich Altersunterschiede zeigen,
 - 3.2 ebenso für das falsche Wiedererkennen nach dem Wortassoziationstest.
- (4) Die korrekte Wiedererkennensrate für präsentierte kritische Items sollte wegen der Zunahme an Wissen und Strategien sich alterskorreliert verbessern:
 - 4.1 sowohl im Wiedererkennen nach dem Ersparnistest
 - 4.2 als auch im Wiedererkennen nach dem Wortassoziationstest.

7.1.1 Vorversuch

Anhand eines Vorversuchs mit 10 Vorschülern, die im Mittel sechs Jahre und fünf Monate alt waren, sollte die Effektivität des Wiederlernparadigmas gesichert werden. Es wurden zwei Listensets (A und B) entworfen, die aus fünf DRM-Wortlisten mit jeweils 10 Items inklusive des kritischen Items an vierter Position bestanden. Vor der Präsentation der Wortlisten erfolgte die Erläuterung der Testdurchführung: Das Experiment wurde den Kindern als Wortspiel vorgestellt, in welchem nacheinander mehrere Listen mit Wörtern vorgelesen werden würden. Es wurde um ihre Aufmerksamkeit gebeten und darum, sich die vorgelesenen Wörter zu mer-

ken. In der ersten Lernphase wurden die fünf DRM-Listen vorgelesen. Danach wurde die Wiederlernphase vorgestellt, indem den Probanden mitgeteilt wurde, dass nun noch eine Liste mit Wörtern vorgelesen werden würde und sofort danach drei Zahlen vorgesprochen werden würden, welche der Proband sofort wiedergeben sollte („Jetzt lese ich dir noch einmal Wörter vor. Du sollst mir sehr gut zuhören und dir die Wörter merken. Wenn ich mit dem Vorlesen fertig bin, sage ich dir drei Zahlen und du sollst sie so schnell wie möglich wiederholen. Danach kommen dann drei neue Zahlen, die du auch so schnell wie möglich wiederholen sollst. Und dann wieder neue Zahlen.“). Insgesamt sollte der Proband dreimal drei Zahlen nachsprechen, um eventuelle Recency-Effekte zu minimieren. In dieser Wiederlernphase wurden 10 kritische Items vorgelesen, von denen fünf aus dem präsentierten und fünf aus dem nicht präsentierten Listenset entstammten. Nachdem die Probanden die Zahlen wiederholt hatten, erfolgte eine Freie Reproduktion, in welcher sie aufgefordert wurden, möglichst viele jener Wörter zu erinnern, welche vor den Zahlen vorgelesen wurden, also in der Wiederlernphase. Abschließend wurde ein Wiedererkennenstest eingesetzt, in welchem 30 Items dargeboten (10 kritische Items, die zur einen Hälfte alt und zur anderen neu waren und jeweils zwei Listenitems) und ein „alt/neu“-Urteil verlangt wurde.

Anhand der Ergebnisse im Test der Freien Reproduktion zeigte sich, dass die präsentierten kritischen Items häufiger genannt wurden (44%) als die nicht präsentierten (12%). Auffällig war aber, dass einige Listenitems genannt wurden (4,2%), so dass davon ausgegangen wurde, dass die hier applizierte Wiederlernphase nicht eindeutig und distinktiv war. Offensichtlich wurden die vorgelesenen Wortlisten mit den in der Wiederlernphase dargebotenen kritischen Items verwechselt. Aus diesem Grund wurde für den Hauptversuch eine Modifikation dahingehend vorgenommen, dass in der Wiederlernphase die kritischen Items von einem männlichen Sprecher auf Band gesprochen dargeboten wurden. Damit sollte die Distinktivität der Phasen gesichert sein.

Im Hauptversuch sollte zudem ein Wortassoziationstest eingesetzt werden. Um sichergestellt zu haben, dass die Cues in einem sinnvollen begrifflichen Zusammenhang zu den Items der Lernphase stehen und dass die Basisrate nicht zu hoch ist, wurden die Cues aus eigenen (Hüsken, 2004) und universitätsinternen Voruntersuchungen dahingehend ausgewählt, dass die Häufigkeit der Assoziation der kritischen Items zwischen 20% und 30% betrug.

7.1.2 Methode

Versuchsplan. Dem ersten Experiment lag für die impliziten Tests ein 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Listenbedingung: mit vs. ohne kritische Items) x 2 (Itemart: alte vs. neue kritische Items) Versuchsplan mit Messwiederholung auf dem letzten Faktor zugrunde. Nach diesem Untersuchungsdesign resultierten sechs Gruppen. Pro Altersgruppe wurden jeweils 16 Probanden den Untersuchungsgruppen per Zufall zugeteilt.

Versuchspersonen. Insgesamt nahmen 96 Versuchspersonen an Experiment 1 teil, von denen 32 Vorschüler eines Kindergartens, 32 Drittklässler einer Grundschule und 32 Zwölftklässler eines Gymnasiums waren.

Für alle der folgenden Experimente gilt, dass die an der Untersuchung partizipierenden Kindergarten- und Grundschul Kinder eine schriftliche Einverständniserklärung der Eltern vorzuweisen hatten, welche in einem Informationsschreiben über die Studie zwei Wochen vor Testbeginn den Eltern ausgehändigt worden war. Der Eltern- und Schülerinformationsbrief befindet sich im Anhang E.

Die Vorschüler, von denen 16 Mädchen und 16 Jungen waren, hatten ein durchschnittliches Alter von fünf Jahren und neun Monaten (zwischen 5;4 und 6;11 Jahren), die Drittklässler, 19 Mädchen und 13 Jungen, waren im Mittel acht Jahre und acht Monate alt (zwischen 8;3 und 10;2 Jahren) und für die 32 Zwölftklässlerinnen belief sich das Durchschnittsalter auf 17 Jahre und sieben Monate (zwischen 16;11 und 19;4 Jahren).

Material. Die in der Lernphase präsentierten DRM-Wortlisten stellen den Ausgangspunkt der Untersuchung dar und wurden sorgfältig zusammengestellt. Bei der Auswahl der kritischen Items und Listenitems musste beachtet werden, dass diese Wörter zum einen sehr konkret (wie z.B. „Hase“) und zum anderen nicht zusammengesetzt waren und aus wenigen Silben bestanden, um die semantische Aktivierung der kritischen Items zu gewährleisten. Die verwendeten Begriffe sollten dem kindlichen Wortschatz entsprechen, aus ihrer Erfahrungswelt stammen und sowohl aktiv produzierbar als auch passiv verstehbar sein. Des Weiteren sollten die Begriffe einmalig in den Listen vorzufinden sein, sich nicht überschneiden oder ähneln, damit eine zuverlässige Trennung richtiger und falscher Erinnerungen in der Testphase gewährleistet werden konnte.

Die von Hasselhorn und Grube (1994) ermittelten Assoziationsnormen stellen die Grundlage dar, aus denen die Listen erstellt wurden. Weitere Listen stammen aus universitätsinternen Untersuchungen. Die Zusammenstellung der Listen richtete sich nach der mittleren Vorwärtsassoziationsstärke, von welcher ausgehend die Listenitems in absteigender Reihenfolge ange-

ordnet wurden, d.h. die ersten Items hatten die höchste Assoziationsstärke zum kritischen Item, die sich kontinuierlich für die folgenden Items reduzierte. Es wurden für alle Altersgruppen vier Listensets (A, B, C, D) entwickelt mit jeweils fünf Wortlisten à 10 Items. Die Liste A oder B wurde in der ersten Lernphase präsentiert, die Liste C oder D in der zweiten Lernphase. Während die Hälfte der Probanden die Listenbedingung mit kritischen Items an vierter Position in der Wortliste erhielt, wurden der anderen Hälfte der Probanden nicht die kritischen Items in der Lernphase präsentiert. Somit resultierten vier Listenbedingungen [(1) A-C mit kritischem Item, (2) B-D mit kritischem Item, (3) A-C ohne kritisches Item, (4) B-D ohne kritisches Item], zu denen die Probanden per Zufall zugeteilt wurden.

Nach der ersten Lernphase erfolgte die Wiederlernphase. Hier wurden zuerst zwei unverbundene Füllitems und dann alle 10 kritischen Items der Listensets A und B auf Tonband dargeboten. Fünf der Items waren alte kritische Items, d.h. die zugehörigen Listen (mit oder ohne kritisches Item) waren in der Lernphase bereits dargeboten worden und fünf Items waren neu, d.h. die Listen waren nicht präsentiert worden. Um Quellenkonfusionen zu vermeiden, las ein männlicher Sprecher im Abstand von 1,5 Sekunden zuerst die beiden Füllitems und die kritischen Items in quasi-randomisierter Folge vor, nach der die Reihenfolge der kritischen Items zufällig war aber darauf geachtet wurde, dass nie mehr als zwei kritische Items einer Listenbedingung aufeinander folgten und die inhaltlichen Kategorien (z.B. Tiere) auch gleichmäßig verteilt waren. Anschließend erfolgte der Test der Freien Reproduktion und der erste Wiedererkennenstest, der aus 30 Items bestand, von denen 10 kritische Items (aus Listenset A und B) und 20 Listenitems (zwei aus jeder Wortliste der Sets A und B) waren, welche in quasi-randomisierter Reihenfolge angeordnet wurden. Alle Items wurden zufällig angeordnet, nur sollten die unmittelbar aufeinander folgenden Items nicht aus derselben Liste sein und damit eine Erinnerungshilfe bieten. Der Wiedererkennenstest bestand für jeden Probanden aus Listenitems, die präsentiert worden waren, Listenitems des anderen Sets, die nicht präsentiert worden, also neu waren und als Distraktoren der Listenitems, kurz Listendistraktoren, bezeichnet werden. Des Weiteren wurden die präsentierten bzw. nicht präsentierten kritischen Items der Liste in den Wiedererkennenstest eingefügt und kritische Items, die nicht in der Lernphase vorkamen, nicht mit den Listen assoziiert waren und im Folgenden kritische Distraktoren genannt werden. Anschließend erfolgte eine nonverbale Ablenkaufgabe.

Nach der Präsentation des Listensets der zweiten Lernphase (C oder D) wurden im Wortassoziationstest 10 Cues dargeboten, die eine geringe Assoziationsstärke zu den kritischen Items der beiden Listensets aufwiesen (s. Abschnitt 7.1.1). Auch diese Cues wurden in eine quasi-randomisierte Reihenfolge gebracht. Der abschließende zweite Wiedererkennenstest ähnelte

dem ersten in der Zusammensetzung der 30 Items. Das vollständige Material der Untersuchung ist dem Anhang A1 zu entnehmen.

Auf die Variation der Abfolge der beiden impliziten Tests wurde verzichtet, da der Ersparnistest wichtiger war.

Durchführung. Alle Experimente der vorliegenden Arbeit wurden als Einzelversuche durchgeführt, die in einem separaten, ruhigen Raum des Kindergartens bzw. der Schulen während der regulären Kindergarten- bzw. Schulunterrichtszeit am Vormittag stattfanden.

An dem ersten Experiment nahmen ein Kindergarten, drei Klassen eines dritten Schuljahres und eine 12. Klasse eines Gymnasiums teil.

Durch die vielen verschiedenen Testphasen mit Ablenktätigkeiten benötigte dieser Versuch ungefähr 20 bis 30 Minuten. Die Versuchsdurchführung gestaltete sich in 11 aufeinander folgenden Phasen [Lernphase (1), Wiederlernphase, Ablenkaufgabe (1), Freie Reproduktion, Ablenkaufgabe (2), Wiedererkennenstest (1), Ablenkaufgabe (3), Lernphase (2), Wortassoziationstest, Ablenkaufgabe (4), Wiedererkennenstest (2)]. Vor der Präsentation der Wortlisten erfolgte die Erläuterung der Testdurchführung: Den Kindern stellte man das Experiment als Wortspiel vor, den Erwachsenen als Gedächtnisexperiment, wobei um ihre Aufmerksamkeit gebeten wurde, wenn ihnen nun Wörter vorgelesen werden würden. In der ersten Lernphase wurden fünf Wortlisten mit 10 bzw. 11 Items nacheinander vorgelesen, wobei die Präsentationsrate 1,5 Sekunden betrug, was auch für die folgenden Experimente galt. Die Instruktionen im Wortlaut befinden sich im Anhang A2. Anschließend wurde den Probanden in der Wiederlernphase mitgeteilt, dass ihnen auf einem Tonbandgerät eine Kassette mit Wörtern vorgespielt werden würde, die sie sich merken sollten. Sofort nachdem die Wörter auf Tonband präsentiert worden waren, erfolgte die erste Ablenkaufgabe von 30 Sekunden, in welcher die Probanden aufgefordert wurden, Zahlen zu wiederholen. Zur Vermeidung des Recency-Effekts sollten Vorschüler drei aufeinander folgende Zahlen wiedergeben, Drittklässler vier und Erwachsene fünf Zahlen. Danach sollten die Probanden in der Freien Reproduktion möglichst viele Wörter erinnern, die sie auf dem Tonband gehört hatten. In einer kurzen Ablenkaufgabe von 30 Sekunden sollten die Probanden erneut Zahlen nachsprechen. Dann wurde eine Standardinstruktion zum ersten Wiedererkennenstest gegeben, in welcher den Probanden mitgeteilt wurde ein „alt/neu“-Urteil zu den am Anfang vorgelesenen Wörtern zu geben. Explizit wurde darauf hingewiesen, dass es nicht um die Beurteilung der auf Tonband gesprochenen Wörter ging. Nach dem Wiedererkennenstest erfolgte eine zweiminütige Ablenkaufgabe, in welcher mit den beiden jüngeren Altersgruppen ein Mikadospiel und mit den Er-

wachsenen einige Aufgaben der Standardform des Progressiven Matrizen-Tests (Raven, 1976) durchgeführt wurden. Nun wurde die zweite Lernphase vorgestellt, mit der gleichen Instruktion wie die erste und der Präsentation von fünf Wortlisten à 10 bzw. 11 Items. Im folgenden Wortassoziationstest wurden die Probanden instruiert zu den Cues möglichst schnell eine Assoziation zu nennen. Wichtig war sie darauf hinzuweisen, dass es keine falschen Antworten gab, sondern dass jede Antwort passend war. Anhand zweier Beispiele wurde ihnen das Vorgehen verdeutlicht („Was fällt dir zu *Sport/Fluss* ein?“). Vor dem abschließenden Wiedererkennenstest sollten die Probanden im Rahmen einer Ablenktätigkeit noch 30 Sekunden Zahlen wiederholen. Im zweiten Wiedererkennenstest wurde von den Probanden ein „alt/neu“-Urteil zu den Items verlangt, wobei darauf hingewiesen wurde, dass es um die Wörter vor dem Assoziationstest, also jene der zweiten Lernphase, ging. Alle Instruktionen im Wortlaut befinden sich im Anhang A2.

7.1.3 Ergebnisse

Folgendes gilt für alle Experimente der vorliegenden Arbeit. Für alle inferenzstatistischen Berechnungen wurde a priori das kritische Fehlerniveau von α auf $p < .05$ festgelegt, weshalb im Folgenden keine p -Werte berichtet werden, mit Ausnahme von tendenziellen Effekten ($p < .10$). Auf Tendenzen von Kontrollfaktoren wird nicht eingegangen. Als a posteriori Tests fungierten Tukey HSD Tests.

Als Maß der Effektstärke wurden partielle multiple Korrelationsquadrate (partielle R^2 -Werte³) berechnet und für die signifikanten Haupt- und Interaktionseffekte angegeben. Der aufgeklärte Varianzanteil der abhängigen Variablen wird auf diese Weise veranschaulicht.

Zudem wurden einseitige t -Tests eingesetzt, um Mittelwerte zu vergleichen und gegen den Zufall zu prüfen.

Die berichteten Mittelwerte im Text und in den Tabellen werden stets in relativen Häufigkeiten angegeben. Werden in Ausnahmefällen andere Werte berechnet, so wird dies deutlich gemacht.

Die Ergebnisse des vorliegenden Experiments werden nach den verschiedenen Phasen der Testdurchführung dargestellt. Zuerst wird auf die impliziten Effekte im Ersparnis- und im Wortassoziationstest eingegangen, anschließend werden die Resultate der beiden expliziten Wiedererkennenstests erläutert.

³ Nach Cohen (1977) gilt: $R^2 = F \times df_{\text{Zähler}} / F \times df_{\text{Zähler}} + df_{\text{Nenner}}$

7.1.3.1 Implizite Tests

Um einen Überblick über die wichtigsten Ergebnisse der impliziten Tests zu geben, sind die entsprechenden Mittelwerte in Abhängigkeit von Altersgruppe und Listenbedingung in Tabelle 7.1.1 zusammengefasst, getrennt nach den beiden impliziten Testverfahren für die kritischen Items, die Basisrate (neue kritische Items) und das Priming. In der Tabelle sind die absoluten Primingwerte („alt minus neu“) angegeben. Nähere Erläuterungen befinden sich im Text.

Tabelle 7.1.1:

Mittelwerte der kritischen Items, der Basisrate und des Priming im Ersparnis- und Wortassoziationstest in Abhängigkeit des Alters und der Listenbedingung (mit und ohne präsenzierte kritische Items).

	Kritische Items (alte)		Basisrate (neue kritische Items)		Priming	
	mit	ohne	mit	ohne	mit	ohne
Ersparnistest						
Vorschüler	.36	.33	.20	.10	.26	.13
Drittklässler	.44	.45	.19	.26	.18	.26
Erwachsene	.64	.56	.21	.24	.40	.35
Wortassoziationstest						
Vorschüler	.34	.31	.13	.06	.21	.25
Drittklässler	.41	.39	.13	.10	.28	.29
Erwachsene	.43	.50	.20	.24	.23	.26

Ersparnistest. Um zu testen, ob es Unterschiede in der Basisrate der kritischen Items gab, wodurch es unklar wäre, was das beste Maß für das Priming ist, wurde eine 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Listenbedingung: mit vs. ohne kritische Items) VA durchgeführt, die ausschließlich jene kritischen Items einschloss, deren Listen in der Lernphase nicht präsentiert wurden und damit für die Probanden neu waren. Im Ersparnistest zeigte sich für die Altersgruppe ein tendenzieller Haupteffekt [$F(2,90) = 2.42$, $R^2 = 0.05$, $p = .09$], welcher jedoch durch eine ebenso tendenzielle Interaktion mit der Listenbedingung überlagert wurde [$F(2,90) = 2.62$, $R^2 = 0.06$, $p = .08$]. Eine Analyse der einfachen Haupteffekte ergab, dass in der Listenbedingung „mit kritischen Items“ alle Altersgruppen eine ähnliche Anzahl neuer kritischer Items nannten ($F < 1$) und damit kein Unterschied in der Basisrate vorzufinden war (zu den Mittelwerten s. Tabelle 7.1.1). In der Listenbedingung „ohne kritische Items“ gab es jedoch Altersunterschiede [$F(2,90) = 4.94$]: Vorschüler nannten weniger neue kritische Items als Drittklässler und Erwachsene (Tukey-Anschlusstests). Wurden die Listenbedingungen innerhalb der Altersgruppen verglichen, dann zeigte sich, dass Vorschüler ohne präsentiertes kritisches Item tendenziell weniger neue kritische Items nannten als mit

präsentierten kritischen Items [$F(1,90) = 3.23, p = .08$; Mittelwerte s. Tabelle 7.1.1]. Für die beiden älteren Gruppen ergab sich keine signifikante Differenz der beiden Listenbedingungen (beide $Fs < 1$).

Anhand einseitiger t -Tests wurden die absoluten Primingwerte aller Altersgruppen überprüft, ob sie signifikant größer als null waren, was zutreffend war [alle $t(15) \geq 2.08$]. Die zugehörigen Werte gehen aus Tabelle 7.1.1 hervor. Damit zeigte sich erwartungsgemäß (Hypothese 1.1) reliables Priming sowohl im Fall der Präsentation der kritischen Items als auch für die nicht in den Listen präsentierten kritischen Items.

Die Ergebnisse der 3×2 VA mit der abhängigen Variablen Priming (Mittelwerte zum Priming für alle Altersgruppen s. Tabelle 7.1.1) zeigten, dass das Priming ohne ($M = .25$) und mit präsentierten kritischen Items ($M = .28$) vergleichbar war ($F < 1$). Die Interaktion mit dem Alter war nicht signifikant [$F(2,90) = 1.71$], so dass keine Altersunterschiede in der Aktivierung vorzuliegen schienen. Ein Haupteffekt ergab sich für die Altersgruppe [$F(2,90) = 5.13, R^2 = 0.10$], wonach sich das Priming von den Kindern (Vorschüler: $M = .19$; Drittklässler: $M = .22$) zu den Erwachsenen ($M = .38$) erhöhte. Nach diesen Ergebnissen gab es also im Einklang mit Hypothese 2.1 alterskorrelierte Unterschiede im Priming im Ersparnistest, aber die Aktivierung der kritischen Items gemessen am Vergleich mit der Bedingung, in der die kritischen Items mit präsentiert wurden, erwies sich als altersinvariant. Von den Mittelwerten her zeigten Vorschüler am wenigsten Priming für die nicht präsentierten kritischen Items, weniger als für die präsentierten und weniger als die anderen Altersgruppen, aber die Unterschiede waren nicht signifikant ($F < 1$), was auch an der Basisrate lag. Drittklässler hatten jedoch kein höheres Priming für präsentierte als für nicht präsentierte kritische Items ($F < 1$). Bei Erwachsenen war das Priming für die präsentierten und die nicht präsentierten kritischen Items vergleichbar ($F < 1$).

Wortassoziationstest. In der VA mit der Basisrate zeigte sich ein Haupteffekt der Altersgruppe [$F(2,90) = 6.37, R^2 = 0.12$]: Erwachsene ($M = .22$) nannten mehr neue kritische Items als die jüngeren Probanden (Vorschüler: $M = .09$; Drittklässler: $M = .11$). Die anderen Werte waren statistisch unbedeutsam (alle $Fs < 1$).

Die vorgefundenen Unterschiede in der Basisrate führten wie im Ersparnistest zu der Entscheidung absolute Primingwerte zu bestimmen. Die t -Tests zeigten, dass auch hier erwartungsgemäß (Hypothese 1.2) signifikantes Priming stattfand, denn alle t -Werte waren größer als null [alle $t(15) \geq 4.58$]. Das Priming im Wortassoziationstest war etwa gleich hoch wie das der Ersparnistestmethode.

Weiterführend wurde eine 3 x 2 VA mit dem Priming gerechnet, welche keine signifikanten Effekte erbrachte (alle F s < 1). Folglich war auch hier die Aktivierung mit präsentierten kritischen Items vergleichbar mit jener der nicht präsentierten (Mittelwerte s. Tabelle 7.1.1). Entgegen der Annahme aus Hypothese 2.2 gab es keine alterskorrelierten Unterschiede in der Aktivierung. Während sich im Ersparnistest wenigstens Unterschiede in den Mittelwerten zeigten, traf dies auf den Wortassoziationstest nicht zu.

7.1.3.2 Explizite Tests: Wiedererkennen

Im Wiedererkennenstest werden die Antworten der Probanden in zwei Kategorien eingeteilt: Es handelt sich um einen „Treffer“ („hit rate“), wenn ein Item, welches in der Lernphase präsentiert wurde, korrekt als „alt“ erkannt wird. „Falsche Alarme“ („false alarm“) beziehen sich auf die Einstufung eines neuen Items als „alt“.

Ausgehend von dieser Klassifizierung entwickelten Snodgrass und Corwin (1988) ein Zweiswellenmodell („two-high-threshold model“), nach welchem es zwei Gedächtnisschwellen geben soll. Eine Schwelle bezieht sich auf alte und eine auf neue Items, wodurch sich drei Gedächtniszustände ergeben: altes Wiedererkennen, neues Wiedererkennen und Unsicherheit. Wenn alte Items die Schwelle zum alten Wiedererkennen übersteigen, werden sie immer als alt bezeichnet (P_0). Ein Item wird dann als neu erkannt, wenn es die Neu-Schwelle (P_n) überschreiten konnte. Es muss darüber hinaus beachtet werden, dass die Schwelle, mit der ein Item als „alt“ eingestuft wird, nur von alten Items und die Schwelle mit der ein Item als „neu“ bezeichnet wird, ausschließlich von neuen Items passiert werden kann. Nach diesem Modell resultieren falsche Alarme aus der Unsicherheit in der Klassifizierung eines Items.

Nach dem Zweiswellenmodell gibt es keine unabhängigen Werte für die Schwellen der „Treffer“- und „falsche Alarmrate“, sondern sie werden anhand der gemeinsamen Schwelle P definiert (Snodgrass & Corwin, 1988). Diese Vorgehensweise wird anhand von Untersuchungen anderer Autoren bestätigt (Glanzer & Adams, 1985), die durch experimentelle Manipulation die „Trefferrate“ erhöhten und die Rate der „falschen Alarme“ reduzierten. Das Zweiswellenmodell wird der Signalentdeckungstheorie bevorzugt, da Untersuchungen zeigten, dass es sensitiver gegenüber Veränderungen im Antwortkriterium und im Diskriminationsmaß ist (Snodgrass & Corwin, 1988).

Die „Trefferrate“ setzt sich aus richtigem Wiedererkennen und aus zufälligem Raten im Fall von Unsicherheit zusammen. Im Gegensatz dazu resultiert die Rate der „falschen Alarme“

ausschließlich aus dem Zustand der Unsicherheit. Damit verhält sich die „Trefferrate“ wie folgt zum Diskriminationsindex Pr :

$$H = Pr + FA$$

Durch Umformung erhält man einen Schätzer des Index Pr:

$$Pr = H - FA$$

Demnach entspricht dieses Sensitivitätsmaß Pr im Zweischwellenmodell den „Treffern“, reduziert um die „falschen Alarme“.

Das Antwortkriterium (Br) ist eine weitere wichtige Komponente im Zweischwellenmodell (Snodgrass & Corwin, 1988) welche die Wahrscheinlichkeit beschreibt, mit der ein Item als alt bezeichnet wird, wenn man sich in einem unsicheren Zustand befindet. Es kommt zu falschen Alarmen, wenn neue Items nicht als solche erkannt werden (mit einer Wahrscheinlichkeit von $1 - Pr$) oder wenn der Proband, aus einem Unsicherheitszustand heraus, die Antwort mit einem „alt“-Urteil errät (mit einer Wahrscheinlichkeit von Br).

Die Beziehung der „falschen Alarmrate“ zum Antwortkriterium Br gestaltet sich wie folgt:

$$FA = (1 - Pr) Br$$

Durch Umformung und Ersetzen des Terminus Pr durch $H - FA$, entsteht folgende Formel:

$$Br = FA / [1 - (H - FA)]$$

Ein neutrales Antwortkriterium liegt vor, wenn der Wert von Br .50 entspricht. Befindet sich Br oberhalb dieser Grenze, handelt es sich um ein liberales Antwortkriterium. Unterschreitet der Wert .50 liegt ein konservatives Kriterium vor.

Anhand zahlreicher Studien konnten Snodgrass und Corwin (1988) die Gültigkeit ihres Modells belegen. Es konnte gezeigt werden, dass Diskriminationsmaß und Antwortkriterium unabhängig voneinander sind und darüber hinaus, dass die Maße des Zweischwellenmodells gegenüber Veränderungen sensitiver sind als andere Modelle.

Aus diesen Gründen orientiert sich die eigene Untersuchung in der Auswertung des Wiedererkennenstests an diesem Modell.

Zusätzlich wurden folgende Formeln zur Berechnung der Parameter wieder erkannter Listenitems und kritischer Items hinsichtlich ihrer jeweiligen Pr- und Br- Werte verwendet:

$$Pr_{\text{Listenitems}} = H - FA_{\text{Listendistraktoren}}$$

$$Br_{\text{Listenitems}} = FA / [1 - (H - FA)]$$

$$Pr_{\text{kritische Items}} = FA_{\text{kritische Items}} - FA_{\text{kritische Distraktoren}}$$

$$Br_{\text{kritische Items}} = FA_{\text{kritische Items}} / [1 - (FA_{\text{kritische Items}} - FA_{\text{Distraktoren}})]$$

Die Befunde lassen sich entsprechend dieser Kategorien dahingehend interpretieren, dass die Sensitivität für das korrekte Wiedererkennen von Listenitems eine Schätzung der itemspezifischen Erinnerung darstellt, die sich aus Treffern für Listenitems reduziert um die falschen Alarme für nicht präsentierte Listendistraktoren (d.h. die Listenitems der nicht präsentierten und damit nicht gelernten Liste) zusammensetzt. Demgegenüber ist die Sensitivität für das falsche Wiedererkennen kritischer Items ein Maß für die Empfänglichkeit für falsche Erinnerungen. Im DRM-Paradigma können in diesem Sinne kritische Items, die als „alt“ erkannt werden, der „Trefferrate“ zugeordnet werden, da die Wortlisten darauf abzielen, das nicht präsentierte aber assoziierte kritische Item zu aktivieren (Seamon et al., 2003). Demnach resultiert das Maß des falschen Wiedererkennens aus den „Treffern“ kritischer Items reduziert um die Rate falscher Alarme zu den nicht präsentierten kritischen Distraktoren (d.h. den kritischen Items der nicht präsentierten und damit nicht gelernten Liste).

Im Folgenden sowie bei der Darstellung aller Ergebnisse der Wiedererkennenstests in den eigenen Experimenten werden immer zwei Maße für korrektes und falsches Erinnern berechnet: Für die korrekten Erinnerungen von Listenitems sind es die Trefferrate und die korrigierten Werte (Pr), für falsche Erinnerungen der kritischen Items die falsche Alarmrate und die korrigierten Werte (Pr). Wenn die Ergebnisse identisch sind, dann werden nur die Pr-Werte dargestellt.

In diesem Abschnitt werden die Ergebnisse für beide Wiedererkennenstests berichtet. Im Einzelnen werden die Resultate der Listenitems, der kritischen Distraktoren, der kritischen Items, der Pr- und Br-Werte aufgeführt. Die Resultate der Listenitems sind in Tabelle 7.1.2 zusammengefasst und jene der kritischen Items in Tabelle 7.1.3.

Tabelle 7.1.2:

Mittelwerte der beiden Wiedererkennentests für die Trefferrate und die Pr-Werte der Listen- items in Abhängigkeit des Alters und der Listenbedingung (mit und ohne präsentierte kritische Items).

	Trefferrate		Pr-Werte	
	mit	ohne	mit	ohne
Wiedererkennen nach dem Ersparnistest				
Vorschüler	.78	.73	.61	.64
Drittklässler	.74	.74	.63	.59
Erwachsene	.79	.80	.76	.76
Wiedererkennen nach dem Wortassoziationstest				
Vorschüler	.68	.67	.48	.51
Drittklässler	.68	.69	.56	.51
Erwachsene	.79	.79	.68	.75

Tabelle 7.1.3:

Mittelwerte der beiden Wiedererkennentests (falsche Alarmrate) für die kritischen Distraktoren, kritischen Items, Pr- und Br-Werte der kritischen Items in Abhängigkeit des Alters und der Listenbedingung (mit und ohne präsentierte kritische Items).

	Kritische Distraktoren		Kritische Items		Pr-Werte		Br-Werte	
	mit	ohne	mit	ohne	mit	ohne	mit	ohne
Wiedererkennen nach dem Ersparnistest								
Vorschüler	.29	.18	.79	.46	.50	.29	.62	.25
Drittklässler	.44	.34	.86	.66	.43	.33	.73	.46
Erwachsene	.29	.16	.93	.71	.64	.55	.84	.32
Wiedererkennen nach dem Wortassoziationstest								
Vorschüler	.20	.10	.73	.38	.53	.28	.41	.11
Drittklässler	.20	.19	.75	.51	.55	.33	.40	.27
Erwachsene	.09	.23	.89	.69	.80	.46	.55	.41

Wiedererkennen nach dem Ersparnistest

Listenitems. Die 3 x 2 VA für die Trefferrate der Listenitems erbrachte im Wiedererkennentest nach dem Ersparnistest keine signifikanten Ergebnisse (alle F s < 1.16, zu den Mittelwerten s. Tabelle 7.1.2). Somit unterschieden sich die Altersgruppen nicht in der Trefferrate der Listenitems.

Die VA mit den Pr-Werten zeigte einen Alterseffekt [$F(2,90) = 8.88, R^2 = 0.16$]. Vorschüler ($M = .63$) und Drittklässler ($M = .61$) unterschieden sich nicht, aber sie hatten eine geringere Rate als die Erwachsenen ($M = .76$). Die unterschiedlichen Ergebnisse gehen darauf zurück, dass die beiden jüngeren Gruppen viele Listendistraktoren falsch wiedererkannten, Erwachse-

ne demgegenüber nicht. Das Antwortkriterium für die Listenitems unterschied sich dementsprechend in Abhängigkeit des Alters [$F(2,85) = 3.64, R^2 = 0.08$]. Vorschüler ($M = .33$) und Drittklässler ($M = .34$) unterschieden sich nicht in ihrem konservativen Kriterium, aber das Kriterium Erwachsener war sehr konservativ ($M = .15$). Die anderen Werte der VA erlangten keine Signifikanz (alle F s < 1.57).

Kritische Items. Die VA für das Antwortkriterium der kritischen Items (Mittelwerte s. Tabelle 7.1.3) erbrachte einen Effekt der Listenbedingung [$F(1,84) = 36.00, R^2 = 0.30$]: Wurden in den Listen die kritischen Items mit präsentiert ($M = .72$), so war das Antwortkriterium liberaler als wenn zuvor die Listen ohne kritische Items präsentiert worden waren ($M = .34$). Zudem ergab sich ein tendenzieller Alterseffekt [$F(2,84) = 2.61, R^2 = 0.06, p = .08$]: Drittklässler und Erwachsene unterschieden sich nicht, urteilten aber beide tendenziell weniger konservativ als Vorschüler. Die Interaktion der Faktoren war nicht bedeutsam ($F < 1.15$).

Wurden die kritischen Distraktoren analysiert, dann ergab sich ein Effekt der Listenbedingung [$F(1,90) = 5.70, R^2 = 0.06$], nach welchem die kritischen Distraktoren eher in der Bedingung mit kritischem Item ($M = .34$) falsch wiedererkannt wurden als ohne kritisches Item ($M = .23$). Die einzelnen Mittelwerte sind Tabelle 7.1.3 zu entnehmen. Zudem zeigte sich ein Alterseffekt [$F(2,90) = 5.09, R^2 = 0.10$], denn Vorschüler ($M = .23$) und Erwachsene ($M = .23$) bezeichneten die kritischen Distraktoren seltener als „alt“ verglichen mit Drittklässlern ($M = .39$). Die Interaktion war statistisch nicht bedeutsam ($F < 1$).

Die Ergebnisse der VA für die „Trefferrate“ und die Pr-Werte sind vergleichbar, weshalb nur letztere berichtet werden (zu den Mittelwerten s. Tabelle 7.1.3). In dieser VA erlangten beide Haupteffekte Signifikanz, die Listenbedingung [$F(1,90) = 6.45, R^2 = 0.07$] und die Altersgruppe [$F(2,90) = 7.11, R^2 = 0.14$], nicht aber deren Interaktion ($F < 1$). Mit präsentem kritischem Item ($M = .52$) konnten erwartungsgemäß mehr kritische Items korrekt wiedererkannt werden als falsch ohne kritisches Item ($M = .39$). Es zeigten sich parallele Alterseffekte bei präsentierten und nicht präsentierten kritischen Items, denn Erwachsene ($M = .59$) erkannten stets mehr kritische Items wieder als Kinder (Vorschüler: $M = .39$; Drittklässler: $M = .38$). Es wurden keine gerichteten Annahmen zum Wiedererkennen der nicht präsentierten kritischen Items formuliert und die Alterseffekte zeigten, dass Erwachsene sich von Kindern unterschieden, indem sie mehr kritische Items falsch wiedererkannten (Hypothese 3.1).

Die Ergebnisse der höheren Pr-Werte Erwachsener sprechen teilweise für Hypothese 4.1 nach der ein Altersunterschied im Wiedererkennen der präsentierten kritischen Items erwartet wur-

de. Erwachsene waren den Kindern im korrekten Wiedererkennen der präsentierten kritischen Items überlegen. Drittklässler unterschieden sich aber nicht von Vorschülern.

Wiedererkennen nach der Wortassoziation

Listenitems. Die VA für die Trefferrate führte zu ähnlichen Ergebnissen wie die VA für die Pr-Werte, weshalb nur letztere erläutert werden (Mittelwerte s. Tabelle 7.1.2). Die Altersgruppe zeigte signifikante Effekte [$F(2,90) = 12.50, R^2 = 0.22$]. Wie im Wiedererkennen nach dem Ersparnistest differierten die beiden jüngeren Gruppen nicht voneinander (Vorschüler: $M = .49$; Drittklässler: $M = .53$), aber sie unterschieden sich von Erwachsenen ($M = .72$), die höhere Pr-Werte hatten. Das Antwortkriterium für die Listenitems zeigte nach der Wortassoziation [$F(2,90) = 2.46, R^2 = 0.05$] nur einen tendenziellen Altersunterschied, Erwachsene ($M = .22$) differierten nur tendenziell von Vorschülern ($M = .38$) in der Wahl eines konservativeren Kriteriums. Drittklässler ($M = .30$) unterschieden sich weder von Erwachsenen noch von Vorschülern.

Kritische Items. Die Analyse des Antwortkriteriums im Wiedererkennen nach der Wortassoziation erbrachte ähnliche Ergebnisse wie im Wiedererkennen nach dem Ersparnistest: Es zeigte sich ein Effekt der Listenbedingung [$F(1,80) = 7.73, R^2 = 0.09$]: Listen, in denen die kritischen Items mit präsentiert wurden ($M = .44$), wurden weniger konservativ beurteilt als Listen, in denen die kritischen Items zuvor nicht mit präsentiert wurden ($M = .26$). Der Alterseffekt [$F(2,80) = 3.38, R^2 = 0.08$] zeigte, dass Erwachsene ($M = .46$) sich von Vorschülern ($M = .26$) in der Wahl eines neutralen Kriteriums unterschieden, während Drittklässler ($M = .34$) eine mittlere Position wählten und sich von keiner der beiden Gruppen unterschieden. Die Interaktion war statistisch unbedeutsam ($F < 1$).

Bei den kritischen Distraktoren erlangten die Haupteffekte keine Signifikanz (alle F s < 1), aber die Interaktion der Altersgruppe mit der Listenbedingung [$F(2,90) = 3.90, R^2 = 0.08$; zu den Mittelwerten s. Tabelle 7.1.3]. In der Analyse der einfachen Haupteffekte zeigte sich, dass sich die Altersgruppen sowohl in den Bedingungen mit [$F(2,90) = 2.22$] als auch ohne präsentierte kritische Items [$F(2,90) = 2.28$] nicht bedeutsam unterschieden. Indem die Listenbedingungen miteinander verglichen wurden ergab sich ein bedeutsamer Effekt für Erwachsene: Wurden die kritischen Items mit präsentiert, dann erkannten sie mehr kritische Distraktoren wieder als bei den Listen ohne kritische Items [$F(1,90) = 5.11$]. Die beiden jüngeren Gruppen hatten keine Unterschiede im Wiedererkennen der kritischen Distraktoren in Abhängigkeit der Listenbedingung [beide $F(1,90) < 2.70$].

Zur Prüfung der Hypothesen für die kritischen Items wurden entsprechende VAn durchgeführt. Die Ergebnisse für die „Trefferrate“ als auch für die korrigierten Werte (Pr-Werte) sind vergleichbar (Mittelwerte s. Tabelle 7.1.3). In der VA mit den Pr-Werten zeigte sich ein Effekt der Listenbedingung [$F(1,90) = 23.34$, $R^2 = 0.21$]: Wurden die kritischen Items mit präsentiert, dann war das korrekte Wiedererkennen höher ($M = .63$) als das falsche Wiedererkennen bei nicht präsentierten kritischen Items ($M = .35$), was im Einklang mit den Ergebnissen des Wiedererkennens nach dem Ersparnistest ist. Gemäß dem Alterseffekt [$F(2,90) = 6.54$, $R^2 = 0.13$] unterschieden sich Vorschüler ($M = .40$) und Drittklässler ($M = .44$) nicht, aber Erwachsene ($M = .63$) gelangten zu höheren Werten. Wie auch nach dem Ersparnistest erkannten Erwachsene mehr nicht präsentierte kritische Items falsch wieder als Kinder (Hypothese 3.2).

Nach dem Wortassoziationstest konnte, ebenso wie nach dem Ersparnistest, teilweise bestätigt werden, dass sich die Altersgruppen im Wiedererkennen der präsentierten kritischen Items unterschieden (Hypothese 4.2; Mittelwerte s. Tabelle 7.1.2). Erwachsene erkannten mehr präsentierte kritische Items wieder als Vorschüler und Drittklässler, die sich nicht unterschieden. Die Interaktion der Faktoren war nicht signifikant ($F < 1$).

7.1.4 Diskussion

Das vorliegende Experiment zu alterskorrelierten Unterschieden falscher Erinnerungen wurde unter der Zielsetzung durchgeführt, den Einfluss der Aktivierung der kritischen Items ohne dagegenwirkende Quellendiskrimination zu erfassen. Zu diesem Zweck wurden implizite Tests eingesetzt, die von allen Altersgruppen mühelos bearbeitet werden konnten. Es wurden zwei konzeptuelle Tests ausgewählt: Während der Wortassoziationstest eindeutig implizite Effekte erfasst, wurde die Ersparnis in der Freien Reproduktion getestet. Ausgehend von Ebbinghaus Ersparnismethode (1885/1964, zit. n. Roediger, 1996) wurden kritische Items in einem Wiederlernparadigma dargeboten. Durch die Freie Reproduktion konnten dann implizite Effekte in einem expliziten Verfahren getestet werden. Der Einsatz dieses Tests hatte den Vorteil, dass er weniger dazu neigte, explizit kontaminiert zu sein, was bei impliziten Tests häufiger vorzufinden ist. Trotz der impliziten Instruktionen kann im Allgemeinen nicht ausgeschlossen werden, dass insbesondere erwachsene Probanden bewusste Abrufbemühungen einsetzen und damit das Ergebnis verfälschen. Der Ersparnistest sollte kaum explizit kontaminiert gewesen sein, da es ein expliziter Test war, in dem explizit erinnert werden sollte. Demzufolge können die Ergebnisse besser interpretiert werden. In zwei anschließenden Wie-

dererkennenstests nach den jeweiligen impliziten Verfahren, sollten falsche Erinnerungen getestet werden, um davon ausgehend alterskorrelierte Unterschiede zu verstehen und in Beziehung zur Aktivierung zu setzen.

Bei den impliziten Tests gab es ein Basisratenproblem, da sich die Altersgruppen in der Basisrate unterschieden. Dieses Ergebnis könnte auf einem Stichprobenfehler beruhen. Damit ist unklar welches das beste Maß für das Priming ist. Im vorliegenden Experiment wurden absolute Werte bestimmt („alt minus neu“), die jedoch die Interpretation der Ergebnisse erschweren. Um vergleichbare Basisraten im Ersparnistest zu erzielen, sollten in zukünftigen Untersuchungen bei den beiden älteren Gruppen zusätzliche Füllitems in die Wiederlernphase des Ersparnistests eingefügt werden. Beispielsweise könnten Drittklässlern drei und Erwachsenen sieben zusätzliche Füllitems präsentiert werden. Um zu vergleichbaren Basisraten im Wortassoziationstest zu gelangen, wären umfangreiche Voruntersuchungen nötig.

Es zeigte sich, dass in beiden impliziten Testverfahren reliables Priming für präsentierte und nicht präsentierte kritische Items resultierte. Beide gewählten Verfahren waren imstande erfolgreich Priming zu erfassen, was nicht selbstverständlich ist, wie aus McDermotts (1997) Studie mit verschiedenen impliziten Verfahren hervorgeht. Im eigenen Experiment war das Priming deutlich höher als bei McDermott. Im Wortassoziationstest ist es nach vorliegenden Befunden (Zeelenberg, Shiffrin, Raaijmakers, 1999) für das Priming entscheidend, dass in der Lernphase bei der Präsentation des kritischen Items an den späteren Cue „gedacht“ wird, wobei mit Denken nicht unbedingt ein bewusstes Denken gemeint ist. Dementsprechend wird Priming gefördert, wenn kritisches Item und Cue bidirektional assoziiert sind. Die Ergebnisse sprechen dafür, dass dies im eigenen Experiment gegeben war. Das signifikante Priming im eigenen Experiment ist grob gesehen im Einklang mit McDermotts (1997) Ergebnissen zum Wortassoziationstest und mit Diliberto-Macaluso (2005) Ergebnissen zur Wortfragmentergänzung. Priming für die kritischen Items resultierte in allen Untersuchungen, wenn auch die Stichprobe bei ihnen größer war und damit das Priming eher signifikant werden konnte.

Das Priming war im eigenen Experiment hoch verglichen mit anderen Untersuchungen (z.B. McDermott, 1997), wobei besonders die Werte erwachsener Probanden im Ersparnistest auffallen. Da das Experiment universitätsintern repliziert wurde, konnten die Werte zum Priming miteinander verglichen werden. In der universitätsinternen Replikation der Untersuchung mit studentischen Probanden fand eine Variation der Präsentationsmodalitäten der Lernphase (auditiv vs. visuell) statt. In der visuellen Bedingung wurden die Listen per Power-Point auf einem Monitor dargeboten, um eine höhere Standardisierung zu erreichen. Es zeigte sich, dass

die Darbietungsmodalität keine Rolle spielte, so dass davon ausgegangen werden kann, dass durch die auditive Listenpräsentation eine hinreichende Standardisierung gewährleistet war. Es sollte geprüft werden, ob der neue Ersparnistest tatsächlich ein konzeptuelles implizites Verfahren ist, was bestätigt wurde, denn die Darbietungsmodalität hatte keinen Effekt. Priming zeigte sich auch dann, wenn die Darbietungsmodalität der Lern- und Testphase nicht übereinstimmte. Das Priming studentischer Probanden der auditiven Lernbedingung war im Ersparnistest (ohne kritisches Item: $M = .34$; mit kritischem Item: $M = .21$) ohne präsentiertes kritisches Item vergleichbar mit dem Priming der erwachsenen Probanden im eigenen Experiment, aber für die Listenbedingung in der das kritische Item mit präsentiert wurde, war es geringer als im eigenen Experiment. Dieses Ergebnis studentischer Probanden ist überraschend, denn das Priming für die Listenbedingung, in der das kritische Item mit präsentiert wird, sollte nicht kleiner sein verglichen mit jener, in der das kritische Item in den Listen nicht mit präsentiert wird. Weitergehend könnte geprüft werden, ob das abweichende Ergebnis studentischer Versuchspersonen stichprobenabhängig ist.

Die Befunde des vorliegenden Experiments deuteten nicht darauf hin, dass die Aktivierung über alle Altersgruppen zunahm. Im Ersparnistest gab es zwar alterskorrelierte Unterschiede im Priming zwischen Kindern und Erwachsenen, aber die Aktivierung der kritischen Items aller Altersgruppen unterschied sich nicht zwischen der Listenbedingung, in der die kritischen Items mit präsentiert wurden, von jener in der sie nicht mit präsentiert wurden. Die fehlenden Unterschiede könnten aber auf die geringe Anzahl von kritischen Items (fünf alte und fünf neue) bei 16 Versuchspersonen pro Bedingung zurückgehen und damit ein Power-Problem sein. Auch Parkin (1997) wies darauf hin, dass altersinvariante Effekte impliziter Behaltensleistungen nicht ohne weiteres angenommen werden können. Geringe Teststärken und damit hohe β -Fehler können die Interpretation verfälschen, die auf nicht signifikante Effekte schließt und zur Annahme der Nullhypothese führt. Es erscheint kaum möglich höhere Teststärken (im Bereich von .90 bis .95) zu erreichen, denn dazu müssten 1000 Versuchspersonen an einem Experiment teilnehmen. Dies ist, insbesondere wenn an die Einzelversuche und die Altersgruppen gedacht wird, unrealistisch, unökonomisch und nicht gerechtfertigt. Eine größere Listenanzahl könnte die Teststärke erhöhen, aber durch die Auswahl der Altersgruppen (Vorschüler und Grundschüler) ist auch dies nicht umsetzbar.

Eine realistische Möglichkeit die Teststärke zu erhöhen könnte darin bestehen mehr Items durch zwei Listen bzw. Lernphasen einzubringen. In der ersten Lernphase könnten nach zwei (Drittklässler drei, Erwachsene sieben) Füllitems zur Vermeidung des Primacy-Effekts, acht bis 10 kritische Items präsentiert werden, von denen die Hälfte alt und die andere neu wären.

Nach einer Distraktoraufgabe könnte ein Ersparnistest erfolgen, dem sich die zweite Lernphase mit acht bis 10 anderen kritischen Items anschließen würde. Zum Abschluss könnte nochmals ein Ersparnistest eingesetzt werden. Auf diese Weise könnten alterskorrelierte Unterschiede in der Aktivierung kritischer Items getestet werden. Zu beachten ist, dass die vorgeschlagene Testprozedur kaum mit Vorschülern oder Erstklässlern durchführbar ist, da die ursprüngliche Testung schon anstrengend für jüngere Kinder war.

Die Wahl der beiden impliziten Verfahren war den Befunden nach sinnvoll, auch wenn Tse und Neely (2005) den Wortassoziationstest kritisieren. Sie vertreten die Meinung, dass bewusste Abrufstrategien in der Latenzzeit eingesetzt werden können bis eine Antwort erteilt wird. Auch kritisieren sie die Test-Cues des Wortassoziationstests, die mit den Items der Lernphase verbunden sind. Durch diese Verknüpfung zur Lernphase könnte ein bewusster Abruf gefördert werden. Diese Kritik erscheint nicht gerechtfertigt, denn gerade bei Kindern erscheint es äußerst unwahrscheinlich, wenn nicht unmöglich, dass sie gezielt kritische Items erinnern. Zudem wurde den Probanden die Instruktion erteilt, möglichst schnell zu assoziieren und dass es keine falschen Antworten gäbe. Wenn der Wortassoziationstest explizit kontaminiert gewesen wäre, dann hätten Erwachsene ein höheres Priming als Kinder zeigen müssen, was nicht der Fall war. Außerdem hätte sich auch im Ergebnis zeigen müssen, dass mit präsentierten kritischen Items ein höheres Priming zustande kam als ohne präsentierte kritische Items. Zusammenfassend erscheint Tse und Neelys (2005) Kritik am Wortassoziationstest für die eigene Untersuchung unbegründet. Sie verwendeten in ihrer Untersuchung eine lexikalische Entscheidungsaufgabe: Diese wäre im eigenen Experiment mit Vorschülern und jungen Grundschulern nicht möglich, da ihre Lesefähigkeiten nicht bzw. unzureichend vorhanden sind.

Sowohl im Wiedererkennen nach dem Ersparnistest als auch nach dem Wortassoziationstest gab es einen alterskorrelierten Anstieg falscher Erinnerungen, denn Erwachsene bezeichneten mehr nicht präsentierte kritische Items falsch als „alt“ als Vorschüler und Drittklässler, die sich nicht unterschieden.

Entsprechend dem „more is less“ Effekt (Toglia et al., 1999) zeigten sich parallele Effekte für richtiges und falsches Wiedererkennen, denn auch für die präsentierten kritischen Items ergab sich nach beiden Tests ein Altersunterschied. Nach dem Ersparnistest und nach dem Wortassoziationstest erkannten Erwachsene mehr präsentierte kritische Items wieder als Vorschüler und Drittklässler, die sich auch hier nicht unterschieden. Dies galt auch für das korrekte Wie-

dererkennen der Listenitems. Dieses Ergebnismuster lässt sich dadurch erklären, dass Erwachsene tiefer verarbeiten, was die richtigen aber auch die falschen Erinnerungen fördert.

Auffallend ist, dass die Ergebnisse der impliziten Tests zur altersinvarianten Aktivierung nicht mit den deutlichen Altersunterschieden im Wiedererkennen in Einklang stehen. Erwachsene hatten einen höheren Anteil falscher Erinnerungen, obwohl nach den Ergebnissen der impliziten Tests ihre Aktivierung mit jener der jüngeren Probanden vergleichbar war. Da keine Theorie vorhersagt, dass die Fähigkeiten zur Quellendiskrimination sich mit zunehmendem Alter verschlechtern und es auch keine dementsprechenden Befunde gibt, können diese nicht als ausschlaggebend für das Ergebnis betrachtet werden. Somit scheinen die Unterschiede im Wiedererkennen auf Unterschieden in der Aktivierung zurückzugehen. Eine mögliche Erklärung könnte sein, dass die hohe Anzahl falscher Erinnerungen Erwachsener darauf zurückgeht, dass sie mehr alte kritische Items als Kinder in den vorangegangenen impliziten Tests genannt und daher auch wiedererkannt haben. Neben dieser Artefaktinterpretation sollten aber trotzdem andere Interpretationen möglich sein, wie z.B. die genaue Betrachtung der eingesetzten Tests.

Obwohl die impliziten Tests sehr gut von allen Altersgruppen bearbeitet werden konnten und es zu reliablem Priming kam, ist nun fraglich, was diese Tests messen oder warum sich die Altersgruppen nicht in der Aktivierung, die in impliziten Tests erfasst wurde, unterschieden, insbesondere weil die Ergebnisse des Wiedererkennenstests dafür sprechen, dass die Aktivierung bei Erwachsenen stärker war als bei Kindern. Dieses Ergebnis wurde auch in der eigenen Diplomarbeit berichtet (Hüsken, 2004) und in diesem Experiment repliziert.

Eine mögliche Erklärung der Befunde könnte darauf hindeuten, dass die impliziten Tests vor allem die automatische Aktivierung erfassen könnten, die in den Altersgruppen vergleichbar war. Im Wiedererkennenstest könnten vor allem bewusste Anteile der Aktivierung erfasst worden sein, die zu Aktivierungsunterschieden führten und damit bei Erwachsenen vermehrt falsche Erinnerungen verursachten.

Um zu prüfen welche Art von Aktivierung der eingesetzte Wortassoziationstest misst, könnten die Instruktionen manipuliert werden: Neben einer gewöhnlichen impliziten (spontan das erste Wort sagen, was zu dem Cue einfällt) Instruktion könnte eine explizite (die zugehörigen Items zu den entsprechenden Cues erinnern) erteilt werden. Damit wäre ein Vergleich der genannten kritischen Items möglich. Eine höhere Anzahl genannter kritischer Items nach der impliziten Instruktion würde für eine eher automatische Aktivierung sprechen. Würden nach der expliziten Instruktion mehr kritische Items genannt werden, dann würde dies auf eine eher

bewusste Aktivierung hindeuten. Im Fall einer vergleichbaren Anzahl genannter kritischer Items nach beiden Instruktionen wären gezielte Aussagen zur primären Art der Aktivierung schwierig.

Eine Möglichkeit die Aktivierung in expliziten und impliziten Tests direkt miteinander zu vergleichen und zu testen, ob es hier Unterschiede zwischen Erwachsenen und Kindern gibt, bestünde darin, einen impliziten Test (z.B. Wortassoziationstest) und einen Standardwiedererkennenstest nach Variation der Verarbeitungstiefe einzusetzen. Der implizite Test sollte ein separates Maß zur Erhebung der Aktivierung ohne entgegenwirkende Quellendiskrimination sein, der Standardwiedererkennenstest sollte Aktivierung und Quellendiskrimination erfassen. In der Lernbedingung könnte eine Variation der Verarbeitungstiefe dahingehend stattfinden, dass die DRM-Listen entweder als Stroop-Test (Farbe der Wörter benennen) vorgegeben werden würden, so dass primär nur eine automatische Aktivierung möglich wäre oder die Wörter selbst gelesen werden würden (automatische und bewusste Aktivierung). In der Testphase würde entweder ein impliziter Test oder ein Standardwiedererkennenstest erfolgen. Durch die unterschiedlichen Verarbeitungsbedingungen könnten Aussagen zur Aktivierung der kritischen Items ermöglicht werden. Wäre die Aktivierung kritischer Items zu einem erheblichen Teil automatisch, dann sollten nach Bearbeitung des Stroop-Tests nicht präsentierte kritische Items überzufällig häufig genannt (implizit) bzw. falsch wiedererkannt (explizit) werden. Würden sich keine Altersunterschiede zeigen, dann wäre die automatische Aktivierung altersinvariant. Gemäß des Effekts der Verarbeitungstiefe sollte nach der sehr oberflächlichen Bearbeitung des Stroop-Tests eine geringere Aktivierung der kritischen Items resultieren als nach dem eher oberflächlichen Lesen der Wörter, was sich in beiden Testarten widerspiegeln sollte. Auch könnte in den verschiedenen Altersgruppen der Effekt der Verarbeitungsbedingungen auf Priming und falsche Erinnerungen geprüft werden und in Bezug zur automatischen bzw. bewussten Aktivierung gesetzt werden. Wenn Kinder bei einer oberflächlichen Verarbeitung mehr Priming bzw. falsche Erinnerungen zeigen würden als bei einer tieferen, wäre bei ihnen die automatische Aktivierung für falsche Erinnerungen entscheidend. An einer solchen Untersuchung könnten aber nur Grundschul Kinder ab dem zweiten Schuljahr teilnehmen, da für die Variation der Verarbeitungstiefe Lesefähigkeiten vorhanden sein müssen. Bei Vorschülern könnte eine primär automatische Aktivierung erreicht werden, indem ihre Aufmerksamkeit abgelenkt werden würde oder die Wortlisten schnell präsentiert werden würden (Darbietungsrate: weniger als eine Sekunde).

Zwei aktuelle Untersuchungen zum Priming in impliziten Tests verdeutlichen die Problematik und geben weitere Aufschlüsse über die Art der Aktivierung in impliziten und expliziten

Tests. Trotz des Einsatzes perzeptueller impliziter Tests ergeben sich interessante Implikationen für die eigenen Ergebnisse. Sowohl bei McBride, Coane und Raulerson (2006) als auch bei Meade, Watson, Balota und Roediger (2007) zeigte sich in perzeptuell impliziten Tests kein Priming bzw. nur unter Bedingungen verstärkter assoziativer Aktivierung. McBride und Mitarbeiter (2006) erteilten implizite und explizite Instruktionen zu zwei perzeptuellen Verfahren: einem Test der Wortstammerngänzung und einem "Graphemic Cued Response" Test [der Testcue ähnelt orthographisch dem kritischen Item und zu diesem soll ein Item generiert werden: entweder ein zuvor gelerntes (explizite Instruktion) oder eines, das ihnen zuerst spontan einfällt (implizite Instruktion)]. Während explizite Instruktionen zu falschen Erinnerungen führten, zeigte sich nach impliziten Instruktionen kein Priming für kritische Items. Die Autoren führen dies auf eine zu geringe Aktivierung zurück und vertreten die Auffassung, dass die Aktivierung für kritische Items in impliziten Tests stärker sein müsste als für falsche Erinnerungen in expliziten Tests. Dies begründen sie mit Forschungsbefunden von Roediger und McDermott (1993), die zeigten, dass nach der Präsentation von semantisch assoziierten Items in der Lernphase explizite Tests zu besseren Erinnerungsleistungen führten, nicht aber implizite Tests.

Meade und Mitarbeiter (2007) spezifizieren diese in impliziten Tests benötigte „stärkere“ Aktivierung. Sie setzten die lexikalische Entscheidungsaufgabe als perzeptuell impliziten Test ein und untersuchten den Einfluss der Testposition des kritischen Items auf das Priming und den Anteil falscher Erinnerungen im expliziten Wiedererkennenstest. Priming zeigte sich nur, wenn das kritische Item beim Test direkt, also an erster Stelle nach dem letzten Listenitem der Lernphase dargeboten wurde, demgegenüber war falsches Wiedererkennen unabhängig von der Testposition des kritischen Items. Diese Befunde sprechen dafür, dass die getestete Aktivierung kurzzeitig war, anders als im Wiedererkennenstest.

Demnach könnte es, bezogen auf die eigenen unterschiedlichen Ergebnisse zur Aktivierung in impliziten und expliziten Tests, sein, dass das Zeitintervall entscheidend ist. Die impliziten Tests erfolgten stets vor den Wiedererkennenstests und waren der Lernphase und damit der Aktivierung zeitlich näher. Folglich könnte die in impliziten Tests erfasste Aktivierung eher kurzzeitig sein im Gegensatz zu jener in den Wiedererkennenstests. Durch die fehlenden Altersunterschiede in impliziten Tests könnte geschlussfolgert werden, dass die kurzfristige Aktivierung kritischer Items also für alle Altersgruppen vergleichbar war. Im Fall einer langfristigen Aktivierung wie sie vermutlich im Wiedererkennenstest erfasst wurde, ergaben sich alterskorrelierte Unterschiede. Es könnte möglich sein, dass die Arbeitsgedächtniskapazität

Einfluss auf die langfristige Aktivierung haben könnte, weshalb Erwachsene eine höhere Aktivierung hatten als Kinder.

Denkbar wäre das Zeitintervall vor impliziten und expliziten Tests anzugleichen. Wenn das Zeitintervall ausschlaggebend wäre, dann sollten sich die Ergebnisse zur Aktivierung in beiden Tests nicht unterscheiden.

Eine andere Möglichkeit bestünde darin das Zeitintervall durch kurze oder lange Ablenkaufgaben nach der Lernphase zu variieren und zu prüfen, ob beide Tests zu vergleichbaren Ergebnissen in der Aktivierung gelangen. Also sollte sich im Fall einer langen Ablenkaufgabe durch eine benötigte langfristige Aktivierung ein geringeres Priming (geringere Aktivierung) zeigen als bei einer kurzen Ablenkaufgabe. Interessant wäre es zu prüfen, ob und inwiefern sich Kinder von Erwachsenen unterscheiden. Wenn keine alterskorrelierten Unterschiede auftraten, dann wäre die getestete Aktivierung altersinvariant.

Ausgehend von den Ergebnissen ist der Nutzen zum weiteren Einsatz impliziter Tests fraglich. Zwar testen sie Aktivierung ohne den Einfluss der entgegenwirkenden Quellendiskrimination, aber es ist unklar welche Aktivierung dies ist.

Da sich aus dem ersten Experiment keine eindeutigen Schlüsse über alterskorrelierte Differenzen im Ausmaß der Aktivierung der kritischen Items ziehen lassen, soll im nächsten Experiment eine andere Möglichkeit geprüft werden, Aktivierung zu erfassen ohne entgegenwirkende Quellendiskrimination: Nach so genannten Inklusionsinstruktionen (s. Abschnitt 5.4) wird die Aktivierung der kritischen Items isoliert erfasst, denn Probanden sollen alle vertrauten (also aktivierten) Items „alt“ nennen. In Voruntersuchungen müssen Instruktionen entwickelt werden, die für alle Altersgruppen verständlich sind und sich von Standardwiedererkennensinstruktionen in der häufigeren Nennung kritischer Items unterscheiden. Anhand dieses Vorgehens könnten sich weitere Rückschlüsse auf mögliche alterskorrelierte Unterschiede in der Aktivierung ergeben. Zudem soll ein separates Maß für die Quellendiskrimination bestimmt werden, das es im ersten Experiment nicht gab. Auf diese Weise können Aktivierung und Quellendiskrimination anhand separater Maße für alle Altersgruppen erfasst und deren Einfluss auf falsche Erinnerungen herausgestellt werden.

8 Alterskorrelierte Differenzen im falschen Reproduzieren und im Wiedererkennen?

8.1 Experiment 2: Falsches Reproduzieren: Einflüsse von Aktivierung und Quellendiskrimination

Es sollte geprüft werden, ob alterskorrelierte Differenzen falscher Erinnerungen beim Freien Reproduzieren von Vorschulkindern, Drittklässlern und Erwachsenen auf Unterschieden in der Quellendiskrimination und/oder in Aktivierungsprozessen gründen. Während im vorangegangenen Experiment die Aktivierung der kritischen Items durch implizite Tests erfasst wurde, aber die Aussagen zur alterskorrelierten Veränderung der Aktivierung nicht eindeutig waren, sollte die Aktivierung im zweiten Experiment durch spezielle Instruktionen in der Freien Reproduktion separat erfasst werden. Neben einer Standardinstruktion zum Freien Reproduzieren wurde eine Inklusionsinstruktion eingesetzt, die gezielt Aktivierung der kritischen Items ohne entgegenwirkende Quellendiskrimination erfassen sollte. Die Probanden wurden in dieser Inklusionsinstruktion nach jeder Liste aufgefordert sowohl die präsentierten Items als auch jene zu nennen, an die sie sehr stark in der Lernphase gedacht hatten (Hedge & Dodson, 2004; Heit et al., 2004; s. Abschnitt 5.4). Es wurde erwartet, dass diese Instruktion zur Nennung des mit der Liste am stärksten assoziierten, also kritischen Item führen sollte, sofern es in der Lernphase aktiviert werden konnte. Alle Versuchspersonen sollten sowohl die Standard- als auch die Inklusionsinstruktion erhalten, um auf diese Weise die Erinnerungsleistungen direkt miteinander zu vergleichen. Im Standardwiedererkennenstest wurde entsprechend die Aktivierung der kritischen Items mit entgegenwirkender Quellendiskrimination getestet und im Inklusionstest ohne entgegenwirkende Quellendiskrimination.

Um ein Maß für die Güte der Quellendiskrimination von Drittklässlern und Erwachsenen zu erhalten, wurden nach der Freien Reproduktion den Probanden alle ihre genannten Items listenweise vorgelesen mit der Aufforderung anzugeben, ob das jeweilige Item in der Lernphase von der Versuchsleiterin tatsächlich vorgelesen wurde. Da die Arbeitsgedächtniskapazität jüngerer Kinder eingeschränkt ist und die zur Aufgabe erforderlichen metakognitiven Fähigkeiten noch nicht ausreichend vorhanden sind, wurde davon ausgegangen, dass sie Probleme mit diesem komplexen Instruktionsverständnis haben würden und folglich wurde Vorschülern diese Aufgabe nicht gestellt.

Die Altersgruppen erhielten unterschiedlich lange Listen, um einerseits zu prüfen, ob bei vergleichbaren korrekten Erinnerungsleistungen alterskorrelierte Unterschiede in der Quellediskrimination geringer ausfallen. In Anlehnung an die Befunde von Seamon et al. (2003, s. Abschnitt 2.2.2 und 2.3.4) sind korrekte und falsche Erinnerungen abhängig von der Güte des Listenlernens und zeigen unterschiedliche Verläufe. Ein Vergleich der drei Altersgruppen erlaubt entsprechend die Prüfung alterskorrelierter Differenzen falscher Erinnerungen bei Konstanzhaltung der korrekten Erinnerungsleistung. Bei gut gelernten Listen sollte die Quellediskrimination leichter fallen. Andererseits sollte untersucht werden, ob wie bei Erwachsenen (Robinson & Roediger, 1997) auch bei Kindern längere Listen mit vermehrt falschen Erinnerungen einhergehen und ob das, wie Robinson und Roediger (1997) schlussfolgern, auf eine höhere Aktivierung zurückgeht und/oder, was ebenso denkbar wäre, auf eine schlechtere Quellediskrimination. Die Listenlängen wurden so ausgewählt, dass die Hälfte der Probanden jeder Altersgruppe gleich lange Listen erhielt (10 Items) und die andere Hälfte eine Listenlänge, die zu vergleichbaren korrekten Erinnerungsleistungen in den Altersgruppen führen sollte. Im vorliegenden Experiment wurden auf der Basis von Vorläuferstudien Vorschülern Listen mit sieben Items, Drittklässlern und Erwachsenen eine gleiche Listenlänge mit 14 Items zugeteilt.

Für den Test der Freien Reproduktion wurden folgende Ergebnismuster für die kritischen Items erwartet:

- (1) Spiegelt die Anzahl reproduzierter kritischer Items nach der Inklusionsinstruktion allein deren Aktivierung ohne den Einfluss von Quellediskrimination wider, dann sollten alle drei Altersgruppen nach dieser bedeutend mehr kritische Items reproduzieren als nach der Standardinstruktion.
- (2) Längere Listen sollten nach den Befunden von Robinson und Roediger (1997) zu einer höheren Aktivierung kritischer Items führen. Daher sollten die Probanden bei längeren Listen nach der Inklusionsinstruktion mehr kritische Items nennen als bei kurzen Listen.
- (3) Erwartet wird eine alterskorrelierte Zunahme der Aktivierung kritischer Items. Deshalb sollten sich die Altersgruppen in der Häufigkeit der Nennung kritischer Items nach der Inklusionsinstruktion unterscheiden.
- (4) Erwartet werden auch alterskorrelierte Verbesserungen der Quellediskrimination. Daher sollte sich bei der sehr schwierigen Quellenunterscheidungsaufgabe kritischer

Items ein Alterseffekt dahingehend zeigen, dass Drittklässler kritische Items eher der falschen externen Quelle zuweisen als Erwachsene.

- (5) Zwar wird angenommen, dass Aktivierung und Quellendiskrimination alterskorreliert zunehmen, aber dennoch lassen sich keine Vorhersagen zu Altersunterschieden in der Häufigkeit falsch reproduzierter kritischer Items nach der Standardinstruktion treffen, einerseits aufgrund der gemischten Befundlage in der Literatur (Brainerd et al., 2002a; Ghetti et al., 2002; vgl. Abschnitt 4.2), andererseits da unklar ist, ob der Einfluss von Aktivierung oder Quellendiskrimination stärker ist (Brainerd et al., 2002a; Ghetti et al., 2002; vgl. Abschnitt 4.2).

8.1.1 Vorversuch

Um zu gewährleisten, dass die Inklusionsinstruktion gegenüber der Standardinstruktion geeignet ist in der Freien Reproduktion die Aktivierung ohne Quellendiskrimination zu erfassen und von der jüngsten Altersgruppe verstanden wird, wurde eine Voruntersuchung mit 16 Vorschülern durchgeführt. Hierbei wurden zwei verschiedene Inklusionsinstruktionen interindividuell und die beiden Sets mit Standard- und Inklusionsinstruktion inter- und intraindividuell variiert, d. h. jede Versuchsperson erhielt die gleiche Standardinstruktion und eine der beiden Inklusionsinstruktionen (A oder B). Dadurch wurde ein Vergleich der beiden Inklusionsinstruktionen und der Standard- mit der Inklusionsinstruktion ermöglicht. Die Reihenfolge der Instruktionen in den Sets war quasi-randomisiert, so dass sie zufällig angeordnet waren aber nie mehr als zwei Listen mit gleichen Instruktionen einander folgten. Die Auswahl der beiden Inklusionsinstruktionen richtete sich an bisherigen Forschungsbefunden (Hege & Dodson, 2004) und an eigenen Überlegungen. Während Instruktion A durch Nennung zusammengehöriger Beispielwörter („Gruppe, Kinder, spielen, singen“) und des kritischen Items („Kindergarten“) die Inklusionsinstruktion dahingehend verdeutlichen sollte, dass sowohl die Beispielwörter genannt werden sollten als auch Wörter an die die Versuchsperson selbst sehr stark während der Präsentation gedacht hatten, zielte Instruktion B darauf ab, ein Rateantwortverhalten hervorzurufen („Du darfst auch raten. Je mehr Wörter du sagst desto besser.“), welches wiederum zu einer häufigeren Wiedergabe kritischer Items führen sollte. Vor der Präsentation von 12 DRM-Wortlisten mit je 10 Items wurde eine ausführliche Anweisung gegeben, wobei Standard- und Inklusionsinstruktion erläutert wurden und deren Unterschiedlichkeit hervorgehoben wurde. Um Standard- und Inklusionsinstruktion klar voneinander zu trennen, wurden zur visuellen Verdeutlichung zwei Handpuppen verwendet, die jeweils die

beiden Instruktionen ankündigten. Die wortgetreuen Instruktionen können im Anhang B2 nachgelesen werden.

Die statistische Auswertung der Voruntersuchung zeigte, dass beide Inklusionsinstruktionen zu einer häufigeren Wiedergabe kritischer Items führten als die Standardinstruktion. Zudem unterschieden sich die Ergebnisse der beiden Inklusionsvarianten nicht bedeutsam voneinander. Für die Hauptuntersuchung wurde Instruktion A ausgewählt, da sie den Instruktionen bisheriger Untersuchungen ähnelt (Hedge & Dodson, 2004).

8.1.2 Methode

Versuchsplan. Experiment 2 basiert auf einem 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Listenlänge: kurz vs. lang) x 2 (Instruktion: Standard vs. Inklusion) Versuchsplan mit Messwiederholung auf dem letzten Faktor. Die Versuchspersonen der drei Altersgruppen wurden zufällig einer der beiden Bedingungen zugeordnet. Nach diesem Untersuchungsdesign resultierten sechs Gruppen mit jeweils 24 Probanden.

Versuchspersonen. Am zweiten Experiment nahmen 144 Probanden teil, die aus 48 Vorschulkindern aus Kindergärten und 48 Drittklässlern einer Grundschule bestanden. Die Gruppe der erwachsenen Probanden setzte sich aus 48 Schülern der 12. Klasse eines Gymnasiums zusammen.

In der Gruppe der Vorschüler nahmen 26 Mädchen und 22 Jungen teil, die durchschnittlich sechs Jahre alt waren (zwischen 5;2 und 6;10 Jahren). Die 25 Mädchen und 23 Jungen der Grundschule hatten ein durchschnittliches Alter von neun Jahren und vier Monaten (zwischen 8;10 und 10;8 Jahren). In der Gruppe der Erwachsenen befanden sich vier Schüler und 24 Schülerinnen, die im Mittel 18 Jahre und drei Monate alt waren (zwischen 17;7 und 19;6 Jahren).

Material. Wie im vorherigen Experiment wurden DRM-Listen eingesetzt, die altersnormiert waren. Es wurden zwei Listensets mit je acht Wortlisten gebildet, die einander in Kategorien (z.B. Tiere) und Assoziationsstärken ähnelten. Alle 16 Listen befanden sich wie im Vorversuch in quasi-randomisierter Reihenfolge mit der Einschränkung, dass nicht mehr als zwei Listen mit gleichen Instruktionen einander folgen durften. Die Listen eines Sets waren der Standardinstruktion zugehörig, die des anderen Sets der Inklusionsinstruktion. Konkret bedeutete dies, dass die Hälfte der Versuchspersonen Set A mit Standard- und Set B mit Inklusionsinstruktion bearbeiteten; für die andere Hälfte gestaltete sich die Präsentation von Set A mit

Inklusions- und Set B mit Standardinstruktion. Allen Probanden wurden die gleichen Listen vorgelesen, die sich zwischen der Gruppe der Vorschüler und jener der Drittklässler und Erwachsenen nur darin unterschieden, dass bei Vorschülern die kurzen Listen aus sieben Items und die langen Listen aus 10 Items bestanden, während in den beiden älteren Gruppen kurze Listen 10 und lange Listen 14 Items beinhalteten. Ausgangspunkt waren Listen mit 14 assoziierten Items, von denen für die Bedingung mit 10 Items die letzten vier Items (mit der geringsten Assoziationsstärke) herausgenommen wurden und für Listensets mit sieben Items wurden wiederum die nächsten drei gering assoziierten Items entnommen. Das verwendete Material ist im Anhang B1 zusammengefasst.

Durchführung. Insgesamt dauerte die Versuchsdurchführung im Kindergarten ca. 20 Minuten, während die beiden älteren Gruppen 25 bis 30 Minuten benötigten, bedingt durch die dem Alter angepassten Listenlängen.

Die Testsituation gestaltete sich gleichermaßen für alle Probanden in drei Phasen, wobei für die beiden älteren Gruppen eine zusätzliche Phase zur Erfassung der Quellendiskriminationsleistung folgte. Zunächst wurde die Standard- und Inklusionsinstruktion erläutert, die für die Gruppe der Vorschüler und Drittklässler identisch war (s. Anhang B2). Eine Handpuppe repräsentierte die Standardinstruktion und teilte den Kindern mit genau nachzudenken, nicht zu raten und nur jene Wörter wiederzugeben, die die Versuchsleiterin vorgelesen hatte, die andere Puppe stellte die Inklusionsinstruktion dar. In dieser wurden den Probanden vier mit einem gemeinsamen Oberbegriff assoziierte Items (Beispielitems für die Vorschüler und Drittklässler: „Gruppe, Kinder, spielen, singen“; kritisches Item: „Kindergarten“) präsentiert, welche wiedergegeben werden sollten. Darüber hinaus wurde den Kindern der Bezug der assoziierten Items zum Oberbegriff erklärt und sie wurden aufgefordert im Fall dieser Instruktion nicht nur die präsentierten Wörter wiederzugeben sondern auch jene, an die sie während der Lernphase stark gedacht hatten. Die Inklusionsinstruktion der Gruppe erwachsener Versuchspersonen beinhaltete dem Alter angepasste Beispielitems („Silber, Schmuck, wertvoll, reich, teuer, Geld, glänzen, Mine, Barren, Bronze“) mit dem zugehörigen kritischen Item („Gold“). Die ausführlichen Instruktionen befinden sich im Anhang B2.

In der folgenden Lernphase wurde je nach Instruktion und Listenbedingung eine Liste mit sieben, 10 oder 14 Items vorgelesen. Sofort danach schloss sich die Testphase an, in welcher den Probanden entweder die Standard- oder Inklusionsinstruktion kurz gegeben wurde. Die jeweilige Instruktion erfolgte erst nach Präsentation einer Liste, womit verhindert werden sollte, dass die Inklusionsinstruktion zur Generierung verwandter Wörter führte. Die Ver-

suchspersonen sollten entsprechend der oben erläuterten Instruktionen ohne zu raten so viele Items nennen wie sie erinnern konnten, mit oder ohne jene Items, an die sie selbst stark gedacht hatten. Während den beiden jüngeren Altersgruppen nach jeder Liste die Instruktionen mit Vorführen der entsprechenden Handpuppe angekündigt wurden, erhielten die Erwachsenen lediglich die Aufforderung zu „Anweisung 1“ (Standard) oder „Anweisung 2“ (Inklusion). Alle 16 Listen wurden auf diese Weise präsentiert und von den Probanden ohne zeitliche Beschränkung frei wiedergegeben.

Nachdem alle Listen bearbeitet worden waren, sollten Drittklässler und Erwachsene in einer weiteren Testphase, sowohl nach der Inklusions- als auch nach der Standardinstruktion, ihre genannten Items einer externen Quelle zuordnen. Liste für Liste wurden alle in der ersten Testphase genannten Items von der Versuchsleiterin vorgelesen und die Probanden sollten bei jedem Item überlegen, ob es der externen Quelle zugeordnet werden konnte [die genaue Instruktion lautete: „Habe ich (d.h. die Versuchsleiterin) das Wort tatsächlich vorgelesen?“ mit den Antwortmöglichkeiten: „ja“ oder „nein“], wobei es auch hier keine zeitliche Begrenzung gab.

8.1.3 Ergebnisse

Zur Prüfung der Hypothesen werden die Ergebnisse der 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Listenlänge: kurz vs. lang) x 2 (Instruktion: Standard vs. Inklusion) VAn mit Messwiederholung zum Freien Reproduktionstest und zur Quelldiskrimination getrennt nach Listenitems und kritischen Items dargestellt. Es wurden verschiedene VAn je nach Listenlänge gerechnet, sowohl eine VA mit allen Listenlängen (kurze und lange; Vorschüler: sieben und 10 Items, Drittklässler und Erwachsene: 10 und 14 Items), eine VA mit dem an das Alter angepasster Listenlänge (Vorschüler: sieben Items, Drittklässler: 10 Items, Erwachsene 14 Items) und eine VA für die gleiche Listenlänge (alle Altersgruppen 10 Items).

8.1.3.1 Listenitems

Freie Reproduktion. In der VA mit allen Listen konnten drei Haupteffekte nachgewiesen werden: Die Variation der Listenlänge führte dazu, dass Listenitems unterschiedlich häufig erinnert wurden [$F(1,138) = 138.27$, $R^2 = 0.50$]. Bei kürzeren Listen wurden erwartungsgemäß prozentual mehr Listenitems erinnert als bei langen Listen ($M = .52$ vs. $.42$).

Zudem zeigten sich Haupteffekte für die Variable Altersgruppe [$F(2,138) = 146.45$, $R^2 = 0.68$] und für die messwiederholte Variable Instruktion [$F(1,138) = 441.87$, $R^2 = 0.76$], welche aber durch eine Interaktion überlagert wurden [$F(2,138) = 58.37$, $R^2 = 0.46$]. Die Mittelwerte zu dieser Interaktion gehen aus Tabelle 8.1.1 hervor. Die Analyse der einfachen Haupteffekte zeigte, dass Vorschüler [$F(1,138) = 76.20$] und Drittklässler [$F(1,138) = 37.01$] nach der Standardinstruktion deutlich weniger Listenitems produzierten als nach der Inklusionsinstruktion, während bei erwachsenen Probanden der Unterschied erinnelter Listenitems nach diesen Bedingungen geringer ausfiel [$F(1,138) = 13.77$]. Die Altersgruppen unterschieden sich sowohl nach der Standard- [$F(2,138) = 109.96$] als auch nach der Inklusionsinstruktion [$F(2,138) = 40.67$]. Nach der Standard- und Inklusionsinstruktion differierten Vorschüler und Drittklässler nicht voneinander, was nicht überraschend ist, denn Drittklässler erhielten längere Listen als Vorschüler. Nur Erwachsene erinnerten deutlich mehr Listenitems als die jüngeren Probanden. Alle anderen Effekte wurden nicht signifikant (alle $F_s < 1$).

In der VA mit angepasster Listenlänge erlangte die Altersgruppe [$F(2,69) = 7.07$, $R^2 = 0.17$] und die Instruktion [$F(1,69) = 193.40$, $R^2 = 0.74$] Signifikanz, die miteinander interagierten [$F(2,69) = 26.37$, $R^2 = 0.43$]. Nach Analyse der einfachen Haupteffekte unterschieden sich die Altersgruppen nur nach der Standardinstruktion [$F(2,69) = 20.95$]. Erwachsene nannten mehr Listenitems als Vorschüler, aber die beiden jüngeren Altersgruppen unterschieden sich nicht. Somit funktionierte die Anpassung der Listenlänge nur nach der Inklusionsinstruktion. Nach der Inklusionsinstruktion nannten die beiden jüngeren Altersgruppen mehr kritische Items als nach der Standardinstruktion [Vorschüler: $F(1,69) = 116.01$, Drittklässler: $F(1,69) = 50.29$], Erwachsene jedoch nicht [$F(1,69) = 3.98$], was verdeutlicht, dass sie weniger als die anderen Altersgruppen zwischen den Instruktionen unterschieden.

Tabelle 8.1.1:

Mittelwerte der reproduzierten Listenitems in Abhängigkeit des Alters, der Instruktion und der unterschiedlichen Listenlänge.

	alle Listen		angepasste Länge		gleiche Länge	
	Standard	Inklusion	Standard	Inklusion	Standard	Inklusion
Vorschüler	.35	.49	.40	.54	.30	.45
Drittklässler	.37	.45	.42	.50	.42	.50
Erwachsene	.55	.59	.50	.53	.61	.64

Allgemein werden in entwicklungspsychologischen Untersuchungen gleiche Listenlängen für alle Altersgruppen verwendet. Bei gleicher Listenlänge von 10 Items zeigte sich ein ähnliches

Muster wie in den vorherigen Analysen: Die Altersgruppe [$F(2,69) = 188.22, R^2 = 0.85$] sowie die Instruktion [$F(1,69) = 209.16, R^2 = 0.75$] und deren Interaktion [$F(2,69) = 29.23, R^2 = 0.46$] erlangten Signifikanz. Der einzige Unterschied zu den vorherigen Analysen zeigte sich im Vergleich zwischen den Altersgruppen [Standard: $F(2,69) = 53.19$; Inklusion: $F(2,69) = 20.30$] darin, dass sich Vorschüler und Drittklässler nun nach der Standard- und der Inklusionsinstruktion voneinander unterschieden.

Zusammenfassend lässt sich festhalten, dass die Anpassung der Listenlänge für Erwachsene nach der Standardinstruktion nicht funktionierte, denn die Altersunterschiede zwischen den Kindern und Erwachsenen blieben bestehen. Vorschüler und Drittklässler unterschieden sich nicht in ihrem Anteil korrekt erinnelter Listenitems.

Quellendiskrimination. Die Güte der Quellenzuordnung wurde für die Listenitems geprüft und setzte sich aus den nach der Inklusionsinstruktion genannten Listenitems zusammen, die in der Nachbefragung zudem als von der Versuchsleiterin vorgelesen bezeichnet wurden. Es zeigte sich kein Alterseffekt ($F < 1$), was bei dieser leichten Quellendiskriminationsaufgabe nicht überrascht: Drittklässler ($M = .97$) und Erwachsene ($M = .98$) konnten die Listenitems gleichermaßen gut zuordnen, wobei die sehr hohen Werte auf Deckeneffekte hinweisen. Auch die anderen Effekte verfehlten das Signifikanzniveau ($F < 1.40$). Entsprechende Analysen der Listen mit dem Alter angepasster Listenlänge und 10 Items pro Liste für alle Altersgruppen erbrachten ein identisches Ergebnismuster.

8.1.3.2 Kritische Items

Freie Reproduktion. Tabelle 8.1.2 fasst die Mittelwerte der genannten kritischen Items aller Altersgruppen nach der Standard- und Inklusionsinstruktion zusammen, woraus die „Güte“ der einzelnen Listen zur Induzierung falscher Erinnerungen hervorgeht. Die Listen sind nach absteigender Stärke falscher Erinnerungen über alle Altersgruppen angeordnet. Die Häufigkeit der Nennung kritischer Items nach der Inklusionsinstruktion, über alle Altersgruppen gemittelt (von $M = .25$ bis $M = .84$), fiel wie beabsichtigt, immer höher aus als nach der Standardinstruktion (von $M = .12$ bis $M = .25$). Generell zeigte sich, dass Listen zum Thema Tiere (wie z.B. „Pferd“, „Vogel“, „Hase“) häufiger zur Nennung des kritischen Items führten. Das kritische Item „Sommer“ wurde am wenigsten falsch reproduziert. Insgesamt kann festgehalten werden, dass alle Listen zur Nennung der kritischen Items nach der Inklusionsinstruktion führten, wenn auch in unterschiedlichem Maß. Auffallend ist, dass nach der Standardinstruktion alle Listen zu weniger falschen Erinnerungen führten als die gleichen Listen von Roedi-

ger und McDermott (1995). Dies könnte an der Sprache liegen, da die deutsche Sprache gegenüber der englischen distinkter ist. Einige der Listen führten auch zu sehr seltenen Nennung der kritischen Items bzw. wurden nicht genannt, wenn die Probanden nicht zur Produktion aufgefordert wurden.

Tabelle 8.1.2:

Mittelwerte der genannten kritischen Items in Abhängigkeit der Instruktion und des Alters.

Kritische Items	Standard				Inklusion			
	Vorsch.	Drittkl.	Erwachs.	Ges.	Vorsch.	Drittkl.	Erwachs.	Ges.
Fenster	.13	.13	.50	.25	.67	.63	.75	.68
Kerze	.09	.04	.59	.24	.38	.63	.75	.59
Pferd	.08	.04	.50	.21	.75	.88	.88	.84
Vogel	.04	.04	.54	.21	.54	.79	.67	.67
Nadel	0	.21	.42	.21	.29	.63	.96	.63
Telefon	.13	.17	.29	.20	.42	.88	.80	.70
Hase	.09	.25	.25	.20	.71	.80	.55	.69
Auto	.04	.17	.38	.20	.58	.79	.75	.71
Hammer	.09	.13	.33	.18	.34	.75	.71	.60
Hund	.04	0	.50	.18	.79	.88	.84	.84
Löwe	.13	.09	.25	.16	.59	.67	.80	.69
Hemd	.04	.04	.38	.15	.17	.21	.38	.25
Bett	0	.04	.42	.15	.42	.67	.50	.53
Brot	0	.04	.33	.12	.46	.75	.50	.28
Fuß	.04	.09	.25	.13	.50	.50	.59	.53
Sommer	.04	0	.33	.12	.29	.25	.63	.39

Die 3 (Altersgruppe) x 2 (Listenlänge) x 2 (Instruktion) VA mit allen Listen für die kritischen Items zeigte, dass das Hinzufügen zusätzlicher Listenitems wider Erwarten (Hypothese 2) nach beiden Instruktionen und in allen Altersgruppen nicht zu einem Anstieg genannter kritischer Items ($F < 1$, Mittelwerte s. Tabelle 8.1.3) und damit der Aktivierung führte, was entgegen der Annahme von Robinson und Roediger (1997) ist.

Tabelle 8.1.3:

Mittelwerte der genannten kritischen Items mit allen Listen in Abhängigkeit des Alters, der Listenlänge und der Instruktion.

	Standard		Inklusion	
	kurz	lang	kurz	lang
Vorschüler	.06	.06	.47	.47
Drittklässler	.08	.10	.66	.68
Erwachsene	.42	.37	.70	.67

Sowohl die Variable Alter [$F(2,138) = 77.32, R^2 = 0.53$] als auch die messwiederholte Variable Instruktion erwiesen sich als signifikant [$F(1,138) = 1067.02, R^2 = 0.89$], wobei diese Haupteffekte durch eine Interaktion überlagert wurden [$F(2,138) = 38.09, R^2 = 0.36$]. Die Analyse der einfachen Haupteffekte (Mittelwerte s. Tabelle 8.1.4) zeigten, dass nach der Inklusionsinstruktion von allen Altersgruppen erwartungsgemäß deutlich mehr kritische Items genannt wurden als nach der Standardinstruktion (Hypothese 1), wobei der Unterschied bei den Erwachsenen am geringsten war [Vorschüler: $F(1,138) = 115.23$; Drittklässler: $F(1,138) = 227.65$; Erwachsene: $F(1,138) = 61.97$].

Nach der Inklusionsinstruktion [$F(2,138) = 109.61$] erinnerten die älteren Gruppen erwartungsgemäß mehr kritische Items als Vorschüler (Hypothese 3), aber Drittklässler und Erwachsene unterschieden sich nicht, was auf eine vergleichbare Aktivierung hinweist (zu den Mittelwerten s. Tabelle 8.1.4).

Es wurde keine gerichtete Annahme für die falsch reproduzierten kritischen Items formuliert (Hypothese 5). Die Ergebnisse zeigen, dass erwachsene Versuchspersonen nach der Standardinstruktion [$F(2,138) = 265.88$] kritische Items häufiger nannten als die beiden jüngeren Altersgruppen, deren Werte auf Bodeneffekte hinweisen (s. Tabelle 8.1.4).

Tabelle 8.1.4:

Mittelwerte der genannten kritischen Items in Abhängigkeit des Alters, der Instruktion und der unterschiedlichen Listenlänge.

	alle Listen		angepasste Länge		gleiche Länge	
	Standard	Inklusion	Standard	Inklusion	Standard	Inklusion
Vorschüler	.06	.47	.06	.47	.06	.47
Drittklässler	.09	.67	.08	.66	.08	.66
Erwachsene	.39	.69	.37	.67	.42	.70

Die statistische Analyse zeigte keine weiteren signifikanten Interaktionseffekte (alle $F_s < 1.09$).

Im Hinblick auf die deutlich geringe Rate falscher Erinnerungen bei Vorschülern (6%), stellte sich die Frage, ob diese als überzufällig angesehen werden konnte. Anhand einseitiger t -Tests zeigte sich, dass Vorschüler sowohl bei kurzen als auch langen Listen eine überzufällige Anzahl falscher Erinnerungen hatten [beide $t(23) > 3.72$].

Wurde eine VA mit dem Alter angepasster Listenlänge gerechnet, so resultierte ein ähnliches Ergebnismuster wie bei der Analyse mit allen Listen. Dies war ebenso der Fall bei gleicher Listenlänge (Mittelwerte s. Tabelle 8.1.4).

Zusammenfassend ist festzuhalten, dass die Analysen der verschiedenen Listenlängen zu vergleichbaren Ergebnissen reproduzierter kritischer Items gelangten, wobei Ergebnisse der beiden jüngeren Probandengruppen aufgrund der Bodeneffekte nach der Standardinstruktion kaum zu interpretieren sind. Nach der Inklusionsinstruktion gab es keine die Interpretation einschränkenden Effekte und die Listenlänge hatte keinen Einfluss auf das Ergebnismuster, womit es fraglich erscheint, ob eine Anpassung der Listenlänge gemäß dem Alter nötig ist.

Quellendiskrimination. Die Mittelwerte für die Fehlattribuierung, also für die „echten“ falschen Erinnerungen⁴, sind nach den verschiedenen Bedingungen der Listenlängen in Tabelle 8.1.5 abgebildet. Erwartungsgemäß ordneten Drittklässler nach beiden Instruktionen ihre genannten kritischen Items häufiger als Erwachsene der falschen Quelle zu, was ihre schlechtere Quellendiskrimination verdeutlicht (Hypothese 4).

Tabelle 8.1.5:

Mittelwerte der falschen (externen) Quellenzuweisung kritischer Items in Abhängigkeit des Alters, der Instruktion und der unterschiedlichen Listenlänge.

	alle Listen		angepasste Länge		gleiche Länge	
	Standard	Inklusion	Standard	Inklusion	Standard	Inklusion
Drittklässler	.85	.59	.93	.64	.93	.64
Erwachsene	.23	.17	.18	.19	.28	.15

Eine 2 (Altersgruppe) x 2 (Instruktion) VA ohne die Listenlänge wurde für die falsche Quellenzuweisung kritischer Items gerechnet, da die Inspektion der Mittelwerte zeigte, dass längere Listen nicht zu einer schlechteren Quellendiskrimination führten. Diese VA zeigte einen Alters- [$F(1,72) = 185.44, R^2 = 0.72$] und Instruktionseffekt [$F(1,72) = 9.61, R^2 = 0.12$] und eine Interaktion der beiden Faktoren [$F(1,72) = 5.49, R^2 = 0.07$]. Die Mittelwerte sind in Ta-

⁴ Anteil der kritischen Items, die reproduziert und im Nachtest als von der Versuchsleiterin als vorgelesen bezeichnet, d.h. extern attribuiert wurden

belle 8.1.5 abgebildet. Nach der Standardinstruktion führten Drittklässler mehr kritische Items auf die falsche Quelle zurück als nach der Inklusionsinstruktion [$F(1,72) = 8.17$] im Gegensatz zu Erwachsenen, bei denen es keinen Unterschied zwischen den Instruktionen gab ($F < 1$). Dies spricht dafür, dass sich letztere anscheinend nicht an die Standardinstruktion gehalten hatten, da sie viele kritische Items falsch reproduzierten, aber anscheinend dennoch die richtige Quelle kannten. Die Altersgruppen unterschieden sich sowohl nach der Standard- [$F(1,72) = 25.08$] als auch nach der Inklusionsinstruktion [$F(1,72) = 10.76$] voneinander, wobei Drittklässler stets einen höheren Anteil falsch attribuerter Items aufwiesen. Dieses Ergebnismuster zeigte sich auch bei einer dem Alter angepassten und bei gleicher Listenlänge.

Um die falschen Erinnerungen in der Inklusion zu prüfen und ob diese sich von jenen in der Standardinstruktion unterschieden, wurden die kritischen Items zur Analyse herangezogen, die in der Freien Reproduktion nach beiden Instruktionen genannt wurden und in der zweiten Testphase falsch extern attribuiert worden waren. Tabelle 8.1.6 fasst die Mittelwerte zusammen.

Tabelle 8.1.6:

Mittelwerte falscher Erinnerungen in Abhängigkeit des Alters, der Listenlänge und der Instruktion.

	Standard		Inklusion	
	kurz	lang	kurz	lang
Drittklässler	.07	.08	.43	.39
Erwachsene	.12	.07	.12	.12

Die entsprechende VA ergab einen Alters- [$F(1,92) = 75.13$, $R^2 = 0.45$] und einen Instruktionseffekt [$F(1,92) = 104.30$, $R^2 = 0.53$] sowie eine Interaktion der beiden Faktoren [$F(1,92) = 75.42$, $R^2 = 0.45$]. Die Analyse der einfachen Haupteffekte zeigte, dass die beiden Altersgruppen sich nach der Standardinstruktion nicht voneinander unterschieden [$F(1,92) < 1$]. Drittklässler hatten 8% und Erwachsene 9% falsche Erinnerungen. Nach der Inklusionsinstruktion zeigten sich Altersdifferenzen dahingehend [$F(1,92) = 135.67$], dass Drittklässler eine höhere Rate falscher Erinnerungen hatten als Erwachsene.

Insgesamt hatten sie 41% falsche Erinnerungen gegenüber 12% bei Erwachsenen. Für Drittklässler war der Anteil falscher Erinnerungen nach der Inklusionsinstruktion höher als nach der Standardinstruktion [$F(1,92) = 176.91$], während sich bei Erwachsenen kein Unterschied zwischen Inklusions- und Standardinstruktion ergab [$F(1,92) = 1.78$]. Die anderen Werte der VA erlangten keine Signifikanz (alle $Fs < 1$). Auffallend ist, dass die Ergebnisse falscher Erinnerungen in der zweiten Testphase sich deutlich von denen der ersten Testphase nach der

Standardinstruktion unterscheiden, denn Erwachsene hatten nun erheblich weniger falsche Erinnerungen.

8.1.4 Diskussion

Im vorliegenden Experiment wurden die Listenlänge und die Testinstruktion variiert.

Das Ziel der Listenlängenvariation war zu prüfen, ob eine Anpassung der Listenlänge an das Alter sinnvoll ist und ob alterskorrelierte Differenzen falscher Erinnerungen und der Quelldiskrimination allein auf Unterschieden in der Güte des Listenlernens zurückgehen. Nach Forschungsbefunden von Seamon und Mitarbeitern (2002c) zur Güte des Listenlernens sollte eine besser gelernte Liste zu einer besseren Quelldiskrimination führen.

Wurde die Listenlänge an das Alter angepasst, so konnten alle Altersgruppen nach der Inklusionsinstruktion eine vergleichbare Anzahl von Listenitems reproduzieren. Mit angepasster Listenlänge konnten nach der Standardinstruktion dennoch nicht die Unterschiede zwischen Kindern und Erwachsenen im Reproduzieren der Listenitems verhindert werden, denn Erwachsene nannten vermutlich aufgrund eines liberalen Antwortverhaltens (s.u.) mehr Listenitems. Eine Anpassung der Länge konnte nur bewirken, dass die Unterschiede zwischen Vorschülern und Drittklässlern aufgehoben wurden. Anders als in Voruntersuchungen ist es nicht vollständig gelungen, dass alle Altersgruppen die Listen vergleichbar gut lernen konnten. Jedoch schränkt dies die weitere Interpretation der Ergebnisse nicht ein, denn es zeigte sich, dass die Listenlänge keinen Einfluss auf das Reproduzieren kritischer Items nach der Inklusionsinstruktion hatte (also auf die Aktivierung), ebenso wenig auf die Quelldiskrimination. Auch nach der Standardinstruktion zeigte die angepasste Listenlänge keine Effekte, aber aufgrund von Bodeneffekten (s. u.) sind diese Ergebnisse nicht interpretierbar. Der Nutzen einer altersangepassten Listenlänge erscheint ausgehend von dieser Befundlage fraglich. Aus ökonomischen Gründen und weil in entwicklungspsychologischen Untersuchungen üblicherweise das gleiche Lernmaterial für alle Altersgruppen verwendet wird, kann in den folgenden Experimenten auf die Anpassung der Listenlänge an das Alter verzichtet werden.

Da Studien an Erwachsenen zeigten, dass falsche Erinnerungen mit steigender Listenlänge zunehmen (Hancock et al., 2003; Robinson & Roediger, 1997), sollte in der eigenen Untersuchung auch geprüft werden, ob lange Listen vermehrt zu falschen Erinnerungen führen als kurze, was durch eine stärkere Aktivierung oder eine schlechtere Quelldiskrimination erklärt werden könnte. Eine vergleichbare Untersuchung mit 10-jährigen Kindern und 21-jährigen Erwachsenen von Sugrue und Hayne (2006) zeigte eine Zunahme falscher Erinne-

rungen in der Freien Reproduktion mit steigender Listenlänge für beide Altersgruppen, was durch eine Zunahme an Aktivierung erklärt wurde. Jedoch war der Anstieg falscher Erinnerungen bei Kindern deutlich geringer als bei Erwachsenen, so dass eine geringere Aktivierung bei ihnen angenommen wurde. Im Fall von kurzen Listen mit sieben Items gab es zudem keinen Altersunterschied falscher Erinnerungen zwischen Kindern und Erwachsenen, jedoch hatten Erwachsene bei langen Listen mit 14 Items deutlich mehr falsche Erinnerungen. Dies führen die Autoren darauf zurück, dass Erwachsene bedingt durch ihr höheres und besser verknüpftes Wissen generell zu einer stärkeren Aktivierung neigen, bei kurzen Listen aber ihre besseren Quellendiskriminationsfähigkeiten der Aktivierung entgegenwirken konnten. Bei langen Listen konnten die Quellendiskriminationsfähigkeiten nicht mehr die stärkere Aktivierung bei Erwachsenen ausgleichen, weshalb Erwachsene in diesem Fall mehr falsche Erinnerungen als Kinder hatten.

Dieser Effekt der Zunahme falscher Erinnerungen mit steigender Listenlänge war in der eigenen Untersuchung nicht vorzufinden. Somit konnten die Befunde von Robinson und Roediger (1997) und Sugrue und Hayne (2006) nicht repliziert werden. Es zeigte sich nur für die korrekten Erinnerungen, dass im Fall der kurzen Listen mehr Listenitems reproduziert werden konnten als bei langen Listen. Dies ist in Einklang mit den Befunden dieser Studien (Robinson & Roediger, 1997; Sugrue & Hayne, 2006) und lässt sich dadurch erklären, dass längere Listen die itemspezifische Verarbeitung erschweren (Gallo & Roediger, 2003).

Ein Ziel des vorliegenden Experiments bestand in der getrennten Erfassung von Aktivierung und Quellendiskrimination. Zu diesem Zweck wurde eine Inklusionsinstruktion erteilt, die in einer anschließenden Freien Reproduktion ausschließlich die Aktivierung der kritischen Items erfassen sollte. Die Aktivierung sollte sich in der Anzahl der genannten kritischen Items widerspiegeln. Erwartungsgemäß wurden nach der Inklusionsinstruktion mehr kritische Items genannt als nach der Standardinstruktion. Dies ist in Einklang mit den bisherigen Befunden (Dehon & Brédart 2004; Hege & Dodson, 2004; Heit et al., 2004) und unterstützt die Annahme, dass die Aktivierung kritischer Items ohne entgegenwirkende Quellendiskrimination durch die Inklusionsinstruktion erfolgreich erhoben werden konnte. Erstmals wurde an jüngeren Altersgruppen die Auswirkung der Inklusionsinstruktion geprüft, wodurch sich zeigte, dass diese auch bei Vorschülern geeignet ist die Aktivierung kritischer Items ohne entgegenwirkende Quellendiskrimination zu erfassen.

Auch sollten alterskorrelierte Differenzen im vorliegenden Design überprüft werden, um zu testen, welche Rolle Aktivierung und Quellendiskrimination in den eventuellen alterskorrelierten Veränderungen falscher Erinnerungen spielen.

Erwachsene und Drittklässler hatten nach der Inklusionsinstruktion bedeutend mehr Erinnerungen an kritische Items als Vorschüler, was für eine alterskorrelierte Zunahme der Aktivierung spricht. Zudem zeigte sich, dass die beiden älteren Gruppen generell mehr zusätzliche Items nach der Inklusionsinstruktion nannten als Vorschüler. Drittklässler sowie Erwachsene nannten mehr semantisch verbundene Items, aber Erwachsene nannten darüber hinaus noch mehr unverbundene Items, die sie zu den Listenitems weitergehend assoziierten [z.B. kritisches Item: *Bett*, Listenitem: *Nacht*, (zu kritischem Item) unverbundenes Item: *Sterne*]. Dieses Ergebnis spricht dafür, dass Erwachsene und Drittklässler scheinbar nicht in der Lernphase an die kritischen Items dachten, sondern erst in der Testphase. Wenn die kritischen Items also erst in der Testphase aktiviert wurden, sollte die zunächst höhere Aktivierung der beiden älteren Gruppen relativiert werden.

Ein Vergleich der eigenen Ergebnisse erwachsener Probanden mit den Befunden von Dehon und Brédart (2004) und Hege und Dodson (2004) ist in diesem Zusammenhang möglich (vgl. auch Abschnitt 5.4). Dehon und Brédart (2004) verglichen nicht eine Standard- mit einer Inklusionsinstruktion. Nach jeder Liste sollte frei reproduziert werden und in einer anschließenden Phase wurden den Probanden Liste für Liste ihre selbst reproduzierten Items vorgelegt. Sie sollten nun jene Items nennen, an die sie während der Listenpräsentation stark gedacht, diese aber nicht in der ersten Testphase genannt hatten, weil sie nicht in der Lernphase vorkamen. Somit entsprach die erste Testphase der Standardbedingung und die Summe der ersten und zweiten Testphase der Inklusionsbedingung. Diese Vorgehensweise erscheint besonders anfällig dafür zu sein, dass erst in der Testphase das kritische Item assoziiert wird und damit nicht die Aktivierung der Lernphase erfasst wird.

Hege und Dodson (2004) erteilten im ersten und zweiten Experiment eine Inklusionsinstruktion, die sich an jener von Brainerd und Reyna (1998b) orientierte (sie kontrastierten die Auswirkungen einer Standardinstruktion mit einer so genannten bedeutungsbasierten, "meaning recognition", Instruktion). Es sollten alle Items reproduziert werden, die zuvor gelernt worden waren und in Bezug zu den gelernten Listenitems standen. Nach der Lernphase erhielt die eine Hälfte der Probanden die Inklusionsinstruktion und die andere die Standardinstruktion zur Freien Reproduktion. In einem abschließenden Nachtest wurde die Quellendiskrimination, wie im eigenen Experiment, separat erhoben, indem die Probanden ihre reproduzierten

Items auf einem Blatt erhielten und jene Items markieren sollten, die ihnen tatsächlich präsentiert worden waren. Da nach der Inklusionsinstruktion mehr kritische Items reproduziert wurden als nach der Standardinstruktion, konnten Hege und Dodson (2004) zeigen, dass die gewählte Inklusionsinstruktion in der Freien Reproduktion erfolgreich die Aktivierung der kritischen Items ohne Quellendiskrimination erfasst hat.

Zusammenfassend ist festzuhalten, dass wie auch im eigenen Experiment nach den Inklusionsinstruktionen von Dehon und Brédart (2004) Hege und Dodson (2004) mehr kritische Items reproduziert wurden als nach den Standardinstruktionen, was dafür spricht, dass tatsächlich nur die Aktivierung kritischer Items erfasst wurde und die Erinnerungsleistung nicht von der gegenläufigen Quellendiskrimination beeinflusst war. Das Vorgehen im eigenen Experiment ähnelt jenem von Hege und Dodson (2004) darin, dass eine Inklusions- und eine Standardinstruktion direkt miteinander verglichen werden, denn die Gefahr, dass das kritische Item in der Testphase assoziiert wird, erscheint anhand der Vorgehensweise von Dehon und Brédart (2004) deutlich größer. Da die wortgetreuen Instruktionen von Hege und Dodson (2004) leider nicht angegeben sind, ist ein genauer Vergleich der Inklusionsinstruktionen schwierig. Im eigenen Experiment wurde darauf Wert gelegt anhand einer beispielhaften Wortliste die Inklusionsinstruktion besonders gut (vor allem für Kinder) verständlich zu machen. Zudem wurden die Probanden im eigenen Experiment angewiesen jene Items zu nennen, an die sie *während* der Präsentation der Wortlisten in der Lernphase stark gedacht hatten. Dies erscheint konkreter als bei Hege und Dodson, die lediglich die Instruktion erteilten, alle Items zu nennen, die in Bezug zu den gelernten Items standen.

Aufschluss über die Güte der Quellendiskrimination gab die zweite Testphase, in der die Probanden ihre genannten Items danach beurteilen sollten, ob sie tatsächlich von der Versuchsleiterin vorgelesen worden waren. Erwartungskonform differierten Drittklässler und Erwachsene nicht in der Quellenzuweisung von Listenitems, was den Forschungsbefunden von Foley und Johnson (1985) entspricht. Hierbei handelte es sich um eine einfache extern-intern Entscheidung, die schon von Drittklässlern sehr gut bewältigt werden konnte.

Erwachsene konnten auch kritische Items erwartungsgemäß besser der korrekten Quelle zuordnen. Sie waren Kindern in der Quellendiskrimination eindeutig überlegen.

Die vielen genannten kritischen Items erwachsener Probanden nach beiden Instruktionen wurden im Nachhinein dennoch nicht der falschen externen Quelle zugeordnet, so dass sie sich augenscheinlich im Klaren waren, dass sie diese selbst genannt hatten.

Die Ergebnisse zu alterskorrelierten Verbesserungen der Quellendiskrimination entsprechen den theoretischen Annahmen, dass Kinder und Senioren Quellendiskriminationsprobleme haben (Roediger & McDermott, 2000), was aber bisher mit Kindern im DRM-Paradigma noch nicht empirisch geprüft wurde. Zusammenfassend lassen sich die Quellendiskriminationsprobleme von Kindern sowohl auf die Lern- als auch auf die Testphase zurückführen.

In der Lernphase können Kinder aufgrund ihrer unzureichenden Frontallappenfunktion (vgl. auch Norman & Schacter, 1997; Schacter et al., 1999; Schacter et al., 1996) und der daraus resultierenden schlechteren Arbeitsgedächtniskapazität weniger quellenrelevante Informationen verarbeiten. Neuere Untersuchungen an jüngeren Erwachsenen und Senioren zeigen, dass letztere wegen der schlechteren Frontallappenfunktion Enkodierdefizite und Probleme bei der Strategiewahl haben (Jacoby, Bishara, Hessels & Toth, 2005; Thomas & Sommers, 2005). Bislang gibt es keine Untersuchungen an Kindern zu alterskorrelierten Veränderungen der Quellendiskrimination, aber durch die ähnliche unzureichende Frontallappenfunktion bei Kindern könnten die Befunde von Senioren auch auf Kinder übertragen werden.

In der Testphase haben Kinder Probleme bei den quellenstrategischen Prozessen aufgrund ihrer geringeren Arbeitsgedächtniskapazität (Johnson et al., 1993). Nach dem Quellendiskriminationsmodell von Johnson et al. (1993, s. Abschnitt 2.2.2) lässt sich schließen, dass der strategische Prozess (und nicht der heuristische) gefordert wurde, der starken alterskorrelierten Verbesserungen unterliegt. Kindern fiel die Aufgabe besonders schwer, denn vermutlich konnten sie nicht genügend itemspezifische Informationen zur Unterscheidung gelernter und nicht gelernter Items enkodieren verglichen mit Erwachsenen, die über bessere Strategien verfügen.

Ein Vergleich der Befunde zur Quellendiskrimination erwachsener Probanden mit jenen der Untersuchung von Hege und Dodson (2004) war auch hier möglich, denn die Probanden sollten dort ihre reproduzierten Items in einer weiteren Testphase nach dem Präsentationsstatus („Kam das genannte Item in der Lernphase tatsächlich vor?“) beurteilen. Die Probanden des eigenen Experiments wiesen kritische Items seltener der falschen Quelle zu als die Versuchspersonen von Hege und Dodson. Dies könnte daran liegen, dass die Probanden des eigenen Experiments, wie bereits erwähnt, zuerst viele kritische Items nannten, dann aber proportional wenige Items der falschen Quelle zuordneten.

Die Ergebnisse zu den genannten kritischen Items nach der Standardinstruktion, also den falschen Erinnerungen, sind aufgrund von Bodeneffekten nicht interpretierbar. Vorschüler erinnerten nur 6% der kritischen Items falsch, Drittklässler 9%, aber Erwachsene immerhin 39%.

Es erscheint wahrscheinlich, dass die Erwachsenen der Standardinstruktion nicht Folge leisteten, sondern so viele Items wie möglich versuchten zu reproduzieren. Dies wird dadurch gestützt, dass sie viele Listenitems reproduzierten und mehr Intrusionen hatten als die jüngeren Probanden, was alles ein Indiz für die unterschiedlichen Verarbeitungsstrategien der Altersgruppen ist. Nicht nur die Inklusions- sondern auch die Standardinstruktion zum Freien Reproduzieren vermochte bei den Erwachsenen den Eindruck erweckt haben, dass die Anzahl erinnerter Items die Güte der eigenen Erinnerungsleistung reflektiert („Je mehr desto besser“) unabhängig von deren tatsächlicher Präsentation, obwohl sie aufgefordert wurden nicht zu raten. Somit sind die hohen falschen Erinnerungen erwachsener Probanden auf ein selbst gewähltes, extrem liberales Antwortverhalten zurückzuführen. Es könnte möglich sein, dass nach der ersten Inklusionsinstruktion bei erwachsenen Probanden der Eindruck entstand, dass auch nach der Standardinstruktion möglichst viele Items genannt werden sollten, was in weiteren Experimenten durch eine bessere Trennung der Instruktionen genau geprüft werden sollte.

Eine Inspektion der Mittelwerte „echter“ falscher Erinnerungen, d.h. der Anteil der kritischen Items, die reproduziert und im Nachtest als von der Versuchsleiterin als vorgelesen bezeichnet wurden, zeigte nach der Standardinstruktion keinen Altersunterschied. Verglichen mit der ersten Testphase reduzierte sich der Anteil falscher Erinnerungen für die erwachsenen Probanden deutlich. Drittklässler hingegen, die von vorneherein weniger falsche Erinnerungen hatten, korrigierten diese vergleichsweise seltener. Nach der Inklusionsinstruktion attribuierten Drittklässler ihre reproduzierten kritischen Items häufiger auf die falsche externe Quelle als die älteren Probanden. Somit können diese Befunde zu den „echten“ falschen Erinnerungen im Zusammenhang mit ihren schlechteren Fähigkeiten zur Quellendiskrimination stehen, sollten aber aufgrund der sehr geringen Werte nach der Standardinstruktion weiter geprüft werden.

Die Ergebnisse des vorliegenden Experiments führen zu neuen Überlegungen. Die geringen falschen Erinnerungen bei Kindern lagen, wie die Leistungen im Inklusionstest zeigten, nicht daran, dass die kritischen Items nicht aktiviert wurden. Im Gegensatz zu Brainerd et al. (2002a) hatten auch schon Vorschüler das Thema der Liste erfasst, denn sie zeigten kaum proaktive Hemmung. Deutschsprachige Listen, wie auch französischsprachige, führen im Allgemeinen zu geringeren falschen Erinnerungen als englischsprachige, vermutlich deshalb, da Wörter in der deutschen und französischen Sprache mehr Silben beinhalten und perzeptuell distinkter sind. Durch die Wahl eines Wiedererkennenstests könnte dem Hauptproblem der

geringen Raten falscher Erinnerungen (Bodeneffekte bei den beiden jüngeren Altersgruppen nach der Standardinstruktion) entgegengewirkt werden, da diese in Wiedererkennentests höher liegen als in der Freien Reproduktion. Indem ein Wiedererkennentest anstatt der Freien Reproduktion eingesetzt wird, können Erwachsene auch nicht mehr nach dem Prinzip „Je mehr desto besser“ antworten, sondern werden sich eher an die Instruktion halten. Somit sollten eventuelle alterskorrelierte Unterschiede in den falschen Erinnerungen besser interpretierbar sein.

Es könnte getestet werden, ob die Abfolge der Instruktionen einen Einfluss auf falsche Erinnerungen hat, also ob unterschiedliche Ergebnisse, je nachdem ob zuerst die Standard- oder die Inklusionsinstruktion erteilt wird, resultieren. Es erscheint sinnvoll in folgenden Experimenten die Auswirkungen der Instruktionsabfolge genauer zu betrachten, da Erwachsene sich möglicherweise von der Inklusionsinstruktion beeinflussen ließen und infolgedessen sich weniger an die Standardinstruktion hielten.

Im nächsten Experiment, das aus zwei Teilexperimenten (3A und 3B) besteht, wird neben einem Standardwiedererkennentest auch ein Inklusionstest eingesetzt (vgl. Heit et al., 2004). Die Altersunterschiede im korrekten Erinnern sollten geringer sein als in der Freien Reproduktion.

Da die gewählte Inklusionsinstruktion von allen Altersgruppen verstanden wurde und die Resultate interpretierbar sind, ist ein weitergehender Einsatz möglich. Dabei muss eine vergleichbare Instruktion für das Wiedererkennen entwickelt werden, die von allen Altersgruppen verstanden wird und die sich hinreichend von der Standardinstruktion unterscheidet.

8.2 Experiment 3A: Falsches Wiedererkennen: Einflüsse von Aktivierung und Quellendiskrimination

Das vorliegende Experiment gliedert sich in zwei Telexperimente: Während das erste Experiment (3A) vorwiegend eine Replikation des zweiten Experiments mit dem Wiedererkennenstest ist, soll im zweiten Telexperiment (3B) zusätzlich eine Aufgabe erteilt werden, die eine semantisch-relationale Verarbeitung der Listen verlangt. Durch den Vergleich beider Experimente soll geprüft werden, welchen Einfluss diese Lerninstruktion auf die Aktivierung, die Quellendiskrimination und falsche Erinnerungen hat.

In Experiment 3A sollte untersucht werden, ob sich alterskorrelierte Differenzen in der Aktivierung und Quellendiskrimination auch im Wiedererkennen zeigen. Indem eine Inklusionsinstruktion eingesetzt wurde, sollte Aktivierung getrennt von der Quellendiskrimination erfasst werden (Heit et al., 2004). Damit unterschied sich dieses Experiment vom zweiten vorrangig im eingesetzten Gedächtnistest. Während im vorherigen Experiment für alle Altersgruppen sehr wenige falsche Erinnerungen in der Freien Reproduktion resultierten, so sollten diese im Wiedererkennenstest häufiger auftreten.

Neben der Instruktion wurde die Abfolge der Instruktionen variiert, da kontrolliert werden sollte, ob diese einen Einfluss auf die Gedächtnisleistung hat (i.S. von "carry over" Effekten), was nicht im zweiten Experiment geprüft wurde. Eine Hälfte der Versuchspersonen erhielt erst die Standard- und dann die Inklusionsinstruktion, während der anderen Hälfte der Probanden zuerst die Inklusions- und danach die Standardinstruktion präsentiert wurde. Im Fall von Abfolgeeffekten werden diese in einem separaten Abschnitt des Ergebnisteils berichtet. Aufgrund dieser Effekte lassen sich die Ergebnisse schlecht interpretieren. Daher wurden separate Analysen nur mit den zuerst erteilten Tests gerechnet, die unbeeinflusst von der Abfolge waren. Die Interpretation der Ergebnisse bezieht sich daher nur auf die zuerst erteilten Tests.

Die Listenlänge wurde in diesem Experiment nicht variiert, denn die Ergebnisse des zweiten Experiments zeigten, dass die Aktivierung, die Quellendiskrimination und falsche Erinnerungen von der Listenlänge unbeeinflusst waren. Für alle Altersgruppen wurde daher eine identische Listenlänge von 10 Items gewählt, die sich aus den Ergebnissen des zweiten Experiments ergab. Zudem ähnelt die Wahl gleich langer Listen anderen entwicklungspsychologischen Studien und ein direkter Vergleich der Ergebnisse zu falschen Erinnerungen wird damit ermöglicht.

Der Versuchsablauf gestaltete sich so, dass in der ersten Phase acht DRM-Listen vorgelesen wurden, gefolgt entweder von einer Standardwiedererkennensinstruktion (jedes Item als „alt“ oder „neu“ beurteilen) oder Inklusionsinstruktion (Items, an die während der Präsentation der Listen gedacht wurde, sollen auch als „alt“ bezeichnet werden). Danach schloss sich die zweite Lern-/Testphase an. Nach dem Inklusionstest wurde ein Quellendiskriminationstest bearbeitet, in dem alle zuvor als „alt“ bezeichneten Items danach beurteilt werden sollten, ob die Versuchsleiterin diese in der Lernphase vorgelesen hatte.

Für den Wiedererkennenstest wurden die Erwartungen 1, 3, 4 und 5 aus Experiment 2 übernommen:

- (1) Spiegelt die Anzahl wieder erkannter kritischer Items nach der Inklusionsinstruktion allein deren Aktivierung ohne den Einfluss von Quellendiskrimination wider, dann sollten alle drei Altersgruppen nach dieser bedeutend mehr kritische Items als „alt“ bezeichnen verglichen mit der Standardinstruktion.
- (2) Es wird von einem alterskorrelierten Anstieg der Aktivierung ausgegangen. Deshalb sollten sich die Altersgruppen in der Häufigkeit wieder erkannter kritischer Items nach der Inklusionsinstruktion unterscheiden.
- (3) Erwartet werden auch alterskorrelierte Verbesserungen der Quellendiskrimination. Daher sollte sich bei der sehr schwierigen Quellenunterscheidungsaufgabe kritischer Items ein Alterseffekt zeigen.
- (4) Zwar wird angenommen, dass Aktivierung und Quellendiskrimination alterskorreliert zunehmen, aber dennoch lassen sich keine Vorhersagen zu Altersunterschieden in der Häufigkeit falsch wieder erkannter kritischer Items nach der Standardinstruktion treffen, einerseits aufgrund der gemischten Befundlage in der Literatur (Brainerd et al., 2002a; Ghetti et al., 2002; vgl. Abschnitt 4.2), andererseits da unklar ist, ob der Einfluss von Aktivierung oder Quellendiskrimination stärker ist (Brainerd et al., 2002a; Ghetti et al., 2002; vgl. Abschnitt 4.2).

8.2.1 Vorversuch

Um die Verständlichkeit der Instruktionen zu sichern und um zu gewährleisten, dass die Inklusions- verglichen mit der Standardinstruktion zu häufigeren „alt“-Urteilen führt, wurden im Rahmen einer Voruntersuchung zwei verschiedene Inklusionsinstruktionen und eine Standardinstruktion an einer Stichprobe von 16 Vorschülern eines Kindergartens getestet. Wie in der Voruntersuchung des ersten Experiments wurden den Probanden 12 DRM-Wortlisten mit

jeweils 10 Items vorgelesen. Anschließend folgte eine Instruktion zum Wiedererkennen. Die drei verschiedenen Instruktionen wurden interindividuell variiert, so dass vier Probanden die Standardinstruktion erhielten und jeweils sechs Kinder eine der Inklusionsinstruktionen. Anhand der Ergebnisse bisheriger Untersuchungen zu der Auswirkung von verschiedenen Inklusionsinstruktionen (Brainerd et al., 2004; Heit et al., 2004) und eigenen Überlegungen wurden die Instruktionen zusammengestellt. Inklusion A ähnelte jener aus der ersten Voruntersuchung darin, dass assoziierte Beispielitems und das zugehörige kritische Item präsentiert und deren Zusammenhang erklärt wurden. Die Versuchspersonen wurden aufgefordert sowohl präsentierte Items als auch jene, an die sie während der Lernphase selbst stark gedacht hatten, als „alt“ zu bezeichnen. Demgegenüber forderte Inklusion B ein liberaleres Antwortkriterium, denn im Fall einer Unsicherheit über den Präsentationsstatus sollten die Kinder das jeweilige Item im Wiedererkennentest als „alt“ bezeichnen. Im anschließenden Test wurden in quasi-randomisierter Reihenfolge 48 Items präsentiert, von denen jeweils 12 kritische Items, 12 Listenitems, 12 schwach assoziierte und 12 nicht assoziierte Items waren.

Die statistische Auswertung ergab einen Unterschied zwischen den drei Instruktionen dahingehend, dass nach der Standardinstruktion weniger kritische und schwach assoziierte Items als „alt“ bezeichnet wurden als nach den beiden Inklusionsvarianten, die sich wiederum nicht voneinander in der Häufigkeit wieder erkannter Items unterschieden. Für die Hauptuntersuchung wurde Inklusionsinstruktion A ausgewählt, da sie analog zur Instruktion aus Experiment 2 ist und in bisherigen Untersuchungen verwendet wurde (Brainerd et al., 2004).

8.2.2 Methode

Versuchsplan. Der Versuch wurde in einem 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Testabfolge: Inklusion-Standard vs. Standard-Inklusion) x 2 (Instruktion: Standard vs. Inklusion) Versuchsplan mit Messwiederholung auf dem letzten Faktor realisiert.

Versuchspersonen. An Experiment 3A nahmen 32 Vorschüler (16 Mädchen und 16 Jungen) eines Kindergartens, 32 Drittklässler (21 Mädchen und 11 Jungen) einer Grundschule und 32 Zwölftklässler (19 Frauen und 13 Männer) eines Gymnasiums teil. Die Vorschüler hatten ein durchschnittliches Alter von sechs Jahren und einem Monat (zwischen 5;3 und 7;2 Jahren), die Drittklässler waren im Mittel neun Jahre und sieben Monate alt (zwischen 8;11 und 9;11 Jahren) und die Zwölftklässler 18 Jahre und sechs Monate (zwischen 18;1 und 19;6 Jahren).

Material. Die verwendeten Wortlisten entsprechen dem vorherigen Experiment (mit 10 Items). Für alle Altersgruppen wurden zwei Listensets (A, B) entworfen mit jeweils acht Wortlisten à 10 Items. Durch die Kreuzung der Listensets mit den beiden Instruktionen (Standard, Inklusion) resultieren vier Lernbedingungen für alle Altersgruppen. Die Versuchspersonen wurden per Zufall den Bedingungen zugeordnet, so dass die Hälfte der Probanden jeder Altersgruppe Set A mit Standardinstruktion und Set B mit Inklusionsinstruktion erhielt. Für die andere Hälfte der Teilnehmer gestaltete sich die Präsentation gegenteilig. Innerhalb der Gruppen wurde die Testabfolge variiert (A-B oder B-A).

Nach der Präsentation von acht Wortlisten folgte der erste Wiedererkennenstest, danach wurden erneut acht Listen vorgelesen und abschließend wurde ein letzter Wiedererkennenstest bearbeitet.

Jeder Wiedererkennenstest bestand aus 40 Items in quasi-randomisierter Reihenfolge (d.h. alle Items wurden zufällig angeordnet, jedoch wurde darauf geachtet, dass nie zwei Items einer selben Liste direkt einander folgten), von denen acht kritische Items, 16 Listenitems (zwei aus jeder Wortliste mit den Assoziationsstärken der Positionen 3 bis 10, jedoch keine aus mehreren Silben bestehenden Wörter), acht schwach assoziierte Items (eins zu jeder Liste mit geringen Assoziationsstärken der Positionen 15 bis 20) und acht nicht assoziierte Items waren. Das verwendete Material kann Anhang C1 entnommen werden.

Durchführung. Die Testsituation dauerte insgesamt ca. 20 Minuten. Die Versuchsdurchführung gestaltete sich in sechs aufeinander folgenden Phasen [Lernphase (1), Wiedererkennen (1), (*Quellendiskrimination*), Ablenkaufgabe (1), Lernphase (2), Wiedererkennen (2), (*Quellendiskrimination*)]. Die Aufgabe zur Quellendiskrimination erfolgte entweder in Anschluss an den ersten oder an den zweiten Wiedererkennenstest.

In der ersten Lernphase wurden acht DRM-Wortlisten nacheinander vorgelesen. Bevor die erste Testphase begann wurde den Versuchspersonen entweder eine Standard- oder eine Inklusionsinstruktion zum Wiedererkennen gegeben (genauer Wortlaut der Instruktionen im Anhang C2). Die Standardinstruktion forderte die Probanden auf zu jedem vorgelesenen Wort ein „alt/neu“-Urteil abzugeben. Dabei sollten alle in der Lernphase präsentierten Wörter als „alt“ bezeichnet werden und alle nicht präsentierten Wörter als „neu“. Analog zum Wiedererkennenstest im ersten Experiment wurden in der Inklusionsinstruktion den verschiedenen Altersgruppen dem Alter angemessene assoziierte Beispielitems (für Vorschüler und Drittklässler: „Lehrer, rechnen, schreiben, lesen“; für Erwachsene: „Silber, Schmuck, wertvoll, reich, teuer, Geld, glänzen, Mine, Barren, Bronze“) mit zugehörigem Zielitem (für Vorschüler und

Drittklässler: „Schule“; für Erwachsene: „Gold“) zur Erklärung der Instruktion präsentiert. Den jüngeren Altersgruppen wurde der Bezug der assoziierten Items zum Oberbegriff erklärt und alle Probanden wurden aufgefordert nicht nur die präsentierten Wörter als „alt“ zu bezeichnen, sondern auch jene, an die sie während der Lernphase stark gedacht hatten. Für die Probanden der Inklusionsinstruktion schloss sich eine zweite Phase an, in welcher die Probanden ihre als „alt“ bezeichneten Items nochmals beurteilen sollten, ob diese tatsächlich von der Versuchsleiterin vorgelesen worden waren. Alle Items wurden in der alten Reihenfolge von der Versuchsleiterin vorgelesen und die Probanden sollten jedes Item einer Quelle zuordnen.

Bevor die nächste Lernphase begann, erfolgte eine nonverbale Ablenkaufgabe. Anschließend wurde die andere Hälfte der Wortlisten vorgelesen und die ausstehende Standard- oder Inklusionsinstruktion erteilt. Jene Probanden, die eine Inklusionsinstruktion erhielten, bearbeiteten auf oben beschriebene Weise den Wiedererkennenstest mit späterem Quellenurteil.

8.2.3 Ergebnisse

Als abhängige Variablen wurden die Raten der „Treffer“ für kritische Items, die der Treffer für Listenitems und die falsche Alarmrate für Distraktoren berechnet, ebenso die Pr- und Br-Werte der beiden Itemtypen.

Die Analysen zeigten deutliche Effekte der Testabfolge. Daher werden im Folgenden die wichtigsten Ergebnisse der 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Abfolge der Instruktionen: Standard-Inklusion vs. Inklusion-Standard) x 2 (Instruktion: Standard vs. Inklusion) VA mit Messwiederholung auf dem letzten Faktor für die verschiedenen abhängigen Variablen getrennt nach Listenitems und kritischen Items vorab dargestellt und diskutiert.

Für den Hauptergebnisteil und zur Prüfung der Hypothesen wurden separate 3 (Altersgruppe) x 2 (Instruktion) VAn ohne Messwiederholung nur mit den in der ersten Phase erteilten Tests (Standard oder Inklusion) gerechnet, so dass die Ergebnisse unbeeinflusst von der anderen zuvor erteilten Instruktion waren und damit eindeutig interpretiert werden können.

Da in Experiment 3B zwei Listen eines jeden Sets (A: Sommer, Kerze; B: Löwe, Hemd) mit sehr seltenen Nennungen der kritischen Items einhergingen (nach der Aufforderung „Das fehlende, mit der Liste zusammenhängende Wort, sollst du mir nach jeder Liste sagen“, wurde das kritische Item selten genannt) und um einen Vergleich der beiden Experimente zu ermög-

lichen, wurden für die statistischen Analysen dieses und des nächsten Experiments nur sechs der insgesamt acht Listen eines jeden Sets herangezogen. In statistischen Analysen zeigte sich, dass es hinsichtlich des Ergebnismusters keinen Unterschied erzeugte, ob die besten sechs oder alle acht Listen verwendet wurden. Bei acht Listen wurden wegen der höheren Power tendenziell signifikante Effekte signifikant. Nur bei Unterschieden werden die Ergebnisse mit acht Listen berichtet, die Mittelwerte der Analysen werden aber zum Überblick in die Tabellen hinzugefügt.

Neben den „alt“-Antworten wurden zur Erfassung des korrekten und falschen Wiedererkennens korrigierte Werte berechnet: „alt“-Antworten für die kritischen Items (bzw. für die Listenitems beim korrekten Wiedererkennen) subtrahiert um die „alt“-Antworten für die Distraktoren (gemittelt über verbundene und unverbundene Distraktoren; vgl. z.B. Dodd & MacLeod, 2004).

8.2.3.1 Einflüsse der Testabfolge

Die Analysen zeigten zahlreiche Testabfolgeeffekte. Zusammenfassend sollen die wichtigsten Ergebnisse der Listenitems und kritischen Items unter Einbezug der Testabfolge nun dargestellt und diskutiert werden.

Listenitems. Die 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Abfolge der Instruktionen: Standard-Inklusion vs. Inklusion-Standard) x 2 (Instruktion: Standard vs. Inklusion) VA mit Messwiederholung auf dem letzten Faktor zum Antwortkriterium (Br-Werte) der Listenitems erbrachte einen tendenziellen Abfolgeeffekt [$F(1,90) = 3.82$, $R^2 = 0.04$, $p = .06$]: Wurde die Inklusions- vor der Standardinstruktion erteilt, dann war das Antwortkriterium tendenziell liberaler ($M = .40$) als in der umgekehrten Abfolge ($M = .32$), was auf den starken Einfluss der Inklusionsinstruktion auf die gesamte Testung hinweist. Die Inklusionsinstruktion könnte die Probanden beeinflusst haben generell liberaler zu antworten und häufiger Listenitems als „alt“ zu bezeichnen.

Die weiteren Analysen für die Listenitems zeigten keine Effekte der Abfolge. Zur Übersicht werden aber die Mittelwerte für Trefferrate und Pr-Werte der Listenitems unter Einbezug der Abfolge in Tabelle 8.2.1 und 8.2.2 zusammengefasst.

Tabelle 8.2.1:

Mittelwerte der „alt“-Antworten für die Listenitems in Abhängigkeit des Alters, Instruktionsabfolge und der Instruktion.

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.55	.56	.56	.60	.65	.62
Drittklässler	.62	.61	.62	.65	.71	.68
Erwachsene	.76	.77	.77	.77	.83	.80
8 Listen						
Vorschüler	.55	.55	.55	.60	.61	.60
Drittklässler	.61	.61	.61	.65	.71	.68
Erwachsene	.75	.75	.75	.79	.81	.80

Tabelle 8.2.2:

Mittlere Pr-Werte der Listenitems in Abhängigkeit des Alters, der Instruktionsabfolge und der Instruktion.

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.42	.39	.40	.39	.44	.42
Drittklässler	.46	.40	.43	.45	.46	.46
Erwachsene	.67	.56	.62	.57	.56	.57
8 Listen						
Vorschüler	.41	.38	.40	.38	.41	.40
Drittklässler	.47	.42	.44	.43	.45	.44
Erwachsene	.67	.54	.60	.57	.54	.55

Kritische Items. Wie bei der Analyse des Antwortkriteriums der Listenitems zeigte sich in der dreifaktoriellen VA ein Abfolgeeffekt für die Br-Werte kritischer Items [$F(1,90) = 6.57$, $R^2 = 0.07$]. Wurde die Inklusionsinstruktion zuerst erteilt, resultierte ein liberaleres Kriterium ($M = .36$) als in der Abfolge Standard-Inklusion ($M = .25$). Die Inklusionsinstruktion hatte somit einen Einfluss auf das Antwortverhalten der gesamten Testung, d.h. liberaler zu antworten und zwar für Listenitems und kritische Items.

Um den Anteil falscher Erinnerungen im Inklusionstest zu prüfen, wurden jene kritischen Items zur Analyse herangezogen, die im Wiedererkennenstest nach der Inklusionsinstruktion als „alt“ bezeichnet wurden und in der zweiten Testphase falsch auf die externe Quelle attribuiert worden waren (d.h. die Versuchsleiterin sollte die kritischen Items vorgelesen haben). Die Mittelwerte sind in Tabelle 8.2.3 in Klammern abgebildet. In dieser Analyse zeigte sich ein Interaktionseffekt der Abfolge mit der Instruktion [$F(1,90) = 9.14$, $R^2 = 0.09$]: Eine Analyse der einfachen Haupteffekte konnte zeigen, dass bei der Abfolge Standard-Inklusion [$F(1,90) = 5.19$] nach der Standardinstruktion ($M = .63$) mehr falsche Erinnerungen zustande

kamen als nach der Inklusionsinstruktion ($M = .56$). Nach der Inklusionsinstruktion [$F(1,90) = 6.46$] war der Anteil falscher Erinnerungen bei der Abfolgebedingung Inklusion-Standard ($M = .63$) bedeutend höher als bei der umgekehrten Abfolge ($M = .51$). Alle anderen einfachen Haupteffekte wurden nicht signifikant (alle $F_s < 1.20$). Somit neigten die Probanden bei beiden Abfolgebedingungen in der ersten Testphase zu mehr falschen Erinnerungen als in der zweiten. Dies kann daran liegen, dass sie die erste Phase als Vorwarnung nutzen konnten und entsprechend in der zweiten Phase ihre falschen Erinnerungen reduzieren konnten.

In den VAn für die als „alt“ bezeichneten kritischen Items (Mittelwerte s. Tabelle 8.2.3) sowie für die korrigierten Werte (Mittelwerte s. Tabelle 8.2.4) zeigten sich deutliche Effekte der Abfolge. Aufgrund der vergleichbaren Ergebnisse der beiden VAn werden nur die Ergebnisse mit den korrigierten Werten berichtet.

Tabelle 8.2.3:

Mittelwerte der „alt“-Antworten für die kritischen Items in Abhängigkeit des Alters, der Instruktionsabfolge und der Instruktion (in Klammern Werte für falsche Erinnerungen nach der Inklusionsinstruktion).

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.55	.54	.55	.74 (.54)	.66 (.57)	.70 (.56)
Drittklässler	.66	.59	.63	.70 (.54)	.81 (.71)	.76 (.63)
Erwachsene	.67	.55	.61	.82 (.45)	.91 (.59)	.87 (.52)
8 Listen						
Vorschüler	.56	.54	.55	.71 (.50)	.66 (.57)	.68 (.54)
Drittklässler	.66	.61	.63	.70 (.52)	.82 (.70)	.76 (.61)
Erwachsene	.69	.62	.65	.83 (.46)	.89 (.62)	.86 (.54)

Tabelle 8.2.4:

Mittlere Pr-Werte der kritischen Items in Abhängigkeit des Alters, der Instruktionsabfolge der Instruktion.

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.42	.38	.40	.53	.45	.49
Drittklässler	.50	.38	.44	.50	.56	.53
Erwachsene	.58	.34	.46	.62	.64	.63
8 Listen						
Vorschüler	.42	.38	.40	.49	.47	.48
Drittklässler	.51	.41	.46	.48	.56	.52
Erwachsene	.61	.41	.51	.61	.62	.61

In der VA mit den Pr-Werten (Mittelwerte s. Tabelle 8.2.4) für die kritischen Items kam es zu einem tendenziellen Effekt der Abfolge [$F(1,90) = 3.20$, $R^2 = 0.03$, $p = .08$], welcher mit der Instruktion interagierte [$F(1,90) = 4.35$, $R^2 = 0.05$]. Die Werte der Interaktion sind Tabelle 8.2.5 zu entnehmen.

Tabelle 8.2.5:

Mittlere Pr-Werte der Interaktion „Abfolge der Instruktionen“ und „Instruktion“ für die kritischen Items.

	Standard	Inklusion
Standard-Inklusion	.50	.55
Inklusion-Standard	.37	.55

Dieser Wechselwirkungseffekt geht darauf zurück, dass nach der Standardinstruktion kritische Items seltener falsch wiedererkannt wurden, wenn diese Instruktion nach der zweiten Lernphase erfolgte als nach der ersten [$F(1,90) = 8.72$]. Werden die Mittelwerte aus Tabelle 8.2.4 betrachtet, dann zeigt sich, dass die Interaktion auf den Werten Erwachsener und tendenziell auch Drittklässler zu beruhen scheint. Vorschüler hatten vergleichbare Pr-Werte nach der Standardinstruktion, unabhängig davon, ob diese nach der ersten Lernphase erteilt wurde oder erst nach der zweiten (s. Tabelle 8.2.4). Dies könnte daran liegen, dass die Inklusionsinstruktion für die beiden älteren Gruppen eine Art Vorwarnung für die zweite Phase gewesen sein könnte, in der die Probanden darauf hingewiesen wurden, auch jene Items, an die sie selbst stark gedacht hatten, als „alt“ zu bezeichnen. In Folge könnten Erwachsene und ansatzweise auch Drittklässler in der zweiten Lernphase Strategien eingesetzt haben, um die kritischen Items zu identifizieren. Möglicherweise könnten die älteren Probanden auch in der zweiten Testphase bei der Quellendiskrimination strategisch vorgegangen sein. Jedoch gab es bei Vorschülern keine Effekte der Testabfolge, weil diese entweder den Zusammenhang von

Lern- und Testphasen nicht erkannt hatten oder ihre Arbeitsgedächtniskapazität für den Strategieeinsatz in der Lernphase nicht ausreichend war und/oder sie zu einem entsprechenden strategischen Vorgehen bei der Quellendiskrimination in der Testphase nicht fähig gewesen sein könnten. Für die Inklusionsinstruktion war die Abfolge der Instruktionen nicht relevant ($F < 1$), da vermutlich die Standardinstruktion nicht als Vorwarnung gesehen werden konnte. Alle anderen Interaktionen verfehlten das Signifikanzniveau (alle F s < 1).

Die Berechnung der Fehlattribution für kritische Items (externe Quelle) erfolgte darüber wie viele der als „alt“ wiedererkannten kritischen Items in der zweiten Phase als vorgelesen bezeichnet wurden. Die entsprechenden Mittelwerte sind Tabelle 8.2.6 zu entnehmen.

Tabelle 8.2.6:

Mittelwerte der Fehlattribution kritischer Items in Abhängigkeit des Alters und der Instruktionsabfolge.

	S-I	I-S	Ges.
6 Listen			
Vorschüler	.72	.89	.81
Drittklässler	.76	.87	.82
Erwachsene	.53	.67	.60
8 Listen			
Vorschüler	.77	.88	.83
Drittklässler	.73	.83	.78
Erwachsene	.54	.63	.59

In der dreifaktoriellen VA erreichte die Abfolge einen signifikanten Haupteffekt [$F(1,90) = 11.38$, $R^2 = 0.11$]: Die falsche Quellenzuweisung war im ersten Test ($M = .81$) stets höher als im zweiten ($M = .67$). Es könnte sein, dass bei Erwachsenen und ggf. auch bei Drittklässlern die Verarbeitung in der zweiten Lernphase von der ersten Phase beeinflusst gewesen war. Die älteren Probanden könnten vor der zweiten Lernphase vorgewarnt gewesen sein und durch ein strategisches Vorgehen in der zweiten Testphase darauf geachtet haben, das kritische Item zu identifizieren. Dadurch könnte die Quellenzuweisung im zweiten Wiedererkennenstest beeinflusst gewesen sein.

8.2.3.2 Ergebnisse der separaten Analysen mit den zuerst erteilten Tests

Aufgrund der Testabfolgeeffekte lassen sich die Ergebnisse nicht eindeutig interpretieren. Daher wurden getrennte VAn für die Listenitems und kritischen Items mit den zuerst erteilten

Tests gerechnet, deren Ergebnisse in den folgenden Abschnitten dargestellt werden und eindeutig interpretierbar sind.

Listenitems

Wiedererkennen. Zur Prüfung des Antwortkriteriums wurde eine 3 (Altersgruppe) x 2 (Instruktion) VA mit Messwiederholung auf dem letzten Faktor für die Br-Werte der Listenitems gerechnet, die einen Instruktionseffekt erbrachte [$F(1,90) = 12.26$, $R^2 = 0.12$]. Nach der Standardinstruktion ($M = .25$) war das Antwortkriterium konservativer als nach der Inklusionsinstruktion ($M = .43$). Auch war die Altersgruppe tendenziell signifikant [$F(2,90) = 2.73$, $R^2 = 0.06$, $p = .07$]: Vorschüler ($M = .28$) wählten ein tendenziell konservativeres Kriterium als Erwachsene ($M = .42$). Drittklässler ($M = .32$) unterschieden sich von beiden Gruppen nicht. Die Interaktion der Faktoren war statistisch nicht bedeutsam ($F < 1$).

In der VA für die Trefferrate der Listenitems war die Altersgruppe signifikant [$F(2,90) = 19.60$, $R^2 = 0.30$]: Erwachsene ($M = .80$) konnten erwartungsgemäß mehr Listenitems wiedererkennen als Drittklässler ($M = .67$) und Vorschüler ($M = .60$), die sich nicht im korrekten Wiedererkennen unterschieden. Für die Instruktion zeigte sich ein Haupteffekt [$F(1,90) = 10.95$, $R^2 = 0.11$]: Nach der Inklusionsinstruktion ($M = .73$) wurden mehr Listenitems wiedererkannt als nach der Standardinstruktion ($M = .64$). Die Interaktion war statistisch nicht bedeutsam ($F < 1$). Die entsprechenden Mittelwerte sind in Tabelle 8.2.7 abgebildet.

Tabelle 8.2.7:

Mittlere Trefferrate und Pr-Werte der Listenitems in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Treffer		Pr-Werte	
	Standard	Inklusion	Standard	Inklusion
6 Listen				
Vorschüler	.55	.65	.42	.44
Drittklässler	.62	.71	.46	.46
Erwachsene	.76	.83	.67	.56
8 Listen				
Vorschüler	.55	.61	.41	.41
Drittklässler	.61	.71	.47	.45
Erwachsene	.75	.81	.67	.54

Für die korrigierte Wiedererkennensleistung (Pr-Werte; Mittelwerte s. Tabelle 8.2.7) ergab sich wie für die Trefferrate ein Haupteffekt der Altersgruppe [$F(2,90) = 9.75$, $R^2 = 0.18$], wo-

nach Erwachsene ($M = .62$) mehr Listenitems wiedererkannten als Kinder (Vorschüler: $M = .43$; Drittklässler: $M = .46$), die sich auch hier nicht unterschieden. Alle anderen Effekte verfehlten das Signifikanzniveau (alle F s < 1.28)

Quellendiskrimination. Die Güte der Quellendiskrimination der Listenitems berechnete sich daraus wie viele der als „alt“ wiedererkannten Listenitems in der zweiten Testphase als vorgelesen bezeichnet wurden. Es gab keinen Alterseffekt ($F < 1$): Alle Altersgruppen konnten in vergleichbarem Maß ihre wiedererkannten Listenitems der korrekten Quelle zuweisen (Mittelwerte s. Tabelle 8.2.8).

Tabelle 8.2.8:

Mittelwerte der Quellenzuordnung von Listenitems in den Altersgruppen (separate Analyse des ersten Tests).

	6 Listen	8 Listen
Vorschüler	.90	.90
Drittklässler	.80	.79
Erwachsene	.84	.84

Die Aufgabe zur Quellendiskrimination war keine schwierige und konnte entsprechend gut von allen Altersgruppen bewältigt werden. Auch die anderen Werte der VA blieben unbedeutend ($F < 1$). Die VA mit allen acht Listen gelangte zu vergleichbaren Ergebnissen (Mittelwerte s. Tabelle 8.2.8).

Kritische Items

Um einen Überblick über die Güte der Listen zu erhalten werden in Tabelle 8.2.9 die Mittelwerte nach Standard- und Inklusionsinstruktion für die einzelnen Listen nach Altersgruppen zusammengefasst. Verglichen mit der Güte der Listen aus Experiment 2 ist zu erkennen, dass die mittleren Werte der kritischen Items im Wiedererkennenstest wie erwartet höher lagen. Nach der Inklusionsinstruktion wurden über alle Altersgruppen gemittelt erwartungs- und instruktionsgemäß mehr kritische Items wiedererkannt (zwischen .63 und .92) als nach der Standardinstruktion (zwischen .42 und .78). Die Listen der Tabellen sind nach mittlerer absteigender falscher Wiedererkennungsrates über alle Altersgruppen geordnet. In allen Listen wurde das kritische Item wiedererkannt. Die Altersgruppen differierten innerhalb der einzelnen Listen weniger voneinander als in der freien Reproduktion. Wie auch im zweiten Experiment waren Tierlisten („Pferd“, „Löwe“, „Hund“) durch höhere mittlere Häufigkeiten ge-

kennzeichnet als andere Listen. Die geringste Wiedererkennensrate zeigte sich in der Liste „Hammer“.

Tabelle 8.2.9:

Mittelwerte der als „alt“ bezeichneten kritischen Items in Abhängigkeit der Instruktion und des Alters.

Kritische Items	Standard				Inklusion			
	Vorsch.	Drittkl.	Erwachs.	Ges.	Vorsch.	Drittkl.	Erwachs.	Ges.
Pferd	.82	.76	.75	.78	.82	.82	.82	.82
Auto	.44	.94	.82	.73	.75	.82	.94	.84
Kerze	.63	.69	.82	.71	.69	.88	.82	.80
Löwe	.63	.69	.82	.71	.75	.69	.94	.79
Nadel	.57	.63	.82	.67	.94	.88	.94	.92
Hund	.63	.69	.69	.67	.88	.82	.88	.86
Sommer	.57	.63	.82	.67	.63	.82	.94	.80
Telefon	.63	.69	.57	.63	.63	.76	.75	.71
Bett	.57	.82	.50	.62	.69	.94	.69	.77
Fenster	.50	.63	.63	.59	.63	.76	.88	.76
Fuß	.57	.57	.51	.55	.69	.44	.76	.63
Vogel	.57	.50	.57	.55	.75	.69	1.00	.81
Hase	.57	.38	.69	.55	.44	.88	.94	.75
Brot	.25	.57	.69	.55	.57	.63	1.00	.73
Hemd	.38	.63	.63	.55	.57	.63	.69	.63
Hammer	.50	.32	.44	.42	.57	.75	.82	.71

Wiedererkennen. In der VA für das Antwortkriterium kritischer Items ergaben sich ein Alters- [$F(2,87) = 3.94$, $R^2 = 0.08$] und ein Instruktionseffekt [$F(1,87) = 19.79$, $R^2 = 0.19$]. Zudem interagierten die Faktoren tendenziell miteinander [$F(2,87) = 2.69$, $R^2 = 0.06$, $p = .07$]. Die Analyse der einfachen Haupteffekte zeigte, dass Erwachsene nach der Inklusionsinstruktion ein liberaleres Kriterium als Vorschüler wählten [$F(2,87) = 6.23$; Mittelwerte s. Tabelle 8.2.10]. Nach der Standardinstruktion unterschieden sich die Altersgruppen nicht ($F < 1$). Erwachsene [$F(1,87) = 19.62$] wählten nach der Inklusionsinstruktion ein deutlich liberaleres Kriterium als nach der Standardinstruktion. Bei Vorschülern ($F < 1$) und Drittklässlern [$F(1,87) = 3.56$] war dies nicht der Fall.

Tabelle 8.2.10:

Mittlere Br-Werte der kritischen Items in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Standard	Inklusion
6 Listen		
Vorschüler	.23	.37
Drittklässler	.31	.51
Erwachsene	.28	.73
8 Listen		
Vorschüler	.23	.38
Drittklässler	.29	.46
Erwachsene	.27	.69

Für das falsche Wiedererkennen im Inklusionstest zeigte sich ein Altersgruppeneffekt [$F(2,90) = 3.25$, $R^2 = 0.07$]: Vorschüler ($M = .56$) hatten weniger falsche Erinnerungen als Drittklässler ($M = .68$), unterschieden sich aber nicht von Erwachsenen ($M = .63$). Es zeigten sich keine weiteren signifikanten Effekte (alle $F_s < 1.78$). Somit waren falsche Erinnerungen nach der Standard- und Inklusionsinstruktion vergleichbar.

In der VA für die „Trefferrate“ der kritischen Items (Mittelwerte s. Tabelle 8.2.11) zeigte sich erwartungskonform (s. Hypothese 1) ein Instruktionseffekt [$F(1,90) = 14.33$, $R^2 = 0.14$]: Nach der Inklusionsinstruktion ($M = .79$) wurden mehr kritische Items als „alt“ bezeichnet als nach der Standardinstruktion ($M = .63$) wie bei der Trefferrate für Listenitems. Zudem ergab sich ein tendenzieller Haupteffekt für die Altersgruppe [$F(2,90) = 2.43$, $R^2 = 0.05$, $p = .09$]: Erwachsene ($M = .79$) hatten eine tendenziell höhere „Trefferrate“ als Vorschüler ($M = .61$). Drittklässler ($M = .74$) unterschieden sich nicht von Vorschülern und Erwachsenen. Die Interaktionen verfehlten das Signifikanzniveau (alle $F_s < 1$).

Tabelle 8.2.11:

Mittlere „Trefferrate“ der kritischen Items in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Standard	Inklusion
6 Listen		
Vorschüler	.55	.66
Drittklässler	.66	.81
Erwachsene	.67	.91
8 Listen		
Vorschüler	.55	.68
Drittklässler	.63	.76
Erwachsene	.65	.86

In der VA mit den Pr-Werten war die Interaktion der Altersgruppe mit der Instruktion nicht signifikant ($F < 1$), aber aufgrund der Hypothesenrelevanz sind die entsprechenden Mittelwerte in Abbildung 8.2.1 zusammengefasst.

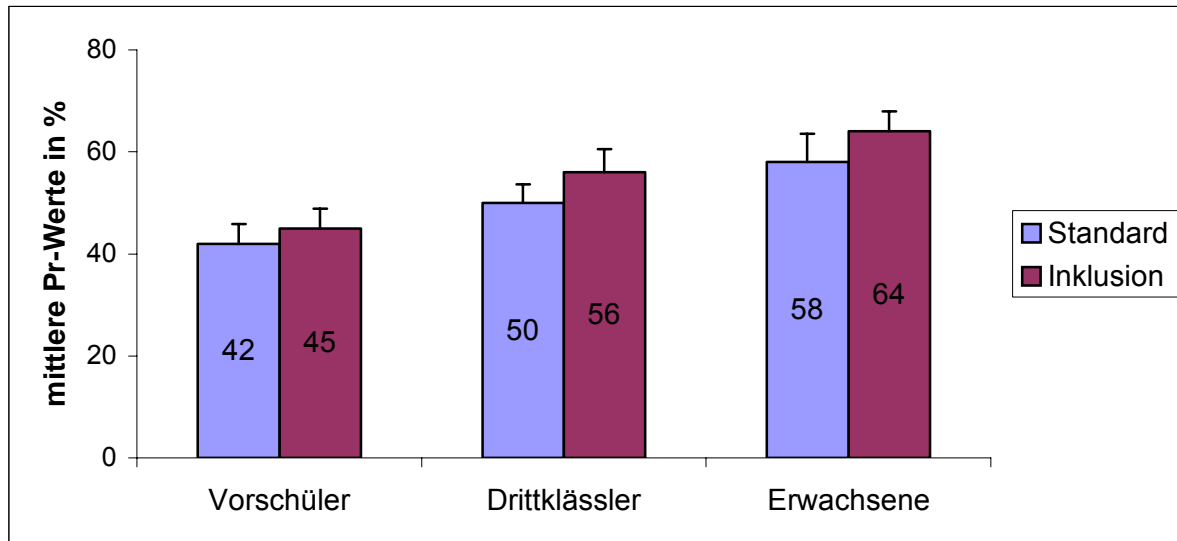


Abbildung 8.2.1: Mittlere Pr-Werte (in Prozent) der Probanden für kritische Items nach Standard- und Inklusionsinstruktion in der separaten Analyse des ersten Tests.

Die VA führte zu einem Alterseffekt [$F(2,90) = 5.29, R^2 = 0.11$]: Erwachsene ($M = .61$) unterschieden sich nur von Vorschülern ($M = .44$) und Drittklässler ($M = .53$) unterschieden sich weder von Erwachsenen noch von Vorschülern. Gemäß des Alterseffekts lässt sich zur zweiten Hypothese festhalten, dass die Aktivierung, gemessen an den mittleren Pr-Werten nach der Inklusionsinstruktion, erwartungskonform mit dem Alter zunahm (s. Abbildung 8.2.1). Ebenso nahmen falsche Erinnerungen (mittlere Pr-Werte nach der Standardinstruktion, s. Abbildung 8.2.1) mit steigendem Alter zu (Hypothese 4).

Quellendiskrimination. Die Variable „Instruktion“ entfällt in dieser Analyse, da die Aufgabe zur Quellenzuordnung nur nach der Inklusionsinstruktion erteilt wurde (Mittelwerte s. Tabelle 8.2.12).

Tabelle 8.2.12:

Mittelwerte der Fehlattribution kritischer Items in den Altersgruppen (separate Analyse des ersten Tests).

	6 Listen	8 Listen
Vorschüler	.89	.88
Drittklässler	.87	.83
Erwachsene	.67	.63

Die dritte Hypothese konnte gestützt werden, denn der Alterseffekt [$F(2,90) = 8.11$, $R^2 = 0.15$] zeigte, dass die Fähigkeiten zur Quellendiskrimination sich mit steigendem Alter verbesserten: Erwachsene ordneten kritische Items seltener der falschen Quelle zu als Kinder. Die beiden jüngeren Altersgruppen differierten nicht in der falschen Quellenzuweisung kritischer Items (Mittelwerte s. Tabelle 8.2.12). Dieses Ergebnismuster wurde auch in der entsprechenden Analyse mit acht Listen vorgefunden (zu den Mittelwerten s. Tabelle 8.2.12) und weicht von den Befunden zur Quellendiskrimination der Listenitems ab in der sich die Altersgruppen nicht unterschieden.

8.2.4 Diskussion

Ausgehend von der Befundlage des zweiten Experiments, in welchem Bodeneffekte die Interpretation alterskorrelierter Differenzen falscher Erinnerungen einschränkten, sollte anhand eines Wiedererkennenstests diesen Bodeneffekten entgegengewirkt werden. Tatsächlich ergaben sich weder Boden- noch Deckeneffekte.

Um die zugrunde liegenden Prozesse falscher Erinnerungen differenziert zu untersuchen wurde wie im vorherigen Experiment eine Variation der Testinstruktion vorgenommen, nach welcher die Aktivierung kritischer Items ohne Quellendiskrimination anhand einer Inklusionsinstruktion erfasst werden sollte.

Auffallend an den vorliegenden Ergebnissen war, dass es zahlreiche "carry over" Effekte gab, weshalb im Fall von diesen Abfolgeeffekten nur der zuerst erteilte Test interpretiert und diskutiert werden kann.

In weiteren Untersuchungen sollte daher ein Augenmerk auf die Auswirkungen der Abfolge der Instruktionen gelegt werden. Um den Einfluss der Testabfolge zu unterbinden, könnte beispielsweise auf die Variation der Instruktionsabfolge verzichtet werden, so dass alle Probanden zuerst die Standard- und dann die Inklusionsinstruktion erhalten. Denkbar wäre auch statt zweier separater Wiedererkennenstests nur einen Test nach der Präsentation aller Listen zu verwenden. Im Wiedererkennenstest könnte zu einem Listenset eine Standard- und zum anderen eine Inklusionsinstruktion erteilt werden. Auf diese Weise würde die erste Instruktion zum Wiedererkennen nicht das Lernen (d.h. die Enkodierbedingung) der zweiten Liste beeinflussen. Es müsste geprüft werden, ob ein solcher Wiedererkennenstest nach der Präsentation aller Listen für jüngere Kindern geeignet ist und nicht ihre Konzentration überfordert. Anderenfalls müssten kürzere Listen verwendet werden.

Die Ergebnisse zu den korrekten Erinnerungsleistungen zeigten, dass diese alterskorreliert zunahm, denn Erwachsene erkannten durch ihre bessere itemspezifische Verarbeitung mehr Listenitems wieder als Kinder. Jedoch gab es keinen Unterschied der Altersgruppen in der korrekten Quellendiskrimination der Listenitems, vermutlich daher, dass diese Aufgabe keine besonders schwierige gewesen sein könnte.

Die gewählte Inklusionsinstruktion konnte die Zielsetzung erfüllen Aktivierung ohne entgegnerwirkende Quellendiskrimination zu erfassen, denn nach der Inklusionsinstruktion wurden mehr kritische Items als „alt“ bezeichnet als nach der Standardinstruktion. Dies entspricht den Annahmen und wurde auch durch Voruntersuchungen geprüft, welche die Unterschiedlichkeit der Instruktionen und das Verständnis vor allem der jüngsten Altersgruppe sichern sollten. Auch ist dieses Ergebnis in Einklang mit anderen Untersuchungen zu Inklusionsinstruktionen (Dehon & Brédart, 2004; Hege & Dodson, 2004; Heit et al., 2004; vgl. Abschnitt 5.4). Jedoch zeigte sich in den um die Antworttendenzen korrigierten Werten kein Unterschied der Instruktionen im Anteil des Wiedererkennens kritischer Items. Somit scheint der Unterschied der Instruktionen darauf zurückzugehen, dass nach der Inklusionsinstruktion insgesamt mehr „alt“-Urteile gegeben wurden, also auch verbundene und unverbundene Distraktoren als „alt“ bezeichnet wurden, als nach der Standardinstruktion.

Die Aktivierung kritischer Items nahm alterskorreliert zu, denn nach der Inklusionsinstruktion erkannten Erwachsene diese häufiger wieder als Vorschüler. Drittklässler nahmen eine mittlere Position ein und unterschieden sich nicht von Vorschülern und nicht von Erwachsenen. Das Ergebnis weicht vom zweiten Experiment dahingehend ab, dass die Unterschiede zwischen Erwachsenen und den jüngeren Gruppen dort noch stärker waren, wobei dieser Effekt auch teilweise daran liegen kann, dass Erwachsene in Experiment 2 insgesamt deutlich mehr Items nach der Inklusionsinstruktion nannten (vgl. Abschnitt 8.1.4). Die jüngeren Altersgruppen hätten sich voneinander unterscheiden können, da Schulkinder über ein stärker vernetztes Wissen verfügen und dementsprechend auch zu einer höheren automatischen Aktivierung gelangen könnten. Diese alterskorrelierten Wissensunterschiede, die für einige Forscher die Haupteinflussvariable für alterskorrelierte Verbesserungen von Gedächtnisleistungen sind (Bjorklund, 1985; 1987; Chi, 1978; Schneider & Büttner, 2002), könnten aber eventuell durch die Auswahl der Listen ausgeschaltet worden sein (vgl. Abschnitt 7.1.2), so dass kein Unterschied in der Aktivierung kritischer Items bei den beiden jüngeren Gruppen resultierte. Die Listen wurden anhand von Voruntersuchungen mit Vorschülern und Dritt-/Viertklässlern so ausgewählt, dass auch bei Vorschülern mit hoher Wahrscheinlichkeit durch die Listenpräsentation das kritische Item aktiviert werden sollte. Zudem waren die kritischen Items Begriffe,

die jüngeren Kindern schon sehr vertraut waren (u.a. einige Tiere wie Hund, Hase). Die Listenitems waren nicht einfach nach Assoziationsnormen für Erwachsene ausgewählt sondern nach altersspezifischen Normen und im Fall von (seltenen) Unterschieden in den Altersnormen hatten die Normen der jüngsten Gruppe Vorrang. Drittklässler unterschieden sich vermutlich nicht bedeutsam von Erwachsenen, da die Analyse mit den zuerst erteilten Tests auf Kosten der Teststärke gerechnet wurde und damit die Versuchspersonenzahl und die Anzahl der Listen sich damit um die Hälfte reduzierte.

Anders als im ersten Experiment, in dem sich keine alterskorrelierten Differenzen in der Aktivierung kritischer Items zeigten, wenn diese mit konzeptuell impliziten Tests gemessen wurden, zeigten sich hier Unterschiede, wenn die Aktivierung mittels so genannter Inklusionstests erfasst wurde. Dies könnte daran liegen, dass Leistungen in impliziten Tests und Inklusionstests auf verschiedene Arten von Aktivierung zurückgehen. In weiteren Experimenten könnte geprüft werden welche Art von Aktivierung für Leistungen im Inklusionstest relevant sind und ob es dabei alterskorrelierte Differenzen gibt. Ein Vorschlag für ein Untersuchungsdesign, in dem explizite und implizite Testverfahren miteinander verglichen werden können, wurde bereits in der Diskussion von Experiment 1 dargestellt (vgl. Abschnitt 7.1.4). So könnte in der Lernphase eine Variation der Verarbeitungstiefe durch den Stroop-Test (nur automatische Aktivierung möglich) oder durch das selbstständige Lesen von Wörtern (automatische und bewusste Aktivierung) erfolgen und in der Testphase könnte ein impliziter Test und ein Inklusionstest (anstelle des Standardwiedererkennenstests) eingesetzt werden. Die Ergebnisse der beiden Tests könnten auf diese Weise nach den verschiedenen Bedingungen der Verarbeitungstiefe und der Aktivierung miteinander verglichen werden.

Die vorliegenden Befunde zur Inklusionsinstruktion lassen sich mit Forschungsbefunden ähnlicher Studien vergleichen. Da Inklusionsinstruktionen bislang nur an Erwachsenen erprobt wurden ist jedoch nur ein Vergleich mit der ältesten Probandengruppe möglich. Wie schon im Abschnitt 5.4 ausführlich beschrieben wurde, analysierten Heit et al. (2004) strategische Prozesse im Wiedererkennen, um den Anteil der bewussten Kontrolle falscher Erinnerungen herauszustellen. Sie variierten die Darbietungsdauer der Items (kurze Zeit: vorwiegend automatische Prozesse, längere Zeit: auch strategische Prozesse) und versuchten durch verschiedene Instruktionen (Standard-, Vorwarnungs- oder Inklusionsinstruktion) entweder die strategischen Prozesse zu stärken oder diese zu minimieren. Durch den Vergleich des Anteils falscher Erinnerungen könnten somit Rückschlüsse auf automatische und strategische Prozesse ermöglicht werden. Mit zunehmender Präsentationsdauer reduzierten sich die falschen Erinnerungen nach der Standard- und Vorwarnungsinstruktion gleichermaßen, jedoch blieben sie nach der

Inklusionsinstruktion konstant. Nach der Inklusionsinstruktion wurden mehr kritische Items als nach den anderen beiden Instruktionen wiedererkannt, so dass geschlussfolgert wurde, dass die Inklusionsinstruktion weniger strategische (Quellendiskriminations-) Prozesse involviert. Ausgehend von den eigenen Ergebnissen kann auch geschlussfolgert werden, dass die Inklusionsinstruktion Aktivierung ohne entgegenwirkende Quellendiskrimination erfasste, denn nach dieser wurden insgesamt mehr „alt“-Urteile gegeben und es zeigten sich höhere Fehlattributionen kritischer Items als nach der Standardinstruktion.

In der Aufgabe zur Quellendiskrimination kritischer Items zeigte sich im Gegensatz zur Quellendiskrimination der Listenitems (s.o.) ein Alterseffekt dahingehend, dass Erwachsene kritische Items erwartungsgemäß seltener der falschen Quelle zuwiesenen als die jüngeren Gruppen. Drittklässler differierten nicht von Vorschülern in der Quellendiskrimination. Die hohe Fehlattribution bei Erwachsenen spricht dafür, dass die geforderte Quellendiskriminationsaufgabe der kritischen Items deutlich schwieriger war als die der Listenitems. Anscheinend waren hier auch in Anlehnung an das Modell von Johnson et al. (1993) Heuristiken nicht ausreichend, sondern Strategien erforderlich über die auch Drittklässler nach den vorliegenden Ergebnissen noch nicht verfügen.

Die Befunde zu den falschen Erinnerungen zeigten Altersunterschiede, denn Erwachsene hatten mehr falsche Erinnerungen als Vorschüler, nicht aber als Drittklässler, die sich auch nicht von Vorschülern unterschieden. Dies spricht dafür, dass die Aktivierung ausschlaggebend für falsche Erinnerungen zu sein scheint. Sie stellt die Grundvoraussetzung dar, damit falsche Erinnerungen überhaupt zustande kommen können. Die Aktivierung war bei Erwachsenen stärker als bei Vorschülern und die bessere Quellendiskrimination Erwachsener schien dieser nicht entgegenwirken zu können. Daher hatten sie auch einen höheren Anteil falsch wiedererkannter kritischer Items, was in Einklang mit den Ergebnissen des ersten Experiments ist. Die beiden jüngeren Altersgruppen unterschieden sich nicht, da zum einen ihre Aktivierung durch die Auswahl der Listen vergleichbar gewesen sein könnte und da es zum anderen keine Unterschiede in den Quellendiskriminationsleistungen der jüngeren Altersgruppen gab.

Aufgrund der Bodeneffekte in Experiment 2 kann kein Vergleich der Ergebnisse falscher Erinnerungen stattfinden.

Ausgehend von den bisherigen Befunden soll im nächsten Teilexperiment geprüft werden, welche Art von Aktivierung für Leistungen im Inklusionstest relevant ist und ob die Aktivierung bei Kindern gefördert werden kann. Zu diesem Zweck soll das vorliegende Experiment repliziert werden, erweitert um eine Lerninstruktion, welche die strategische Verarbeitung

und damit die bewusste Aktivierung erhöhen sollte. Eine solche Lerninstruktion muss entwickelt und insbesondere an der jüngsten Altersgruppe hinsichtlich des Verständnisses erprobt werden, da Vorschüler nicht spontan fähig sind semantisch-relationale Verarbeitungsstrategien gezielt einzusetzen (vgl. Kapitel 3.2). Würden nach erfolgreicher Förderung der strategischen Verarbeitung dennoch Altersunterschiede auftreten, so würde dies dafür sprechen, dass die automatische Aktivierung eine bedeutende Rolle bei den alterskorrelierten Unterschieden einnimmt.

8.3 Experiment 3B: Falsches Wiedererkennen: Effekte von Aktivierung fördernden Lerninstruktionen

Während in Experiment 1 sich keine alterskorrelierten Unterschiede in der Aktivierung der kritischen Items in konzeptuell impliziten Tests zeigten, deuteten die Ergebnisse von Experiment 2 und 3A aber auf eine alterskorrelierte Zunahme der Aktivierung hin, wenn diese anhand von Inklusionstests getestet wurde. Ausgehend von diesen widersprüchlichen Ergebnissen stellt sich die Frage, auf welche Art von Aktivierung die Unterschiede zurückgehen und, was entscheidend ist, welche Aktivierung der Inklusionstest misst.

Um die Frage zu beantworten, welche Art von Aktivierung der kritischen Items bei Erwachsenen stärker ist als bei fünf- bis 10-jährigen Kindern, wurde in Experiment 3B zusätzlich zu einer Lerninstruktion eine Aufgabe gegeben, die eine semantisch-relationale Verarbeitung der Listen verlangt und mit hoher Wahrscheinlichkeit zur Generierung der kritischen Items führen sollte. Im vorliegenden Experiment sollten die Versuchspersonen nach jeder Liste das Wort (kritische Item) suchen, das mit allen Items etwas gemeinsam hat, sie also zusammenfasst. Gehen die beobachteten Unterschiede in der Aktivierung hauptsächlich auf fehlende elaborative Verarbeitungsstrategien der Kinder zurück, dann sollten diese besonders von der Aufgabe profitieren und Altersunterschiede minimiert werden. In vielen Untersuchungen konnte gezeigt werden, dass Kinder im Vorschulalter auf eine Lerninstruktion hin keine effizienten Verarbeitungsstrategien spontan einsetzen (z.B. Bjorklund, 1985; 1989; Paris, 1978). Wird ihnen aber eine geeignete Strategie durch die Instruktion oder die Art der Präsentation des Materials nahegelegt, so können sie in der Regel von der Strategie profitieren (z.B. Schneider & Büttner, 2002). Wenn die Altersunterschiede auf automatischen Aktivierungsprozessen gründen, dann sollten sie trotz fördernder Lerninstruktion bestehen bleiben.

Das vorliegende Experiment entspricht dem vorherigen, ergänzt um die Generierungsinstruktion. Es wird festgehalten, ob die kritischen Items generiert werden oder nicht, um diese mit den wiedererkannten kritischen Items zu vergleichen. Aufschlüsse darüber, ob der Inklusionstest eher eine automatische oder eine bewusste Aktivierung erfasst und ob falsche Erinnerungen im Standardtest auf automatischer oder bewusster Aktivierung beruhen, ergeben sich durch den Vergleich wiedererkannter generierter und nicht generierter kritischer Items. Es gibt konträre Befunde über den Zusammenhang generierter und wiedererkannter kritischer Items, die Aussagen über eine bewusste oder automatische Aktivierung ermöglichen. Marsh und Bower (2004) erteilten eine Generierungsinstruktion und stellten fest, dass in der Lernphase generierte kritische Items häufiger falsch wiedererkannt wurden als nicht generierte.

Daher wurde davon ausgegangen, dass die kritischen Items bewusst aktiviert worden waren. Wenn der Inklusionstest primär eine bewusste Aktivierung erfasst und wenn falsche Erinnerungen im eigenen Standardtest auf bewusster Aktivierung gründen, dann sollten die generierten kritischen Items eher wiedererkannt werden als die nicht generierten. Hingegen sind Seamon et al. (2002a) Vertreter von automatischen Prozessen. Sie erteilten Probanden eine Lerninstruktion laut zu memorieren (vgl. Abschnitt 2.2.2). Wurde das kritische Item memoriert, dann wurde dieses in der freien Reproduktion häufiger genannt, aber nicht häufiger im Wiedererkennenstest als „alt“ bezeichnet. Sollten sich keine Unterschiede in der Anzahl der „alt“-Urteile zwischen den in der Lernphase generierten und nicht generierten kritischen Items zeigen oder eher nicht generierte kritische Items wiedererkannt werden, dann würde dies dafür sprechen, dass der Inklusionstest primär eine automatische Aktivierung erfasst bzw. dass falsche Erinnerungen im Standardtest auf automatischer Aktivierung zu gründen scheinen. Diese Ergebnisse basieren alle auf Untersuchungen mit Erwachsenen. Welche Aktivierung (eher automatisch oder eher bewusst strategisch) für die Altersgruppen im Inklusionstest erfasst wird und auf welcher Art von Aktivierung (eher automatisch oder eher bewusst strategisch) falsche Erinnerungen im Standardtest basieren ist vollkommen unklar. Ebenso ist unklar, ob sich Unterschiede zwischen den einzelnen Altersgruppen zeigen.

Des Weiteren soll im vorliegenden Experiment geprüft werden, wie sich die Generierungsinstruktion auf die Aktivierung, die Quellendiskrimination der kritischen Items und auf falsche Erinnerungen auswirkt. Durch die semantisch-relationale Verarbeitung sollte sich die Aktivierung erhöhen (vgl. Marsh & Bower, 2004). Aufgrund der ohnehin schlechteren Quellendiskriminationsleistung von Kindern ist zu erwarten, dass sich die Generierungsbedingung negativ auf die beiden jüngeren Gruppen auswirkt, denn durch das Generieren sollte die Quellendiskriminationsaufgabe zusätzlich erschwert werden. Auch bei Erwachsenen sollte in Anlehnung an die Befunde von Marsh und Bower (2004) die Quellendiskrimination erschwert werden. Durch die höhere Aktivierung und die schlechtere Quellendiskrimination sollte das Generieren zu einer Erhöhung falscher Erinnerungen führen. Inwiefern sich die Altersgruppen unterscheiden lässt sich kaum vorhersehen.

Wie auch im vorherigen Experiment wird die Abfolge der Instruktionen variiert. Ergeben sich Abfolgeeffekte, dann werden diese in einem separaten Abschnitt des Ergebnisteils dargestellt und interpretiert. Aufgrund der zahlreichen Abfolgeeffekte werden Analysen nur mit den zuerst erteilten Tests gerechnet, die eindeutig, ohne „carry over“ Effekte, interpretiert werden können. Der Hauptegebnisteil stützt sich auf diese Analysen. Auch für den Vergleich der Experimente 3A und 3B werden nur die zuerst erteilten Tests analysiert und interpretiert.

Das vorliegende Experiment entspricht also dem vorherigen abgesehen von der Generierungsinstruktion.

Zur Untersuchung der Aktivierung in Experiment 3B wurden folgende Erwartungen aufgestellt:

- (1) Es werden keine gerichteten Annahmen zu alterskorrelierten Unterschieden in der Aktivierung nach der Inklusionsinstruktion formuliert, da unklar ist, auf welcher Art von Aktivierung diese beruhen. Falls die Aktivierungsunterschiede auf Problemen der jüngeren Kinder in der semantisch-relationalen Verarbeitung basieren, sollten nach der Generierungsinstruktion keine alterskorrelierten Unterschiede auftreten, demgegenüber sollten sie bei einer primär automatischen Aktivierung bestehen bleiben.
- (2) Da unklar ist, welche Art von Aktivierung der Inklusionstest misst, werden keine gerichteten Annahmen über den Vergleich generierter und wiedererkannter kritischer Items für die Altersgruppen im Inklusionstest formuliert.
- (3) Da unklar ist auf welcher Art von Aktivierung falsche Erinnerungen im Standardtest gründen (s. Marsh & Bower, 2004; Seamon et al., 2002a), werden keine gerichteten Annahmen über den Vergleich generierter und wiedererkannter kritischer Items für die Altersgruppen im Standardtest formuliert.

Für die Effekte des Generierens im Vergleich der Experimente 3A und 3B wurden folgende Erwartungen formuliert:

- (4) Es wird angenommen, dass die Generierungsinstruktion, welche die semantisch-relationale Verarbeitung fördern soll, die Aktivierung in allen Altersgruppen erhöht. Dementsprechend sollten alle Altersgruppen nach der Inklusionsinstruktion mit Generieren mehr kritische Items wiedererkennen als ohne Generieren.
- (5) Ausgehend von den Befunden des zweiten Experiments, in welchem Drittklässler Probleme bei der Quellendiskrimination der genannten kritischen Items (nach der Inklusionsinstruktion) hatten, wird davon ausgegangen, dass sich die Generierungsbedingung negativ auf die Quellendiskrimination bei Kindern auswirkt und weniger negativ bei Erwachsenen verglichen mit der Bedingung ohne Generieren.
- (6) Da die Aktivierung durch das Generieren in allen Altersgruppen erhöht und die Quellendiskrimination erschwert werden soll, wird angenommen, dass falsche Erinnerungen mit Generieren höher sind als ohne Generieren.

8.3.1 Methode

Versuchsplan. Dem Versuch lag der identische Versuchsplan des Experiments 3A zugrunde.

Versuchspersonen. An Experiment 3B nahmen 32 Vorschüler (15 Mädchen und 17 Jungen) eines Kindergartens, 32 Drittklässler (18 Mädchen und 14 Jungen) einer Grundschule und 32 Zwölftklässler (17 Frauen und 15 Männer) eines Gymnasiums teil. Die Vorschüler hatten ein durchschnittliches Alter von sechs Jahren (zwischen 5;0 und 7;3 Jahren), die Drittklässler waren im Mittel neun Jahre und zwei Monate alt (zwischen 8;1 und 10;4 Jahren) und die Zwölftklässler 18 Jahre und vier Monate (zwischen 17;4 und 21;4 Jahren).

Material und Durchführung. Das Material und die Durchführung des Versuchs entsprechen Experiment 3A mit dem Zusatz der Generierungsaufgabe (s. Anhang C1).

Demzufolge wurde eine andere Instruktion gegeben: Allen Probanden wurden in der Lernphase vor Präsentation der DRM-Listen eine beispielhafte Wortliste präsentiert („Krone, Palast, Schloss, Diener, Prinzessin, Macht, Zepter, reich, Thron, Kaiser“). Vorschülern und Drittklässlern wurde mitgeteilt, dass in allen Listen ein bestimmtes Wort fehlt, welches die einzelnen Listenwörter zusammenfasst. Dieses fehlende Wort sollten sie nach dem Vorlesen einer jeden Liste benennen. Erwachsene Probanden erhielten eine dem Alter angepasste Instruktion, wobei das fehlende zugehörige Wort verdeutlicht wurde („Nun kann dir auffallen, dass alle Wörter mit dem Wort *König* zusammenhängen, das aber nicht vorgelesen wurde. Das fehlende zusammenhängende Wort sollst du mir nach jeder Liste sagen.“). Die gesamten ausführlichen Instruktionen können im Anhang C2 nachgelesen werden.

8.3.2 Ergebnisse

Wie auch im vorherigen Experiment ergaben sich zahlreiche Abfolgeeffekte, weshalb zunächst in einem separaten Abschnitt die wichtigsten Ergebnisse der 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Abfolge der Instruktionen: Standard/Inklusion vs. Inklusion/Standard) x 2 (Instruktion: Standard vs. Inklusion) VA für die Listenitems und die kritischen Items dargestellt werden.

Für den Hauptergebnisteil und zur Prüfung der Hypothesen wurden separate 3 (Altersgruppe) x 2 (Instruktion) VAn ohne Messwiederholung nur mit den in der ersten Phase erteilten Tests (Standard oder Inklusion) gerechnet, was einer interindividuellen Manipulation entspricht. Somit waren die Ergebnisse unbeeinflusst von der anderen zuvor erteilten Instruktion und können damit eindeutig interpretiert werden.

Als entscheidendes Maß für falsches Wiedererkennen (im Standardwiedererkennenstest) und Aktivierung (im Inklusionstest) sowie für korrektes Wiedererkennen der Listenitems wurden wie auch in Experiment 3A korrigierte Werte berechnet (vgl. Abschnitt 3.2.3). Nach Darstellung der Ergebnisse für die wiedererkannten Listenitems und kritischen Items und die Quellediskrimination beider Itemtypen sowie die Ergebnisse zum Generieren kritischer Items wird auf den Vergleich zwischen den generierten und wiedererkannten kritischen Items eingegangen und abschließend auf den Vergleich der Telexperimente 3A und 3B. Ausgehend davon soll der Einfluss der Generierungsinstruktion auf Aktivierung und Quellediskrimination und auch auf falsche Erinnerungen herausgestellt werden.

Falls die VAn mit acht Listen zu abweichenden Ergebnissen für die hypothesenrelevanten Faktoren führen, so werden sie ergänzend hinzugefügt. Die Mittelwerte der Analysen mit acht Listen werden aber zum Überblick und zur Vollständigkeit in die Tabellen eingefügt.

8.3.2.1 Einflüsse der Testabfolge

Die Analysen zeigten wiederum zahlreiche Testabfolgeeffekte. Zusammenfassend sollen wie in Experiment 3A die wichtigsten Ergebnisse der Listenitems und kritischen Items unter Einbezug der Testabfolge nun dargestellt und diskutiert werden.

Listenitems. Die Trefferrate und die Pr-Werte gelangten zu vergleichbaren Ergebnissen. Die Mittelwerte beider Analysen sind in Tabelle 8.3.1 und 8.3.2 zusammengefasst.

Tabelle 8.3.1:

Mittelwerte der „alt“-Antworten für die Listenitems in Abhängigkeit des Alters, der Instruktionsabfolge und der Instruktion.

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.82	.73	.78	.76	.74	.75
Drittklässler	.80	.70	.75	.76	.72	.74
Erwachsene	.85	.84	.84	.79	.88	.83
8 Listen						
Vorschüler	.82	.70	.76	.76	.74	.75
Drittklässler	.80	.68	.74	.74	.72	.73
Erwachsene	.84	.83	.83	.79	.85	.82

Tabelle 8.3.2:

Mittlere Pr-Werte der Listenitems in Abhängigkeit des Alters, der Instruktionsabfolge und der Instruktion.

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.68	.65	.66	.63	.59	.61
Drittklässler	.71	.57	.64	.62	.58	.60
Erwachsene	.72	.66	.69	.59	.77	.68
8 Listen						
Vorschüler	.73	.65	.69	.64	.65	.64
Drittklässler	.73	.61	.67	.64	.64	.64
Erwachsene	.72	.67	.69	.64	.74	.69

Wie auch in den vorangehenden Experimenten wurde anhand der Pr-Werte die korrigierte Wiedererkennensleistung getestet. Die entsprechende dreifaktorielle VA zeigte konform mit der Trefferrate der Listenitems einen, wenn hier auch nur tendenziellen Haupteffekt der Altersgruppe [$F(2,90) = 2.56$, $R^2 = 0.05$, $p = .08$]. Zudem erlangte die messwiederholte Variable Instruktion Signifikanz [$F(1,90) = 3.35$, $R^2 = 0.04$]. Jedoch wurden beide Haupteffekte durch eine Interaktion mit der Abfolge überlagert [Altersgruppe und Abfolge: $F(2,90) = 3.49$, $R^2 = 0.07$; Instruktion und Abfolge: $F(1,90) = 9.11$, $R^2 = 0.09$]. Die Mittelwerte sind Tabelle 8.3.2 zu entnehmen.

Außerdem interagierten alle drei Faktoren miteinander [$F(2,90) = 3.34$, $R^2 = 0.07$]. Die Mittelwerte sind in Tabelle 8.3.2 abgebildet. Es gab für die beiden älteren Gruppen, nicht aber für Vorschüler “carry over“ Effekte: Nach der Standardinstruktion schien die Abfolge bei Drittklässlern relevant zu sein, die in der Bedingung Standard-Inklusion höhere Pr-Werte als in der umgekehrten Abfolge [$F(1,90) = 5.32$] hatten, während nach der Inklusionsinstruktion die

Abfolge nur bei Erwachsenen einen Einfluss hatte [$F(1,90) = 5.41$]. Hier erkannten sie Listenitems besser wieder, wenn zuerst die Inklusionsinstruktion erteilt wurde. Es erscheint möglich, dass die Inklusionsinstruktion, sofern sie als Erste gegeben wurde, Erwachsenen die Erinnerung an die Listenitems erleichtern konnte. Möglicherweise konnten Erwachsene, wenn sie aufgefordert wurden, kritische Items als „alt“ zu bezeichnen, auch Strategien zum Wiedererkennen der Listenitems einsetzen, indem sie z.B. zu den kritischen Items assoziierten. Die anderen Effekte zeigten keine Signifikanz ($F < 1$).

Kritische Items. Wie in Experiment 3A wurde der Anteil falscher Erinnerungen im Inklusionstest berechnet, indem jene kritischen Items zur Analyse herangezogen wurden, die im Wiedererkennenstest nach der Inklusionsinstruktion als „alt“ bezeichnet wurden und in der zweiten Testphase auf eine externe Quelle fehlattribuiert worden waren. In der dreifaktoriellen VA interagiert die Abfolge mit der Instruktion [$F(1,90) = 6.98$, $R^2 = 0.07$], deren Mittelwerte Tabelle 8.3.3 zu entnehmen sind.

Tabelle 8.3.3:

Mittelwerte der Interaktion „Abfolge der Instruktionen“ und „Instruktion“ für falsches Wiedererkennen kritischer Items.

	Standard	Inklusion
Standard-Inklusion	.61	.48
Inklusion-Standard	.54	.55

Eine Analyse der einfachen Haupteffekte ergab nur einen signifikanten Effekt der Instruktion für die Abfolge Standard-Inklusion [$F(1,90) = 9.93$]: Hier wurden nach der Standardinstruktion mehr kritische Items falsch wiedererkannt als nach der Inklusionsinstruktion (Mittelwerte s. Tabelle 8.3.3). Alle anderen Werte differierten nicht bedeutsam voneinander (alle F s < 3.80) und die anderen Effekte blieben insignifikant ($F < 1$).

Die VAn mit der „Trefferrate“ der kritischen Items und den Pr-Werten zeigten keine signifikanten Testabfolgeeffekte. Zur Übersicht und Vollständigkeit werden die Mittelwerte der Analysen in Tabelle 8.3.4 und 8.3.5 zusammengefasst.

Tabelle 8.3.4:

Mittelwerte der „alt“-Antworten für die kritischen Items in Abhängigkeit des Alters, der Instruktionsabfolge und der Instruktion (in Klammern Werte für falsches Wiedererkennen nach Inklusion).

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.72	.71	.71	.83 (.60)	.81 (.70)	.82 (.65)
Drittklässler	.76	.60	.68	.85 (.58)	.89 (.66)	.87 (.62)
Erwachsene	.34	.31	.33	.96 (.25)	.98 (.29)	.97 (.27)
8 Listen						
Vorschüler	.70	.67	.68	.81 (.63)	.77 (.67)	.79 (.65)
Drittklässler	.76	.63	.70	.85 (.59)	.84 (.60)	.85 (.60)
Erwachsene	.33	.34	.33	.72 (.27)	.73 (.31)	.73 (.29)

Tabelle 8.3.5:

Mittlere Pr-Werte der kritischen Items in Abhängigkeit des Alters, der Instruktionsabfolge und der Instruktion.

	Standard			Inklusion		
	S-I	I-S	Ges.	S-I	I-S	Ges.
6 Listen						
Vorschüler	.57	.63	.60	.70	.67	.69
Drittklässler	.67	.47	.57	.72	.74	.73
Erwachsene	.21	.13	.17	.76	.87	.82
8 Listen						
Vorschüler	.58	.63	.60	.66	.60	.63
Drittklässler	.71	.46	.58	.71	.72	.71
Erwachsene	.19	.08	.14	.70	.87	.78

8.3.2.2 Ergebnisse der separaten Analysen mit den zuerst erteilten Tests

Aufgrund der Testabfolgeeffekte lassen sich die Ergebnisse nicht eindeutig interpretieren. Daher wurden getrennte VAn für die Listenitems und kritischen Items mit den zuerst erteilten Tests gerechnet, deren Ergebnisse in den folgenden Abschnitten dargestellt werden und eindeutig interpretierbar sind. Ergänzend soll hinzugefügt werden, ob die vorliegenden Ergebnisse mit denen von Experiment 3A übereinstimmen oder von diesen abweichen.

Listenitems

Wiedererkennen. Die Berechnung des Antwortkriteriums (Br-Werte) in der 3 (Altersgruppe) x 2 (Instruktion) VA für die Listenitems führte zu keinen signifikanten Effekten (alle $F_s < 1.97$). Anders als in Experiment 3A unterschieden sich die Altersgruppen nicht in ihrem kon-

servativen bis neutralen Antwortverhalten (Vorschüler: $M = .42$, Drittklässler: $M = .31$; Erwachsene: $M = .47$).

In der VA mit den Pr-Werten (die VA mit der Trefferrate gelangte zu vergleichbaren Ergebnissen; alle Mittelwerte in Tabelle 8.3.6) ergab sich ein Effekt der Altersgruppe [$F(2,90) = 5.05$, $R^2 = 0.10$], wie in Experiment 3A, und ein tendenzieller der Instruktion [$F(1,90) = 3.59$, $R^2 = 0.04$, $p = .06$], zudem interagierten die Faktoren [$F(2,90) = 3.21$, $R^2 = 0.07$].

Tabelle 8.3.6:

Mittlere Trefferrate und Pr-Werte der Listenitems in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Treffer		Pr-Werte	
	Standard	Inklusion	Standard	Inklusion
6 Listen				
Vorschüler	.82	.74	.68	.59
Drittklässler	.80	.72	.71	.58
Erwachsene	.85	.88	.72	.77
8 Listen				
Vorschüler	.80	.68	.66	.57
Drittklässler	.79	.73	.69	.57
Erwachsene	.87	.84	.70	.73

Die Analyse der einfachen Haupteffekte zeigte, dass sich die Altersgruppen nach der Inklusionsinstruktion deutlich unterschieden [$F(2,90) = 7.89$]: Erwachsene konnten mehr Listenitems wiedererkennen als Kinder. Nach der Standardinstruktion ergab sich kein Altersunterschied ($F < 1$). Demnach ist die korrekte Wiedererkennensleistung für alle Altersgruppen vergleichbar und es gibt, was unerwartet ist, trotz gleicher Listenlänge und gleichem Lernmaterial keine Unterschiede in der Güte des Listenlernens. Nach der Standardinstruktion konnten nur Drittklässler mehr Listenitems wiedererkennen als nach der Inklusionsinstruktion [$F(1,90) = 6.67$]. Für Vorschüler [$F(1,90) = 2.54$] und Erwachsene ($F < 1$) war die Instruktion zum Wiedererkennen der Listenitems unbedeutend.

Quellendiskrimination. Die VA für die Listenitems führte zu keinem signifikanten Effekt (alle F s < 1). Die Mittelwerte dieser Analyse für sechs und acht Listen sind Tabelle 8.3.7 zu entnehmen.

Tabelle 8.3.7:

Mittelwerte der Quellenzuordnung von Listenitems in den Altersgruppen (separate Analyse des ersten Tests).

	6 Listen	8 Listen
Vorschüler	.96	.98
Drittklässler	.94	.92
Erwachsene	.94	.97

Wie in Experiment 3A konnten alle Altersgruppen die Listenitems in ähnlichem Maß der korrekten Quelle zuordnen (Mittelwerte s. Tabelle 8.3.7), was sich dadurch erklären lässt, dass die Aufgabe offensichtlich nicht schwierig war. Die insgesamt sehr hohen Werte deuten jedoch auf nicht interpretierbare Deckeneffekte hin.

Kritische Items

Generierung. Tabelle 8.3.8 stellt die mittleren Häufigkeiten der kritischen Items absteigend nach ihrer Generierungshäufigkeit dar, getrennt nach Altersgruppen und Listenset A und B. Diese Liste spiegelt die Identifizierbarkeit (vgl. Neuschatz et al., 2003) der kritischen Items wieder, nach welcher es den Probanden leichter oder schwerer fällt, kritische Items aus der Liste zu entdecken.

Tabelle 8.3.8:

Mittlere Häufigkeiten der generierten kritischen Items in Abhängigkeit des Alters.

Kritische Items	Generierung			
	Vorschüler	Drittklässler	Erwachsene	
Listenset A	1. Hase	.72	.78	.88
	2. Pferd	.69	.66	1.00
	3. Auto	.75	.69	.72
	4. Brot	.66	.66	.66
	5. Hammer	.50	.56	.63
	6. Kerze	.34	.50	.81
	7. Nadel	.44	.31	.59
	8. Sommer	.22	.28	.50
Listenset B	1. Hund	.81	.75	.84
	2. Vogel	.53	.72	1.00
	3. Bett	.69	.69	.84
	4. Telefon	.53	.69	.91
	5. Löwe	.38	.72	.97
	6. Fenster	.66	.50	.81
	7. Fuß	.41	.50	.72
	8. Hemd	.06	.19	.34

Wie aus Tabelle 8.3.8 ersichtlich wird, waren die kritischen Items der Tierlisten am leichtesten identifizierbar (Set A: „Hase“, „Pferd“; Set B: „Hund“, „Vogel“). Die beiden jüngeren Altersgruppen ähnelten sich, für Erwachsene waren kritische Items leichter identifizierbar.

Tabelle 8.3.8 zeigte, dass zwei Wortlisten der beiden Sets sehr selten zur Nennung der kritischen Items führten (Set A: „Sommer“; Set B: „Hemd“) und zusätzlich die Listen „Kerze“ und „Löwe“ äußerst selten von der jüngsten Altersgruppe genannt wurden. Die Werte der mittleren Generierungshäufigkeit für die beiden Kindergruppen entsprachen zwar denen der Pilotstudie zur Normierung der DRM-Listen an Kindern, waren aber bei den Erwachsenen höher als bei den Kindern. Um die Homogenität der Listen zu erhöhen und um die jüngeren Kinder nicht zu benachteiligen, wurden für die statistischen Analysen dieses Experiments (und des vorherigen) nur sechs der insgesamt acht Listen eines jeden Sets einbezogen. Sogar in den Analysen mit den drei besten Listen (Set A: „Brot“, „Hase“, „Auto“; Set B: „Bett“, „Hund“, „Fenster“) ergaben sich Unterschiede in der Nennung der kritischen Items. Nach der Generierungsinstruktion sollten die Versuchspersonen kritische Items der zuvor präsentierten Listen nennen. Besonders Vorschüler neigten dazu Listenitems zu wiederholen. Es gab keine bedeutenden Unterschiede in den Ergebnissen mit drei, sechs und acht Listen in der Generierung der kritischen Items.

Eine 3 (Altersgruppe) x 2 (Instruktion) VA mit den generierten kritischen Items für sechs Listen wurde gerechnet: Diese führte zu einem Haupteffekt der Altersgruppe [$F(2,90) = 12.16, R^2 = 0.21$]: Erwachsene ($M = .80$) generierten mehr kritische Items als die jüngeren Probanden, die sich nicht voneinander unterschieden (Drittklässler: $M = .63$; Vorschüler: $M = .62$). Die anderen Effekte erlangten keine Signifikanz (alle $F_s < 1.98$). Demnach waren die Listen kaum vergleichbar für Kinder und Erwachsene. Erwachsene konnten die kritischen Items trotz der kindgerecht normierten Listen besser identifizieren als die jüngeren Probanden. Auf damit verbundene Probleme wird in der Diskussion eingegangen.

Wiedererkennen. Anhand von Tabelle 8.3.9 wird ein Überblick über die Mittelwerte der wiedererkannten kritischen Items getrennt nach Altersgruppen und Instruktionen gegeben, wobei die kritischen Items absteigend nach ihren mittleren Häufigkeiten falscher Erinnerungen angeordnet sind.

Tabelle 8.3.9:

Mittelwerte der als „alt“ bezeichneten kritischen Items in Abhängigkeit der Instruktion und des Alters.

Kritische Items	Standard				Inklusion			
	Vorsch.	Drittkl.	Erwachs.	Ges.	Vorsch.	Drittkl.	Erwachs.	Ges.
Löwe	.69	.75	.50	.65	.81	.75	.94	.83
Hase	.81	.69	.38	.63	.88	.88	1.00	.92
Fuß	.81	.81	.25	.62	.94	.88	1.00	.94
Brot	.50	.81	.56	.62	.69	.75	.88	.77
Hemd	.75	.69	.38	.61	.88	1.00	1.00	.96
Hund	.81	.50	.50	.60	.69	.88	.100	.86
Sommer	.75	.81	.25	.60	.88	.81	1.00	.90
Vogel	.75	.88	.13	.59	.81	.81	1.00	.87
Pferd	.63	.75	.38	.59	.88	.93	.94	.92
Kerze	.63	.69	.38	.57	.75	.81	.88	.81
Hammer	.63	.75	.25	.54	.69	.88	.88	.82
Nadel	.56	.75	.31	.54	.94	.94	1.00	.96
Auto	.75	.69	.06	.50	.81	.94	.94	.90
Fenster	.75	.50	.25	.50	.81	1.00	1.00	.94
Bett	.56	.75	.19	.50	.63	.81	1.00	.81
Telefon	.56	.31	.56	.48	.56	.63	1.00	.73

Auch hier wurden Tierlisten besonders häufig falsch von Kindern wiedererkannt, wobei auffallend ist, dass Erwachsene diese eher nicht als „alt“ bezeichneten (z.B. „Hase“ und „Pferd“). Erwartungsgemäß waren die mittleren Häufigkeiten der wiedererkannten kritischen Items nach der Inklusionsinstruktion höher als nach der Standardinstruktion. Verglichen mit Experiment 3A (zwischen .42 und .78) waren die Werte für falsches Wiedererkennen etwas geringer (zwischen .48 und .65), jedoch waren die Werte für die Inklusionsinstruktion im vorliegenden Experiment etwas höher (zwischen .73 und .96) als im vorherigen (zwischen .63 und .92).

Die VA für das Antwortkriterium kritischer Items ergab einen Instruktionseffekt [$F(1,81) = 34.13$, $R^2 = 0.30$]: Nach der Inklusionsinstruktion ($M = .66$) wurde liberaler geantwortet als nach der Standardinstruktion ($M = .28$). Zudem interagierten die Altersgruppen und die Instruktion [$F(2,81) = 6.44$, $R^2 = 0.14$]: Die Analyse der einfachen Haupteffekte zeigte, dass sich die Altersgruppen nach der Inklusionsinstruktion unterschieden [$F(2,81) = 6.68$; Mittelwerte s. Tabelle 8.3.10], denn Erwachsene wählten wie in Experiment 3A ein liberaleres Kri-

terium als Vorschüler. Ebenso wie in Experiment 3A unterschieden sich die Altersgruppen nach der Standardinstruktion nicht ($F < 1$). Sowohl Drittklässler [$F(1,81) = 11.99$] als auch Erwachsene [$F(1,81) = 34.41$] wählten nach der Inklusionsinstruktion ein deutlich liberaleres Kriterium als nach der Standardinstruktion. Bei Vorschülern war dies nicht der Fall ($F < 1$).

Tabelle 8.3.10:

Mittlere Br-Werte der kritischen Items in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Standard	Inklusion
6 Listen		
Vorschüler	.35	.44
Drittklässler	.26	.65
Erwachsene	.21	.87
8 Listen		
Vorschüler	.36	.42
Drittklässler	.22	.64
Erwachsene	.19	.86

Für das falsche Wiedererkennen im Inklusionstest ergab sich wie in Experiment 3A ein Altersgruppeneffekt [$F(2,90) = 34.55$, $R^2 = 0.43$]: Vorschüler ($M = .71$) und Drittklässler ($M = .71$) unterschieden sich nun zwar nicht, aber Erwachsene ($M = .32$) hatten hier auch deutlich weniger falsche Erinnerungen. Es zeigten sich keine weiteren signifikanten Effekte (alle F s < 1.78). Somit waren falsche Erinnerungen nach der Standard- und Inklusionsinstruktion vergleichbar.

Die Resultate für die als „alt“ bezeichneten kritischen Items stimmen mit der korrigierten Wiedererkennensleistung überein, so dass nur die Ergebnisse der VA mit den korrigierten Werten dargestellt werden. Die Mittelwerte der „alt“-Urteile für die kritischen Items können aus Tabelle 8.3.11 entnommen werden.

Tabelle 8.3.11:

Mittlere „Trefferrate“ der kritischen Items in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Standard	Inklusion
6 Listen		
Vorschüler	.72	.81
Drittklässler	.76	.89
Erwachsene	.34	.98
8 Listen		
Vorschüler	.73	.80
Drittklässler	.71	.88
Erwachsene	.31	.97

In der VA mit den Pr-Werten für kritische Items war die Altersgruppe [$F(2,90) = 5.02$, $R^2 = 0.10$] bedeutsam, die Instruktion [$F(1,90) = 41.55$, $R^2 = 0.32$], und anders als in Experiment 3A, die Interaktion der beiden Faktoren [$F(2,90) = 20.60$, $R^2 = 0.31$], deren Mittelwerte aus Abbildung 8.3.1 hervorgehen.

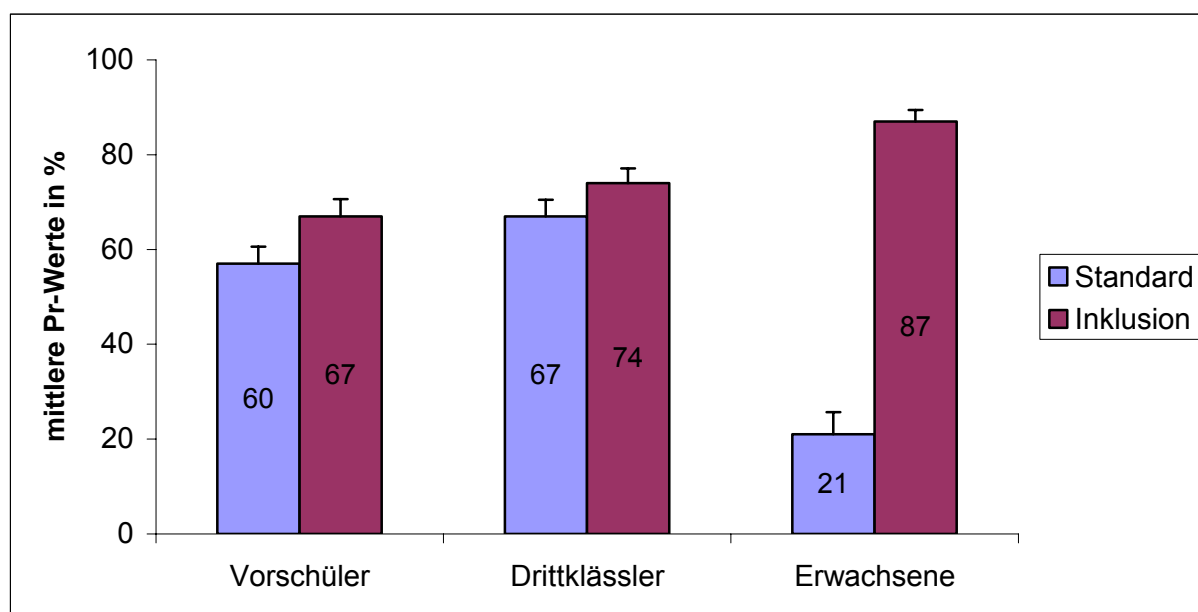


Abbildung 8.3.1: Mittlere Pr-Werte (in Prozent) der Probanden für kritische Items nach Standard- und Inklusionsinstruktion in der separaten Analyse des ersten Tests.

Ein Vergleich innerhalb der Altersgruppen zeigte, dass nur Erwachsene [$F(1,90) = 80.26$] nach der Inklusionsinstruktion mehr kritische Items wiedererkannten als nach der Standardinstruktion, Drittklässler ($F < 1$) und Vorschüler nicht [$F(1,90) = 1.64$; Mittelwerte s. Abbildung 8.3.1]. Nach der Inklusionsinstruktion zeigte sich ein Altersunterschied [$F(2,90) = 3.95$]: Trotz Förderung der semantisch-relationalen Verarbeitung differierten die jüngste und älteste Gruppe in der Aktivierung (Hypothese 1). Die Altersgruppen unterschieden sich auch nach

der Standardinstruktion [$F(2,90) = 21.69$]: Erwachsene erkannten weniger kritische Items falsch wieder als Vorschüler und Drittklässler, die nicht in ihren Pr-Werten differierten, was anders als in Experiment 3A ist, wo sich nur die jüngste und älteste Gruppe signifikant unterschieden.

Vergleich generierter und wieder erkannter kritischer Items. Um zu prüfen, welche Art von Aktivierung der Standard- und der Inklusionstest primär erfassen und auf welcher Art von Aktivierung falsche Erinnerungen primär basieren, wurden Analysen durchgeführt, wie viele von den generierten kritischen Items wiedererkannt wurden. Zu diesem Zweck wurden die „alt“-Antworten für die kritischen Items, die in der Lernphase generiert worden waren mit denen verglichen, die nicht generiert worden waren. Diese Analyse basierte auf allen acht Listen, da es ansonsten zu wenige nicht generierte Items gegeben hätte. Für eine differenziertere Auswertung (Leistung im Inklusionstest stehen als Maß für die Aktivierung, die Leistung im Standardtest als Maß für falsche Erinnerungen) wurden Analysen für die Inklusions- und Standardinstruktion getrennt gerechnet.

In der separaten VA für die als „alt“ bezeichneten kritischen Items zeigte sich nach der Inklusionsinstruktion, dass sowohl die Altersgruppe [$F(2,87) = 10.33$, $R^2 = 0.19$] als auch der Generierungsstatus [$F(1,87) = 50.91$, $R^2 = 0.37$] signifikant waren, die wiederum miteinander interagierten [$F(2,87) = 5.15$, $R^2 = 0.11$]. Die Mittelwerte sind Tabelle 8.3.12 zu entnehmen. Waren die kritischen Items in der Lernphase generiert worden, dann wurden sie auch von allen Altersgruppen im hohen Maß im Inklusionstest als „alt“ bezeichnet ($F < 1$). Bei den nicht generierten kritischen Items zeigte die Analyse der einfachen Haupteffekte einen Altersunterschied [$F(2,87) = 34.05$]: Erwachsene erkannten auch die nicht generierten kritischen Items sehr häufig als „alt“, jedoch gaben Vorschüler und Drittklässler weitaus weniger „alt“-Urteile für diese kritischen Items und unterschieden sich signifikant von den Erwachsenen. Innerhalb der Altersgruppen zeigte sich zu der ungerichteten Annahme aus Hypothese 2, dass Erwachsene ähnlich viele generierte und nicht generierte kritische Items wiedererkannten [$F(1,87) = 1.07$], aber Vorschüler [$F(1,87) = 29.73$] und Drittklässler [$F(1,87) = 28.23$] dazu neigten, eher selbst generierte kritische Items als „alt“ zu bezeichnen.

Tabelle 8.3.12:

Mittelwerte wiedererkannter generierter und nicht generierter kritischer Items in Abhängigkeit des Alters und der Instruktion (separate Analyse des ersten Tests).

	Inklusion		Standard	
	generiert	nicht generiert	generiert	nicht generiert
Vorschüler	.92	.65	.75	.61
Drittklässler	.95	.69	.66	.74
Erwachsene	.98	.92	.21	.79

Die separate VA für die als „alt“ bezeichneten kritischen Items nach der Standardinstruktion führte zu einem Effekt der Altersgruppe [$F(2,86) = 12.23$, $R^2 = 0.22$] und des Generierungsstatus [$F(1,86) = 12.29$, $R^2 = 0.13$], deren Effekte ebenso durch eine Interaktion überlagert wurden [$F(2,86) = 21.50$, $R^2 = 0.33$]. Die Mittelwerte sind Tabelle 8.3.12 zu entnehmen. Sowohl bei den generierten [$F(2,86) = 51.38$] als auch bei den nicht generierten kritischen Items [$F(2,86) = 6.20$] zeigte die Analyse der einfachen Haupteffekte Altersunterschiede: Erwachsene bezeichneten weniger selbst generierte kritische Items als „alt“ verglichen mit Vorschülern und Drittklässlern, die sich nicht unterschieden. Im Wiedererkennen der nicht generierten kritischen Items unterschieden sich nur Vorschüler von Erwachsenen, Drittklässler unterschieden sich von keiner der beiden Gruppen. Innerhalb der Altersgruppen fielen unterschiedliche Ergebnisse auf: Entsprechend kann zu Hypothese 3 festgehalten werden, dass Vorschüler selbst generierte kritische Items eher falsch wiedererkannten als nicht generierte [$F(1,86) = 4.95$]. Jedoch war es für falsche Erinnerungen von Drittklässlern unbedeutend, ob kritische Items generiert wurden oder nicht [$F(1,86) = 1.08$]. Erwachsene erkannten kritische Items seltener wieder, wenn sie diese generiert hatten als wenn sie diese nicht generiert hatten [$F(1,86) = 63.66$].

Quellendiskrimination. Die Mittelwerte der VA für die Quellendiskrimination kritischer Items sind in Tabelle 8.3.13 zusammengefasst. Die Berechnung des jeweiligen Quellendiskriminationsmaßes erfolgte auf die gleiche Weise wie in Experiment 3A.

Tabelle 8.3.13:

Mittelwerte der falschen Quellenzuordnung von kritischen Items in den Altersgruppen (separate Analyse des ersten Tests).

	6 Listen	8 Listen
Vorschüler	.86	.88
Drittklässler	.72	.72
Erwachsene	.30	.27

Es zeigte sich ein Effekt für die Altersgruppe [$F(2,45) = 25.88$, $R^2 = 0.37$]: Erwartungsgemäß unterschieden sich Erwachsene von Drittklässlern und Vorschülern, die wiederum nicht voneinander differierten.

8.3.3 Vergleich der Experimente 3A und 3B

Indem die Experimente 3A und 3B miteinander verglichen werden, lassen sich Rückschlüsse über den Einfluss des Generierens auf falsche Erinnerungen unmittelbar ziehen. Beide Experimente sind vollkommen identisch aufgebaut und unterscheiden sich nur in der Generierungsinstruktion, die in Experiment 3B zusätzlich gefordert wurde.

Um beide Experimente miteinander zu vergleichen wurde eine 3 (Altersgruppe: Vorschüler vs. Drittklässler vs. Erwachsene) x 2 (Generieren: mit vs. ohne) x 2 (Instruktion: Standard vs. Inklusion) VA mit Messwiederholung auf dem letzten Faktor gerechnet. Im Folgenden werden die Ergebnisse der VAn nur mit den hypothesenrelevanten abhängigen Variablen für die Listenitems und kritischen Items erläutert, wobei es in der Darstellung primär um die Effekte des Generierens geht, um Wiederholungen der Ergebnisse zu vermeiden. Auch werden die VAn nur mit sechs Listen durchgeführt, um die Homogenität der Listen zu erhöhen und um die jüngeren Kinder nicht zu benachteiligen. Da es Abfolgeeffekte gab, werden nur die Ergebnisse der VA ohne Messwiederholung für die in der ersten Testphase erteilten Tests (Standard oder Inklusion) separat gerechnet, damit die Ergebnisse interpretiert werden können.

8.3.3.1 Listenitems

Wiedererkennen. Die Mittelwerte der 3 (Altersgruppe) x 2 (Generieren) x 2 (Instruktion) VA für die Trefferrate der Listenitems sowie für die Pr-Werte gehen aus Tabelle 8.3.14 hervor. Die Ergebnisse der VAn sind vergleichbar, weshalb nur die korrigierten Werte dargestellt werden.

In der VA mit den Pr-Werten erlangten alle Haupteffekte Signifikanz: die Altersgruppe [$F(2,180) = 14.74$, $R^2 = 0.14$], das Generieren [$F(1,180) = 51.47$, $R^2 = 0.22$] und tendenziell die Instruktion [$F(1,180) = 3.03$, $R^2 = 0.02$, $p = .08$]. Zudem zeigte sich eine Tripel-Interaktion dieser Faktoren [$F(2,180) = 3.86$, $R^2 = 0.04$].

Tabelle 8.3.14:

Mittlere Trefferrate und Pr-Werte der Listenitems in Abhängigkeit des Alters, der Instruktion und des Generierens.

	Treffer				Pr-Werte			
	Standard		Inklusion		Standard		Inklusion	
	ohne	mit	ohne	mit	ohne	mit	ohne	mit
Vorschüler	.55	.82	.65	.74	.42	.68	.44	.59
Drittklässler	.62	.80	.71	.72	.46	.71	.46	.58
Erwachsene	.76	.85	.83	.88	.67	.72	.56	.77

Anmerkung: ohne = ohne Generieren; mit = mit Generieren

Eine Analyse der einfachen einfachen Haupteffekte (auch „Einfacheffekte“ genannt) zeigte, dass nach der Standardinstruktion Vorschüler [$F(1,180) = 19.70$] und Drittklässler [$F(1,180) = 18.91$] mit Generieren höhere Pr-Werte hatten als ohne Generieren, was bei Erwachsenen jedoch nicht der Fall war ($F < 1$). Nach der Inklusionsinstruktion führte das Generieren bei Vorschülern [$F(1,180) = 6.62$] und Erwachsenen [$F(1,180) = 11.98$] zu höheren Pr-Werten, nicht jedoch bei Drittklässlern [$F(1,180) = 3.82$; Mittelwerte s. Tabelle 8.3.14]. Dies verdeutlicht die Abhängigkeit des Generierungseffekts von der Instruktion und dem Alter beim korrigierten Wiedererkennen der Listenitems.

Quellendiskrimination. Die Probanden, unabhängig vom Alter, hatten keine Schwierigkeiten mit der Quellenzuweisung von Listenitems. Die Mittelwerte gehen aus Tabelle 8.3.15 hervor.

Tabelle 8.3.15:

Mittelwerte der Quellendiskrimination von Listenitems in Abhängigkeit des Alters und des Generierens.

	ohne Generieren	mit Generieren
Vorschüler	.90	.96
Drittklässler	.80	.94
Erwachsene	.84	.94

Nur das Generieren zeigte einen Effekt [$F(1,90) = 15.05$, $R^2 = 0.14$] dahingehend, dass mit diesem ($M = .95$) Listenitems besser der externen Quelle zugeordnet werden konnten als ohne Generieren ($M = .84$). Die anderen Werte der VA waren nicht bedeutsam (beide F s < 1.56).

8.3.3.2 Kritische Items

Wiedererkennen. Beide Experimente lassen sich im Anteil falscher Erinnerungen nach den Instruktionen vergleichen: Zu diesem Zweck werden die „alt“-Urteile nach der Standardin-

struktion betrachtet und die „alt“-Urteile inklusive vermeintlicher Präsentation nach der Inklusionsinstruktion, also die falschen Erinnerungen in der Inklusion. Die Altersgruppe [$F(2,180) = 20.30$, $R^2 = 0.18$] zeigte einen Haupteffekt, welcher durch eine Interaktion mit dem Generieren überlagert wurde [$F(2,180) = 21.86$, $R^2 = 0.20$]. Die Mittelwerte gehen aus Tabelle 8.3.16 hervor. Vorschüler hatten mit Generieren einen höheren Anteil falscher Erinnerungen als ohne Generieren [$F(1,180) = 5.37$], während sich dieser Effekt bei Erwachsenen umkehrte [$F(1,180) = 20.17$]. Bei Drittklässlern zeigte sich kein Unterschied zwischen den Generierungsbedingungen [$F(1,180) = 2.10$]. Die Analyse der einfachen Haupteffekte zeigte, dass sich die Altersgruppen nur in der Bedingung ohne Generieren unterschieden [$F(2,180) = 20.34$]: Vorschüler und Drittklässler hatten einen ähnlich hohen Anteil falscher Erinnerungen, aber Erwachsene einen deutlich geringeren. In der Bedingung mit Generieren zeigte sich kein Altersunterschied [$F(2,180) = 2.06$]. Alle anderen Werte der VA erlangten keine Signifikanz (alle $F_s < 1.55$). Die Mittelwerte sind Tabelle 8.3.16 zu entnehmen.

Neben den falschen Erinnerungen nach der Inklusionsinstruktion wurde für die kritischen Items auch eine VA nur für die „alt“-Urteile nach der Standard- und Inklusionsinstruktion gerechnet, deren Mittelwerte aus Tabelle 8.3.16 zu entnehmen sind. Die vorgefundenen Ergebnisse stimmen mit denen der korrigierten Werte überein, die im Folgenden dargestellt werden.

Tabelle 8.3.16:

Mittlere „alt“-Urteile für kritische Items in Abhängigkeit des Alters, des Generierens und der Instruktion.

	Standard		Inklusion falsch		Inklusion	
	ohne	mit	ohne	mit	ohne	mit
Vorschüler	.55	.72	.57	.70	.66	.81
Drittklässler	.66	.76	.71	.66	.81	.89
Erwachsene	.67	.34	.59	.29	.91	.98

Anmerkung: ohne = ohne Generieren; mit = mit Generieren

In der VA für die Pr-Werte erlangten alle Faktoren Signifikanz [Altersgruppe: $F(2,180) = 2.95$, $R^2 = 0.03$, $p = .06$; Generieren: $F(1,180) = 10.72$, $R^2 = 0.06$; Instruktion: $F(1,180) = 29.14$, $R^2 = 0.14$; Interaktion der Altersgruppe und des Generierens: $F(2,180) = 7.37$, $R^2 = 0.08$; Interaktion der Altersgruppe und der Instruktion: $F(2,180) = 10.25$, $R^2 = 0.10$; Interaktion des Generierens und der Instruktion: $F(1,180) = 13.10$, $R^2 = 0.07$; Tripel-Interaktion der Altersgruppe, des Generierens und der Instruktion: $F(2,180) = 9.95$, $R^2 = 0.10$]. Die Mittelwerte der VA sind in Tabelle 8.3.17 zusammengefasst.

Tabelle 8.3.17:

Mittlere Pr-Werte der kritischen Items in Abhängigkeit des Alters, des Generierens und der Instruktion.

	Standard		Inklusion	
	ohne	mit	ohne	mit
Vorschüler	.42	.57	.45	.67
Drittklässler	.50	.67	.56	.74
Erwachsene	.58	.21	.64	.87

Anmerkung: ohne = ohne Generieren; mit = mit Generieren

Die Analyse der Einfacheffekte zeigte, dass sich das Generieren unterschiedlich auf die Pr-Werte je nach Altersgruppe und Instruktion auswirkte. Nach der Inklusionsinstruktion hatten alle Altersgruppen übereinstimmend mit der Annahme aus Hypothese 4 mit Generieren höhere Pr-Werte als ohne Generieren [Vorschüler: $F(1,180) = 8.32$; Drittklässler: $F(1,180) = 5.73$, Erwachsene: $F(1,180) = 10.01$], so dass sogar Vorschüler das Generieren zur Stärkung der Aktivierung nutzen konnten. Erwartungsgemäß (Hypothese 6) führte das Generieren nach der Standardinstruktion bei den beiden jüngeren Altersgruppen zu mehr falschen Erinnerungen [Vorschüler: $F(1,180) = 4.45$; Drittklässler: $F(1,180) = 5.71$]. Bei Erwachsenen aber wurden diese entgegen der Erwartungen reduziert [$F(1,180) = 24.23$]. Erwachsene konnten also das Generieren nutzen, um falschen Erinnerungen entgegenzuwirken.

Quellendiskrimination. In der VA für die Fehlattribution der kritischen Items erlangten die Altersgruppe [$F(2,90) = 33.21$, $R^2 = 0.43$] und das Generieren [$F(1,90) = 19.81$, $R^2 = 0.18$] Signifikanz, welche miteinander interagierten [$F(2,90) = 5.68$, $R^2 = 0.11$].

Tabelle 8.3.17:

Mittelwerte der Fehlattribution von kritischen Items in Abhängigkeit des Alters und des Generierens.

	ohne Generieren	mit Generieren
Vorschüler	.89	.86
Drittklässler	.87	.72
Erwachsene	.67	.30

Entgegen der Erwartungen aus Hypothese 5 attribuierten Vorschüler häufig falsch, egal ob zum Generieren des kritischen Items aufgefordert wurde oder nicht ($F < 1$) und Drittklässler konnten sogar leicht vom Generieren profitieren [$F(1,90) = 4.44$]. Erwachsene hatten sehr großen Nutzen vom Generieren [$F(1,90) = 26.52$; Mittelwerte s. Tabelle 8.3.17]. Wie bereits dargestellt unterschieden sich die Altersgruppen sowohl mit [$F(2,90) = 33.01$] als auch ohne

Generieren [$F(2,90) = 5.88$] darin, dass Erwachsene kritische Items seltener der falschen Quelle zuwiesen, was deren besseren Fähigkeiten zur Quellendiskrimination entspricht.

8.3.4 Diskussion

Experiment 3B wurde unter der Zielsetzung durchgeführt zu prüfen, welche Art von Aktivierung (eher eine automatische oder eine bewusste) der Inklusionstest erfasst. Darüber hinaus sollte untersucht werden, auf welcher Art von Aktivierung die beobachteten Altersunterschiede gründen. Daher wurde in Experiment 3B zusätzlich eine Generierungsinstruktion erteilt, welche eine semantisch-relationale Verarbeitung bei allen Probanden fördern sollte. Basierten die beobachteten Altersunterschiede in der Aktivierung auf unzureichenden elaborativen Verarbeitungsstrategien der Kinder, dann hätten diese besonders von der Aufgabe profitieren und Altersunterschiede minimiert werden sollen. Des Weiteren sollte im vorliegenden Experiment untersucht werden, auf welcher Art von Aktivierung (eher einer automatischen oder einer bewussten) falsche Erinnerungen im Standardtest gründen und ob sich Altersunterschiede zeigen. Von besonderer Wichtigkeit war der Vergleich der Experimente 3A und 3B, der Aufschlüsse zur Auswirkung der Generierungsinstruktion auf Aktivierung, Quellendiskrimination und auch falsche Erinnerungen ermöglichte.

Wie in Experiment 3A ergaben sich in Experiment 3B zahlreiche Abfolgeeffekte, indem der zuerst erteilte Test die Leistungen im zweiten Test beeinflusste. Aus diesem Grund wurden separate Analysen mit den in der ersten Testphase erteilten Tests gerechnet. Die Ergebnisse der Analysen unter Beachtung der Abfolge wurden im zugehörigen Ergebnisteil kurz diskutiert.

Anders als in Experiment 3A zeigten Erwachsene in Experiment 3B kein besseres Wiedererkennen der Listenitems nach der Standardlerninstruktion als Drittklässler und Vorschüler. Dies könnte daran liegen, dass verglichen mit der Standardlerninstruktion in Experiment 3A die Kinder, die spontan keine oder kaum elaborative Verarbeitungsstrategien einsetzen, erwartungsgemäß von der verlangten elaborativen Verarbeitung in Experiment 3B profitierten. Erwachsene setzen vermutlich diese Verarbeitungsstrategie spontan ein, daher förderte das Generieren nicht ihr korrektes Wiedererkennen. Das Generieren hatte also in Abhängigkeit des Alters unterschiedliche Effekte auf korrekte und falsche Erinnerungen: Bei Kindern zeigte sich der “more is less“ Effekt, denn durch das Generieren hatten sie vermehrt richtige aber auch tendenziell mehr falsche Erinnerungen. Bei Erwachsenen war dies anders: Das Generie-

ren verbesserte nicht ihre korrekte Erinnerungsleistung, aber es reduzierte falsche Erinnerungen.

In der Aufgabe zum Generieren zeigte sich, dass Erwachsene sich von Kindern in der Generierungshäufigkeit der kritischen Items unterschieden. Die Wortlisten führten trotz ihrer Altersnormiertheit dazu, dass Erwachsene die kritischen Items leichter identifizieren konnten als Kinder, womit die Listen nicht altersäquivalent waren. Dies ließe sich damit erklären, dass Erwachsene über ein höheres Wissen verfügen als Kinder (s.u.).

In den um die Antworttendenzen korrigierten Werten des Wiedererkennens kritischer Items von Experiment 3A und 3B konnte festgestellt werden, dass die Aktivierung der kritischen Items in allen Altersgruppen durch das Generieren gleichermaßen gesteigert werden konnte. Nach der Inklusionsinstruktion konnten Erwachsene mit und ohne Generieren mehr kritische Items wiedererkennen als Vorschüler. Drittklässler unterschieden sich nicht von beiden Gruppen. Bei Erwachsenen konnte eine stärkere Aktivierung nachgewiesen werden als bei Kindern, welche darauf zurückgehen kann, dass Erwachsene das kritische Item häufiger zu den DRM-Listen generierten verglichen mit Kindern. Möglicherweise könnten die Aktivierungsunterschiede Effekte der Identifizierbarkeit der kritischen Items widerspiegeln (vgl. Neuschatz et al., 2003), da in Abhängigkeit des Alters unterschiedliche mittlere Werte für die Generierungshäufigkeit, d.h. die Identifizierbarkeit kritischer Items resultierten. Aber Kinder waren auch imstande, das Thema der Liste zu erfassen, was sich an der hohen Anzahl an Nennungen der kritischen Items als fehlendes Wort zu den Listen in der Lernphase von Experiment 3B zeigt und an der hohen Anzahl von „alt“-Antworten für die kritischen Items im Inklusionstest in Experiment 3A und 3B. Die Altersunterschiede in der Identifizierbarkeit können auf Wissensunterschieden gründen, denn wie aus den Annahmen zu semantischen Netzwerken hervorgeht, wird durch mehr Wissen die Verbindungsstärke zwischen den Knoten gestärkt (vgl. Abschnitt 2.2). Mit zunehmendem Alter nimmt die Vernetzung innerhalb und zwischen den Konzepten zu, so dass die resultierende Aktivierung stärker ist. Die Bedeutung des Wissens zeigt sich darin, dass sich übliche Alterseffekte umdrehen lassen, wenn Kinder über mehr Wissen über das Lernmaterial als Erwachsene verfügen, was z.B. Chi (1978) am Vergleich der Rekonstruktionsleistung von 10-jährigen Schachexperten und erwachsenen Schachnovizen demonstrierte (s. Abschnitt 3.4 Bereichsspezifisches Wissen). Um altersäquivalente Listen zu erstellen, müssten derartige Wissensunterschiede ausgeglichen werden, was eine sehr schwierige Aufgabe darstellt, da es kaum Gebiete gibt, in denen Kinder (abgesehen von kindlichen Expertengruppen) über gleiches bzw. mehr Wissen wie Erwachsene verfügen. Es könnte versucht werden, spezielle Kategorien für Kinder zu erstellen, in de-

nen ihr Wissen ähnlich jenem von Erwachsenen ist. Problematisch ist, dass Erwachsene sich in den Gebieten, in denen Kinder über mehr Wissen als sie verfügen (z.B. Comics), nicht oder unzureichend auskennen. Gelänge es derartige altersadäquate Listen zu erstellen, dann könnte geprüft werden, ob die Altersvariable immer noch einen Einfluss auf die Identifizierbarkeit kritischer Items und die Aktivierung hat.

Um die Aktivierung getrennt nach Altersgruppen zu prüfen, können Listen mit hohen und niedrigen Generierungshäufigkeiten für die verschiedenen Altersgruppen konstruiert werden. Zu diesem Zweck wurden die Listen getrennt für alle Altersgruppen in gut und schlecht identifizierbare unterteilt, d.h. von den acht Listen eines jeden Sets wurden die vier höchsten und die vier geringsten mittleren Häufigkeiten der generierten kritischen Items aus Tabelle 8.3.8 bestimmt. Für diese wurden die Mittelwerte der als „alt“ bezeichneten kritischen Items aus Tabelle 8.3.9 entnommen und getrennt für die gut und schlecht identifizierbaren Listen separat berechnet. Daraus resultieren neue Mittelwerte wiedererkannter kritischer Items, die aus Tabelle 8.3.18 hervorgehen.

Tabelle 8.3.18:

Mittelwerte der wiedererkannten kritischen Items in Abhängigkeit des Alters, der Identifizierbarkeit, den Experimenten und der Instruktion.

	Standard				Inklusion			
	Experiment 3A		Experiment 3B		Experiment 3A		Experiment 3B	
	gut iden- tifiz.	schlecht identifiz.	gut iden- tifiz.	schlecht identifiz.	gut iden- tifiz.	schlecht identifiz.	gut iden- tifiz.	schlecht identifiz.
Vorschüler	.54	.56	.70	.67	.69	.68	.78	.81
Drittklässler	.67	.60	.73	.66	.79	.74	.84	.87
Erwachsene	.73	.62	.36	.31	.79	.71	.96	.97

Auffallend war, dass sich diese Werte nicht in Abhängigkeit der Identifizierbarkeit nach der Standard- und Inklusionsinstruktion unterschieden, d.h. Listen mit gut identifizierbaren kritischen Items wurden nicht häufiger wiedererkannt als Listen mit schlechter identifizierbaren. Die einzig auffallende Differenz zeigte sich bei Erwachsenen in Experiment 3A: Nach der Standardinstruktion wurden Listen mit gut identifizierbaren kritischen Items etwas häufiger falsch wiedererkannt als jene mit schlecht identifizierbaren. Jedoch ist diese Differenz nicht sehr groß.

Trotz der Unterteilung der Listen für jede Altersgruppe in gut und schlecht identifizierbare Listen blieben die Altersunterschiede in der Aktivierung bestehen. Nach diesen Resultaten scheint die Rolle der Identifizierbarkeit eher gering und nicht die Haupteinflussvariable auf

Altersunterschiede in der Aktivierung zu sein. Festzuhalten bleibt, dass Erwachsene kritische Items besser identifizieren konnten und eine stärkere Aktivierung der kritischen Items hatten als Kinder. An welchen Faktoren dies liegt muss weitergehend untersucht werden.

Um zu prüfen, welche Art von Aktivierung der Inklusionstest erfasst, wurden die „alt“-Antworten für die kritischen Items, die in der Lernphase generiert wurden mit denen, die nicht generiert wurden, verglichen. Nach der Inklusionsinstruktion bezeichneten Erwachsene nahezu gleich viele selbst generierte und nicht generierte kritische Items als „alt“, was dafür spricht, dass der Inklusionstest bei ihnen in hohem Maß die automatische Aktivierung erfasst hat. Jedoch sollten diese Ergebnisse mit Vorsicht interpretiert werden, da es bei Erwachsenen wenig nicht generierte kritische Items gab. Die Ergebnisse der Kinder weisen darauf hin, dass der Inklusionstest bei ihnen eher die bewusste Aktivierung erfasst hat, denn sie erkannten eher die generierten kritischen Items wieder. Durch die noch hohe Wiedererkennungsraten nicht generierter kritischer Items scheint es, dass auch die automatische Aktivierung getestet wurde. Die entgegen den Erwartungen ausbleibende Reduktion alterskorrelierter Unterschiede in der Aktivierung durch das Generieren könnte sich darauf zurückführen lassen, dass Kinder die kritischen Items weniger häufig in der Lernphase generierten als Erwachsene und dass der Inklusionstest bei Kindern weniger die automatische Aktivierung zu erfassen schien. Nach der Standardinstruktion zeigten sich auch unterschiedliche Ergebnisse in Abhängigkeit des Alters: Vorschüler erkannten hier ebenso eher die selbst generierten kritischen Items falsch wieder als die nicht generierten, bei Drittklässlern war das Wiedererkennen unbeeinflusst vom Generieren und Erwachsene bezeichneten deutlich mehr nicht generierte kritische Items falsch als „alt“ als selbst generierte. Demnach schienen falsche Erinnerungen bei Vorschülern eine Folge bewusster Aktivierung in der Lernphase zu sein, bei Erwachsenen eher eine Folge automatischer Aktivierung.

Die Befunde zu bewusst aktivierten kritischen Items bei Kindern sind in Einklang mit der Auffassung von Howe (2005), dass jüngere Kinder kritische Items nur bewusst unter kognitiver Anstrengung aktivieren. Diese Annahme beruht darauf, dass jüngere Kinder einen gerichteten Vergessenseffekt sowohl für Listenitems als auch für kritische Items zeigen, Erwachsene jedoch nicht für kritische Items, was bedeutet, dass beide Arten von Aktivierung nicht vergleichbar sind. Also sollten Kinder im Gegensatz zu Erwachsenen die kritischen Items bewusst hemmen und sie deshalb auch bewusst aktiviert haben können. Dadurch sollten sie aber weniger falsche Erinnerungen als Erwachsene haben. In einem Paradigma zum gerichteten Vergessen wurden fünf-, sieben- und 11-jährige Kinder zu falschen Erinnerungen untersucht. Kindern verschiedener Altersgruppen wurden Instruktionen zum gerichteten Vergessen und

zum Erinnern gegeben. Es wurde angenommen, dass Kinder durch die bewusste Aktivierung kritischer Items gerichtete Vergessenseffekte falscher Erinnerungen hätten, nicht aber Erwachsene. Die Ergebnisse zeigten, dass nach der Vergessensinstruktion Kinder falsche Erinnerungen unterdrücken und damit verhindern konnten. Der Anteil falscher Erinnerungen war bei Fünf- und Siebenjährigen vergleichbar, nur 11-Jährige hatten mehr falsche Erinnerungen.

Darüber hinaus wurden die Ergebnisse mit der Studie von Kimball und Bjork (2002) verglichen, die Erwachsene mit dieser Prozedur untersuchten. Kinder hatten weniger falsche und korrekte Erinnerungen als Erwachsene, sowohl nach der Vergessens- als auch nach der Erinnerungsinstruktion. Im Gegensatz zu Erwachsenen konnten Kinder kritische Items hemmen und zeigten einen gerichteten Vergessenseffekt.

Davon ausgehend schlussfolgert Howe (2005), dass Personen, bei denen ein geringes Maß an automatischer Aktivierung vorliegt, eher fähig sein sollten die falsche Wiedergabe kritischer Items zu unterdrücken. Da die Aktivierungsausbreitung bei Kindern im assoziativen Netzwerk kaum automatisch erfolgt sondern mit einem erhöhten Maß an kognitiver Anstrengung, erinnern sie weniger kritische Items falsch als ältere Kinder und Erwachsene. Die größere kognitive Anstrengung soll nach Howe (2005) bei den jüngeren Kindern, im Gegensatz zu älteren und Erwachsenen, dazu führen, dass ihnen das kritische Item stärker bewusst wird und dadurch eher wie ein Listenitem behandelt wird. Mit zunehmendem Alter sollte die Verarbeitung automatischer und weniger kognitiv anstrengend werden und falsche Erinnerungen wahrscheinlicher machen (Howe, 2005).

Es ist kritisch anzumerken, dass die bewusste Aktivierung jüngerer Kinder wohl kaum die Folge semantisch-relationaler Verarbeitungsstrategien sein kann. Diese Annahme ist unvereinbar mit den Befunden zur Gedächtnisentwicklung (vgl. Kapitel 3). Vielmehr könnte es sein, dass das kritische Item durch eine automatische Aktivierungsausbreitung bewusst werden konnte. Demnach sollte zwischen drei Arten von Aktivierung unterschieden werden:

- (1) der bewussten Aktivierung als Folge semantisch-relationaler Verarbeitungsstrategien,
- (2) der Aktivierung als Folge automatischer Aktivierungsausbreitung im Netzwerk, durch welche das kritische Item bewusst wird und
- (3) wie (2), jedoch wird den Versuchspersonen das kritische Item nicht bewusst.

Die beiden letzten Arten der Aktivierung lehnen sich an das Modell der Aktivierungsquelle ("source of activation model" SAC) von Reder, Nhouyvanisvong, Schunn, Ayers, Angstadt und Hiraki (2000) an. Das SAC Modell beschreibt neuronale Netzwerke: In der Lernphase aktiviert ein Wort einen Konzept- und einen Ereignisknoten, wobei der Ereignisknoten Kon-

textinformationen zu der Enkodierepisode beinhaltet. Während des Abrufs verteilt sich die Aktivierung von dem Konzeptknoten zu den verbundenen Ereignisknoten. Es ist nun relevant, ob es sich um Konzeptknoten häufiger oder seltener Wörter handelt: Die Konzeptknoten häufiger Wörter werden stärker aktiviert, aber durch die Verbindung mit vielen Ereignisknoten erhält der einzelne nur wenig Aktivierung. Demgegenüber werden die Konzeptknoten seltener Wörter weniger aktiviert, dennoch erhalten die wenigen verbundenen Ereignisknoten jeweils mehr Aktivierung. Übertragen auf die eigene Untersuchung bedeutet dies, dass nicht relevant ist, ob die Wörter häufig oder selten sind, sondern wie die Aktivierungsausbreitung in den verschiedenen Altersgruppen abläuft. Erwachsene verfügen über mehr Wissen, das auch besser strukturiert, also vernetzt ist. Demnach werden im DRM-Paradigma ihre Konzeptknoten stärker aktiviert, allerdings erhalten die einzelnen Ereignisknoten eine vergleichbar starke Aktivierung wie die der jüngeren Kinder, deren Konzeptknoten weniger aktiviert werden, aber weniger Verbindung zu Ereignisknoten haben. Bewusst kann den Versuchspersonen das kritische Item dann werden, wenn der Prozess noch nicht automatisiert abläuft, was bei Kindern der Fall sein sollte.

Die eigenen Ergebnisse weisen darauf hin, dass falsche Erinnerungen bei Erwachsenen hauptsächlich eine Folge automatischer Aktivierung zu sein scheinen, was in Einklang mit verschiedenen Befunden ist (Dodd, Sheard & MacLeod, 2006; Gallo & Seamon, 2004; Gallo & Seamon, 2002a; Seamon et al., 1998). Wie in Abschnitt 2.2.2 erläutert wurde, zeigte sich bei Gallo und Seamon (2002a), dass das Memorieren der kritischen Items deren falsche Reproduktion erhöhte, aber nicht deren falsches Wiedererkennen. Falsche Erinnerungen entstanden unabhängig vom Memorieren, so dass davon ausgegangen werden kann, dass keine bewusste Aktivierung in der Enkodierphase notwendig war, was im Gegensatz zu den Befunden von Marsh und Bower (2004) steht.

Für automatisch aktivierte kritische Items scheint die Quellendiskrimination, also die Zuordnung zu der korrekten internen Quelle, besonders schwierig zu sein verglichen mit bewusst aktivierten kritischen Items, da für erstere z.B. die mit internen Quellen verbundenen enkodierten Attribute fehlen (vgl. Johnson & Raye, 1981). Allgemein gehen Johnson und Raye (1981) davon aus, dass die Quelle einer Information nicht zusammen mit dieser abgespeichert, sondern in der Testphase aus enkodierten Attributen rekonstruiert wird (s. Abschnitt 2.2.2). Informationen aus internen Quellen weisen Attribute über kognitive Operationen auf, die, bezogen auf das vorliegende Experiment, bei automatisch aktivierten Items fehlen.

Die Quellediskrimination wurde in der Nachbefragung erfasst und es zeigte sich, dass Erwachsene über bessere Quellediskriminationsleistungen verfügen als Kinder, die sich nicht voneinander unterscheiden.

Auch aus dem Vergleich der Quellediskrimination mit vs. ohne Generieren ging hervor, dass die Quellediskriminationsleistungen bei dieser Aufgabe altersabhängig waren. Vorschüler und Drittklässler ordneten mit und ohne Generieren kritische Items häufiger der falschen Quelle zu als Erwachsene. Während Vorschüler mit sowie ohne Generieren kritische Items der falschen Quelle zuwiesen, konnten Drittklässler etwas vom Generieren profitieren und Erwachsene hatten großen Nutzen vom Generieren und bezeichneten deutlich weniger kritische Items falsch als „alt“. Das Generieren könnte bei den beiden älteren Gruppen zu mehr quellenrelevanten, in diesem Fall internalen Informationen geführt haben, die die Quellediskrimination erleichterten. Durch die alterskorrelierte Verbesserung der strategischen Quellediskrimination, die vom Metagedächtnis abhängig ist, lassen sich die Altersunterschiede erklären.

Durch das Generieren zu jeder Liste sollte die semantisch-relationale Verarbeitung gefördert werden, welche eine verstärkte semantische Aktivierung der kritischen Items bewirken und damit zu vermehrt richtigen aber auch falschen Erinnerungen führen sollte (McDermott & Watson, 2001; Payne et al., 1996; Thapar & McDermott, 2001). Das Generieren bewirkte bei Erwachsenen und Drittklässlern eine stärkere Aktivierung und bessere Quellediskrimination, wirkte sich jedoch unterschiedlich auf ihre falschen Erinnerungen aus. Nur Erwachsene konnten das Generieren zur Reduktion falscher Erinnerungen durch ihre bessere Quellediskrimination nutzen, nicht Kinder.

Andere Untersuchungen zur Generierungsinstruktion zeigten eine Zunahme falscher Erinnerungen bei Erwachsenen (Dewhurst, Barry & Holmes, 2005; Marsh & Bower, 2004). Diese Autoren führen ihre Ergebnisse darauf zurück, dass das Generieren die bewusste Aktivierung bei der Enkodierung erhöht (Dewhurst et al., 2005), die Vertrautheit steigert und in Folge Quellediskriminationsprobleme auftreten (Marsh & Bower, 2004).

Die eigenen Ergebnisse wiesen darauf hin, dass die Förderung der semantisch-relationalen Verarbeitung falsche Erinnerungen bei Kindern erhöhte, was im Gegensatz zu den Befunden einer aktuellen Untersuchung von Howe (2006) ist. Hier wurde geprüft, ob sich bei semantischen (kategorialen) Wortlisten falsche Erinnerungen bei Kindern verändern, wenn ihnen entsprechende relationale Verarbeitungsstrategien gegeben werden. Der Hälfte der Probanden wurde die jeweilige Kategorie der Liste explizit genannt (Tierkategorie: „Alle Wörter der

Liste sind Namen von Tieren“), der anderen Hälfte nicht. Die relationale Verarbeitung hatte hier keinen Einfluss auf falsche Erinnerungen bei Kindern, so dass Howe (2006) schlussfolgert, dass die assoziative Aktivierung bei Kindern von größerer Bedeutung zu sein scheint als die semantische Aktivierung. In der eigenen Untersuchung wurden assoziative und nicht kategoriale Listen verwendet, die zusammen mit einer anderen Instruktion zur semantisch-relationalen Verarbeitung die abweichenden Ergebnisse erklären können.

Im Gegensatz zu den Befunden Howes (2006) und teilweise in Einklang mit den eigenen Ergebnissen, konnten Lampinen, Leding, Reed und Odegard (2006) zeigen, dass Drittklässler durch Förderung des Strategieeinsatzes (das gemeinsame Thema der Liste suchen) *vermehrt* falsche Erinnerungen produzieren (Thema suchen: $M = .66$; Kontrollbedingung: $M = .45$). Bei Erstklässlern (Thema suchen: $M = .55$; Kontrollbedingung: $M = .44$) und Erwachsenen (Thema suchen: $M = .59$; Kontrollbedingung: $M = .45$) war dies aber nicht der Fall. Dieses Ergebnis ähnelt dem der eigenen Untersuchung für die *beiden* jüngeren Altersgruppen, die durch das Generieren vermehrt falsche Erinnerungen hatten als ohne dieses. Ausgehend von der FTT führen Lampinen und Mitarbeiter (2006) ihr Ergebnis darauf zurück, dass die Fähigkeit den „gist“ der Liste zu erfassen (bzw. nach dem AMA die Aktivierung) sich mit steigendem Alter verbessert und Drittklässler die entsprechenden dargebotenen Strategien nutzen können, die sie noch nicht selbstständig verwenden. Erwachsene sollten diese Strategien aber ohnehin spontan einsetzen, so dass sich ihre falschen Erinnerungen nicht veränderten und Erstklässler könnten von der dargebotenen Strategie noch keinen Nutzen haben, da diese wissensabhängig zu sein scheinen.

Während bei Lampinen und Mitarbeitern (2006) durch die Instruktion zur Förderung des Strategieeinsatzes falsche Erinnerungen nur bei Drittklässlern erhöht wurden, hatten in der eigenen Untersuchung Drittklässler *und* Vorschüler vermehrt falsche Erinnerungen, denn die Instruktion zur Förderung des Strategieeinsatzes (also das Generieren) stärkte ihre Aktivierung und die Quellendiskrimination konnte dieser nicht entgegenwirken. Auch Erwachsene hatten durch die Instruktion eine stärkere Aktivierung, aber zudem eine bessere Quellendiskrimination, wodurch sie *weniger* falsche Erinnerungen hatten. Die von Lampinen und Mitarbeiter (2006) abweichenden Ergebnisse der eigenen Untersuchung für Vorschüler und Erwachsene könnten an den unterschiedlichen Instruktionen zur Förderung des Strategieeinsatzes liegen. Lampinen und Mitarbeiter (2006) wiesen ihre Probanden an, das gemeinsame Thema der Liste zu suchen, jedoch sollen sie dieses nicht laut sagen, sondern nur daran denken. Es erscheint fraglich, ob die jüngste Altersgruppe bereits fähig war, das Thema der Liste zu suchen, ohne es zu nennen. Es könnte sein, dass sie der Instruktion nicht Folge leisten *konnten* und damit

ihre Aktivierung geringer war gegenüber den anderen Altersgruppen. Zu den abweichenden Befunden erwachsener Probanden kann angemerkt werden, dass die älteste Gruppe (studentische Versuchspersonen) der Instruktion das gemeinsame Thema der Liste zu suchen nicht Folge leisten *wollten*, da sie wenig motiviert gewesen sein könnten sich anzustrengen. Daher könnte es sein, dass sie eine geringere Aktivierung und entsprechend weniger falsche Erinnerungen hatten. Erfahrungen mit studentischen Versuchspersonen unterstützen die Annahme und zeigen, dass diese Personengruppe im Allgemeinen weniger als jüngere Altersgruppen motiviert ist, an Experimenten teilzunehmen.

Das vorgefundene Ergebnismuster wird durch eine universitätsinterne Replikation der Untersuchung an studentischen Probanden unterstützt. Auch in dieser Untersuchung an 64 Probanden konnte demonstriert werden, dass mit Generieren nach der Inklusionsinstruktion mehr „alt“-Urteile für die kritischen Items gegeben wurden als ohne Generieren, was für eine höhere Aktivierung spricht. Auch wurde durch das Generieren der falschen Quellenzuordnung entgegengewirkt. Diese Ergebnisse sind in Einklang mit den eigenen und deuten darauf hin, dass bei Erwachsenen die Aktivierung durch die Generierungsinstruktion gefördert werden konnte und die Quellendiskrimination vereinfacht wurde. Da sich mit Generieren insgesamt deutlich weniger falsche Erinnerungen als ohne dieses zeigten, schien die durch das Generieren verbesserte Quellendiskrimination ausschlaggebend zu sein.

Zusammenfassend weisen die Ergebnisse von Experiment 3B auf eine Überlegenheit Erwachsener sowohl in der bewussten als auch automatischen Aktivierung der kritischen Items hin sowie in ihren besseren Fähigkeiten zur Quellendiskrimination. Drittklässler und Vorschüler unterschieden sich nicht in der Generierungshäufigkeit kritischer Items, in der geringeren Aktivierung kritischer Items und schlechteren Quellendiskrimination verglichen mit Erwachsenen und dementsprechend gab es keinen Unterschied der beiden jüngeren Gruppen in ihren falschen Erinnerungen. Der Inklusionstest erfasste bei Erwachsenen eher die automatische Aktivierung und bei Kindern eher die bewusste, denn Erwachsene erkannten gleichermaßen generierte und nicht generierte kritische Items wieder, Kinder hingegen eher die generierten kritischen Items. Durch den Vergleich der Experimente 3A und 3B waren Rückschlüsse auf die Bedeutung der Generierungsinstruktion möglich. Die Ergebnisse nach der Inklusionsinstruktion für den zuerst erteilten Test zeigten für alle Probanden einen Generierungseffekt, denn durch das Generieren resultierte eine stärkere Aktivierung. Der Effekt nahm alterskorreliert zu. In der Quellendiskrimination gab es nur Generierungseffekte für Drittklässler und Erwachsene, die mit Generieren weniger kritische Items der falschen Quelle zuwiesen. Das

Generieren erhöhte bei Kindern falsche Erinnerungen, jedoch reduzierte es sie bei Erwachsenen.

Anhand der Ergebnisse lassen sich vielfältige Schlüsse ziehen und neue Überlegungen zu weiteren Untersuchungen treffen. Die Instruktion zur semantisch-relationalen Verarbeitung konnte die Aktivierung der kritischen Items in allen Altersgruppen erhöhen, aber die Altersunterschiede blieben bestehen. Es soll in einem weiteren Experiment geprüft werden, ob eine andere Form des Generierens wie bei Marsh und Bower (2004) die Altersunterschiede in der Aktivierung aufheben kann. Anstatt die Instruktion zu erteilen nach jeder Liste zu generieren, sollen die Probanden zu jedem Item ein Wort generieren, wodurch neben der semantisch-relationalen möglicherweise auch die itemspezifische Verarbeitung gefördert wird.

In Experiment 3A und 3B ergaben sich zwar alterskorrelierte Unterschiede in der Aktivierung, aber entsprechende Unterschiede in der Quellendiskriminationsleistung fielen deutlich stärker aus. Im nächsten Experiment sollte die Quellendiskrimination differenzierter erfasst werden, da sich die beiden jüngeren Altersgruppen nicht voneinander unterschieden. Es sollte geprüft werden, ob die Quellendiskriminationsleistungen in der sehr schweren Aufgabe sich erst in der frühen Adoleszenz verbessern.

Im Quellendiskriminationstest in Experiment 3A und 3B wurde nach der Inklusionsinstruktion nur die Zuordnung zur externen Quelle geprüft („ja/nein“). Dadurch wird die Quellendiskrimination nicht sehr differenziert geprüft, denn nach Befunden aus der Zeugenaussagenforschung mit dem Falschinformationsparadigma (Paradigma nachträglicher irreführender Informationen) ist es möglich, dass insbesondere jüngere Kinder Informationen zugleich der falschen und richtigen Quelle zuordnen (Ackil & Zaragoza, 1995; 1998). Kinder könnten also wissen, dass sie die kritischen Items selbst generiert haben (insbesondere nach der Annahme, dass falsche Erinnerungen auf bewusster Aktivierung basieren), aber sie können auch meinen, diese seien präsentiert worden. Anhand eines anderen Antwortformats, d.h. einer Unterscheidung der externen *und* internen Quelle, könnten somit genauere Aussagen ermöglicht werden.

Es sollte geprüft werden, ob sich die Ergebnisse von Experiment 3A und 3B mit einem solchen Quellendiskriminationstest replizieren lassen und ob dieser Test, der verdeutlicht, dass die Vertrautheit kritischer Items auf zwei Quellen zurückgehen kann, zu besserer Quellendiskrimination und entsprechend weniger falschen Erinnerungen führt.

9 Alterskorrelierte Differenzen im Quellendiskriminationstest?

9.1 Experiment 4

Ausgehend von dem Vergleich der Experimente 3A und 3B, welcher zeigte, dass das Generieren zu einer Steigerung der Aktivierung der kritischen Items in allen Altersgruppen führte, wurde für das vierte Experiment die Überlegung aufgestellt die vorherigen Experimente zu replizieren, aber mit anderen Aktivierungs- und Quellendiskriminationsmaßen. Außerdem sollte die Auswirkung einer anderen Generierungsinstruktion geprüft werden, wobei die von Marsh und Bower (2004) entwickelte Generierungsinstruktion übernommen wurde (vgl. Abschnitt 5.3). Damit können die eigenen Ergebnisse besser mit ihren verglichen werden. Jedoch sollten die Probanden nicht zu jedem Listenitem zwei Wörter generieren, sondern nur ein Wort, da Erstklässler häufig nur ein Wort nannten. Diese Art der Verarbeitung sollte semantisch-relational sein aufgrund des Generierens kritischer Items, aber auch itemspezifisch, da zu jedem Item assoziiert wird.

Die Instruktion sollte intraindividuell variiert werden, so dass die Bedingungen mit und ohne Generieren direkt miteinander verglichen werden können. In Experiment 3B sollten die Probanden nach der Präsentation *aller* Items einer Liste das fehlende zugehörige Wort finden. Sollen die Probanden hingegen zu jedem Listenitem ein Wort assoziieren, so steigt die Wahrscheinlichkeit, dass das kritische Item mindestens einmal genannt wird.

Da die alterskorrelierten Unterschiede in der Quellendiskriminationsleistung stark waren, die beiden jüngeren Altersgruppen sich aber nicht unterschieden, sollte im vorliegenden Experiment die Quellendiskrimination differenzierter getestet werden. In Experiment 3A und 3B wurde nach der Inklusionsinstruktion im Wiedererkennenstest generell zu den Items gefragt, ob diese in der Lernphase präsentiert worden waren („Kam das Wort vor?“) und dazu sollten die Probanden ein „alt/neu“-Urteil abgeben. Die Antwort „alt“ konnte entweder bedeuten, dass die Versuchsleiterin das Item vorgelesen hatte oder dass die Probanden an das Item gedacht hatten. Die anschließende Frage zu den als „alt“ bezeichneten Items lautete „Hat die Versuchsleiterin das Wort in der Lernphase vorgelesen?“. Dadurch konnte nur die externe Quellenzuordnung, nicht aber die interne Quellenzuordnung geprüft werden, was zu einer eingeschränkten Interpretierbarkeit der Befunde führte. Da Vorschüler in Experiment 3A und 3B häufig auf die Frage, ob das Item vorgelesen wurde mit „ja“ antworteten, lässt sich nicht eindeutig die Quellenzuordnung prüfen, denn es erscheint auch möglich, dass die Probanden gleichzeitig an das Item gedacht hatten. Problematisch ist also die eindeutige Bestimmung, ob

die Probanden das Item nur der externen Quelle (vorgelesen) zuordneten oder gleichzeitig auch der internen (gedacht). Zur Prüfung der Quellendiskriminationsleistungen ist es daher aufschlussreicher ein differenziertes Urteil zu verlangen.

Das eigene Vorgehen basierte auf den Untersuchungen von Ackil und Zaragoza (1995, 1998), die prüfen wollten, ob jüngere Kinder zu stärkeren Suggestibilitätseffekten im Falschinformationsparadigma neigen als ältere Kinder und Erwachsene. Die Autoren zeigten den Probanden vierer Altersgruppen (Erst-, Dritt-, Fünftklässler und Erwachsene) zunächst einen achtminütigen Videoausschnitt, dem folgte eine vorgelesene Zusammenfassung mit suggerierten Fehlinformationen über das Video, die dort hätten vorkommen können (1995). In der Untersuchung von 1998 erfanden die Versuchspersonen diese Informationen auf Fragen hin selber, was eher dem DRM-Paradigma entspricht (zumindest wenn es um die bewusste Aktivierung geht). In einem Nachtest (entweder sofort oder eine Woche später) wurde anhand von zwei aufeinander folgenden Fragen nach jedem Item eine Quellendiskrimination über verschiedene Details erhoben: (1) Kam das Item im Video vor? (2) Wurde das Item in der Zusammenfassung vorgelesen (1995)? bzw. Wurde selbst das Item genannt (1998)? Es zeigte sich, dass es einen deutlichen Altersunterschied in der Fehlattribution der suggerierten Informationen gab. Erstklässler wiesen kritische Items häufiger der falschen externen Quelle zu als alle anderen Gruppen und sie antworteten auf beide Fragen häufig mit „ja“. Demnach wussten sie, dass die suggerierten Informationen in der Zusammenfassung vorgekommen waren bzw. dass sie diese selber konfabuliert hatten, aber waren zugleich der Meinung, diese Information auch im Video gesehen zu haben. Dritt- und Fünftklässler unterschieden sich nicht, aber hatten auch häufigere Fehlattributionen als Erwachsene, die am wenigsten Probleme mit der Aufgabe hatten. Hingegen unterschieden sich die Altersgruppen nicht in der korrekten internen Quellenzuweisung. Die Ergebnismuster beider Experimente waren vergleichbar, obwohl es sich in der Untersuchung von 1995 um eine extern-extern Quellendiskrimination handelte und in der Studie von 1998 um eine extern-intern Unterscheidung wie beim DRM-Paradigma.

Im vorliegenden Experiment sollten die Probanden aller Altersstufen also die Items nicht als „alt“ oder „neu“ beurteilen, sondern anhand zweier getrennter Fragen unterteilen, ob diese vorgelesen wurden und/oder ob an diese selbst gedacht (ohne Generieren) bzw. ob diese gesagt (mit Generieren) wurden oder ob nichts von beidem zutrifft. Auf diese Weise lässt sich differenzierter die Quellendiskrimination prüfen, d.h. anhand der separaten Fragen kann jeweils die externe und die interne Quellenzuordnung geprüft werden, während im vorherigen Experiment nur Aussagen zur externen Quellenzuordnung möglich waren.

Das vorliegende Experiment unterscheidet sich von den bisherigen in der Wahl der Altersgruppen. Anstelle von Vorschülern sollten Erstklässler die jüngste Altersgruppe sein, da das Schuljahr zum Erhebungszeitpunkt erst am Anfang stand und die Erstklässler somit vor kurzem eingeschult worden waren. Demnach waren Erstklässler nur einige Monate älter als die Vorschüler der vorherigen Experimente. Um einen gleichen Altersabstand zur nächst älteren Gruppe zu haben, wurden anstelle von Drittklässlern nun Viertklässler gewählt.

Der Grob Ablauf des vorliegenden Experiments stellte sich folgendermaßen dar: Nach der ersten Lernphase, in der die Listen vorgelesen wurden (ohne Generieren), erfolgte der erste Wiedererkennenstest, in dem zu jedem Item die Fragen nacheinander gestellt wurden „Wurde das Wort vorgelesen?“ und „Hast du an das Wort während des Vorlesens der Listen gedacht?“. Nach einer Ablenkaufgabe wurde das zweite Listenset vorgelesen (mit Generieren) und es sollte nach jedem vorgelesenen Item ein Wort generiert werden. Abschließend wurde der letzte Wiedererkennenstest bearbeitet, in dem zu jedem Item wieder die beiden vorherigen Fragen gestellt wurden. Als Kontrollfaktor wurde die Abfolge der beiden Fragen variiert (Frage nach der externen oder internen Quelle), um zu prüfen, ob diese einen Einfluss auf die Quellenzuweisung hat, was Ackil und Zaragoza (1995, 1998) jedoch nicht prüften.

Es gab eine feste Abfolge des Generierens: Dieses wurde stets in der zweiten Testhälfte verlangt um auszuschließen, dass die Kinder auch ohne Anweisung im nächsten Listenset weiter generieren.

Unter Beachtung der theoretischen Annahmen sowie der Forschungsergebnisse vergleichbarer Studien werden folgende Hypothesen formuliert:

- (1) Es werden keine gerichteten Annahmen zur Auswirkung der Generierungsinstruktion auf die Aktivierung kritischer Items in den verschiedenen Altersgruppen aufgestellt. Es ist unklar, ob die Aktivierung bei Erwachsenen stärker ist oder ob eine altersinvariante Aktivierung erreicht werden kann.
- (2) Da Erwachsene auf bessere Fähigkeiten zur Quellendiskrimination zurückgreifen können als Viertklässler und diese auf elaboriertere Fähigkeiten als Erstklässler, sollte sich bei der sehr schwierigen Quellenunterscheidungsaufgabe kritischer Items ein Alterseffekt zeigen, wobei sich die Formulierung der Hypothesen an den Befunden aus der Untersuchung von Ackil und Zaragoza (1995, 1998) orientieren.
 - 2.1 Erstklässler neigen eher zur Fehlattribution kritischer Items als Viertklässler und diese wiederum als Erwachsene.
 - 2.2 Bei der internen Quellenzuweisung sollte sich kein Altersunterschied zeigen.

2.3 Erstklässler ordnen kritische Items häufiger beiden Quellen zu als die älteren Probanden.

9.1.1 Methode

Versuchsplan. Das Experiment basierte, von Kontrollfaktoren abgesehen, auf einem 3 (Altersgruppe: Erstklässler vs. Viertklässler vs. Erwachsene) x 2 (Abfolge der Quellenzuordnung: extern-intern vs. intern-extern) x 2 (Generieren: mit vs. ohne) Versuchsplan mit Messwiederholung auf dem letzten Faktor.

Versuchspersonen. Die Versuchspersonen setzten sich aus 20 Erstklässlern (11 Mädchen und neun Jungen), 20 Viertklässlern (11 Mädchen und neun Jungen) einer Grundschule und 20 Erwachsenen aus dem persönlichen Bekanntenkreis (12 Frauen und acht Männer) zusammen, die jedoch nicht das Ziel der Untersuchung kannten. Alle hatten eine vergleichbare Schulbildung (Abitur). Die Erstklässler waren im Mittel sieben Jahre und zwei Monate alt (zwischen 6;5 und 7;9 Jahren), die Viertklässler hatten ein durchschnittliches Alter von 10 Jahren und zwei Monaten (zwischen 9;11 und 10;11 Jahren) und die Erwachsenen 24 Jahre und sechs Monate (zwischen 18;2 und 30;2 Jahren).

Material. Die Auswahl der Listen des vorliegenden Experiments entsprach jenen der Experimente 3A und 3B, wobei nur 12 (zwei Sets à sechs Listen) der insgesamt 16 Listen verwendet wurden. Die zwei Listen eines jeden Sets mit den am schlechtesten identifizierbaren kritischen Items wurden herausgenommen (Set A: *Sommer, Kerze*; Set B: *Löwe, Hemd*). Zu Übungszwecken wurde eine Zusatzliste als siebte Liste mit sieben Items hinzugefügt. Anhand dieser sollen die Instruktionen des Wiedererkennentests mit der Unterscheidung der internen (*ohne Generieren*: selbst gedacht oder *mit Generieren*: selbst gesagt) und externen Quelle (*vorgelesen*) verdeutlicht werden.

Nach der Präsentation eines jeden Listensets wurde ein Quellendiskriminationstest bearbeitet, der in der Auswahl der Items dem Wiedererkennens-/Inklusionstest der Experimente 3A und 3B entsprach.

Wurde in der ersten Untersuchungshälfte Listenset A dargeboten mit der Fragenabfolge extern-intern, so wurde in der zweiten Untersuchungshälfte Listenset B präsentiert mit derselben Fragenabfolge. Die Variation erfolgte so, dass die Reihenfolge der Listensetdarbietung umgedreht wurde (B-A), ebenso die Fragenabfolge (interne-externe Quelle). Das verwendete Material kann Anhang D1 entnommen werden.

Durchführung. Die Dauer des vorliegenden Experiments belief sich auf ca. 30 Minuten, wobei die Schnelligkeit des Generierens der Probanden in der zweiten Testhälfte für die Testdauer ausschlaggebend war. Die Versuchsdurchführung gestaltete sich in sieben aufeinander folgenden Phasen [Lernphase (1), Ablenkaufgabe (1), Testphase (1), Ablenkaufgabe (2), Lernphase (2), Ablenkaufgabe (3), Testphase (2)].

In der ersten Lernphase wurde eine Instruktion zum Wortspiel bzw. Gedächtnisexperiment gegeben, die mit den vorherigen Experimenten übereinstimmte. Die erste Ablenkaufgabe hatte das Ziel die Probanden mit den Antwortmöglichkeiten des folgenden Quellendiskriminationstests vertraut zu machen. Das Vorgehen richtete sich nach Ackil und Zaragoza (1995, 1998), die auch zur Einführung des Quellentests Beispielaufgaben verwendeten (s. o.). Beispielsweise wurden die Kategorien „Obst“ und „rote Farbe“ gewählt. Dann wurden konkrete Gegenstände gezeigt (Apfel, Orange, roter Spielzeugteufel, Teddybär), die beiden Kategorien zugeordnet werden sollten. Es wurden nacheinander die Fragen gestellt: „Ist XY ein Obst?“ und „Ist XY rot?“. Somit resultierten die Antwortmöglichkeiten „ja/ja“, „ja/nein“, „nein/ja“ und „nein/nein“. Eben diese Antwortmöglichkeiten sollten auch im folgenden Quellendiskriminationstest gegeben werden.

Im Quellendiskriminationstest wurde die übliche Instruktion erteilt und die Probanden sollten anhand zweier aufeinander folgenden Fragen zu allen Items diese danach beurteilen, ob sie in der Lernphase vorgelesen wurden und/oder ob sie an die Items selbst gedacht hatten. Zu diesem Zweck wurde die letzte Liste der Lernphase, also die Übungsliste, vorgelesen. Zu vier Items (ein Listenitem, ein nicht assoziiertes Item, ein präsentiertes kritisches Item und ein nicht präsentiertes kritisches Item) wurde die Frage gestellt, ob sie vorgelesen wurden und/oder ob an sie gedacht wurde. Die Items wurden so ausgewählt, dass alle vier verschiedenen Antwortmöglichkeiten gegeben waren.

Nach dieser ersten Testphase erfolgte die zweite Ablenkaufgabe. Hier sollten Aufgaben der Standardform des Progressiven Matrizen-Tests (Raven, 1976) für fünf Minuten bearbeitet werden.

Anschließend wurde die zweite Lernphase vorgestellt, in welcher die Listen Item für Item vorgelesen wurden und zu jedem Item ein Wort generiert werden sollte. Dies wurde anhand zweier Beispielwörter verdeutlicht (für die beiden jüngeren Gruppen: „Lehrer-Schule“; „schreiben-rechnen“; für Erwachsene: „Universität-Student“; „lesen-schreiben“). Zudem wurde darauf hingewiesen, dass dieselben Wörter mehrmals in einer Liste genannt werden könnten und auch dass bereits vorgelesene Wörter aufgeführt werden könnten, sofern dies den

eigenen Assoziationen entsprechen sollte. Anhand einer beispielhaften Wortliste zu der die Probanden generieren sollten, wurde das Verständnis getestet.

Bevor der Quellendiskriminationstest erfolgte wurde eine letzte Ablenkaufgabe durchgeführt, welcher der ersten Ablenkaufgabe entsprach, nur mit anderen Kategorien und Beispielwörtern.

Die Instruktion zum abschließenden Quellendiskriminationstest stimmte mit der vorherigen überein. Die Items der letzten Wortliste der Lernphase (Übungsliste) wurden vorgelesen und die generierten Items notiert. Wie auch im Fall ohne Generieren wurden nun zu vier verschiedenen Itemtypen zwei Fragen gestellt: „Habe ich XY vorgelesen?“ und „Hast du XY gesagt?“, wobei die Fragenabfolge der Quelle (extern-intern oder intern-extern) in Abhängigkeit des ersten Tests beibehalten wurde. Wenn das Verständnis gesichert war, begann die Bearbeitung des Quellendiskriminationstests. Die Instruktionen im Wortlaut befinden sich im Anhang D2.

9.1.2 Ergebnisse

Für das Generieren von Listenitems und kritischen Items wurde eine 3 (Altersgruppe: Erstklässler vs. Viertklässler vs. Erwachsene) x 2 (Listenset A vs. Listenset B) VA ohne Messwiederholung gerechnet. Für das Wiedererkennen und die Quellendiskrimination von Listenitems und kritischer Items sowie für die Aktivierung kritischer Items werden die Ergebnisse der 3 (Altersgruppe: Erstklässler vs. Viertklässler vs. Erwachsene) x 2 (Abfolge der Quellenzuordnung: extern-intern vs. intern-extern) x 2 (Generieren: mit vs. ohne) VA mit Messwiederholung auf dem letzten Faktor dargestellt.

Als Maß für falsches Wiedererkennen und Aktivierung wie auch für korrektes Wiedererkennen der Listenitems wurden wie auch in den vorherigen Experimenten korrigierte Werte berechnet.

9.1.2.1 Listenitems

Generierung. Zu jedem Item sollte ein Wort generiert werden. Es soll geprüft werden, ob sich die Altersgruppen in der Generierung von Listenitems unterscheiden. In der zugehörigen 3 (Altersgruppe) x 2 (Listenset) VA gab es nur einen Effekt der Altersgruppe [$F(2,54) = 5.19$, $R^2 = 0.16$], nach welchem sich Erstklässler ($M = .12$) nicht von Viertklässlern ($M = .09$) unterscheiden, nur Erwachsene generierten weniger Listenitems ($M = .07$). Viertklässler unter-

schieden sich weder von Erstklässlern noch von Erwachsenen. Alle anderen Werte waren statistisch unbedeutsam (alle $F_s < 1.99$).

Wiedererkennen. Die Ergebnisse der Analyse der Trefferrate, deren Mittelwerte aus Tabelle 9.1.1 hervorgehen, stimmen mit denen der Pr-Werte überein, weshalb nur letztere berichtet werden.

Tabelle 9.1.1:

Mittelwerte der Trefferrate sowie der korrigierten Wiedererkennensleistung für Listenitems in Abhängigkeit des Alters und des Generierens.

	Treffer		Pr-Werte	
	ohne Generieren	mit Generieren	ohne Generieren	mit Generieren
Erstklässler	.79	.87	.71	.77
Viertklässler	.70	.95	.62	.93
Erwachsene	.85	.89	.81	.85

In der 3 (Altersgruppe) x 2 (Abfolge) x 2 (Generieren) VA für die Pr-Werte erlangte die Altersgruppe [$F(2,54) = 4.02$, $R^2 = 0.13$] und das Generieren Signifikanz [$F(1,54) = 30.32$, $R^2 = 0.36$], welche miteinander interagierten [$F(2,54) = 12.50$, $R^2 = 0.32$]. Die Mittelwerte sind Tabelle 9.1.1 zu entnehmen. Die Altersgruppen unterschieden sich in der Bedingung ohne Generieren [$F(2,54) = 9.02$]: Erstklässler unterschieden sich von Viertklässlern und auch von Erwachsenen, Viertklässler hatten geringere Werte als Erwachsene. Mit Generieren zeigte sich auch ein Altersunterschied [$F(2,54) = 6.96$], nach welchem sich die beiden älteren Gruppen nicht unterschieden, aber Viertklässler hatten eine höhere Wiedererkennensleistung als Erstklässler. Viertklässler profitierten deutlich von der durch die Zusatzaufgabe verlangten elaborativen Verarbeitungsstrategie, denn sie hatten mit Generieren höhere Pr-Werte als ohne Generieren [$F(1,54) = 25.70$]. Bei Erstklässlern führte das Generieren nicht zu höheren Pr-Werten, möglicherweise dadurch, dass sie diese Strategie noch nicht ausreichend nutzen konnten. Erwachsene profitierten nicht, da sie ohne explizite Anweisung effektive Lernstrategien einsetzen können (beide $F_s < 1$). Die anderen Werte der VA waren nicht signifikant (alle $F_s < 1.79$).

Quellendiskrimination. Die Ergebnisse der Quellenattribution von Listenitems werden im einzelnen berichtet, unterteilt nach externer, interner Quellenattribution und Aufschlüsselung der Antworten.

Die externe Quellenattribution der Listenitems (Anzahl der „ja“-Antworten auf die Frage, ob die Listenitems vorgelesen wurden) war abhängig vom Alter [$F(2,57) = 4.39$, $R^2 = 0.13$] und

tendenziell vom Generieren [$F(1,57) = 3.34, R^2 = 0.06, p = .07$], welche wiederum tendenziell miteinander interagierten [$F(2,57) = 3.01, R^2 = 0.10, p = .06$].

Tabelle 9.1.2:

Mittelwerte der korrekten (externen) Quellenattribution von Listenitems in Abhängigkeit des Alters und des Generierens.

	ohne Generieren	mit Generieren
Erstklässler	.94	.95
Viertklässler	.93	.99
Erwachsene	.99	.99

Die Analyse der einfachen Haupteffekte zeigte, dass sich die Altersgruppen ohne Generieren ($F < 1$) nicht unterschieden, aber mit Generieren [$F(2,57) = 5.43$] attribuierten Viertklässler und Erwachsene Listenitems etwas häufiger auf die korrekte Quelle als Erstklässler (Mittelwerte s. Tabelle 9.1.2). Aufgrund der sehr hohen Werte ist von einem Deckeneffekt auszugehen. Innerhalb der Altersgruppen ergab sich kein bedeutsamer Unterschied mit und ohne Generieren (alle F s < 1).

Wurden Listenitems mindestens einmal der internen Quelle zugeordnet („ja“-Antwort auf die Frage, ob an die Listenitems gedacht bzw. ob diese selbst genannt wurden), was nicht unbedingt eine falsche Quellenzuordnung war, weil die Probanden auch an die Listenitems gedacht haben bzw. diese selbst genannt haben könnten, dann führte die Altersgruppe zu bedeutsamen Unterschieden [$F(2,57) = 17.15, R^2 = 0.38$]. Mit zunehmendem Alter reduzierte sich die interne Quellenattribution (Erstklässler: $M = .28$ vs. Viertklässler: $M = .18$ vs. Erwachsene: $M = .07$). Alle anderen Werte waren statistisch nicht bedeutsam ($F < 1.38$).

Zur Aufschlüsselung der Antworten auf die externe und interne Quelle werden die einzelnen Werte in Abhängigkeit des Generierens in Tabelle 9.1.3 abgebildet.

Die Werte der ausschließlich internen und damit falschen Quellenattribution deuten auf Bodeneffekte hin.

Eine ausschließlich externe und damit korrekte Quellenattribution zeigte Alterseffekte [$F(2,57) = 14.67, R^2 = 0.34$]: Erwachsene ($M = .82$) wiesen bessere Quellendiskriminationsfähigkeiten auf als die beiden jüngeren Altersgruppen (Erstklässler: $M = .62$ vs. Viertklässler: $M = .70$). Der Generierungseffekt [$F(1,57) = 16.10, R^2 = 0.22$] verdeutlichte, dass das Generieren ($M = .77$) die Quellenattribution verbesserte (ohne Generieren: $M = .66$). Die Interaktion verfehlte knapp das Signifikanzniveau [$F(2,57) = 2.39$].

Tabelle 9.1.3:

Verteilung der „ja“- und „nein“-Antworten in Mittelwerten auf die beiden Quellen (extern, intern) für die Listenitems in Abhängigkeit des Alters und des Generierens (* = korrekte Antwort).

	extern	intern	extern	intern	extern	intern
	nein	ja	ja	nein*	ja	ja
	ohne	mit	ohne	mit	ohne	mit
Erstklässler	.05	.05	.59	.65	.20	.22
Vierklässler	.06	.01	.60	.80	.10	.15
Erwachsene	.01	.01	.78	.86	.07	.03

Anmerkung: ohne = ohne Generieren; mit = mit Generieren

Wurden Listenitems sowohl der internen als auch der externen Quelle zugeordnet, dann war die Altersgruppe bedeutsam [$F(2,57) = 13.4$, $R^2 = 0.32$]. Mit zunehmendem Alter reduzierte sich entsprechend der Anteil der doppelten Quellenattribution (Erstklässler: $M = .21$; Viertklässler: $M = .13$; Erwachsene: $M = .05$).

9.1.2.2 Kritische Items

Generierung. In der 3 (Altersgruppe) x 2 (Listenset) VA konnte ein Alterseffekt in der Generierung kritischer Items festgestellt werden [$F(2,54) = 4.27$, $R^2 = 0.14$], nach welchem Erst- ($M = .87$) und Viertklässler ($M = .86$) sich nicht unterschieden, aber Erwachsene ($M = .97$) generierten häufiger mindestens einmal das kritische Item zu den Listenitems, wie in Experiment 3B, nur dass hier die Altersunterschiede geringer ausfielen. Insgesamt wurden die kritischen Items aber häufiger genannt als in Experiment 3B, so dass eine getrennte Auswertung nach generierten und nicht generierten kritischen Items nicht möglich war. Die anderen Werte der VA waren nicht bedeutsam (alle $F_s < 1$).

Aktivierung. Als Aktivierungsmaß für die kritischen Items diente die Proportion der überhaupt als „alt“ erkannten Items, also der Items, die bei den beiden Fragen nach den Quellen mindestens einmal mit „ja“ beantwortet worden waren. Die Mittelwerte der unkorrigierten und korrigierten Rate aktivierter kritischer Items sind in Tabelle 9.1.4 für die verschiedenen Altersgruppen zusammengefasst.

Tabelle 9.1.4:

Mittelwerte der unkorrigierten und korrigierten Aktivierung kritischer Items in Abhängigkeit des Alters und des Generierens.

	unkorrigierte Aktivierung		korrigierte Aktivierung	
	ohne Generieren	mit Generieren	ohne Generieren	mit Generieren
Erstklässler	.88	.85	.58	.67
Viertklässler	.87	.88	.61	.80
Erwachsene	.91	.93	.67	.85

In der dreifaktoriellen VA mit den unkorrigierten Werten der aktivierten kritischen Items ergab sich ein tendenzieller Alterseffekt [$F(2,54) = 2.66$, $R^2 = 0.09$, $p = .08$], nach welchem sich Erst- ($M = .86$) und Viertklässler ($M = .87$) nicht unterschieden, aber Erwachsene ($M = .92$) tendenziell eine stärkere Aktivierung hatten. Auch führte die Analyse zu einem tendenziellen Abfolgeeffekt [$F(1,54) = 3.47$, $R^2 = 0.07$, $p = .06$], der mit der Generierungsbedingung interagierte [$F(1,54) = 5.12$, $R^2 = 0.09$]. Die Analyse der einfachen Haupteffekte zeigte in der Bedingung ohne Generieren [$F(1,54) = 7.57$], dass eine höhere Aktivierung resultierte, wenn zuerst nach der externen Quelle ($M = .93$) gefragt wurde als nach der internen ($M = .83$).

Die Ergebnisse der unkorrigierten Aktivierung für die kritischen Items weichen etwas von jenen der korrigierten Aktivierung ab, in welcher die Aktivierung der verbundenen und unverbundenen Distraktoren von der Aktivierung kritischer Items subtrahiert wurde. In der entsprechenden VA mit den korrigierten Werten hatte das Generieren einen Effekt [$F(1,57) = 22.22$, $R^2 = 0.28$]: Die Aktivierung mit Generieren war höher ($M = .77$) als ohne Generieren ($M = .62$). Auch die Altersgruppe war bedeutsam [$F(2,57) = 3.77$, $R^2 = 0.12$]: Wie in Experiment 3B hatten Erwachsene ($M = .76$) eine stärkere Aktivierung als Erstklässler ($M = .63$) und unterschieden sich nicht von Viertklässlern ($M = .70$). Auch in diesem Experiment konnte keine altersinvariante Aktivierung kritischer Items durch das Generieren erreicht werden (Hypothese 1). Die alterskorrelierten Unterschiede fielen insgesamt gering aus wie in Experiment 3A und 3B. Die Interaktion von Alter und Generieren wurde auch hier nicht signifikant [$F(2,57) = 1.15$].

Quellendiskrimination. Zur Ermittlung der Fehlattribution wurde gerechnet, wie viele der als „alt“ erkannten kritischen Items der falschen, d.h. der externen Quelle, zugeordnet wurden. In der VA (Mittelwerte s. Tabelle 9.1.5) zeigte sich ein Alterseffekt [$F(2,57) = 53.11$, $R^2 = 0.65$].

Tabelle 9.1.5:

Mittelwerte der Fehlattribution kritischer Items in Abhängigkeit des Alters und des Generierens.

	ohne Generieren	mit Generieren
Erstklässler	.78	.85
Viertklässler	.37	.50
Erwachsene	.28	.24

Erwartungsgemäß (Hypothese 2.1) hatten Erstklässler ($M = .82$) eine höhere Fehlattribution kritischer Items als Viertklässler ($M = .44$) und diese wiederum als Erwachsene ($M = .26$). Alle anderen Werte der VA erlangten keine Signifikanz (alle F s < 2.14). Dementsprechend hatte das Generieren keine Effekte auf die Fehlattribution in den Altersgruppen (Mittelwerte s. Tabelle 9.1.5).

Bei der korrekten Quellendiskrimination der aktivierten kritischen Items (Zuordnung zur internen Quelle) ergaben sich anders als bei Ackil und Zaragoza (1995, 1998) und entgegen der Annahme (Hypothese 2.2) Altersunterschiede [$F(2,54) = 13.45$, $R^2 = 0.37$]: Erwachsene ($M = .96$) waren besser als Viertklässler ($M = .85$) und Erstklässler ($M = .76$), die sich nicht unterschieden. Auch Erstklässler zeigten schon eine gute korrekte Quellenattribution. Auch das Generieren [$F(1,54) = 12.77$, $R^2 = 0.19$] war bedeutsam: Mit Generieren ($M = .81$) wurden kritische Items seltener der korrekten Quelle zugewiesen als ohne Generieren ($M = .91$). Die Interaktion der beiden Faktoren verfehlte das Signifikanzniveau nur knapp [$F(2,54) = 2.39$], die Mittelwerte gehen aus Tabelle 9.1.6 hervor. Wichtig war, dass es bei Erwachsenen für die korrekte Attribution unbedeutend war, ob sie generiert hatten oder nicht. Alle anderen Werte waren nicht signifikant (alle F s < 2.03).

Tabelle 9.1.6:

Mittelwerte der korrekten (internen) Attribution von aktivierten kritischen Items in Abhängigkeit des Alters und des Generierens.

	ohne Generieren	mit Generieren
Erstklässler	.84	.68
Viertklässler	.92	.79
Erwachsene	.97	.96

Zur differenzierteren Analyse wurden die Antworten nach beiden Quellendiskriminationsfragen aufgeschlüsselt (s. Tabelle 9.1.7). Von besonderem Interesse war es, hier zu prüfen, ob Erstklässler wie bei Ackil und Zaragoza (1995, 1998) auch häufig zweimal „ja“ auf die Fra-

gen antworteten und damit wussten, dass sie die kritischen Items selbst gesagt bzw. selbst an sie gedacht hatten, aber auch der Meinung waren, dass sie vorgelesen worden waren.

Wurden kritische Items nur der internen und damit korrekten Quelle zugeordnet, dann erzielte die Altersgruppe Signifikanz [$F(2,54) = 53.90$, $R^2 = 0.67$], welche mit dem Generieren tendenziell interagiert [$F(2,54) = 2.78$, $R^2 = 0.09$, $p = .07$]. Die Mittelwerte sind Tabelle 9.1.7 zu entnehmen. Die Altersgruppen unterschieden sich in der Bedingung ohne Generieren [$F(2,54) = 78.27$] dahingehend, dass Erstklässler kritische Items seltener nur der internen Quelle zuordneten als die beiden älteren Gruppen. In der Bedingung mit Generieren erhöhte sich die interne Quellenattribution mit zunehmendem Alter [$F(2,54) = 113.41$]. Innerhalb der Altersgruppen ergab sich keine bedeutsame Differenz mit und ohne Generieren (alle $F_s < 1.20$). Des Weiteren interagiert die Abfolge mit dem Generieren [$F(1,54) = 4.56$, $R^2 = 0.08$]: Mit Generieren [$F(1,54) = 4.48$] wurden in der Abfolge intern-extern ($M = .47$) etwas höhere Werte erzielt als in der umgekehrten Abfolge ($M = .38$). Dies zeigt, dass diese Abfolgebedingung, in der zuerst nach der korrekten, internen Quelle gefragt wurde, die Quellendiskrimination erleichterte. Diese Fragenabfolge könnte eine Erinnerungshilfe für die selbstständig generierten kritischen Items darstellen. Alle anderen Werte differierten nicht bedeutsam voneinander (alle $F_s < 3.82$).

Tabelle 9.1.7:

Verteilung der „ja“- und „nein“-Antworten in Mittelwerten auf die beiden Quellen (extern, intern) für die kritischen Items in Abhängigkeit des Alters und des Generierens (* = korrekte Antwort).

	extern	intern	extern	intern	extern	intern
	nein	ja*	ja	nein	ja	ja
	ohne	mit	ohne	mit	ohne	mit
Erstklässler	.19	.13	.13	.25	.55	.47
Vierklässler	.54	.43	.08	.18	.25	.26
Erwachsene	.64	.71	.03	.04	.24	.18

Anmerkung: ohne = ohne Generieren; mit = mit Generieren

Bei einer ausschließlich externen und damit völlig falschen Attribuierung zeigten sich nur Haupteffekte. Die Mittelwerte sind nach Altersgruppen und Generieren in Tabelle 9.1.7 abgebildet. Zum einen erlangte die Altersgruppe Signifikanz [$F(2,54) = 14.70$, $R^2 = 0.35$]: Erstklässler ($M = .19$) unterschieden sich nicht von Viertklässlern ($M = .13$), aber beide Gruppen attribuierten kritische Items häufiger ausschließlich extern als Erwachsene ($M = .03$). Auch die Abfolge war bedeutsam [$F(1,54) = 4.83$, $R^2 = 0.08$]: Wurde zuerst nach der externen Quelle gefragt ($M = .14$), so resultierten höhere Werte als bei der umgekehrten Abfolge ($M = .09$).

Ein dritter Haupteffekt resultierte für die Generierungsbedingung [$F(1,54) = 9.79, R^2 = 0.15$]: Mit Generieren ($M = .16$) wurden mehr kritische Items nur der externen Quelle zugeordnet als ohne Generieren ($M = .08$). Die Interaktionen erlangten keine Signifikanz (alle $F_s < 2.53$).

Die Mittelwerte der verschiedenen Altersgruppen in Abhängigkeit des Generierens für die doppelte Quellenattribution sind in Tabelle 9.1.7 abgebildet. Wurden kritische Items sowohl der internen als auch der externen Quelle zugeordnet, dann erlangte die Altersgruppe Signifikanz [$F(2,57) = 16.72, R^2 = 0.37$]. Erwartungskonform (Hypothese 2.3) neigten Erstklässler ($M = .51$) stärker zur doppelten Quellenattribution als Viertklässler ($M = .25$) und Erwachsene ($M = .21$), die sich nicht unterschieden. Alle anderen Werte waren nicht signifikant (alle $F_s < 1.52$).

Da die alterskorrelierten Unterschiede in der Aktivierung eher gering ausfielen, die in der Quellendiskrimination aber deutlich, sollte sich eine alterskorrelierte Abnahme im falschen Wiedererkennen der kritischen Items zeigen. Im Folgenden werden nur die Ergebnisse der korrigierten falschen Wiedererkennungsraten dargestellt. Die Mittelwerte der VA mit der „Trefferrate“ der falsch wiedererkannten kritischen Items sind in Tabelle 9.1.8 abgebildet.

Tabelle 9.1.8:

Mittlere „Trefferrate“ der falsch wiedererkannten kritischen Items in Abhängigkeit des Alters und des Generierens.

	ohne Generieren	mit Generieren
Erstklässler	.68	.72
Viertklässler	.33	.44
Erwachsene	.27	.23

Die Berechnung der Pr-Werte in der VA führte sowohl zu einem Haupteffekt der Altersgruppe [$F(2,57) = 27.64, R^2 = 0.49$] als auch zu einem Interaktionseffekt mit dem Generieren [$F(2,57) = 3.43, R^2 = 0.11$]. Abbildung 9.1.1 stellt die Interaktion dar.

Die Altersgruppen unterschieden sich in den Bedingungen ohne [$F(2,57) = 17.75$] und mit Generieren [$F(2,57) = 20.42$]. Ohne Generieren differierten die beiden älteren Gruppen nicht, hatten aber weniger falsche Erinnerungen als Erstklässler. Sollte generiert werden, dann zeigte sich eine alterskorrelierte Abnahme, denn Erstklässler waren sensitiver für falsche Erinnerungen als Viertklässler und diese wiederum als Erwachsene (s. Abbildung 9.1.1).

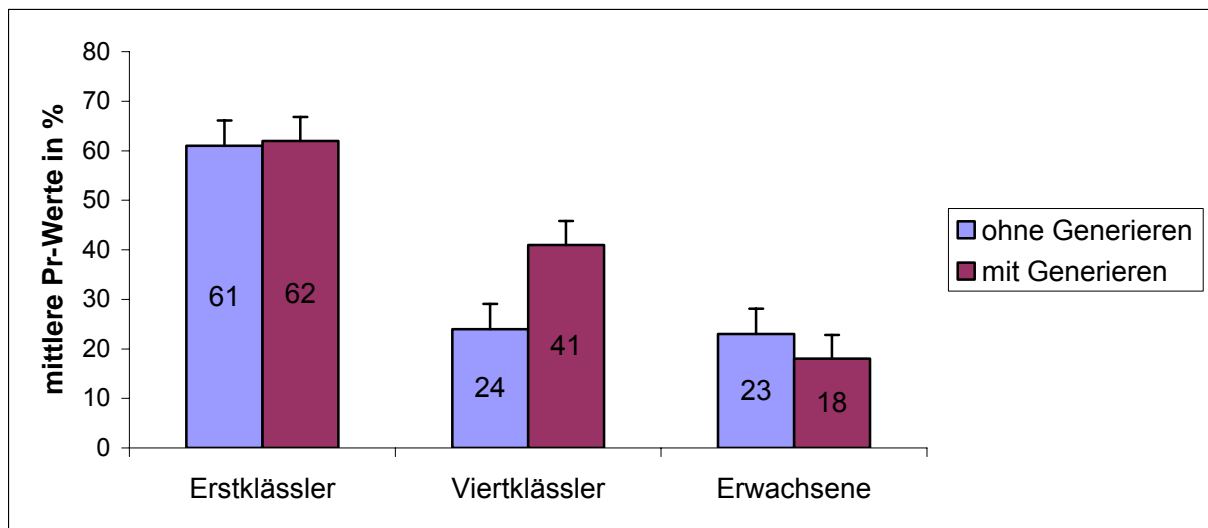


Abbildung 9.1.1: Mittlere Pr-Werte (in Prozent) falsch wieder erkannter kritischer Items in Abhängigkeit des Generierens und des Alters.

Der Vergleich des Generierens innerhalb der Altersgruppen zeigte, dass nur Viertklässler mit Generieren mehr kritische Items falsch wiedererkannten als ohne Generieren [$F(1,57) = 8.20$]. Bei Erstklässlern und Erwachsenen waren die Pr-Werte unbeeinflusst vom Generieren (beide $F_s < 1$). Eine getrennte Auswertung nach generierten und nicht generierten kritischen Items war, wie bereits erwähnt, durch die geringe Anzahl nicht generierter kritischer Items nicht möglich. Somit wurden die fälschlicherweise als präsentiert erkannten kritischen Items in der Lernphase auch in der Regel mindestens einmal generiert.

9.1.3 Diskussion

Zielsetzung des vorliegenden Experiments war es, die beiden vorangegangenen Experimente zu replizieren, wobei andere Aktivierungs- und Quellendiskriminationsmaße gewählt wurden. In Anlehnung an die Befunde zur Generierungsinstruktion von Experiment 3B, welche durch den Vergleich mit Experiment 3A eine Erhöhung der Aktivierung der kritischen Items durch die Förderung von semantisch-relationalen Verarbeitungsstrategien bei Vorschülern zeigte, aber nicht zu einer altersinvarianten Aktivierung gelangte, sollte im vorliegenden Experiment eine andere Generierungsinstruktion entwickelt und eingesetzt werden, die sowohl die semantisch-relationale (es sollten häufig die kritischen Items assoziiert werden) als auch die item-spezifische (zu jedem Item wurde assoziiert) Verarbeitung fördern sollte. Damit sollte geprüft werden, ob Erwachsene tatsächlich stets eine höhere Aktivierung haben, bedingt durch ein komplexeres Wissensnetzwerk und elaboriertere Verarbeitungsstrategien. Die Untersuchung von Marsh und Bower (2004) wurde als Grundlage gewählt, um ausgehend von dieser einen

Vergleich mit den eigenen Ergebnissen zu ermöglichen. Während die Probanden bei Marsh und Bower (2004) zu jedem Item einer Wortliste zwei Items generieren sollten, wurde im eigenen Experiment die Instruktion erteilt nur ein Item zu jedem Listenitem zu nennen. Darüber hinaus sollte im vorliegenden Experiment die Quellenattribution differenzierter erfasst werden, wobei die Vorgehensweise im Falschinformationsparadigma von Ackil und Zaragoza (1995, 1998) als Vorlage diente. Anstelle des „alt/neu“-Urteils sollten die Probanden die jeweiligen Items hinsichtlich der externen/internen Quelle beurteilen. Konkret bedeutete dies, dass zwei aufeinander folgende Fragen gestellt wurden, (1) ob die Items vorgelesen wurden und/oder (2) ob an diese selbst gedacht (ohne Generieren) bzw. ob diese gesagt (mit Generieren) wurden oder ob nichts von beidem zutreffend war. Die erste Frage stellte die externe Quelle dar und die zweite die interne, die in der Formulierung abhängig vom Generieren war. Hier ging es darum einen Vergleich mit dem Falschinformationsparadigma zu ermöglichen, nach dem deutliche Unterschiede in der Quellendiskriminationsleistung zwischen allen Altersgruppen resultierten.

Ähnlich wie in Experiment 3 profitierten die Kinder beim korrekten Wiedererkennen der Listenitems von der elaborativen Verarbeitungsstrategie, allerdings in erheblichem Maße nur die Viertklässler. Die Erstklässler profitierten nur tendenziell, möglicherweise bedingt durch ein Nutzungsdefizit, und die Erwachsenen, die selbstständig effektive Lernstrategien einsetzen sollten, wie in Experiment 3 kaum.

Um die Experimente 3 und 4 besser miteinander vergleichen zu können, werden sie anhand ihrer Ergebnisse zur Aktivierung, der Fehlattribution und dem falschen Wiedererkennen in Tabelle 9.1.9 zusammengefasst, auf die im Folgenden öfter verwiesen wird. Für Experiment 3 wurden nur die Werte der separaten Analyse mit den in der ersten Phase erteilten Tests ausgewählt, da es zahlreiche Testabfolgeeffekte gab und daher nur der erste Test interpretiert werden kann.

Tabelle 9.1.9:

Vergleich der Experimente 3 und 4 in den Mittelwerten der Aktivierung (Pr-Werte), der Fehl-attribution und des falschen Wiedererkennens (Pr-Werte) in Abhängigkeit des Alters und des Generierens.

		Aktivierung (Pr)		Fehl-attribution		Falsches Wiederer- kennen (Pr)	
		Generieren		Generieren		Generieren	
		ohne	mit	ohne	mit	ohne	mit
Experiment 3	Vorschüler	.45	.67	.89	.86	.42	.57
	Drittklässler	.56	.74	.87	.72	.50	.67
	Erwachsene	.64	.87	.67	.30	.58	.21
Experiment 4	Erstklässler	.58	.67	.78	.85	.61	.62
	Viertklässler	.61	.80	.37	.50	.24	.41
	Erwachsene	.67	.85	.28	.24	.23	.18

Die Ergebnisse der vorliegenden Untersuchung zur Aktivierung zeigten, dass in beiden Experimenten die Aktivierung durch die Generierungsinstruktionen gefördert werden konnte, denn die Aktivierung mit Generieren war höher als ohne Generieren. Jedoch konnte in Experiment 4 wie in Experiment 3B keine altersinvariante Aktivierung der kritischen Items durch das Generieren erreicht werden. Die zugehörigen Mittelwerte sind Tabelle 9.1.9 zu entnehmen. Die Generierungsinstruktion förderte wie in Experiment 3B die Aktivierung, allerdings entgegen den Erwartungen am wenigsten für die jüngste Gruppe, während deutliche Steigerungen bei Viertklässlern und Erwachsenen zu verzeichnen waren. Erwachsene hatten eine stärkere Aktivierung als Erstklässler, aber Viertklässler unterschieden sich nicht von Erwachsenen und nicht von Erstklässlern. Erwachsene verfügen über mehr Wissen, wodurch ihre automatische Aktivierung gestärkt werden kann und auch über elaboriertere Verarbeitungsstrategien, die ihre bewusste Aktivierung fördern können.

Die Befunde zum Quellendiskriminationstest konnten bestätigen, dass kritische Items mit zunehmendem Alter seltener fehl-attribuiert wurden, was in Einklang mit den Ergebnissen der Untersuchungen von Ackil und Zaragoza (1995, 1998) ist, da die Fähigkeiten zur Quellenattribution sich mit dem Alter verbessern sollten. Erwachsene hatten in Experiment 4 sehr wenige Fehl-attributionen sowohl ohne als auch mit Generieren (s. Tabelle 9.1.9). Da Erwachsene kritische Items bewusst generieren konnten, war es für sie leichter diese der korrekten internen Quelle zuzuordnen und in Folge resultierten kaum falsche Erinnerungen. Während die Ergebnisse der Experimente 3 und 4 in der Bedingung mit Generieren vergleichbar sind (alterskorrelierte Abnahme der Fehl-attribution), zeigen sich deutliche Unterschiede in der Bedingung ohne Generieren: Im dritten Experiment hatten Erwachsene weniger Fehl-attributionen als Kinder, aber im vierten Experiment hatten bereits ältere Schulkinder ohne Generie-

ren weniger Fehlattritionen als die jüngste Gruppe. Darüber hinaus unterschieden sich die älteren Schulkinder nicht von Erwachsenen, die beide in Experiment 4 ohne Generieren geringere Fehlattritionen als im dritten Experiment hatten.

Die Ergebnisse weisen darauf hin, dass ein Quellendiskriminationstest, der deutlich macht, dass die Vertrautheit eines Items nicht nur auf die tatsächliche Präsentation, sondern auch auf die selbständige Generierung (gedacht oder selbst genannt) zurückgehen kann, bei älteren Grundschulkindern und Erwachsenen zu deutlich weniger Fehlattritionen und damit auch zu weniger falschen Erinnerungen führt als ein Wiedererkennenstest oder auch ein nachträglicher Quellendiskriminationstest. Der Quellendiskriminationstest könnte als Vorwarnung für die zweite Lernphase fungiert haben, so dass spezielle Strategien zur Identifikation des kritischen Items eingesetzt wurden. Ausgehend von dieser Annahme stimmen die vorliegenden Ergebnisse mit der generellen Befundlage nicht überein nach der eine Vorwarnung nur zur Reduktion von falschen Erinnerungen führt, wenn sie vor der Lernphase stattfindet (z.B. Gallo et al., 1997; Gallo et al., 2001b; Multhaup & Connor, 2002). Diese Ergebnisse sind stimmig mit der Studie von Multhaup und Connor (2002), die ebenfalls einen, wenn auch etwas anderen Quellendiskriminationstest verwendeten, und eine Verringerung von falschen Erinnerungen beobachten konnten. Allerdings erfolgte der Test bei Multhaup und Connor nach jeder Liste, so dass ab der zweiten Liste von einer Vorwarnung vor dem Lernen auszugehen ist.

Die jüngste Gruppe hatte in Experiment 3 und 4 vergleichbar viele Fehlattritionen (Mittelwerte Tabelle 9.1.9), was mit ihren schlechteren Quellendiskriminationsfähigkeiten zusammenhängt. Es ist unklar und sollte geprüft werden, ob diese Defizite in der Lernphase liegen, in welcher die für eine korrekte Quellenattribuierung relevanten Merkmale nicht enkodiert wurden und/oder in der Testphase zu lokalisieren sind, in der das Metagedächtnis und Entscheidungsstrategien gefordert sind (Jacoby et al., 2005).

In Experiment 3 gab es im Gegensatz zu Experiment 4 Generierungseffekte auf die Quellendiskrimination, denn Drittklässler konnten leicht und Erwachsene stark vom Generieren profitieren, um die kritischen Items nicht der falschen Quelle zuzuweisen. Die fehlenden Generierungseffekte in Experiment 4 sollten aber mit Vorsicht interpretiert werden, da diese Bedingung stets als zweite erfolgte und nicht ausgeschlossen werden kann, dass insbesondere die beiden älteren Gruppen durch den ersten Quellendiskriminationstest vor der zweiten Lernphase vorgewarnt waren. Um dies auszuschließen könnten in einer weiteren Untersuchung erst die Listen präsentiert werden, die ersten ohne und die zweiten mit Generieren und daran anschließend könnte der Test erfolgen.

Die interne und damit korrekte Quellenattribution führte entgegen der Erwartungen und bisherigen Befunde zu Altersunterschieden. Erwachsene waren besser als Viertklässler, die sich nicht von Erstklässlern unterschieden. Erstaunlicherweise hatten Erstklässler eine recht hohe korrekte Quellenattribution. Wie bei Ackil und Zaragoza (1995, 1998) zeigte die jüngste Gruppe ein hohes Ausmaß an Fehlattri-bution, aber demgegenüber auch eine gute Attribution auf die korrekte Quelle.

Erwartungskonform ordneten Erstklässler im vorliegenden Experiment kritische Items häufiger gleichzeitig der internen und externen Quelle zu als die beiden älteren Gruppen. Dies entspricht auch den Befunden von Ackil und Zaragoza (1995, 1998). Erstklässler wussten, dass sie die kritischen Items selbst generiert hatten, waren aber zugleich der Meinung, dass sie auch vorgelesen worden waren. Das Ergebnis lässt sich nicht auf einen generellen Bias, zweimal „ja“ zu antworten, zurückführen, denn sie ordneten nicht in vergleichbarem Maß die Listenitems ($M = .21$) und verbundenen Distraktoren ($M = .04$) beiden Quellen gleichzeitig zu (kritische Items: $M = .51$). Die doppelten „ja“-Antworten sind mit der Auffassung kompatibel, dass falsche Erinnerungen bei Kindern auf bewusster Aktivierung basieren (vgl. Diskussion Experiment 3B), da die kritischen Items auch der internen Quelle zugeordnet werden.

Alle diese Befunde zeigen, dass die Fähigkeiten zur Quellendiskrimination sich mit dem Alter verbessern, bedingt durch die noch reifende Frontallappenfunktion bei Kindern. Die notwendigen Strategien zur Unterscheidung der Quellen sind noch nicht ausreichend vorhanden (Foley & Johnson, 1985). In diesem Experiment scheinen Viertklässler bessere Quellendiskriminationsleistungen als Erstklässler zu haben, nicht aber in Experiment 3B, in dem die beiden jüngeren Gruppen gleichgestellt waren.

Ausgehend von den geringen alterskorrelierten Unterschieden in der Aktivierung gegenüber den deutlichen Unterschieden in der Quellendiskrimination wurde eine alterskorrelierte Abnahme falscher Erinnerungen erwartet, was die Ergebnisse auch bestätigten. Ohne Generieren hatten Erstklässler mehr falsche Erinnerungen als die älteren Gruppen, mit Generieren unterschieden sich alle Altersgruppen (Mittelwerte s. Tabelle 9.1.9), was mit der erhöhten Sensitivität der Viertklässler nach dem Generieren zusammenhängt, die sowohl auf eine höhere falsche Alarmrate für die kritischen Items als auch auf eine sehr niedrige für die Distraktoren zurückging. Anknüpfend an Marsh und Bowers (2004) Annahme, dass das Generieren falsche Erinnerungen erhöht, zeigten die Ergebnisse des vorliegenden Experiments, dass dies nur für Viertklässler eintrat.

Das Resultat der Viertklässler entspricht auch Experiment 3 (Mittelwerte s. Tabelle 9.1.9), während im Gegensatz zu Experiment 3, die jüngste Gruppe nicht mehr falsche Erinnerungen nach dem Generieren zeigte als ohne. Während die Erwachsenen in Experiment 3 von der Generierung eines Items zur gesamten Liste dahingehend profitierten, dass sie kaum sensitiv für falsche Erinnerungen waren, hatten sie in Experiment 4 generell wenige falsche Erinnerungen, unabhängig von den Lernbedingungen (Mittelwerte s. Tabelle 9.1.9).

Wie auch in Experiment 3 zeigte sich ein “more is less“ Effekt bei Viertklässlern, nicht aber bei Erstklässlern. Viertklässler profitierten vom Generieren und hatten mehr richtige, aber auch mehr falsche Erinnerungen. Zunächst mag es unlogisch erscheinen, dass Erstklässler nicht mehr falsche Erinnerungen nach dem Generieren hatten als ohne, weil sie eine mögliche Vorwarnung entweder nicht erkannt haben könnten oder aufgrund ihrer geringen Arbeitsgedächtniskapazität nicht von ihr profitiert haben könnten und/oder zu einem entsprechenden strategischen Vorgehen bei der Quellendiskrimination in der Testphase nicht fähig gewesen sein könnten. Die Aufschlüsselung der Antworten zeigte, dass Erst- und Viertklässler nach dem Generieren insgesamt mehr Items der völlig falschen Quelle zuordneten (nur extern), aber Erstklässler numerisch gesehen weniger Doppelzuordnungen hatten. Die vorgefundenen Unterschiede scheinen demzufolge mit den unterschiedlichen Quellentests der Experimente 3 und 4 zusammenzuhängen. Der vorgefundene Generierungseffekt wurde von Marsh und Bower (2004) damit erklärt, dass durch das Generieren in der Lernphase die Gedächtnisspur stärker wird und die Vertrautheit der Items steigt, so dass in der Testphase Quellendiskriminationsprobleme auftreten.

Die erwachsenen Probanden des eigenen Experiments hatten bei einer vergleichbaren Anzahl von Listenitems (Marsh und Bower: 9 Items pro Liste vs. eigenes Experiment: 10 Items pro Liste) insgesamt weniger falsche Erinnerungen als die Probanden bei Marsh und Bower.

In der Untersuchung von Marsh und Bower erhöhten sich mit Generieren deutlich falsche Erinnerungen, während die Generierungsbedingungen keinen Einfluss in der eigenen Untersuchung bei Erwachsenen hatten. Da bei Marsh und Bower zu jedem Listenitem zwei Items generiert werden sollten, erhöhte sich die Wahrscheinlichkeit, dass das kritische Item mehrmals genannt wurde. Je häufiger das kritische Item generiert wird, desto wahrscheinlicher ist es nach der Auffassung der Autoren, dass die Probanden durch die stärkere Gedächtnisspur ein Quellendiskriminationsproblem haben und der Meinung sind, dass sie das kritische Item auch gehört haben (Marsh & Bower, 2004).

Eine weitere mögliche Ursache könnte darin liegen, dass Marsh und Bower eine interindividuelle Variation des Generierens vornahmen, d.h. von ihren insgesamt 60 Versuchspersonen wurden 40 der Generierungsbedingung zugewiesen und 20 der Kontrollbedingung. Hingegen fand im eigenen Experiment eine intraindividuelle Variation statt, so dass alle Probanden eine Lernbedingung mit und eine ohne Generieren hatten. Die erste Lernphase (ohne Generieren) könnte eine Vorwarnung, i. S. von Gallo et al. (1997) und Neuschatz et al. (2003, vgl. Abschnitt 2.3) für die folgende Phase sein (s.o.). Die Vorwarnung könnte falsche Erinnerungen bei Erwachsenen konstant halten, so dass trotz stärkerer Aktivierung der kritischen Items in der Lernphase diese nicht häufiger in der Testphase falsch wiedererkannt wurden, was auch den Ergebnissen von Experiment 3B entspricht, nach denen die bewusst generierten kritischen Items besonders gut der richtigen Quelle zugeordnet werden konnten.

Die fehlenden Auswirkungen des Generierens auf falsche Erinnerungen bei erwachsenen Probanden im vorliegenden Experiment ähnelt den Befunden von Gallo und Seamon (2004) und Seamon et al. (2002a), wo im Wiedererkennenstest keine Zunahme falscher Erinnerungen in Abhängigkeit des Generierens festgestellt werden konnte, so dass automatische Prozesse zu vermuten waren. Demgegenüber fanden Marsh und Bower (2004) und Dewhurst et al. (2005) heraus, dass eine Generierungsinstruktion zu einer Zunahme falscher Erinnerungen führte. Von den Autoren wurden aber unterschiedliche Prozesse als ausschlaggebend hervorgehoben: Während Marsh und Bower vorwiegend auf Quellendiskriminationsprobleme hinweisen, verweisen Dewhurst und Mitarbeiter auf eine Verstärkung der automatischen Aktivierung bei der Enkodierung. Es ist auffällig, dass die Probanden der eigenen Untersuchung einen deutlich geringeren Anteil wieder erkannter kritischer Items hatten gegenüber den Probanden der Studie von Marsh und Bower. Somit bleibt zu prüfen, ob das Generieren zweier Items (bei Marsh und Bower) vermehrt falsche Erinnerungen verursacht haben kann. Die Art des Tests scheint entscheidend zu sein, denn sie verwendeten einen Standardwiedererkennenstest, wie auch in Experiment 3, und keinen differenzierten Quellendiskriminationstest.

Die Ergebnisse von Experiment 3 konnten nur zum Teil repliziert werden und zwar für die Aktivierung, die in beiden Experimenten trotz Förderung der semantisch-relationalen Verarbeitung nicht altersinvariant war. Bei der Quellendiskrimination zeigten die beiden älteren Gruppen in Experiment 4 deutlich bessere Leistungen im Quellendiskriminationstest und eine alterskorrelierte Abnahme in der Sensitivität für falsches Wiedererkennen der kritischen Items. Das Generieren führte nur bei Viertklässlern zu vermehrt falschen Erinnerungen, wie auch in Experiment 3, aber nicht bei Erstklässlern. Bei Erwachsenen reduzierte das Generieren nicht falsche Erinnerungen (entgegen der Ergebnisse von Experiment 3), was allem An-

schein nach testbedingt ist. Die vorgefundenen Ergebnisse zur Quellendiskrimination stimmen gut mit den berichteten Befunden von Ackil und Zaragoza (1995, 1998) überein, obwohl deren Untersuchungen ein anderes Paradigma (Falschinformationsparadigma vs. DRM-Paradigma) zugrunde lag, sie anderes Material (Videoausschnitte vs. Wortlisten) verwendeten und in ihrer ersten Studie eine extern-extern Quellenunterscheidung verlangt wurde.

Das vorliegende Experiment konnte in Übereinstimmung mit Ackil und Zaragoza (1995, 1998) zeigen, dass schon Erstklässler fähig sind, zwischen den beiden vorgegebenen Quellen zu unterscheiden. Daher sollte in weiteren Untersuchungen auf ein grobes „alt/neu“-Urteil verzichtet werden, denn der gewählte Test bietet mehr Rückschlüsse auf die Quellendiskrimination.

Um einen Gesamtüberblick über die Ergebnisse der eigenen Experimente im Hinblick auf die alterskorrelierten Veränderungen der Aktivierung, der Quellendiskrimination und den falschen Erinnerungen zu erlangen, werden diese im folgenden Kapitel zusammenfassend diskutiert.

10 Gesamtdiskussion

In den vorliegenden Experimenten wurden mögliche Veränderungen falscher Erinnerungen im Vorschul-/bzw. frühem Grundschulalter, späterem Grundschulalter und im jungen Erwachsenenalter untersucht. Es sollte geprüft werden, ob jüngere Kinder für falsche Erinnerungen empfänglicher sind als ältere und Erwachsene wie Befunde aus der Zeugenaussageforschung zeigen. Ausgehend vom AMA sollte geprüft werden inwieweit mögliche alterskorrelierte Differenzen falscher Erinnerungen auf Unterschiede in der Wahrscheinlichkeit der Aktivierung der kritischen Wörter und/oder auf Unterschiede in der Quellendiskriminationsleistung zurückgehen. Anhand unterschiedlicher Prozeduren, die in den verschiedenen Experimenten realisiert wurden, wurde das Ziel verfolgt, alterskorrelierte Differenzen in Aktivierung und Quellendiskrimination getrennt voneinander zu untersuchen.

Nachdem nun alle Experimente mit den Ergebnissen berichtet wurden, sollen diese in einen Zusammenhang gesetzt werden. Zunächst werden die statistisch-methodischen Probleme erörtert mit denen die eigene Arbeit konfrontiert war. Daran anschließend werden die eigenen Befunde zusammenfassend theoretisch eingeordnet. Abschließend wird ein Ausblick auf zukünftige Forschungsmöglichkeiten gegeben und die Relevanz des DRM-Paradigmas für Anwendungszwecke herausgestellt.

10.1 Statistisch-methodische Probleme

Teststärkeproblem. Insgesamt war die Teststärke in den eigenen Experimenten gering. Geringe Teststärken und damit hohe β -Fehler können die Interpretation verfälschen, die auf nicht signifikante Effekte schließt und zur Annahme der Nullhypothese führt.

Um höhere Teststärken (im Bereich von .90 bis .95) zu erzielen, müssten 1000 Versuchspersonen an einem Experiment teilnehmen. Dies ist unrealistisch, unökonomisch und nicht gerechtfertigt. Auch könnte eine größere Listenanzahl die Teststärke verbessern, was aber durch die Auswahl der Altersgruppen nicht möglich erscheint. Demzufolge lässt sich das Teststärkeproblem kaum beheben.

Altersinäquivalente Listen. Alle verwendeten DRM-Listen wurden sorgfältig zusammengestellt, wobei darauf geachtet wurde, dass die Items dem kindlichen Wortschatz entsprachen und dass sie sowohl aktiv produzierbar als auch passiv verstehbar waren. Die ermittelten Normen von Hasselhorn und Grube (1994) stellen den Ausgangspunkt der Listenkonstruk-

tion dar, wobei die eigenen Listen an universitätsinternen Untersuchungen an Kinderstichproben normiert wurden. Daher konnte angenommen werden, dass viele Vorbereitungen getroffen wurden, um Ungleichheiten in der Listenkonstruktion auszuschließen. Dennoch ergaben sich Altersunterschiede in der Identifizierbarkeit der kritischen Items, was an dem höheren Wissen Erwachsener zu liegen scheint. Dadurch ist eine Konstruktion altersäquivalenter Listen erschwert und auch kaum möglich

Es erscheint notwendig und sinnvoll aktuelle umfangreiche Normierungen der Listen mit verschiedenen Altersgruppen durchzuführen und die Assoziationsstärken zu berechnen, insbesondere die RAS. Die bisherigen deutschsprachigen Normen von Hasselhorn und Grube (1994) könnten veraltet sein.

10.2 Theoretische Einordnung der Befunde

Die eigene Arbeit zur Untersuchung falscher Erinnerungen orientierte sich am AMA, um ausgehend von dieser Aktivierung und Quellendiskrimination anhand separater Maße zu prüfen. Eine Herausforderung der eigenen Arbeit bestand darin anhand des ursprünglich aus der Kognitionspsychologie stammenden AMA entwicklungspsychologische Befunde zu erklären. Daher ist die andere, häufig diskutierte und populäre FTT, eher nebensächlich und soll in diesem Abschnitt nur zur ergänzenden Erklärung der Befunde erläutert werden. Die eigene Arbeit hatte nicht die Zielsetzung beide Theorien gegeneinander zu prüfen, da die Gemeinsamkeiten größer sind als die Unterschiede.

10.2.1 Aktivierungs-Monitoring Ansatz (AMA)

Bislang gibt es keine eindeutigen theoretischen Annahmen des AMA zu alterskorrelierten Veränderungen falscher Erinnerungen bei Kindern, da dieser aus der Kognitionspsychologie stammt und andere Themenschwerpunkte hat. Die Vorhersagen zu alterskorrelierten Veränderungen gründen auf Befunden der Gedächtnisentwicklung (vgl. Kapitel 3). Wie bereits im Abschnitt 4.1 ausführlich erläutert wurde wird davon ausgegangen, dass Aktivierung und Quellendiskrimination alterskorreliert zunehmen und sich mit dem Alter erhöhen bzw. verbessern. In den folgenden Abschnitten werden die Ergebnisse getrennt nach Aktivierung, Quellendiskrimination und den resultierenden falschen Erinnerungen zusammengefasst.

Alterskorrelierte Unterschiede in der Aktivierung. Die Ergebnisse der verschiedenen Experimente wiesen darauf hin, dass es alterskorrelierte Differenzen in der Aktivierung kriti-

scher Items gab, die aber schwach ausfielen und auch nur zwischen der jüngsten (Vorschüler bzw. Erstklässler) und der ältesten Probandengruppe (Erwachsene) bedeutsam waren. Dies kann daran liegen, dass auf eine altersgerechte Listenkonstruktion geachtet wurde, so dass auch schon bei jüngeren Kindern die kritischen Items aktiviert werden konnten. Die alterskorrelierte Zunahme der Aktivierung ist darauf zurückzuführen, dass Erwachsene über mehr bereichsspezifisches und allgemeines Wissen verfügen und daher können die kritischen Items bei ihnen stärker automatisch aktiviert werden. Auch ist die bewusste Aktivierung erwachsener Probanden durch Kompetenzen zur strategischen Verarbeitung höher.

Die Aktivierung wurde anhand verschiedener Tests erfasst: Neben impliziten Verfahren (Experiment 1) wurde ein Inklusionstest (Experiment 2, 3A und 3B) eingesetzt und ein Quelldiskriminationstest (Frage nach beiden möglichen Quellen für die kritischen Items; Experiment 4). Da der Inklusionstest und der Quelldiskriminationstest zu vergleichbaren Aussagen über alterskorrelierte Unterschiede in der Aktivierung gelangten, wurde vermutet, dass beide Tests eine vergleichbare Aktivierung messen. Der einzig überraschende Befund zeigte sich in den impliziten Tests, wo sich keine alterskorrelierten Unterschiede ergaben, jedoch im falschen Wiedererkennen kritischer Items. Die Aktivierung, die in den impliziten Tests erfasst wurde, scheint daher nicht mit der Aktivierung im Wiedererkennenstest überein zu stimmen. Somit könnte das gewählte Testverfahren ausschlaggebend für Aktivierungsunterschiede sein.

Auffallend ist, dass in vielen unveröffentlichten Forschungsarbeiten der Universität Trier mit impliziten konzeptuellen und perzeptuellen Tests kein signifikantes Priming resultierte. Auch bei McDermott (1997) waren die berichteten Primingeffekte sehr gering und nur aufgrund der Stichprobengröße signifikant. In einigen aktuellen Untersuchungen mit perzeptuellen Tests zeigte sich kein Priming bzw. nur unter Bedingungen verstärkter assoziativer Aktivierung (McBride et al., 2006; Meade et al., 2007). Das signifikante Priming im eigenen Experiment fällt auf und führt zu der Vermutung, dass diesem Ergebnis ein Stichprobenfehler zugrunde gelegen haben könnte. Insgesamt scheint es, dass implizite Tests wenig geeignete Verfahren sind, um alterskorrelierte Unterschiede in der Aktivierung im DRM-Paradigma zu erfassen, da sich selten signifikantes Priming zeigt und wenn doch dann unklar ist, welche Art von Aktivierung getestet wird.

Ausgehend von den Befunden der Studie von Meade et al. (2007) wurde vermutet, dass das Zeitintervall, das zwischen Lern- und Testphase lag, einen Einfluss auf die unterschiedlichen Ergebnisse der impliziten und expliziten Tests gehabt haben könnte.

Im Gegensatz zu den impliziten Tests war die Interpretation der Ergebnisse des Inklusions- und Quellendiskriminationstests weniger problematisch, da sich in diesen Tests nur die jüngste und älteste Gruppe in der Aktivierung unterschieden. Die gewählten Inklusionsinstruktionen konnten allem Anschein nach die Aktivierung der kritischen Items erfassen und wurden auch von der jüngsten Altersgruppe verstanden, was sich an der höheren Anzahl genannter bzw. wiedererkannter kritischer Items nach der Inklusions- gegenüber der Standardinstruktion zeigte.

McDermott (2006) stellte ein anderes Verfahren vor, das wie die Inklusionsinstruktionen geeignet ist Aktivierung separat zu erfassen. Probanden wurden DRM-Listen vorgelesen, die listenweise reproduziert werden sollten. Es gab zwei verschiedene Testinstruktionen zur Reproduktion: Entweder sollten die Probanden die Wörter ohne zu raten reproduzieren oder sie sollten innerhalb von 90 Sekunden die präsentierten Items reproduzieren und in den anschließenden 90 Sekunden so viele Items erraten wie es ihnen möglich war (d.h. ihnen bekannte erscheinende Items reproduzieren). Die Probanden der Ratebedingung nannten mehr kritische Items als jene, die nicht raten sollten, was den eigenen Befunden zum Vergleich der Inklusions- mit der Standardinstruktion entspricht.

Diese Instruktion zum Raten wurde auch in der eigenen Voruntersuchung erteilt. Jedoch wurde eine andere Inklusionsinstruktion ausgewählt (das Wort, an das stark gedacht wurde, nennen bzw. „alt“ bezeichnen, s. Anhang B2 und C2), die anderen bisherigen Untersuchungen ähnelte, die zum Zeitpunkt der Konzeptualisierung der eigenen Experimente vorlagen (Brainerd et al., 2004; Hege & Dodson, 2004; Heit et al., 2004). Dadurch sollten die Ergebnisse miteinander verglichen werden können. Das vorgestellte Verfahren von McDermott (2006) bietet aber auch eine Möglichkeit die Aktivierung kritischer Items ohne entgegenwirkende Quellendiskrimination zu erfassen.

Der Einsatz von Inklusions- und Quellendiskriminationstests lässt sich problematisieren. Durch die verschiedenen Testphasen (*Inklusionstest*: Freie Reproduktion/Wiedererkennen mit anschließender Quellendiskrimination; *Quellendiskriminationstest*: Wiedererkennen mit zwei aufeinander folgenden Quellendiskriminationsfragen) könnte es sein, dass nicht die Aktivierung in der Lernphase erfasst wurde sondern zum Teil die der Testphase. Dies wird deutlich an der recht hohen Anzahl der wiedererkannten verbundenen Distraktoren dieser Experimente. Da im Quellendiskriminationstest weniger verbundene Distraktoren wiedererkannt wurden als im Inklusionstest, erscheint dieser besser geeignet zu sein, die Aktivierung der Lern- und nicht der Testphase zu erfassen. Für weitere Untersuchungen wäre es denkbar einen Inklusi-

onstest einzusetzen ohne anschließende Quellendiskrimination. Damit würde zwar nur die Aktivierung erfasst werden, aber es wäre weniger wahrscheinlicher, dass diese der Testphase entspräche.

Die alterskorrelierten Aktivierungsunterschiede ließen sich auch nicht durch Instruktionen zur Förderung der semantisch-relationalen Verarbeitung aufheben, wovon besonders die jüngsten Altersgruppen profitieren sollten, die diese Strategien nicht spontan einsetzen. Die aktivierungsfördernden Instruktionen konnten in allen Altersgruppen die Aktivierung erhöhen, wodurch Erwachsene immer noch eine stärkere Aktivierung als Kinder hatten, was mit ihrem höheren und besser vernetzten Wissen zusammenhängen könnte. Insgesamt war auffallend, dass schon die jüngste Altersgruppe in allen Experimenten zu einer hohen Aktivierung der kritischen Items gelangen konnte, womit die Bedeutung der semantisch-relationalen Verarbeitungsstrategien in Frage gestellt wird.

Eine weitergehende Analyse möglicher qualitativer Aktivierungsunterschiede zwischen den Altersgruppen zeigte, dass falsche Erinnerungen bei Kindern hauptsächlich auf bewusster Aktivierung zu gründen schienen, bei Erwachsenen eher auf nicht bewusster Aktivierung⁵.

Zusammenfassend ist festzuhalten, dass die alterskorrelierten Unterschiede in der Aktivierung mit den eingesetzten Verfahren eher gering waren. Die gewählten Verfahren zur separaten Erfassung der Aktivierung müssen kritisch betrachtet werden. Sowohl für die impliziten Tests als auch für den Inklusions- und Quellendiskriminationstest ist es unklar, welche Aktivierung erfasst wurde. Während sich bei impliziten Tests die Frage automatische oder bewusste Aktivierung stellte und weitere Einflussfaktoren untersucht werden sollten (Zeitintervall), war es bei dem Inklusions- und Quellendiskriminationstest möglich, dass auch die Aktivierung der Test- und nicht nur der Lernphase geprüft wurde.

Alterskorrelierte Unterschiede in der Quellendiskrimination. Die Aktivierung ist Grundvoraussetzung für falsche Erinnerungen, wird aber von der Quellendiskrimination beeinflusst, die falschen Erinnerungen wiederum entgegenwirkt. Auch in der Quellendiskrimination sollten Erwachsene Kindern überlegen sein, da die zugrunde liegenden Fähigkeiten alterskorrelierte Verbesserungen zeigen.

Die Befunde zur Quellendiskrimination kritischer Items wiesen erwartungskonform in allen Experimenten auf alterskorrelierte Verbesserungen hin, wobei sich die jüngeren Altersgruppen nicht immer unterschieden.

⁵ Wenn kritische Items bei Erwachsenen bewusst aktiviert werden, dann können gleichzeitig Informationen zur Quellendiskrimination genutzt werden, so dass falschen Erinnerungen wiederum entgegenwirkt wird.

Folgte die Aufgabe zur Quellendiskrimination auf den Inklusionstest und wurde nur nach einer Quelle gefragt (Experiment 3B), dann ergaben sich Unterschiede zwischen Erwachsenen und Kindern. Die Schwierigkeit der Aufgabe spiegelte sich im recht hohen Anteil der Fehlattribuierungen bei jungen Erwachsenen wieder und darin, dass die beiden jüngeren Altersgruppen sich nicht unterschieden. Durch eine Zuordnung zu beiden Quellen (Experiment 4) zeigten sich Unterschiede zwischen allen Altersgruppen und dieser Quellentest führte zu weniger Fehlattribuierungen bei älteren Grundschulkindern und Erwachsenen.

In zukünftigen Studien sollte der differenzierte Quellendiskriminationstest weiter angewendet werden um herauszustellen was zu den Unterschieden der beiden jüngsten Altersgruppen geführt haben könnte. Demnach ist es von besonderem Interesse, die Quellendiskrimination genauer zu untersuchen und zu prüfen, ob der heuristische und systematische Prozess spezifiziert und in Bezug zu alterskorrelierten Veränderungen gesetzt werden können. In einer Untersuchung mit Manipulation der zeitlichen Begrenzung könnte beispielsweise nur der heuristische Prozess untersucht werden, denn unter Zeitdruck ist eine strategische Quellendiskrimination nicht möglich. Fraglich ist hier aber, ob eine solche Aufgabe schon von jüngeren Kindern bewältigt werden kann.

Die vorgefundenen alterskorrelierten Unterschiede in den Quellendiskriminationsleistungen können mit Unterschieden in der Lern- und Testphase erklärt werden, wobei die höhere Arbeitsgedächtniskapazität Erwachsener ausschlaggebend zu sein scheint. In der Lernphase könnten sich die Altersgruppen darin unterscheiden, dass Erwachsene durch ihre höhere Arbeitsgedächtniskapazität mehr quellenrelevante Informationen verarbeiten können als Kinder. Dies wird durch Befunde zur semantisch-relationalen Verarbeitungsinstruktion (zu den Wortlisten ein Item generieren) in Experiment 3 gestützt, von welcher Erwachsene stark, Drittklässler leicht profitieren und zu einer besseren Quellendiskrimination gelangen konnten. In der Testphase können Erwachsene zu besseren Leistungen in der Quellendiskrimination gelangen, da ihre Arbeitsgedächtniskapazität die Nutzung von quellenstrategischen Prozessen erleichtert.

Ungeklärt ist aber ob die vorgefundenen Altersunterschiede hauptsächlich auf die Lern- oder die Testphase oder gleichermaßen auf beide zurückgehen und ob die Quellendiskriminationsleistung von Kindern gefördert werden kann. Dies könnte wie folgt in weiteren Untersuchungen geprüft werden: Vor der Lernphase könnte eine Vorwarnungsinstruktion erteilt werden darauf zu achten, ob die kritischen Items wirklich in der Liste vorgelesen werden. Damit könnte die Enkodierung quellenrelevanter Informationen gefördert werden und es würde sich

zeigen, ob jüngere Kinder diese nutzen können. Um die in der Testphase geforderten Prozesse zu untersuchen könnten Entscheidungsstrategien gefördert werden. Dies wäre auch mit einem differenzierten Quellendiskriminationstest möglich, der verdeutlicht, dass aktivierte Items mehreren Quellen zugeordnet werden können. Beide Prozesse könnten unabhängig voneinander mit Standardbedingungen verglichen werden, um Rückschlüsse auf die Wirkweise der einzelnen Variablen zu ermöglichen und auf eine eventuelle Wechselwirkung beider Variablen.

Um die Bedeutung der Arbeitsgedächtniskapazität innerhalb der Altersgruppen bei der Quellendiskrimination weitergehend zu prüfen, könnten Tests zur Arbeitsgedächtniskapazität [wie beispielsweise der Wisconsin Card Sorting Test (WCST); Grant & Berg, 1993] eingesetzt werden, welche die Frontallappenfunktion testen, die für die Quellendiskrimination wichtig zu sein scheint (Norman & Schacter, 1997; Schacter et al., 1999). Dadurch ließen sich die Probanden aller Altersgruppen in zwei Gruppen aufteilen: jene mit hoher und jene mit niedriger Arbeitsgedächtniskapazität. Es könnte so geprüft werden, ob Probanden mit hoher Arbeitsgedächtniskapazität zu besseren Leistungen in der Quellendiskrimination gelangen. Interessant wäre es auch zu untersuchen, ob die Ergebnisse zur Quellendiskrimination von Kindern mit hoher Arbeitsgedächtniskapazität sich weniger von denen Erwachsener mit niedriger Arbeitsgedächtniskapazität unterscheiden. Jedoch muss einschränkend hinzugefügt werden, dass die Arbeitsgedächtniskapazität kein unabhängiges Personenmerkmal zu sein scheint, sondern stark mit der Intelligenz korreliert (Oberauer, Wilhelm, Schulze & Süß, 2005).

Zusammenfassend ist zur Quellendiskrimination zu sagen, dass es wie erwartet alterskorrelierte Verbesserungen gab, wobei die Altersunterschiede zwischen den beiden jüngeren Gruppen abhängig von dem eingesetzten Quellentest waren.

Alterskorrelierte Unterschiede in falschen Erinnerungen. Die Befunde zu den alterskorrelierten Unterschieden falscher Erinnerungen in den eigenen Experimenten sind nicht einheitlich.

In den Experimenten 1 und 3A zeigte sich eine alterskorrelierte Zunahme falscher Erinnerungen von Kindern zu Erwachsenen.

Die Ergebnisse der freien Reproduktion (Experiment 2), in welcher Erwachsene auf den ersten Blick mehr falsche Erinnerungen als Kinder hatten, sind aufgrund von Bodeneffekten der beiden jüngeren Probandengruppen nicht interpretierbar. Es war zudem sehr wahrscheinlich, dass Erwachsene der Instruktion nicht Folge leisteten und insgesamt sehr viele kritische Items nannten.

In Experiment 3B und 4 jedoch hatten Erwachsene eindeutig weniger falsche Erinnerungen als Kinder. Vermutlich konnte durch den nachträglichen Quellentest der stärkeren Aktivierung entgegengewirkt werden. Während sich in Experiment 3B nur Erwachsene von Kindern unterschieden, was vermutlich daran liegen könnte, dass nur nach einer Quelle gefragt wurde (s.o.), konnte in Experiment 4 eine alterskorrelierte Abnahme falscher Erinnerungen von der jüngsten Gruppe bis zur ältesten beobachtet werden.

Die vorgefundenen Unterschiede zwischen Experiment 3 und 4 könnten also mit den verschiedenen Quellentests zusammenhängen. Sowohl bei Erwachsenen als auch bei Viertklässlern war der Anteil falscher Erinnerungen in Experiment 4 deutlich geringer als in Experiment 3, da in Experiment 4 beide mögliche Quellen genannt wurden und die beiden älteren Gruppen wahrscheinlich fähig waren diese Information zu nutzen, um strategisch bei der Quellediskrimination vorzugehen.

Falsche Erinnerungen ließen sich durch die semantisch-relationale Verarbeitungsinstruktion in Experiment 3 (zu der Wortliste ein Item generieren) beeinflussen. Erwachsene konnten ihre falschen Erinnerungen reduzieren (nicht aber in Experiment 4, in welchem der Anteil falscher Erinnerungen bei Erwachsenen ohnehin schon gering war), bei älteren Grundschulkindern (Experiment 3 und 4) und Vorschülern (Experiment 3) wurden diese erhöht. Die Aktivierung kritischer Items wurde bei allen Probanden durch die stärkere semantisch-relationale (in Experiment 4 auch die itemspezifische) Verarbeitung erhöht und war bei den beiden jüngeren Gruppen anscheinend stärker als die entgegenwirkende Quellediskrimination. Für sie zeigte sich ein "more is less" Effekt (Toglia et al., 1999), denn sie hatten nicht nur vermehrt falsche Erinnerungen, sondern auch richtige.

Zusammenfassend ist festzuhalten, dass die vorliegenden Befunde zu alterskorrelierten Veränderungen falscher Erinnerungen nicht einheitlich dafür sprechen, dass jüngere Kinder empfänglicher für falsche Erinnerungen sind als ältere und Erwachsene, aber alle Befunde sind gut vereinbar mit den Annahmen des AMA. Zur Erklärung der eigenen Ergebnisse ist es notwendig, Befunde der Gedächtnisentwicklung hinzuzuziehen und Aktivierungs- und Quellediskriminationsprozesse separat zu betrachten, da der AMA alleine nicht genügend Aussagen zu alterskorrelierten Veränderungen falscher Erinnerungen trifft.

10.2.2 Fuzzy-Trace Theorie (FTT)

Wie bereits angedeutet sollen die eigenen Befunde ergänzend mit der FTT erklärt werden. Die FTT entstand aus entwicklungspsychologischen Forschungsrichtungen, genauer gesagt, dem Konstruktivismus und der Schematheorie (vgl. Abschnitt 2.2.3).

Nach der FTT gibt es unterschiedliche Entwicklungstrends, die sich aber auf Befunde außerhalb des DRM-Paradigmas beziehen. Im Rahmen von Untersuchungen nach dem DRM-Paradigma nimmt der führende Vertreter der FTT, Brainerd, an, dass sich falsche Erinnerungen mit dem Alter erhöhen.

Mit steigendem Alter sollten sich die den "verbatim"- und "gist"-Repräsentationen zugrunde liegenden Fähigkeiten verbessern, wodurch der Anteil wahrer aber auch falscher Erinnerungen mit dem Alter zunehmen soll (Brainerd et al., 1995b; Brainerd & Mojardin, 1998; Brainerd & Reyna, 2002b; Brainerd & Reyna, 2004). Jüngeren Kindern fällt es aufgrund des mangelnden Zugriffs auf die "gist"-Spur schwer den Bedeutungskern von Ereignissen zu erfassen. Im DRM-Paradigma wird durch die Listenkonstruktion (semantische Assoziiertheit der Listenelemente) vor allem die "gist"-Spur angesprochen, was dazu führt, dass besonders Erwachsene falsche Erinnerungen produzieren sollten, da Kinder sich auf die einzelnen Items konzentrieren und nicht die zugrunde liegende Bedeutung erkennen können (Brainerd et al., 2002a; Brainerd & Reyna, 2004).

Eine differenzierte Betrachtung der eigenen Befunde nach Aktivierung und Quellendiskrimination ist nach der FTT nicht möglich, da sie keine äquivalenten Prozesse annimmt.

Die Ergebnisse zum Einfluss der semantisch-relationalen Verarbeitungsinstruktion auf falsche Erinnerungen in Experiment 3B lassen sich mit der FTT gut in Einklang bringen. Durch diese Instruktion sollte die "gist"-Verarbeitung gefördert werden, welche entscheidend für die alterskorrelierte Zunahme falscher Erinnerungen sein soll (Brainerd et al., 1995b; Brainerd & Mojardin, 1998; Brainerd & Reyna, 2002b; Brainerd & Reyna, 2004). Da Kinder die "gist"-Spur schlechter als Erwachsene nutzen können, erhöhten sich bei ihnen nach der Verarbeitungsinstruktion falsche Erinnerungen.

Werden die eigenen Ergebnisse nur nach den resultierten falschen Erinnerungen betrachtet (Standardwiedererkennenstest: alterskorrelierte Zunahme bzw. keine Altersunterschiede vs. differenzierter Quellendiskriminationstest: alterskorrelierte Abnahme), dann lassen sich diese grob gesehen nicht mit der von Brainerd postulierten Annahme vereinbaren. Jedoch ist eine Interpretation nach dem aus der FTT entwickelten "recollection rejection" Prozess möglich (vgl. Abschnitt 2.2.3). Erwachsene können die "verbatim"-Spur besser nutzen als Kinder, um

ausgehend von dieser das mit dem “gist“-konsistenten kritische Item abzulehnen entgegen der empfundenen Vertrautheit. Einige aktuelle Studien verweisen auf die Bedeutung des “recollection rejection“ Prozesses in der Unterdrückung falscher Erinnerungen (Lampinen, Meier, Arnal, & Leding, 2005; Odegard & Lampinen, 2005; Starns & Hicks, 2004). Die benötigten kognitiven Prozesse übersteigen die Fähigkeiten von jüngeren Kindern. Diese Interpretation gilt für die Ergebnisse der alterskorrelierten Abnahme falscher Erinnerungen. Das Ergebnis der alterskorrelierten Zunahme falscher Erinnerungen lässt sich gut mit der FTT in Einklang bringen. Erwachsene konnten durch den besseren “gist“ das kritische Item eher als Listenthema extrahieren und es falsch erinnern.

Die Annahme der FTT, dass sich korrekte Erinnerungen mit dem Alter erhöhen, konnte anhand der erinnerten Listenitems bestätigt werden. Dies ist laut der FTT auf eine bessere Nutzung der “verbatim“- und auch der “gist“-Spur zu erklären (Brainerd & Reyna, 2004; Brainerd et al., 2004).

Auch die Vertreter der FTT untersuchten den Einfluss verschiedener Instruktionen auf falsche Erinnerungen. Brainerd et al. (2004) entwickelten eine der Inklusionsinstruktion vergleichbare Instruktion, indem sie die Probanden verschiedener Altersgruppen (Fünf-, Sieben-, 11- und 14-Jährige) anwies entweder nur die präsentierten Items im Wiedererkennenstest als „alt“ zu bezeichnen (“verbatim“) oder nur die nicht präsentierten Items, die der Kategorie zugehörig waren (“gist“) oder aber sowohl die präsentierten Items als auch die nicht präsentierten aber zugehörigen Items (“verbatim und gist“). Die “verbatim“-Instruktion ist vergleichbar mit der eigenen Standardinstruktion und die kombinierte “verbatim und gist“-Instruktion entspricht im Groben der eigenen Inklusionsinstruktion. Wie auch in den eigenen Experimenten zeigte sich eine Zunahme der wiedererkannten kritischen Items von der “verbatim“- zu der “verbatim und gist“-Instruktion. Ein Vergleich der Altersgruppen führte ebenso zu einer alterskorrelierten Zunahme falscher Erinnerungen, im Einklang mit der FTT und zum Teil mit den eigenen Ergebnissen.

Die insgesamt deutlichen Werte falscher Erinnerungen werden nach der FTT auch in Bezug zur Listenkonstruktion gesetzt. Obwohl die Untersuchung des Zusammenhangs von Assoziationsstärken auf falsche Erinnerungen bislang hauptsächlich von Vertretern des AMA vorgenommen wurde (Gallo & Roediger, 2002; Roediger et al., 2001), prüften dies auch Brainerd und Wright (2005) und bezogen diese auf das Konzept der “recollection rejection“. Eine hohe RAS und VAS sollte zu weniger “recollection rejection“ führen und in Folge vermehrt falsche Erinnerungen hervorrufen. Dies sollte bedingt sein durch eine schlechtere Speicherung der

“verbatim“-Spur und/oder deren eingeschränkter Verfügbarkeit beim Abruf im Wiedererkennenstest.

Schlussfolgernd ist zu der FTT zu sagen, dass die eigenen Befunde zu falschen Erinnerungen sich auch mit dieser unter Beachtung des “recollection rejection“ Prozesses erklären lassen.

10.3 Ausblick

Studien, die zu Erkenntnissen über alterskorrelierte Veränderungen falscher Erinnerungen gelangen können, liefern einen entscheidenden Beitrag zur Evaluation und Weiterentwicklung der Theorien. Ziel des AMA war es ursprünglich falsche Erinnerungen zu erklären. Erweiterungen fanden bei der Erforschung der Einflussfaktoren statt, die falsche Erinnerungen erhöhen bzw. reduzieren sollten. In der eigenen Studie sollte ausgehend von dem AMA Aktivierung und Quellendiskrimination in verschiedenen Altersgruppen separat voneinander getestet werden, um Aussagen zu alterskorrelierten Veränderungen der einzelnen Komponenten treffen zu können. Ein Hauptziel war es also nicht nur die alterskorrelierten Veränderungen falscher Erinnerungen zu untersuchen sondern vor allem die zugrunde liegenden Prozesse. Da der AMA eine rein kognitionspsychologische Theorie ist, profitiert dieser von Erweiterungen um entwicklungspsychologische Aspekte, weil dadurch dessen Aussagekraft und Anwendungsgebiet größer werden. Viele der eigenen Hypothesen wurden durch entwicklungspsychologische Befunde zur Gedächtnisentwicklung gestützt oder sie wurden ungerichtet formuliert aufgrund fehlender Aussagen des AMA zur entwicklungsabhängigen Veränderung der Aktivierung und der Quellendiskrimination. Die eigenen Befunde tragen somit zur Spezifizierung des AMA bei.

Die Rückschlüsse zu alterskorrelierten Veränderungen der Aktivierung, der Quellendiskrimination und falscher Erinnerungen bei der Erwachsenenstichprobe sind in den Experimenten 1 bis 3 unter Vorbehalt zu interpretieren, denn es handelte sich um eine jüngere Erwachsenenstichprobe, um nicht zu sagen um eine Stichprobe jugendlicher Probanden. Besonders die Ergebnisse des zweiten Experiments überraschten aufgrund ihres hohen Anteils falscher Erinnerungen nach der Standardinstruktion in der Freien Reproduktion. Dies wurde dadurch erklärt, dass die Gymnasiasten sich nicht an die Instruktion hielten und nach dem selbst gewählten Kriterium „Je mehr desto besser“ möglichst viele Items (und auch kritische Items) reproduzierten. Daher könnte besonders dieses Experiment mit einer „richtigen“ Erwachsenenstichprobe repliziert werden, die nicht den Ehrgeiz der Gymnasiasten haben, sondern sich an die Instruktion nicht zu raten halten.

Aufgrund dessen dass es bisher keine Untersuchung zur Veränderung falscher Erinnerungen über die gesamte Lebensspanne gibt, wäre es äußerst aufschlussreich, alterskorrelierte Veränderungen der Aktivierung und Quellendiskrimination anhand einer Stichprobe bestehend aus Vorschülern, Grundschulern, jungen Erwachsenen und Senioren zu überprüfen. Indem innerhalb einer Studie den Probanden altersadäquate und -normierte Anforderungen gestellt, Aktivierung und Quellendiskrimination auf eine der möglichen Weisen getrennt voneinander erfasst werden würden, könnte ein großer Erkenntnisgewinn erreicht werden, der das Verständnis von falschen Erinnerungen und den zugrunde liegenden Prozessen entscheidend bereichern würde. Es ließen sich direkte Vergleiche zwischen allen Altersgruppen ziehen und es könnte beispielsweise geprüft werden, ob und wie sich die Defizite von Kindern und Senioren in der Quellendiskrimination beeinflussen lassen.

Wie Steffens und Mecklenbräuer (2007) in ihrem Überblick zu Studien falscher Erinnerungen herausstellen, gab es in den vergangenen Jahren ein explosionsartig wachsendes Interesse an diesem Gebiet aufgrund der Praxisrelevanz. Hier ist zum einen im juristischen Bereich die Glaubwürdigkeit von Zeugenaussagen hervorzuheben und zum anderen im psychotherapeutischen Bereich die wiederentdeckten/-belebten Erinnerungen.

Auch die eigenen Ergebnisse zu alterskorrelierten Unterschieden falscher Erinnerungen im DRM-Paradigma lassen sich beispielsweise mit den Befunden zur Zeugenaussageforschung vergleichen.

In diesem Zusammenhang ist das bekannte Sam-Stone Experiment zu erwähnen (Leichtman & Ceci, 1995), in welchem der Einfluss des Alters, der Stereotype und des Suggestierens auf falsche Erinnerungen untersucht wurde. In dieser Untersuchung nahmen zwei Altersgruppen teil: vier- und sechsjährige Kinder, bei denen interindividuell eine Variation der Variablen Stereotypinduzierung und Suggestion stattfand. Den Probanden der Bedingung Stereotypinduzierung wurde vor dem Besuch des Mannes Sam Stone dieser als ungeschickte und plumpe Person beschrieben. Als er dann kurz auftauchte, grüßte er die Kinder, ging umher und verabschiedete sich wieder. Nach seinem Besuch wurden ein beschädigtes Buch und ein schmutziger Teddybär gefunden. Den Kindern der Suggestierungsbedingung wurden in den folgenden Wochen mehrfach befragt, wobei ihnen Suggestivfragen gestellt wurden („War Sam Stone glücklich oder traurig, dass er den Teddybär schmutzig gemacht hat?“). Im abschließenden Interview wurden alle Kinder befragt, ob Sam Stone den Teddybär beschmutzt und das Buch beschädigt hat. Die Ergebnisse zeigten, dass sowohl die Induzierung von Stereotypen und die

Suggestivfragen falsche Erinnerungen erhöhten. Zudem waren jüngere Kinder empfänglicher für falsche Erinnerungen als ältere.

Auch Ackil und Zaragoza (1995, 1998) konnten zeigen, dass jüngere Kinder stärker auf suggerierte Falschinformationen ansprechen als ältere Kinder und Erwachsene. Auf diese Studien wurde bereits in Abschnitt 9.1 im Zusammenhang mit Experiment 4 eingegangen. Jüngere Kinder hatten große Probleme bei der externen Quellendiskrimination (Kam das Item im Film vor oder in der nachträglichen Zusammenfassung?) nicht aber bei der internen (Wurde das Item selbst genannt?). Sie konnten sich selbst als Quelle der Information benennen aber schienen empfänglicher als ältere für Fehlinformationen zu sein, wenn es sich um externe Quellen handelte und diese einander ähnlich waren.

Die Befunde, dass jüngere Kinder mehr falsche Erinnerungen als ältere und Erwachsene haben, stimmen nicht generell mit den eigenen Ergebnissen im DRM-Paradigma überein. Abgesehen von Experiment 4 konnte kein Altersunterschied zwischen jüngeren und älteren Kindern in falschen Erinnerungen festgestellt werden. Die Ergebnisse von zwei vorliegenden eigenen Experimenten (1 und 3A) sowie einiger anderer Studien im DRM-Paradigma zeigten, dass Kinder sogar weniger falsche Erinnerungen als Erwachsene hatten (Brainerd et al., 2002a; Howe, 2005), was an einer geringeren Aktivierung der kritischen Items liegen müsste, denn es gibt keine Befunde darüber, dass Erwachsene eine schlechtere Quellendiskrimination als Kinder haben. Eine stärkere Aktivierung hingegen führt bei Kindern vermehrt zu falschen Erinnerungen als bei Erwachsenen, da ihre Fähigkeiten zur Quellendiskrimination schlechter sind. Im Gegensatz zum DRM-Paradigma werden in der Zeugenaussageforschung Kindern Falschinformationen suggeriert und/oder Stereotype induziert. Jüngere Kinder wollen konsistent mit Erwartungen sein und neigen dazu zu behaupten, dass sie alle erdenklichen Ereignisse erlebt haben (Steffens & Mecklenbräuker, 2007). Das DRM-Paradigma kann allenfalls als Selbstsuggestion betrachtet werden. Die Listenitems sind mit dem nicht präsentierten kritischen Item assoziiert, aber dieses kritische Item wird nicht von einer anderen Person suggeriert.

Bei der Prüfung von Zeugenaussagen auf ihre Verlässlichkeit zeigt sich die Praxisrelevanz der Befunde aus dem DRM-Paradigma. Sowohl Kinder als auch Erwachsene können falsche Erinnerungen haben ohne wissentliche, bewusste Täuschung. Es ist nicht generell davon auszugehen dass Kinder, auch jüngere, weniger geeignete d.h. zuverlässige Zeugen sind als Erwachsene. Neben dem Verzicht auf Suggestivfragen sollte beachtet werden, dass für jüngere Kinder die Aufgabe der Unterscheidung zweier interner Quellen (selbst gesagt bzw. gedacht)

oder zweier sehr ähnlicher Quellen eine schwierige ist, nicht aber die Unterscheidung zweier externer Quellen (Hat Person A oder Person B etwas gesagt?). Es sollten insbesondere bei Kindern keine mehrfachen Befragungen stattfinden, da dies die Quellendiskrimination zusätzlich erschweren würde.

Zusammengefasst lässt sich festhalten, dass das noch junge Forschungsfeld der falschen Erinnerungen im DRM-Paradigma viele interessante Fragestellungen bietet. Sorgfältig geplante, durchgeführte und ausgewertete Studien mit Prozeduren zur separaten Erfassung der Aktivierung kritischer Items und der entgegenwirkenden Quellendiskrimination können zu einer bedeutsamen Weiterentwicklung und Differenzierung des AMA führen. Im Hinblick auf die allgemein gesellschaftliche und speziell juristische Relevanz der Empfänglichkeit gegenüber falschen Erinnerungen ist es offensichtlich, dass weitere Studien nicht nur erwünscht, sondern auch äußerst notwendig sind.

11 Zusammenfassung

Die vorliegende Arbeit hatte die Zielsetzung einen Beitrag zum Verständnis von alterskorrelierten Veränderungen falscher Erinnerungen zu leisten und ausgehend von dem AMA zu prüfen, inwieweit diese auf entsprechende Unterschiede in der Quellendiskrimination und/oder auf Unterschiede in der Wahrscheinlichkeit der Aktivierung der kritischen Wörter zurückgehen. Zu diesem Zweck wurden Prozeduren entwickelt, anhand derer die beiden Prozesse in allen Altersgruppen getrennt voneinander untersucht werden konnten und der Einfluss auf falsche Erinnerungen herausgestellt werden konnte. Damit fand eine Erweiterung des ursprünglich aus der Kognitionspsychologie stammenden AMA dahingehend statt auch entwicklungspsychologische Befunde zu erklären.

An allen Experimenten nahmen drei Altersgruppen teil: Vorschüler/Erstklässler, Dritt-/ Viertklässler und junge Erwachsene. Im ersten Experiment wurde die Aktivierung der kritischen Items anhand zweier impliziter Testverfahren („Ersparnistest“ und Wortassoziationstest) geprüft, die ausschließlich die Aktivierung erfassten. Um falsche Erinnerungen zu untersuchen erfolgte nach beiden impliziten Tests jeweils ein Wiedererkennenstest, der Aktivierung und Quellendiskrimination zusammen testete. Die Ergebnisse zur Aktivierung waren nicht eindeutig. In beiden impliziten Tests zeigten sich zwar übereinstimmend keine alterskorrelierten Veränderungen der Aktivierung, aber im Wiedererkennenstest kam es zu einem alterskorrelierten Anstieg falscher Erinnerungen, der nicht auf schlechtere Fähigkeiten erwachsener Probanden in der Quellendiskrimination zurückgeführt werden kann, sondern auf Aktivierungsunterschiede der Probanden. Als mögliche Einflussvariable wurde das Zeitintervall zwischen Lern- und Testphasen betrachtet, so dass bei den impliziten Tests eher eine kurzfristige Aktivierung, bei den expliziten Tests eine langfristige Aktivierung erfasst wurde.

In den nächsten drei Experimenten wurden falsche Erinnerungen mit expliziten Verfahren untersucht und so genannte Inklusionsinstruktionen erteilt, anhand derer nur die Aktivierung der kritischen Items getestet werden sollte ohne entgegenwirkende Quellendiskrimination. Diese Inklusionsinstruktionen wurden in der Freien Reproduktion (Experiment 2) und im Wiedererkennenstest (Experimente 3A und 3B) mit Standardinstruktionen verglichen, die Aktivierung und Quellendiskrimination erfassten. Das dritte Experiment bestand aus zwei Teilen: Experiment 3A und 3B waren im Versuchsaufbau und im Einsatz des Wiedererkennenstests identisch, aber in Experiment 3B wurde darüber hinaus eine Instruktion zur semantisch-relationalen Verarbeitung erteilt (zu jeder DRM-Liste das fehlende, die Listenitems zusammenfassende Wort benennen), welche die Aktivierung fördern und Altersunterschiede

minimieren sollte. Durch den Vergleich beider Experimente (3A und 3B) konnten Aussagen zu den Möglichkeiten der Aktivierungsförderung und zu dem Einfluss der Verarbeitungsinstruktion auf falsche Erinnerungen getroffen werden.

In Experiment 2, 3A und 3B zeigte sich eine alterskorrelierte Zunahme der Aktivierung, aber nur zwischen Vorschülern und Erwachsenen. Die Lerninstruktionen zur strategischen Verarbeitung in Experiment 3B führten in allen Altersgruppen zu einer Erhöhung der Aktivierung. Auch die Quellendiskriminationsleistungen verbesserten sich mit zunehmendem Alter. Sowohl im zweiten als auch im dritten Experiment, in welchen nur nach einer Quelle gefragt wurde, unterschieden sich die beiden jüngeren Altersgruppen nicht, was die Schwierigkeit der Aufgabe verdeutlichte.

Ausgehend von den eher geringen Aktivierungsunterschieden, aber den deutlichen alterskorrelierten Verbesserungen der Quellendiskrimination, wäre eine alterskorrelierte Abnahme falscher Erinnerungen erwartet worden. Dies war nicht der Fall: In Experiment 3A nahmen falsche Erinnerungen mit dem Alter zu, weil vermutlich der Einfluss der Aktivierung stärker war als jener der Quellendiskrimination. Die Ergebnisse zu falschen Erinnerungen in Experiment 2 waren aufgrund von Bodeneffekten nicht interpretierbar. In Experiment 3B kam es zu einer alterkorrelierten Abnahme falscher Erinnerungen, was mit den Ergebnissen zur Aktivierung und Quellendiskrimination stimmig ist. Die Instruktion zur Aktivierungsförderung in Experiment 3B hatte Einfluss auf falsche Erinnerungen aller Altersgruppen: Erwachsene konnten ihren Anteil falscher Erinnerungen reduzieren, bei Kindern erhöhten sich falsche aber auch richtige Erinnerungen wie nach dem "more is less" Effekt (Toglia et al., 1999).

Im letzten Experiment sollte eine ähnliche Förderung der Aktivierung wie im vorherigen stattfinden, aber anstatt nur ein Wort zu einer DRM-Liste zu generieren, sollte zu jedem Listenitem ein Wort generiert werden, wodurch auch die itemspezifische Verarbeitung gefördert wurde. Anhand eines differenzierten Quellendiskriminationstests wurden Aktivierung und Quellendiskrimination getrennt voneinander erfasst. Die Probanden ordneten die Items einer internen (Person hat an das Item gedacht bzw. es selbst gesagt) und/oder externen Quelle (Item wurde vorgelesen) zu. Die Aktivierung erhöhte sich im letzten Experiment alterskorreliert, wobei sich auch hier nur die älteste und die jüngste Gruppe unterschieden. Die Lerninstruktion konnte wie in Experiment 3 in allen Altersgruppen zu einer Steigerung der Aktivierung führen. Es zeigte sich wie erwartet eine alterskorrelierte Verbesserung der Quellendiskriminationsleistungen, aber im Gegensatz zu den vorherigen Experimenten unterschieden sich nun auch die beiden jüngeren Altersgruppen in der falschen Quellenzuordnung kritischer

Items. In diesem Experiment reduzierten sich falsche Erinnerungen mit zunehmendem Alter, was mit der alterskorrelierten Verbesserung der Quellendiskrimination zusammenzuhängen schien. Die Lerninstruktion hatte nur bei Viertklässlern einen Einfluss auf falsche Erinnerungen indem sie diese erhöhte.

Zusammenfassend wurde in allen Experimenten eine geringe alterskorrelierte Zunahme der Aktivierung festgestellt und deutliche alterskorrelierte Verbesserungen der Quellendiskrimination. Es war möglich die Aktivierung durch spezielle Lerninstruktionen zu erhöhen, nicht nur bei Kindern sondern auch bei Erwachsenen. Die Befunde zu falschen Erinnerungen in den verschiedenen Experimenten waren je nach Lern- und Testbedingung unterschiedlich: Die Ergebnisse deuteten sowohl auf eine alterskorrelierte Zunahme als auch auf eine alterskorrelierte Abnahme hin. Die aktivierungsfördernden Instruktionen hatten unterschiedliche oder auch keine Effekte auf falsche Erinnerungen je nach Instruktion und Altersgruppe.

In zukünftigen Studien zu alterskorrelierten Veränderungen falschen Erinnerungen sollte darauf geachtet werden nicht nur Aktivierung und Quellendiskrimination getrennt voneinander zu erfassen, sondern auch zu versuchen, diese beiden Komponenten zu spezifizieren. Weitere Untersuchungen mit dem differenzierten Quellendiskriminationstest wären sinnvoll, da dieser vielfältige Aussagen getrennt zu Aktivierung und Quellendiskrimination ermöglicht.

Die praktische Relevanz der vorliegenden Untersuchung betrifft unter anderem das juristische Gebiet der Zeugenaussagen. Im Gegensatz zu den Befunden der Zeugenaussageforschung, in welchen an der Verlässlichkeit kindlicher Zeugenaussagen gezweifelt wird, konnte die vorliegende Studie nicht bestätigen, dass jüngere Kinder generell mehr falsche Erinnerungen als ältere und Erwachsene haben. In Abhängigkeit von bestimmten Lern- und Testbedingungen hatten Kinder sogar weniger falsche Erinnerungen als Erwachsene.

12 Literaturverzeichnis

- Ackil, J.K. & Zaragoza, M.S. (1995). Developmental differences in eyewitness suggestibility and memory for source. *Journal of Experimental Child Psychology*, 60, 57-83.
- Ackil, J.K. & Zaragoza, M.S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology*, 34, 1358-1372.
- Andrews, B. (1997). Can a survey of British False Memory Society members reliably inform the recovered memory debate? *Applied Cognitive Psychology*, 11, 19-23.
- Anisfeld, M. & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 77, 171-179.
- Atkinson, R.C. & Shiffrin, R.M. (1968). Human Memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation; advances in research and theory* (Vol. 2, pp. 89-159). New York: Academic Press.
- Baddeley, A. (1990). *Human Memory: Theory and Practice*. Hillsdale, NJ: Erlbaum.
- Balota, D.A., Cortese, M.J., Duchek, J.M., Adams, D., Roediger, H.L., III, McDermott, K.B. & Yerys, B.E. (1999). Veridical and false memories in healthy older adults and in dementia of Alzheimers type. *Cognitive Neuropsychology*, 16, 312-384.
- Balota, D.A. & Paul, S.T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 827-845.
- Balwin, J.M. & Shaw, W.J. (1895). Memory for square size. *Psychological Review*, 2, 236-239.
- Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Benjamin, A.S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 941-947.
- Bjorklund, D.F. (1985). The role of conceptual knowledge in the development of organization in children's memory. In C.J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development* (pp. 103-142). New York: Springer.

- Bjorklund, D.F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretative review. *Developmental Review*, 7, 93-130.
- Bjorklund, D.F. (1989). Memory. In D.F. Bjorklund (Ed.), *Children's thinking: Developmental function and individual differences* (pp. 152-172). Pacific Grove, California: Brooks/Cole Publishing Company.
- Bjorklund, D.F. & Coyle, T.R. (1995). Utilization deficiencies in the development of memory strategies. In F.E. Weinert & W. Schneider (Eds.), *Memory Performance and Competencies: Issues in Growth and Development* (pp. 161-180). Mahwah, NJ: Erlbaum.
- Bjorklund, D.F. & Douglas, R.N. (1997). The development of memory strategies. In N. Cowan (Ed.), *The Development of Memory in Childhood* (pp. 201-246). Hove: Psychology Press.
- Bjorklund, D.F. & Jacobs, J.W. (1985). Associative and categorial processes in children's memory: The role of automaticity in the development of organization in free recall. *Journal of Experimental Child Psychology*, 39, 599-617.
- Blair, I.V., Lenton, A.P. & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review*, 9, 590-596.
- Brainerd, C.J., Holliday, R.E. & Reyna, V.F. (2004). Behavioral measurement of remembering phonologies: So simple a child can do it. *Child Development*, 75, 505-522.
- Brainerd, C.J. & Mojardin, A.H. (1998). Children's spontaneous false memories for narrative statements: Longterm persistence and mere-testing effects. *Child Development*, 69, 1361-1377.
- Brainerd, C.J. & Reyna, V.F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology*, 32, 467-478.
- Brainerd, C.J. & Reyna, V.F. (1998a). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology*, 71, 81-129.
- Brainerd, C.J. & Reyna, V.F. (1998b). When things that never happened are easier to remember than things that did. *Psychological Science*, 9, 484-489.
- Brainerd, C.J., & Reyna, V.F. (2002a). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11, 164-169.
- Brainerd, C.J., & Reyna, V.F. (2002b). Recollection rejection: How children edit their false memories. *Developmental Psychology*, 38, 156-172.
- Brainerd, C.J., & Reyna, V.F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, 24, 396-439.

- Brainerd, C.J., Reyna, V.F. & Brandse, E. (1995a). Are children's false memories more persistent than their true memories? *Psychological Science*, 6, 359-364.
- Brainerd, C.J., Reyna, V.F. & Forrest, T.J. (2002a). Are young children susceptible to the false-memory illusion? *Child Development*, 73, 1363-1377.
- Brainerd, C.J., Reyna, V.F. & Kneer, R. (1995b). False recognition reversal: When similarity is distinctive. *Journal of Memory and Language*, 34, 157-185.
- Brainerd, C.J. & Wright, R. (2005). Forward association, backward association, and the false-memory illusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 554-567.
- Brainerd, C.J., Wright, R., Reyna, V.F. & Payne, D.G. (2002b). Dual-retrieval processes in free and associative recall. *Journal of Memory and Language*, 46, 120-152.
- Bransford, J.D., Franks, J.J., Morris, C.D. & Stein, B.S. (1979). Some general constraints on learning and memory research. In L.S. Cermac & F.I.M. Craik (Eds.), *Levels of processing in human memory* (pp. 331-354). Hillsdale, NJ: Erlbaum.
- Cabeza, R. & Lennartson, E.R. (2005). False memories across language : Implicit associative response vs fuzzy trace view. *Memory*, 13, 1-5.
- Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.
- Ceci, S.J. & Bruck, M. (1993). The suggestibility of child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403-439.
- Ceci, S.J. & Howe, M.J.A. (1978). Semantic knowledge as a determinant of developmental differences in recall. *Journal of Experimental Child Psychology*, 26, 230-245.
- Chaiken, S. Lieberman, A. & Eagly, A.H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J.S. Uleman & J.A. Bargh (Eds.), *unintended thought* (pp. 212-252). New York: Guilford Press.
- Challis, B.H. & Sidhu, R. (1993). Dissociative effects of massed repetition on implicit and explicit measures of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 115-127.
- Chan, J.K., McDermott, K.B., Watson, J.M. & Gallo, D. (2005). The importance of material processing interactions in inducing false memories. *Memory & Cognition*, 33, 389-395.
- Chi, M.T.H. (1978). Knowledge structure and memory development. In R.S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum.
- Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing. A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 4, 671-684.

- Craik, F.I.M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.
- Collins, A.M. & Loftus, E.F. (1975). A spreading-activation theory of semantic memory. *Psychological Review*, 82, 407-428.
- Deese, J. (1959a). Influence of interitem associative strength upon immediate free recall. *Psychological Reports*, 5, 235-241.
- Deese, J. (1959b). On the prediction of occurrence of particular verbal intrusions in immediate free recall. *Journal of Experimental Psychology*, 58, 17-22.
- Dehon, H. & Brédart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging*, 19, 191-197.
- Dempster, F.N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, 89, 63-100.
- Dewhurst, S.A., Barry, C & Holmes, S.J. (2005). Exploring the false recognition of category exemplars: Effects of divided attention and explicit generation. *European Journal of Cognitive Psychology*, 17, 803-819.
- Diliberto-Macaluso, K.A. (2005). Priming and false memories from Deese-Roediger-McDermott lists on a fragment completion test with children. *American Journal of Psychology*, 118, 13-28.
- Dodd, M.D., Sheard, E.D. & MacLeod, C.M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory*, 14, 115-126.
- Dodson, C.S. & Schacter, D.L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychosomatic Bulletin & Review*, 8, 155-161.
- Dodson, C.S. & Schacter, D.L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, 46, 782-803.
- Dufresne, A. & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, 47, 274-296.
- Echterhoff, G. (2001). False memory. In N. Pethes & J. Ruchatz (Hrsg.), *Gedächtnis und Erinnerung. Ein interdisziplinäres Lexikon* (S. 165-166). Hamburg: Rowohlt.
- Erdfelder, E. (2001). Gedächtnistäuschungen. In N. Pethes & J. Ruchatz (Hrsg.), *Gedächtnis und Erinnerung. Ein interdisziplinäres Lexikon* (S. 207-208). Hamburg: Rowohlt.
- Fivush, R. (1997). Event memory in early childhood. In N. Cowan (Ed.), *The Development of Memory in Childhood* (pp. 136-161). Hove: Psychology Press.

- Flavell, J.H. (1971). First discussion's comments: What is memory development the development of? *Human Development*, 14, 272-278.
- Flavell, J.H., Beach, D.H. & Chinsky, J.M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, 37, 283-299.
- Flavell, J.H. & Wellman, H.M. (1977). Metamemory. In R.V. Kail & J.W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3-33). Hillsdale, NJ: Erlbaum.
- Foley, M.A. & Johnson, M.K. (1985). Confusion between memories for performed and imagined actions. *Child Development*, 56, 1145-1155.
- Foley, M.A., Johnson, M.K. & Raye, C.L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development*, 54, 51-60.
- Gallo, D.A., McDermott, K.B., Percer, J.M. & Roediger, H.L. III (2001a). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 339-353.
- Gallo, D.A., Roberts, M.J. & Seamon, J.G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, 4, 271-276.
- Gallo, D.A. & Roediger, H.L., III (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, 47, 469-497.
- Gallo, D.A. & Roediger, H.L., III (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition*, 31, 1036-1044.
- Gallo, D.A., Roediger, H.L., III & McDermott, K.B. (2001b). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, 8, 579-586.
- Gallo, D.A. & Seamon, J.G. (2004). Are nonconscious processes sufficient to produce false memories? *Consciousness and Cognition*, 13, 158-168.
- Ghetti, S., Qin, J. & Goodman, G. (2002). False memories in Children and Adults: Age, Distinctiveness, and subjective experience. *Developmental Psychology*, 38, 705-718.
- Glanzer, M. & Adams, J. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13, 8-20.
- Graf, P. & Schacter, D.L. (1985). Priming across modalities and priming across category levels: Extending the domain of preserved function in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 501-518.
- Grant, D.A. & Berg, E.A. (1993). *Wisconsin Card Sorting Test (WCST)*. Göttingen: Hogrefe.
- Green, D.M. & Swets, J.A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.

- Gudjonsson, G.H. (1997). Accusations by adults of childhood sexual abuse: a survey of the members of the British False Memory Society (BFMS). *Applied Cognitive Psychology, 11*, 3-18.
- Hamann, S.B. (1990). Level-of-processing effects in conceptually driven implicit tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 970-977.
- Hancock, T.W., Hicks, J.L., Marsh, R.L. & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese-Roediger-McDermott paradigm. *American Journal of psychology, 116*, 1-14.
- Hasselhorn, M. (1995). Beyond production of deficiency and utilization inefficiency. Mechanisms of the emergence of strategic categorization in episodic memory tasks. In F.E. Weinert & W. Schneider (Eds.), *Memory Performance and Competencies: Issues in Growth and Development* (pp. 141-160). Mahwah, NJ: Erlbaum.
- Hasselhorn, M. (1996). *Kategoriales Organisieren bei Kindern: Zur Entwicklung einer Gedächtnisstrategie*. Göttingen: Hogrefe.
- Hasselhorn, M. & Grube, D. (1994). Erstassoziationen von Kindern und Erwachsenen zu 53 konkreten Substantiven. In W. Hager & M. Hasselhorn (Hrsg.), *Handbuch deutschsprachiger Wortnormen* (S. 59-64). Göttingen: Hogrefe.
- Hege, A.C.G. & Dodson, C.S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 787-795.
- Heit, E., Brockdorff, N. & Lamberts, K. (2004). Strategic processes in false recognition memory. *Psychonomic Bulletin & Review, 11*, 380-386.
- Hicks, J.L. & Hancock, T.W. (2002). Backward associative strength determines source attributions given to false memories. *Psychonomic Bulletin & Review, 9*, 807-815.
- Howe, M.L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science, 16*, 927-931.
- Howe, M.L. (2006). Developmentally invariant dissociations in children's true and false memories: Not all relatedness is created equal. *Child Development, 77*, 1112-1123.
- Howe, M.L., Cichetti, D., Sheree, L.T. & Cerrito, B.M. (2004). True and false memories in maltreated children. *Child development, 75*, 1402-1417.
- Hunt, R.R. & McDaniel, M.A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421-445.

- Hupbach, A. (2000). *Alterskorrelierte Differenzen in konzeptuellen impliziten Gedächtnisleistungen: Eine experimentelle Studie an Kindern*. Unveröffentlichte Dissertation, Universität Trier.
- Hüsken, A. (2004). "False memories"- Wie empfänglich sind Kinder für falsche Erinnerungen? Unveröffentlichte Diplomarbeit, Universität Trier.
- Hutchison, K.A. & Balota, D.A. (2005). Decoupling semantic and associative information in false memories : Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language*, 52, (1-28).
- Israel, L. & Schacter, D.L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, 4, 577-581.
- Jacoby, L.L., Bishara, A.J., Hessels, S & Toth, J.P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General*, 134, 131-148.
- Johnson, M.K., Hashtroudi, S. & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.
- Johnson, M.K. & Raye, C.L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Kawasaki, Y. & Yama, H. (2006). The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory*, 14, 68-78.
- Keniston, A.H. & Flavell, J.H. (1979). A developmental study of intelligent retrieval. *Child Development*, 50, 1144-1152.
- Kimball, D.R. & Bjork, R.A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General*, 131, 116-130.
- Kreutzer, M.A., Leonard, C. & Flavell, J.H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development*, 40, Serial No. 159, 1-58.
- Lampinen, J.M., Leding, J.K., Reed, K.B. & Odegard, T.N. (2006). Global gist extraction in children and adults. *Memory*, 14, 952-964.
- Lampinen, J.M., Meier, C.R., Arnal, J.D. & Leding, J.K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 954-963.
- Lampinen, J.M., Neuschatz, J.S. & Payne, D.G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*, 6, 130-135.

- Leichtman, M.D. & Ceci, S.J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology, 31*, 568-578.
- Liben, L.S. (1982). The developmental study of children's memory. In T.M. Field, A. Huston, H.C. Quay, T. Troll & G.E. Finley (Eds.), *Review of Human Development* (pp. 269-284). New York: Wiley.
- Lindsay, D.S., Johnson, M.K. & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*, 297-318.
- Lockhart, R.S. (2000). Methods of memory research. *Current Directions in Psychological Science, 9*, 45-57.
- Loftus, E.F. (1993). The reality of repressed memories. *American Psychologist, 48*, 518-537.
- Marsh, E.J. & Bower, G.H. (2004). The role of rehearsal and generation in false memory creation. *Memory, 12*, 748-761.
- McBride, D.M., Coane, J.H. & Raulerson, B.A., III. (2006). An investigation of false memory in perceptual implicit tasks. *Acta Psychologica, 123*, 240-260.
- McCabe, D.P. & Smith, A.D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition, 30*, 1065-1077.
- McDermott, K.B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*, 212-230.
- McDermott, K.B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review, 4*, 582-586.
- McDermott, K.B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*, 261-267.
- McDermott, K.B. & Roediger, H.L., III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*, 508-520.
- McDermott, K.B. & Watson, J.M. (2001). The rise and fall of false recall: the impact of presentation duration. *Journal of Memory and Language, 45*, 160-176.
- McEvoy, C.L., Nelson, D.L. & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1177-1194.
- McKone, E. & Murphy, B. (2000). Implicit false memory: Effects of modality and multiple study presentations on long-lived semantic priming. *Journal of Memory and Language, 43*, 89-109.

- Meade, M.L., Watson, J.M., Balota, D.A. & Roediger, H.L., III. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language*, 56, 305-320.
- Miguelles, M. & Garcia-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology*, 13, 257-268.
- Miller, G.A. (1956). The magical number seven plus or minus two : Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, M.B. & Wolford, G.L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, 106, 398-405.
- Multhaup, K.S. & Conner, C.A. (2002). The effects of considering nonlist sources on the Deese-Roediger-McDermott memory illusion. *Journal of Memory and Language*, 47, 214-228.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Appeltion-Century-Crofts.
- Neuschatz, J.S., Benoit, G.E. & Pane, D.G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: The role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 35-41.
- Norman, K. & Schacter, D.L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, 25, 838-848.
- Oberauer, K., Wilhelm, O., Schulze, R & Süß, H.M. (2005). Working memory and intelligence- their correlation and their relation: Comment on Ackerman, Baier, and Boyle (2005). *Psychological Bulletin*, 131, 61-65.
- Odegard, T.M. & Lampinen, J.M. (2005). Recollection rejection: Gist cuing of verbatim memory. *Memory & Cognition*, 33, 1422-1430.
- Opwis, K., Gold, A, Gruber, H. & Schneider, W. (1990). Zum Einfluss von Expertise auf Gedächtnisleistungen sowie deren Selbsteinschätzungen bei Kindern und Erwachsenen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22, 207-224.
- Ornstein, P.A., Baker-Ward, L. & Naus, M.J. (1988). The development of mnemonic skill. In F.E. Weinert & M. Perlmutter (Eds.), *Memory development: Universal changes and individual differences* (pp. 31-50). Hillsdale, NJ: Erlbaum.
- Ornstein, P.A., Naus, M.J. & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child development*, 46, 818-830.
- Paris, S.G. (1978). The development of inference and transformation as memory operations. In P.A. Ornstein (Ed.), *Memory and development in children* (pp. 129-156). Hillsdale, NJ: Erlbaum.

- Paris, S.G. & Lindauer, B.K. (1976). The role of inference in children's comprehension and memory for sentence. *Cognitive Psychology*, 8, 217-227.
- Parkin, A.J. (1997). The development of procedural and declarative memory. In N. Cowan (Ed.), *The Development of Memory in Childhood* (pp. 113-138). Hove: Psychology Press.
- Payne, D.G., Elie, C.J., Blackwell, J.M. & Neuschatz, J.S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, 35, 261-285.
- Pesta, B.J., Murphy, M.D. & Sanders, R.E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 328-338.
- Piaget, J. & Inhelder, B. (1973). *Memory and Intelligence*. New York: Basic Books.
- Pressley, M. (1982). Elaboration and memory development. *Child development*, 53, 296-309.
- Rabinowitz, M. & Chi, M.T.H. (1978). An interactive model of strategic processing. In S.J. Ceci (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities* (Vol. 2, pp. 83-102). Hillsdale, NJ: Erlbaum.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, 21, 89-102.
- Raven, J. (1976). *Raven Standard Progressive Matrices (SPM)*. Oxford: Oxford Psychologists Press. Deutsche Überarbeitung: Heller, K.A., Kratzmeier, H. & Lengfelder, A. (1998). *Matrizen-Test Manua, Band 1. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices von J.C. Raven*. Weinheim: Beltz.
- Read, J.D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, 3, 105-111.
- Reder, L.M., Nhouyvanisvong, A., Schunn, C.D., Ayers, M.S., Angstadt, P. & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgement in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 294-320.
- Reed, A.V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, 181, 574-576.
- Reyna, V.F. (1992). Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In M.L. Howe, C.L. Brainerd & V.F. Reyna (Eds.), *Development of long-term retention* (pp. 103-127). New York: Springer-Verlag.
- Reyna, V.F. (1995). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F.N. Dempster & J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 29-61). San Diego, CA: Academic Press.

- Reyna, V.F. & Brainerd, C.J. (1995). Fuzzy-trace theory: An interitem synthesis. *Learning and Individual Differences*, 7, 1-75.
- Reyna, V.F. & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions and retention interval. *Developmental Psychology*, 30, 178-191.
- Rhodes, M.G. & Anastasi, J.S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review*, 7, 158-162.
- Robinson, K.J. & Roediger, H.L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, 8, 231-237.
- Roediger, H.L., III (1996). Memory illusions. *Journal of Memory and Language*, 35, 76-100.
- Roediger, H.L., III & McDermott, K.B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 8, pp. 63-131).
- Roediger, H.L., III & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803-814.
- Roediger, H.L., III & McDermott, K.B. (1999). False alarms about false memories. *Psychological Review*, 106, 406-410.
- Roediger, H.L., III & McDermott, K.B. (2000). Tricks of memory. *Current Directions in Psychological Science*, 9, 123-127.
- Roediger, H.L., III, McDermott, K.B., & Goff, L.M. (1997). Recovery of true and false memories: Paradoxical effects of repeated testing. In M.A. Conway (Ed.), *Recovered memories and false memories* (pp. 118-149). Oxford, Oxford University Press.
- Roediger, H.L., III, McDermott, K.B., & Robinson, K.J. (1998). The role of associative processes in creating false memories. In M.A. Conway, S.E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory, Vol. II* (pp. 187-245). Hove, U.K.: Psychology Press.
- Roediger, H.L., III, Watson J.M., McDermott, K.B., & Gallo, D.A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychological Bulletin & Review*, 8, 308-407.
- Russel, W.A. & Jenkins, J.J. (1954). The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word-Association Test. *Technical Report*, 11. University of Minnesota, Office of Naval Research.
- Schacter, D.L. (1999). *Wir sind Erinnerung*. Reinbek: Rowohlt-Verlag.
- Schacter, D.L., Israel, L. & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, 40, 1-24.

- Schacter, D.L., Verfaillie, M. & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language*, 35, 319-334.
- Schneider, W. (1985). Developmental trends in the metamemory-memory behaviour relationship: An integrative review. In D.L. Forrest-Pressley, G.E. McKinnon & T.G. Waller (Eds.), *Cognition, metacognition, and human performance* (Vol. I, pp. 57-109). New York: Academic Press.
- Schneider, W. (1989). *Zur Entwicklung des Meta-Gedächtnisses bei Kindern*. Bern: Huber.
- Schneider, W. & Bjorklund, D.F. (1998). Memory. In W. Damon (General Ed.), *Handbook of child psychology* (5th ed.), Vol. 2: Cognition, perception and language. New York: Wiley.
- Schneider, W. & Büttner, G. (1998). Entwicklung des Gedächtnisses. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie, 4. Aufl.* (S. 654-704). Weinheim: Psychologie Verlags Union.
- Schneider, W. & Büttner, G. (2002). Entwicklung des Gedächtnisses. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie, 5. Aufl.* (S. 495-516). Weinheim: Psychologie Verlags Union.
- Seamon, J.G., Goodkind, M.S., Dumey, A.D., Dick, E., Aufsesser, M.S., Strickland, S.E., Woulfin, F.R., & Fung, N.S. (2003). "If I didn't write it, why would I remember it?" Effects of encoding, attention, and practice on accurate and false memory. *Memory & Cognition*, 31, 445-457.
- Seamon, J.G., Lee, I.A., Tonner, S.K., Wheeler, R.H., Goodkind, M.S. & Birch, A.D. (2002a). Thinking of critical words during study is unnecessary for false memories in the Deese, Roediger, and McDermott procedure. *Psychological Science*, 13, 526-531.
- Seamon, J.G., Luo, C.R. & Gallo, D.A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, 9, 20-26.
- Seamon, J.G., Luo, C.R., Kopecky, J.J., Price, C.A., Rothschild, L., Fung, N.S. & Schwartz, M.A. (2002b). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition*, 30, 1054-1064.
- Seamon, J.G., Luo, C.R., Schlegel, S.E., Greene, S.E. & Goldenberg, A.B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language*, 42, 120-146.

- Seamon, J.G., Luo, C.R., Schwartz, M.A., Jones, K.J., Lee, D.M. & Jones, S.J. (2002c). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language*, 46, 323-340.
- Smith, R.E. & Hunt, R.R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, 5, 710-715.
- Smith, S.M., Ward, T.B., Tindell, D.R., Sifonis, C.M. & Wilkenfeld, M.J. (2000). Category structure and created memories. *Memory & Cognition*, 28, 386-395.
- Snodgras, J.G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.
- Soraci, S.A., Carlin, M.T., Toglia, M.P., Chechile, R.A. & Neuschatz, J.S. (2003). Generative processing and false memories: When there is no cost. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 511-523.
- Stadler, M.A., Roediger, H.L., III, & McDermott, K.B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, 27, 494-500.
- Starns, J.J. & Hicks, J.L. (2004). Episodic generation can cause semantic forgetting: Retrieval induced forgetting of false memories. *Memory & Cognition*, 32, 602-609.
- Steffens, M.C. & Mecklenbräuker, S. (2007). False memories: Phenomena, theories, and implication. *Zeitschrift für Psychologie/Journal of Psychology*, 215, 12-24.
- Sugrue, K. & Hayne, H. (2006). False memories produced by children and adults in the DRM Paradigm. *Applied Cognitive Psychology*, 20, 625-631.
- Thapar, A. & McDermott, K.B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention intervall and levels of processing. *Memory & Cognition*, 29, 424-432.
- Thomas, A.K. & Sommers, M.S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language*, 52, 71-86.
- Toglia, M.P., Neuschatz, J.S. & Goodwin, K.A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, 7, 233-256.
- Tse, C.-S. & Neely, J.H. (2005). Assessing activation without source monitoring in the DRM false memory paradigm. *Journal of Memory and Language*, 53, 532-550.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, 26, 1-12.
- Underwood, B.J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 70, 122-129.

- Watson, J.M., Bunting, M.F., Poole, B.J. & Conway, A.R.A (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 76-85.
- Wickens, T.D. & Hirshman, E. (2000). False memories and statistical decision theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review*, 107, 377-383.
- Wippich, W. (1984). *Lehrbuch der angewandten Gedächtnispsychologie*. Stuttgart: Kohlhammer.
- Wixted, J.T. & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, 107 (2), 368-376.
- Wulf, F. (1922). Über Veränderungen von Vorstellungen. *Psychologische Forschung*, 1, 333-373.
- Zeelenberg, R., Plomp, G. & Raaijmakers, J.G.W. (2003). Can false memories be created through nonconscious processes? *Consciousness and Cognition*, 12, 403-412.
- Zeelenberg, R., Shiffrin, R.M. & Raaijmakers, J.G.W. (1999). Priming in a free association task as a function of association directionality. *Memory & Cognition*, 27, 956-961.
- Zelazo, P., Carter, S., Resnick, K. & Frye, C. (1996). The development of executive functions in children. *Review of General Psychology*, 1, 198-226.

Anhang

Anhang A: Experiment 1

Anhang B: Experiment 2

Anhang C: Experimente 3A und 3B

Anhang D: Experiment 4

Anhang E: Eltern- und Schülerinformationsbrief

Anhang A: Experiment 1

A1 Material

Lernphase: DRM-Listensets A, B, C, D mit oder ohne kritische Items

Kritische Items	Listenitems
<i>Liste A</i>	
Hase	Ohren, Ostern, Möhre, (<i>Hase</i>), hoppeln, Igel, Kaninchen, Fuchs, Braten, Feld, Jäger
Auto	fahren, Straße, Reifen, (<i>Auto</i>), Lenkrad, Bus, Führerschein, Benzin, Porsche, Motor, Unfall
Pferd	reiten, Sattel, galoppieren, (<i>Pferd</i>), Zügel, wiehern, Hengst, Fohlen, Weide, Cowboy, Kutsche
Hammer	Nagel, Säge, Daumen, (<i>Hammer</i>), Zange, Lärm, Werkzeug, Holz, Wand, Arbeit, schlagen
Bett	Kopfkissen, Decke, Schlafanzug, (<i>Bett</i>), träumen, Matratze, Nacht, schlafen, müde, Kuscheltier, Möbel
<i>Liste B</i>	
Hund	bellen, Katze, Leine, (<i>Hund</i>), Fell, beißen, Schnauze, Schwanz, Dackel, Gassi, Welpen
Fuß	Schuh, Zehen, Bein, (<i>Fuß</i>), laufen, gehen, Socken, Hand, Sohle, Ball, Strumpf
Löwe	Mähne, Tiger, Raubtier, (<i>Löwe</i>), Afrika, Zoo, brüllen, Tatze, Fleisch, König, gefährlich
Telefon	klingseln, sprechen, Hörer, (<i>Telefon</i>), anrufen, wählen, Handy, Tasten, Kabel, Leitung, Rechnung
Sommer	Sonne, Freibad, Eis, (<i>Sommer</i>), Frühling, Ferien, Winter, Sonnencreme, Strand, Sandalen, warm
<i>Liste C</i>	
Kuh	Milch, Kalb, Wiese, (<i>Kuh</i>), Euter, Flecken, Stier, melken, muhen, Herde, Stall
Nadel	nähen, Faden, pieksen, (<i>Nadel</i>), spitz, stechen, Öhr, Knopf, Garn, Schneider, Spritze
Vogel	Nest, fliegen, Amsel, (<i>Vogel</i>), Federn, Adler, Spatz, Schnabel, Papagei, zwitschern, Baum
König	Krone, Palast, Schloss, (<i>König</i>), Diener, Prinzessin, Macht, Zep-ter, reich, Thron, Kaiser
Fenster	Glas, putzen, Licht, (<i>Fenster</i>), Rahmen, Gardine, Tür, Fliegengit-ter, hinausschauen, Rollladen, Vorhang

Liste D

Schlange	Gift, Zunge, Biss, (<i>Schlange</i>), Kobra, Dschungel, Reptil, Wüste, Gebüsch, Muster, Leder
Mädchen	Junge, klein, Kleid, (<i>Mädchen</i>), jung, Frau, Kind, Barbie, rosa, Rock, Puppe
Kerze	Wachs, Leuchter, Docht, (<i>Kerze</i>), Weihnachten, anzünden, Streichhölzer, Biene, brennen, Advent, Kirche
Brot	essen, Butter, Käse, (<i>Brot</i>), Bäcker, Messer, Mehl, Nutella, Brötchen, Marmelade, Krümel
Hemd	Hose, bügeln, Kragen, (<i>Hemd</i>), zuknöpfen, Anzug, Ärmel, Mann, Kleidung, kariert, schick

Wiederlernphase (auf Band gesprochene Wörter)

-
1. singen (Füllitem)
 2. Schiff (Füllitem)
 3. Pferd (Liste A)
 4. Hammer (Liste A)
 5. Fuß (Liste B)
 7. Auto (Liste A)
 8. Löwe (Liste B)
 9. Bett (Liste A)
 10. Telefon (Liste B)
 11. Hase (Liste A)
 12. Sommer (Liste B)
-

Testphase: Wiedererkennenstest 1

Zu beurteilende Items
(in Klammern Itemart und welcher Liste zugehörig)

1. Hund (kritisches Item)	16. Auto (kritisches Item)
2. Motor (Listenitem, <i>Auto</i>)	17. Säge (Listenitem, <i>Hammer</i>)
3. Holz (Listenitem, <i>Hammer</i>)	18. Socken (Listenitem, <i>Fuß</i>)
4. Hase (kritisches Item)	19. Telefon (kritisches Item)
5. laufen (Listenitem, <i>Fuß</i>)	20. Möhre (Listenitem, <i>Hase</i>)
6. Dackel (Listenitem, <i>Hund</i>)	21. Tatze (Listenitem, <i>Löwe</i>)
7. Afrika (Listenitem, <i>Löwe</i>)	22. Sommer (kritisches Item)
8. Hammer (kritisches Item)	23. Nacht (Listenitem, <i>Bett</i>)
9. träumen (Listenitem, <i>Bett</i>)	24. Fuß (kritisches Item)
10. Bus (Listenitem, <i>Auto</i>)	25. Sattel (Listenitem, <i>Pferd</i>)
11. Löwe (kritisches Item)	26. Schnauze (Listenitem, <i>Hund</i>)
12. Handy (Listenitem, <i>Telefon</i>)	27. Ostern (Listenitem, <i>Hase</i>)
13. Strand (Listenitem, <i>Sommer</i>)	28. Ferien (Listenitem, <i>Sommer</i>)
14. Bett (kritisches Item)	29. Pferd (kritisches Item)
15. Zügel (Listenitem, <i>Pferd</i>)	30. Hörer (Listenitem, <i>Telefon</i>)

Wiedererkennenstest 2

Zu beurteilende Items
(in Klammern Itemart und welcher Liste zugehörig)

1. Hemd (kritisches Item)	16. zwitschern (Listenitem, <i>Vogel</i>)
2. Amsel (Listenitem, <i>Vogel</i>)	17. Käse (Listenitem, <i>Brot</i>)
3. Kuh (kritisches Item)	18. Fenster (kritisches Item)
4. Palast (Listenitem, <i>König</i>)	19. Anzug (Listenitem, <i>Hemd</i>)
5. Mädchen (kritisches Item)	20. Bäcker (Listenitem, <i>Brot</i>)
6. Garn (Listenitem, <i>Nadel</i>)	21. Schlange (kritisches Item)
7. Kobra (Listenitem, <i>Schlange</i>)	22. melken (Listenitem, <i>Kuh</i>)
8. Kerze (kritisches Item)	23. Leuchter (Listenitem, <i>Kerze</i>)
9. Stier (Listenitem, <i>Kuh</i>)	24. stechen (Listenitem, <i>Nadel</i>)
10. Kleid (Listenitem, <i>Mädchen</i>)	25. König (kritisches Item)
11. Nadel (kritisches Item)	26. Kragen (Listenitem, <i>Hemd</i>)
12. Zepter (Listenitem, <i>König</i>)	27. Kind (Listenitem, <i>Mädchen</i>)
13. Brot (kritisches Item)	28. Wüste (Listenitem, <i>Schlange</i>)
14. Rollladen (Listenitem, <i>Fenster</i>)	29. Vogel (kritisches Item)
15. brennen (Listenitem, <i>Kerze</i>)	30. Gardine (Listenitem, <i>Fenster</i>)

Wortassoziationstest

Cue	Zielitem
1. Krawatte	Hemd
2. Wurm	Vogel
3. öffnen	Fenster
4. Zöpfe	Mädchen
5. Flamme	Kerze
6. Bauer	Kuh
7. Urwald	Schlange
8. stricken	Nadel
9. Körner	Brot
10. Herrscher	König

A2 Instruktionen

Vorschüler und Drittklässler

Lernphase 1

Ich möchte gerne mit dir ein Wortspiel spielen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. Wenn ich mit dem Vorlesen aller Listen fertig bin, sage ich dir wie es weitergeht.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Wiederlernphase und Ablenkaufgabe 1

Hier habe ich eine Kasette, auf die jemand Wörter gesprochen hat. Jetzt lege ich die Kasette in das Tonbandgerät ein. Du sollst sehr gut zuhören und dir die Wörter, die du nun hörst, merken. Wenn die Kasette zu Ende gelaufen ist, sage ich dir drei (*Drittklässler: vier*) Zahlen und du sollst sie so schnell wie möglich wiederholen. Danach kommen drei (*Drittklässler: vier*) neue Zahlen, die du auch wieder so schnell wie möglich wiederholen sollst. Und dann wieder neue Zahlen.

Freie Reproduktion

Jetzt sage mir möglichst viele Wörter, die du eben auf der Kasette gehört hast. Weißt du noch Wörter?

Ablenkaufgabe 2

Jetzt sage ich dir wieder Zahlen, die du wiederholen sollst.

Wiedererkennenstest 1

Nun frage ich dich nach den Wörtern, die ich dir ganz am Anfang vorgelesen habe. Es geht aber nicht um die Wörter, die du auf der Kasette gehört hast. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es am Anfang schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Wenn ich dir das Wort vorgelesen habe, dann sage „alt“. Wenn ich dir das Wort nicht vorgelesen habe, dann sage „neu“.

Ablenkaufgabe 3

Nun wollen wir ein anderes Spiel anfangen. Hier sind bunte Mikadohölzer...

Lernphase 2

Ich lese dir nacheinander andere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. Wenn ich mit dem Vorlesen dieser Listen fertig bin, sage ich dir wie es weitergeht.

Wortassoziationstest

Jetzt lese ich dir Wörter vor und du sollst mir sagen was dir dazu einfällt. Das sollst du ganz schnell machen, ohne zu lange nachzudenken. Es gibt keine falschen Antworten. Alles was dir zu dem Wort einfällt ist gut. Ich zeige dir, wie das Wortspiel geht: Wenn ich jetzt zum Beispiel „Sport“ sage, welche Wörter fallen dir ein?

Zum Beispiel kannst du „Fußball“ oder „laufen“ sagen. Und welche Wörter fallen dir ein, wenn ich „Fluss“ sage?

Zum Beispiel kannst du „Wasser“ oder „nass“ sagen. Hast du das Spiel nun verstanden? Jetzt sage ich dir andere Wörter.

Ablenkaufgabe 4

Jetzt sage ich dir wieder Zahlen, die du wiederholen sollst.

Wiedererkennenstest 2

Jetzt frage ich dich nach den Wörtern, die ich dir vor deinen genannten Wörtern vorgelesen habe. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es am Anfang schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Wenn ich dir das Wort vorgelesen habe, dann sage „alt“. Wenn ich dir das Wort nicht vorgelesen habe, dann sage „neu“.

Erwachsene

Lernphase 1

Ich möchte gerne mit dir ein Gedächtnisexperiment durchführen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. Wenn ich mit dem Vorlesen aller Listen fertig bin, sage ich dir wie es weitergeht.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Wiederlernphase und Ablenkaufgabe 1

Hier habe ich eine Kassette, auf der Wörter aufgenommen sind. Jetzt lege ich die Kassette in das Tonbandgerät ein. Du sollst sehr gut zuhören und dir die Wörter, die du nun hörst merken. Wenn die Kassette zu Ende gelaufen ist, sage ich dir fünf Zahlen und du sollst sie so schnell wie möglich wiederholen. Danach kommen dann fünf neue Zahlen, die du auch wieder so schnell wie möglich wiederholen sollst. Und dann wieder neue Zahlen.

Freie Reproduktion

Jetzt sage mir möglichst viele Wörter, die du eben auf der Kassette gehört hast. Weißt du noch Wörter?

Ablenkaufgabe 2

Jetzt sage ich dir wieder Zahlen, die du wiederholen sollst.

Wiedererkennenstest 1

Nun frage ich dich nach den Wörtern, die ich dir ganz am Anfang vorgelesen habe. Es geht aber nicht um die Wörter, die du auf der Kassette gehört hast. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es am Anfang schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Wenn ich dir das Wort vorgelesen habe, dann sage „alt“. Wenn ich dir das Wort nicht vorgelesen habe, dann sage „neu“.

Ablenkaufgabe 3: Standardform des Progressiven Matrizen Tests (Raven, 1976)

Nun wollen wir uns etwas anderes anschauen. Hier habe ich ein Muster das unvollständig ist. Es gibt immer verschiedene Stücke, die in die Lücke passen würden, aber nur eines ist richtig. Finde bitte das passende Stück.

Lernphase 2

Ich lese dir nacheinander andere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. Wenn ich mit dem Vorlesen dieser Listen fertig bin, sage ich dir wie es weitergeht.

Wortassoziationstest

Jetzt lese ich dir Wörter vor und du sollst mir sagen was dir dazu einfällt. Das sollst du ganz schnell machen, ohne zu lange nachzudenken. Es gibt keine falschen Antworten. Alles was dir zu dem Wort einfällt ist gut. Ich zeige dir an einem Beispiel wie es geht: Wenn ich jetzt zum Beispiel „Sport“ sage, welche Wörter fallen dir ein?

Zum Beispiel kannst du „Fußball“ oder „laufen“ sagen. Und welche Wörter fallen dir ein, wenn ich „Fluss“ sage?

Zum Beispiel kannst du „Wasser“ oder „nass“ sagen. Hast du es nun verstanden? Jetzt sage ich dir andere Wörter.

Ablenkaufgabe 4

Jetzt sage ich dir wieder Zahlen, die du wiederholen sollst.

Wiedererkennenstest 2

Jetzt frage ich dich nach den Wörtern, die ich dir vor deinen genannten Wörtern vorgelesen habe. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es am Anfang schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Wenn ich dir das Wort vorgelesen habe, dann sage „alt“. Wenn ich dir das Wort nicht vorgelesen habe, dann sage „neu“.

Anhang B: Experiment 2

B1 Material

Lernphasen: 4 Lernbedingungen der Listensets A und B mit 7, 10 oder 14 Listenitems

- (1) A Inklusion, B Standard; 7 bzw. 10 Listenitems
- (2) A Standard, B Inklusion; 7 bzw. 10 Listenitems
- (3) B Inklusion, A Standard; 10 bzw. 14 Listenitems
- (4) A Standard, B Inklusion; 10 bzw. 14 Listenitems

Kritische Items	Listenitems
Vorschüler: 7 Listenitems; Drittklässler und Erwachsene 10 Listenitems (<u>unterstrichen: zusätzliche Items für Drittklässler und Erwachsene</u>)	
Hase (A)	Ohren, Ostern, Möhre, hoppeln, Igel, Kaninchen, Fuchs, <u>Braten, Feld, Jäger</u>
Auto (A)	fahren, Straße, Reifen, Lenkrad, Bus, Führerschein, Benzin, <u>Porsche, Motor, Unfall</u>
Hund (B)	bellen, Katze, Leine, Fell, beißen, Schnauze, Schwanz, <u>Dackel, Gassi, Welpen</u>
Brot (A)	essen, Butter, Käse, Bäcker, Messer, Mehl, Frühstück, <u>Marmelade, Nutella, Körner</u>
Fenster (B)	Glas, putzen, Licht, Rahmen, Gardine, öffnen, Tür, <u>Ausblick, hinausschauen, Rollläden</u>
Löwe (B)	Mähne, Tiger, Raubtier, Afrika, Zoo, brüllen, Tatze, <u>Fleisch, König, gefährlich</u>
Pferd (A)	reiten, Sattel, Galopp, Zügel, wiehern, Hengst, Fohlen, <u>Weide, Koppel, Nüstern</u>
Fuß (B)	Schuh, Zehen, Bein, laufen, gehen, Socken, Hand, <u>Sohle, Ball, Strumpf</u>
Vogel (B)	fliegen, Nest, Federn, Amsel, Flügel, Spatz, Adler, <u>Wurm, Papagei, Schnabel</u>
Sommer (A)	Sonne, Freibad, Eis, Frühling, Ferien, Winter, Sonnencreme, <u>Strand, Sandalen, warm</u>
Bett (B)	Kopfkissen, Decke, Schlafanzug, träumen, Matratze, Nacht, schlafen, <u>müde, Kuscheltier, Möbel</u>
Nadel (A)	nähen, Faden, pieksen, spitz, Knopf, stechen, stricken, <u>Garn, Heuhaufen, Spritze</u>
Kerze (A)	Wachs, Leuchter, Flamme, Weihnachten, Biene, Kirche, anzünden, <u>Streichhölzer, Docht, Advent</u>
Telefon (B)	klingseln, sprechen, Hörer, anrufen, wählen, Handy, Tasten, <u>Kabel, Leitung, Rechnung</u>

Hemd (B)	Hose, Anzug, bügeln, Krawatte, Kragen, gestreift, Mann, <u>Ärmel, Kleidung, kariert</u>
Hammer (A)	Nagel, Säge, Daumen, Zange, Lärm, Werkzeug, Holz, <u>Wand, Arbeit, schlagen</u>

Vorschüler: 10 Listenitems; Drittklässler und Erwachsene: 14 Items (unterstrichen: zusätzliche Items für Drittklässler und Erwachsene)

Hase (A)	Ohren, Ostern, Möhre, hoppeln, Igel, Kaninchen, Fuchs, Braten, Feld, Jäger, <u>Bau, Käfig, Wald, Stall</u>
Auto (A)	fahren, Straße, Reifen, Lenkrad, Bus, Führerschein, Benzin, Porsche, Motor, Unfall, <u>Abgas, Hupe, Ampel, tanken</u>
Hund (B)	bellen, Katze, Leine, Fell, beißen, Schnauze, Schwanz, Dackel, Gassi, Welpen, <u>Knochen, Pudel, sabbern, Freund</u>
Brot (A)	essen, Butter, Käse, Bäcker, Messer, Mehl, Frühstück, Marmelade, Nutella, Körner, <u>Scheibe, Wurst, Roggen, Korb</u>
Fenster (B)	Glas, putzen, Licht, Rahmen, Gardine, öffnen, Tür, Ausblick, hinausschauen, Rollladen, <u>durchsichtig, Vorhang, Luft, Haus</u>
Löwe (B)	Mähne, Tiger, Raubtier, Afrika, Zoo, brüllen, Tatze, Fleisch, König, gefährlich, <u>Savanne, Zirkus, Zähne, Safari</u>
Pferd (A)	reiten, Sattel, Galopp, Zügel, wiehern, Hengst, Fohlen, Weide, Koppel, Nüstern, <u>Kutsche, Cowboy, Stiefel, Pony</u>
Fuß (B)	Schuh, Zehen, Bein, laufen, gehen, Socken, Hand, Sohle, Ball, Strumpf, <u>Gelenk, Pilz, verstauchen, stinken</u>
Vogel (B)	fliegen, Nest, Federn, Amsel, Flügel, Spatz, Adler, Wurm, Papagei, Schnabel, <u>Baum, Eier, zwitschern, Taube</u>
Sommer (A)	Sonne, Freibad, Eis, Frühling, Ferien, Winter, Sonnencreme, Strand, Sandalen, warm, <u>schwitzen, Urlaub, hell, Palmen</u>
Bett (B)	Kopfkissen, Decke, Schlafanzug, träumen, Matratze, Nacht, schlafen, müde, Kuscheltier, Möbel, <u>schnarchen, Ruhe, erholen, Sandmännchen</u>
Nadel (A)	nähen, Faden, pieksen, spitz, Knopf, stechen, stricken, Garn, Heuhaufen, Spritze, <u>Öhr, stopfen, Loch, Blut</u>
Kerze (A)	Wachs, Leuchter, Flamme, Weihnachten, Biene, Kirche, anzünden, Streichhölzer, Docht, Advent, <u>brennen, Schein, romantisch, gemütlich</u>
Telefon (B)	klingeln, sprechen, Hörer, anrufen, wählen, Handy, Tasten, Kabel, Leitung, Rechnung, <u>Nummer, Telekom, Kontakt, Gespräch</u>
Hemd (B)	Hose, Anzug, bügeln, Krawatte, Kragen, gestreift, Mann, Ärmel, Kleidung, kariert, <u>elegant, Jackett, Baumwolle, Seide</u>
Hammer (A)	Nagel, Säge, Daumen, Zange, Lärm, Werkzeug, Holz, Wand, Arbeit, schlagen, <u>Handwerker, klopfen, Bild, verletzen</u>

B2 Instruktionen

Vorschüler und Drittklässler

Lernphase

Ich möchte gerne mit dir ein Wortspiel machen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. Wenn ich mit einer Liste fertig bin, sage ich dir wie es weitergeht.

Bevor ich mit dem Vorlesen anfangen, stelle ich dir noch meine zwei schönen Puppen vor.

Das hier ist die Prinzessin Anna. Sie denkt immer ganz genau nach und sagt nur die Wörter, die ich vorgelesen habe.

Wenn ich dir Prinzessin Anna zeige, sollst du wie sie nachdenken und nur die Wörter sagen, bei denen du ganz sicher bist, dass ich sie dir auch vorgelesen habe. Du sollst nicht raten.

Das hier ist der Kasper. Er hört auch genau zu und sagt die Wörter, die ich vorgelesen habe. Ich lese z.B. „Gruppe, Kinder, spielen, singen“ vor und du sagst wie der Kasper...

Aber er sagt noch „Kindergarten“, denn er hat, als ich die Wörter vorgelesen habe, ganz stark an Kindergarten gedacht. Im Kindergarten gibt es verschiedene Gruppen, viele Kinder sind da, sie spielen und singen. Aber der Kasper sagt nicht „Frisör“, denn daran hat er nicht gedacht als ich die Wörter gesagt habe.

Immer wenn ich dir den Kasper zeige, sollst du wie er die Wörter sagen, die ich vorgelesen habe und noch die Wörter, an die du selber stark gedacht hast, als ich die Liste vorgelesen habe.

Nach jeder Liste zeige ich dir entweder die Prinzessin Anna oder den Kasper und dann sage ich dir, ob du nur wie Prinzessin Anna die Wörter sagen sollst, bei denen du ganz sicher bist, dass ich sie dir vorgelesen habe oder ob du wie der Kasper auch noch die Wörter sagen sollst, an die du selber stark gedacht hast, als ich die Liste vorgelesen habe.

Alles klar? Dann fange ich jetzt an und lese die Wörter vor.

Testphase (Testphase 1 für Drittklässler): Freie Reproduktion

Nun frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe.

Standardinstruktion: Du sollst wie Prinzessin Anna antworten und nur die Wörter sagen, bei denen du ganz sicher bist, dass ich sie dir auch vorgelesen habe. Du sollst nicht raten.

(Inklusionsinstruktion: Du sollst wie der Kasper antworten und die Wörter sagen, die ich vorgelesen habe und noch die Wörter, an die du selber stark gedacht hast, als ich die Liste vorgelesen habe.)

Erwachsene

Lernphase

Ich möchte mit dir ein Gedächtnisexperiment durchführen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor, die du dir gut einprägen sollst. Wenn ich mit einer Liste fertig bin, sage ich dir wie es weitergeht. Es gibt zwei verschiedene Anweisungen nach denen du die Wörter erinnern sollst.

Wenn ich die Liste vorgelesen habe sollst du gemäß Anweisung 1 genau nachdenken und nur die Wörter sagen, bei denen du ganz sicher bist, dass ich sie dir auch vorgelesen habe. Du sollst nicht raten.

Wenn ich dir nach der Liste Anweisung 2 gebe, sollst du ebenfalls alle Wörter sagen, bei denen du ganz sicher bist, dass ich sie dir vorgelesen habe. Zusätzlich dazu sollst du die Wörter nennen, an die du selber stark gedacht hast als ich die Liste vorgelesen habe. Zum Beispiel lese ich „Silber, Schmuck, wertvoll, reich, teuer, Geld, glänzen, Mine, Barren, Bronze“ vor. Nun sollst du diese Wörter erinnern und auch noch Wörter sagen an die du selber stark gedacht hast. Im Beispiel kannst du etwa „Gold“ sagen, wenn du daran stark gedacht hast.

Nach jeder Liste sage ich dir, ob du entsprechend Anweisung 1 oder 2 antworten sollst und demzufolge nur die Wörter sagen sollst bei denen du ganz sicher bist, dass ich sie dir vorgelesen habe oder ob du auch noch die Wörter sagen sollst an die du selber stark gedacht hast, als ich die Liste vorgelesen habe. Alles klar? Dann fange ich jetzt an und lese die Wörter vor.

Testphase 1: Freie Reproduktion

Nun frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe.

Standardinstruktion: Du sollst gemäß Anweisung 1 antworten und nur die Wörter sagen bei denen du ganz sicher bist, dass ich sie dir auch vorgelesen habe. Du sollst nicht raten.

(Inklusionsinstruktion: Du sollst gemäß Anweisung 2 antworten und die Wörter sagen, die ich sicher vorgelesen habe und noch die Wörter an die du selber stark gedacht hast, als ich die Liste vorgelesen habe.)

Drittklässler und Erwachsene

Testphase 2: Quellendiskrimination

Jetzt lese ich dir Liste für Liste alle Wörter vor die du mir gesagt hast. Du sollst dich nun entscheiden, ob jedes von dir genannte Wort tatsächlich von mir vorgelesen wurde oder ob du selber daran gedacht hast.

Wenn ich dir deine genannten Wörter vorlese frage ich dich „Habe ich das Wort gesagt?“ und du antwortest mit „ja“ oder „nein“.

Anhang C: Experimente 3A und 3B

C1 Material (Experimente 3A und 3B)

Lernphasen: 4 Lernbedingungen der Listensets A und B:

- (1) A Inklusion, B Standard
- (2) B Standard, A Inklusion
- (3) B Inklusion, A Standard
- (4) A Standard, B Inklusion

Kritische Items	Listenitems
<i>Liste A</i>	
Hase	Ohren, Ostern, Möhre, hoppeln, Igel, Kaninchen, Fuchs, Braten, Feld, Jäger
Auto	fahren, Straße, Reifen, Lenkrad, Bus, Führerschein, Benzin, Porsche, Motor, Unfall
Brot	essen, Butter, Käse, Bäcker, Messer, Mehl, Frühstück, Marmelade, Nutella, Körner
Pferd	reiten, Sattel, Galopp, Zügel, wiehern, Hengst, Fohlen, Weide, Koppel, Nüstern
Sommer	Sonne, Freibad, Eis, Frühling, Ferien, Winter, Sonnencreme, Strand, Sandalen, warm
Nadel	nähen, Faden, pieksen, spitz, Knopf, stechen, stricken, Garn, Heuhaufen, Spritze
Kerze	Wachs, Leuchter, Flamme, Weihnachten, Biene, Kirche, anzünden, Streichhölzer, Docht, Advent
Hammer	Nagel, Säge, Daumen, Zange, Lärm, Werkzeug, Holz, Wand, Arbeit, schlagen
<i>Liste B</i>	
Hund	bellen, Katze, Leine, Fell, beißen, Schnauze, Schwanz, Dackel, Gassi, Welpen
Fenster	Glas, putzen, Licht, Rahmen, Gardine, öffnen, Tür, Ausblick, hinausschauen, Rollladen
Löwe	Mähne, Tiger, Raubtier, Afrika, Zoo, brüllen, Tatze, Fleisch, König, gefährlich
Fuß	Schuh, Zehen, Bein, laufen, gehen, Socken, Hand, Sohle, Ball, Strumpf
Vogel	fliegen, Nest, Federn, Amsel, Flügel, Spatz, Adler, Wurm, Papagei, Schnabel
Bett	Kopfkissen, Decke, Schlafanzug, träumen, Matratze, Nacht, schlafen, müde, Kuscheltier, Möbel

Telefon	klingeln, sprechen, Hörer, anrufen, wählen, Handy, Tasten, Kabel, Leitung, Rechnung
Hemd	Hose, Anzug, bügeln, Krawatte, Kragen, gestreift, Mann, Ärmel, Kleidung, kariert

 Testphase: Wiedererkennenstest Liste A

Zu beurteilende Items
(in Klammern Itemart und welcher Liste zugehörig)

1. Bild (verbunden <i>Hammer</i>)	21. Flamme (Listenitem <i>Kerze</i>)
2. Pferd (kritisches Item)	22. Zügel (Listenitem <i>Pferd</i>)
3. Butter (Listenitem <i>Brot</i>)	23. Säge (Listenitem <i>Hammer</i>)
4. Hase (kritisches Item)	24. Schiff (unverbunden <i>Kerze</i>)
5. Strand (Listenitem <i>Sommer</i>)	25. Auto (kritisches Item)
6. Messer (Listenitem <i>Brot</i>)	26. Kirche (Listenitem <i>Kerze</i>)
7. Ampel (verbunden <i>Auto</i>)	27. Hengst (Listenitem <i>Pferd</i>)
8. Uhr (unverbunden <i>Sommer</i>)	28. Nadel (kritisches Item)
9. Motor (Listenitem <i>Auto</i>)	29. Klee (verbunden <i>Hase</i>)
10. Möhre (Listenitem <i>Hase</i>)	30. Ferien (Listenitem <i>Sommer</i>)
11. schwitzen (verbunden <i>Sommer</i>)	31. Musik (unverbunden <i>Brot</i>)
12. Loch (verbunden <i>Nadel</i>)	32. Hammer (kritisches Item)
13. Brot (kritisches Item)	33. gemütlich (verbunden <i>Kerze</i>)
14. Fuchs (Listenitem <i>Hase</i>)	34. Becher (unverbunden <i>Auto</i>)
15. Bus (Listenitem <i>Auto</i>)	35. Kerze (kritisches Item)
16. Wurst (verbunden <i>Brot</i>)	36. Holz (Listenitem <i>Hammer</i>)
17. Boot (unverbunden <i>Nadel</i>)	37. Knopf (Listenitem <i>Nadel</i>)
18. Sommer (kritisches Item)	38. Geige (unverbunden <i>Hase</i>)
19. Tee (unverbunden <i>Pferd</i>)	39. Kutsche (verbunden <i>Pferd</i>)
20. stricken (Listenitem <i>Nadel</i>)	40. Parfüm (unverbunden <i>Hammer</i>)

Wiedererkennenstest Liste B

Zu beurteilende Items
(in Klammern Itemart und welcher Liste zugehörig)

1. Pfanne (unverbunden <i>Hund</i>)	21. Löwe (kritisches Item)
2. Fenster (kritisches Item)	22. Socken (Listenitem <i>Fuß</i>)
3. Bier (unverbunden <i>Löwe</i>)	23. Adler (Listenitem <i>Vogel</i>)
4. melken (unverbunden <i>Fuß</i>)	24. Zucker (unverbunden <i>Bett</i>)
5. Amsel (Listenitem <i>Vogel</i>)	25. stinken (verbunden <i>Fuß</i>)
6. Hund (kritisches Item)	26. Fell (Listenitem <i>Hund</i>)
7. Licht (Listenitem <i>Fenster</i>)	27. träumen (Listenitem <i>Bett</i>)
8. Baum (verbunden <i>Vogel</i>)	28. Hemd (kritisches Item)
9. Fuß (kritisches Item)	29. Hügel (unverbunden <i>Telefon</i>)
10. Handy (Listenitem <i>Telefon</i>)	30. Mosel (unverbunden <i>Fenster</i>)
11. Bett (kritisches Item)	31. Zirkus (verbunden <i>Löwe</i>)
12. Feuer (unverbunden <i>Hemd</i>)	32. Nacht (Listenitem <i>Bett</i>)
13. Dackel (Listenitem <i>Hund</i>)	33. Krawatte (Listenitem <i>Hemd</i>)
14. Mann (Listenitem <i>Hemd</i>)	34. Knochen (verbunden <i>Hund</i>)
15. brüllen (Listenitem <i>Löwe</i>)	35. Zöpfe (unverbunden <i>Vogel</i>)
16. laufen (Listenitem <i>Fuß</i>)	36. Afrika (Listenitem <i>Löwe</i>)
17. Vogel (kritisches Item)	37. Telefon (kritisches Item)
18. Hörer (Listenitem <i>Telefon</i>)	38. Seide (verbunden <i>Hemd</i>)
19. Ruhe (verbunden <i>Bett</i>)	39. Gespräch (verbunden <i>Telefon</i>)
20. Tür (Listenitem <i>Fenster</i>)	40. Vorhang (verbunden <i>Fenster</i>)

C2 Instruktionen

Vorschüler und Drittklässler

Lernphase Experiment 3A

Ich möchte gerne mit dir ein Wortspiel spielen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. Wenn ich mit dem Vorlesen aller Listen fertig bin, sage ich dir wie es weitergeht.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Lernphase Experiment 3B

Ich möchte gerne mit dir ein Wortspiel machen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst. In allen Listen habe ich Wörter vergessen. Nach dem Vorlesen jeder Liste frage ich dich, ob dir Wörter einfallen, die noch in die Liste gehören. Ich sage z.B. „Krone, Palast, Schloss, Diener, Prinzessin, Macht, Zepter, reich, Thron, Kaiser“. Nun kann dir auffallen, dass ich das Wort „König“ vergessen habe, denn ein König trägt eine Krone, lebt in einem Palast oder Schloss und hat viele Diener. Das Wort, das fehlt, sollst du mir nach jeder Liste sagen.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Testphase: Wiedererkennen 1 und 2 (Experimente 3A und 3B)

Standardinstruktion: Jetzt frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es eben schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Wenn ich dir das Wort vorgelesen habe, dann sage „alt“. Wenn ich dir das Wort nicht vorgelesen habe, dann sage „neu“.

Inklusionsinstruktion: Jetzt frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es eben schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Du sollst sagen, dass das Wort „alt“ ist, wenn ich es dir vorgelesen habe. Du sollst zu einem Wort aber auch dann „alt“ sagen, wenn ich es dir nicht vorgelesen habe, du aber selbst stark an dieses Wort gedacht hast.

Wenn ich z.B. „Lehrer, rechnen, schreiben, lesen“ vorgelesen habe, dann sollst du mir sagen, dass diese Wörter schon vorkamen, also „alt“ sind. Wenn ich dich frage, ob ich das Wort „Schule“ vorgelesen habe, dann kannst du auch „alt“ sagen, wenn du daran gedacht hast als ich die Wörter vorgelesen habe. Denn in der Schule gibt es Lehrer, die Kinder lernen rechnen,

schreiben und lesen. Zu „Schokolade“ sollst du aber nicht „alt“ sagen, denn daran hast du sicher nicht gedacht als ich dir die Wörter gesagt habe. Zu „Schokolade“ sagst du „neu“.

Ablenkaufgabe (nach Wiedererkennenstest 1)

Nun wollen wir ein anderes Spiel anfangen. Hier sind bunte Mikadohölzer...

Testphase: Quellendiskrimination (nur nach Inklusionsinstruktion; Experimente 3A und 3B)

Jetzt lese ich dir alle Wörter vor, zu denen du „alt“ gesagt hast. Du sollst dich nun entscheiden, ob jedes dieser Wörter tatsächlich von mir vorgelesen wurde oder ob du selbst daran gedacht hast.

Wenn ich dir deine „alt“ genannten Wörter sage, frage ich dich: „Habe ich das Wort am Anfang vorgelesen?“ und du antwortest mit „ja“ oder „nein“.

Erwachsene

Lernphase Experiment 3A

Ich möchte mit dir ein Gedächtnisexperiment durchführen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor, die du dir gut einprägen sollst. Wenn ich mit dem Vorlesen aller Listen fertig bin sage ich dir wie es weitergeht.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Lernphase Experiment 3B

Ich möchte mit dir ein Gedächtnisexperiment durchführen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor, die du dir gut einprägen sollst. In allen Listen fehlen bestimmte zugehörige Wörter. Nach dem Vorlesen jeder Liste frage ich dich, ob dir Wörter einfallen, die noch in die Liste gehören. Ich sage z.B. „Krone, Palast, Schloss, Diener, Prinzessin, Macht, Zepter, reich, Thron, Kaiser“. Nun kann dir auffallen, dass alle Wörter mit dem Wort „König“ zusammenhängen, das aber nicht vorgelesen wurde. Das fehlende zusammenhängende Wort sollst du mir nach jeder Liste sagen.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Testphase: Wiedererkennen 1 und 2 (Experimente 3A und 3B)

Standardinstruktion: Jetzt frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe. Bei jedem Wort sollst du genau überlegen, ob es ein neues Wort ist oder ob es eben schon einmal vorkam. Viele der Wörter habe ich dir vorgelesen, viele aber auch nicht. Wenn ich dir das Wort vorgelesen habe, dann sage „alt“. Wenn ich dir das Wort nicht vorgelesen habe, dann sage „neu“.

Inklusionsinstruktion: Jetzt frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe. Du sollst alle Wörter als „alt“ bezeichnen bei denen du ganz sicher bist, dass ich sie dir vorgelesen habe. Zusätzlich dazu sollst du die Wörter als „alt“ bezeichnen an die du selbst stark gedacht hast, als ich die Listen vorgelesen habe. Zum Beispiel lese ich „Silber, Schmuck, wertvoll, reich, teuer, Geld, glänzen, Mine, Barren, Bronze“ vor. Nun kannst du diese Wörter „alt“ nennen und auch noch andere Wörter an die du selbst stark gedacht hast. Im Beispiel kannst du etwa „Gold“ als „alt“ bezeichnen, wenn du daran stark gedacht hast.

Ablenkaufgabe (nach Wiedererkennenstest 1): Zahlennachsprechen

Jetzt sage ich dir fünf Zahlen und du sollst sie so schnell wie möglich wiederholen. Danach kommen dann fünf neue Zahlen, die du auch wieder so schnell wie möglich wiederholen sollst. Und dann wieder neue Zahlen.

Testphase: Quellendiskrimination (nur nach Inklusionsinstruktion; Experimente 3A und 3B)

Jetzt lese ich dir alle Wörter vor, die du als „alt“ bezeichnet hast. Du sollst dich nun entscheiden, ob jedes dieser Wörter tatsächlich von mir vorgelesen wurde oder ob du selbst daran gedacht hast.

Wenn ich dir deine als „alt“ bezeichneten Wörter sage, frage ich dich: „Habe ich das Wort am Anfang vorgelesen?“ und du antwortest mit „ja“ oder „nein“.

Anhang D: Experiment 4

D1 Material

Lernphasen: 2 Lernbedingungen der Listensets A und B:

(1) A ohne Generieren, B mit Generieren

(2) B ohne Generieren, A mit Generieren

Kritische Items	Listenitems
<i>Liste A</i>	
Hase	Ohren, Ostern, Möhre, hoppeln, Igel, Kaninchen, Fuchs, Braten, Feld, Jäger
Hammer	Nagel, Säge, Daumen, Zange, Lärm, Werkzeug, Holz, Wand, Arbeit, schlagen
Brot	essen, Butter, Käse, Bäcker, Messer, Mehl, Frühstück, Marmelade, Nutella, Körner
Auto	fahren, Straße, Reifen, Lenkrad, Bus, Führerschein, Benzin, Porsche, Motor, Unfall
Pferd	reiten, Sattel, Galopp, Zügel, wiehern, Hengst, Fohlen, Weide, Koppel, Nüstern
Nadel	nähen, Faden, pieksen, spitz, Knopf, stechen, stricken, Garn, Heuhaufen, Spritze
Löwe (<i>Übungsliste</i>)	Mähne, Raubtier, Zoo, Tiger, gefährlich, Tatze, Fleisch
<i>Liste B</i>	
Hund	bellen, Katze, Leine, Fell, beißen, Schnauze, Schwanz, Dackel, Gassi, Welpen
Fenster	Glas, putzen, Licht, Rahmen, Gardine, öffnen, Tür, Ausblick, hinausschauen, Rollladen
Bett	Kopfkissen, Decke, Schlafanzug, träumen, Matratze, Nacht, schlafen, müde, Kuscheltier, Möbel
Fuß	Schuh, Zehen, Bein, laufen, gehen, Socken, Hand, Sohle, Ball, Strumpf
Telefon	klingseln, sprechen, Hörer, anrufen, wählen, Handy, Tasten, Kabel, Leitung, Rechnung
Vogel	fliegen, Nest, Federn, Amsel, Flügel, Spatz, Adler, Wurm, Papagei, Schnabel
Mädchen (<i>Übungsliste</i>)	Zöpfe, klein, Puppe, Kleid, Kind, rosa, Barbie

Testphase: Wiedererkennenstest Liste A

Zu beurteilende Items
(in Klammern Itemart und welcher Liste zugehörig)

1. Bild (verbunden <i>Hammer</i>)	16. Boot (unverbunden <i>Nadel</i>)
2. Pferd (kritisches Item)	17. Tee (unverbunden <i>Pferd</i>)
3. Butter (Listenitem <i>Brot</i>)	18. stricken (Listenitem <i>Nadel</i>)
4. Hase (kritisches Item)	19. Hammer (kritisches Item)
5. Messer (Listenitem <i>Brot</i>)	20. Zügel (Listenitem <i>Pferd</i>)
6. Ampel (verbunden <i>Auto</i>)	21. Auto (kritisches Item)
7. Parfüm (unverbunden <i>Hammer</i>)	22. Nadel (kritisches Item)
8. Möhre (Listenitem <i>Hase</i>)	23. Hengst (Listenitem <i>Pferd</i>)
9. Motor (Listenitem <i>Auto</i>)	24. Klee (verbunden <i>Hase</i>)
10. Loch (verbunden <i>Nadel</i>)	25. Musik (unverbunden <i>Brot</i>)
11. Brot (kritisches Item)	26. Säge (Listenitem <i>Hammer</i>)
12. Holz (Listenitem <i>Hammer</i>)	27. Becher (unverbunden <i>Auto</i>)
13. Fuchs (Listenitem <i>Hase</i>)	28. Knopf (Listenitem <i>Nadel</i>)
14. Bus (Listenitem <i>Auto</i>)	29. Geige (unverbunden <i>Hase</i>)
15. Wurst (verbunden <i>Brot</i>)	30. Kutsche (verbunden <i>Pferd</i>)

Wiedererkennenstest Liste B

Zu beurteilende Items
(in Klammern Itemart und welcher Liste zugehörig)

1. Pfanne (unverbunden <i>Hund</i>)	16. Tür (Listenitem <i>Fenster</i>)
2. Fenster (kritisches Item)	17. Socken (Listenitem <i>Fuß</i>)
3. melken (unverbunden <i>Fuß</i>)	18. Adler (Listenitem <i>Vogel</i>)
4. Gespräch (verbunden <i>Telefon</i>)	19. Zucker (unverbunden <i>Bett</i>)
5. Hund (kritisches Item)	20. stinken (verbunden <i>Fuß</i>)
6. Licht (Listenitem <i>Fenster</i>)	21. Fell (Listenitem <i>Hund</i>)
7. Baum (verbunden <i>Vogel</i>)	22. träumen (Listenitem <i>Bett</i>)
8. Fuß (kritisches Item)	23. Hügel (unverbunden <i>Telefon</i>)
9. Handy (Listenitem <i>Telefon</i>)	24. Mosel (unverbunden <i>Fenster</i>)
10. Bett (kritisches Item)	25. Nacht (Listenitem <i>Bett</i>)
11. Dackel (Listenitem <i>Hund</i>)	26. Knochen (verbunden <i>Hund</i>)
12. laufen (Listenitem <i>Fuß</i>)	27. Papier (unverbunden <i>Vogel</i>)
13. Vogel (kritisches Item)	28. Telefon (kritisches Item)
14. Hörer (Listenitem <i>Telefon</i>)	29. Amsel (Listenitem <i>Vogel</i>)
15. Ruhe (verbunden <i>Bett</i>)	30. Vorhang (verbunden <i>Fenster</i>)

D2 Instruktionen

Erst- und Viertklässler

Lernphase 1

Ohne Generieren: Ich möchte gerne mit dir ein Wortspiel spielen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst, die ich dir vorlese. Später frage ich dich nach den Wörtern. Wenn ich mit dem Vorlesen aller Listen fertig bin, sage ich dir wie es weitergeht.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Lernphase 2

Mit Generieren: Ich möchte gerne mit dir ein Wortspiel spielen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor. Ich möchte, dass du mir ganz genau zuhörst und dir die Wörter merkst, die ich dir vorlese. Später frage ich dich nach den Wörtern.

Ich sage dir einen Trick wie man sich die Wörter merken kann. Du sollst über die Wörter nachdenken. Dazu sollst du zu jedem Wort, das ich dir vorlese, ein Wort sagen, das dir als erstes einfällt. Ich zeige dir, wie das Wortspiel geht. Ich sage z.B. „Lehrer“, dann kannst du „Schule“ sagen. Wenn ich „schreiben“ sage, kannst du „rechnen“ sagen. Manchmal kann es sein, dass dir ein Wort einfällt, das ich schon vorgelesen habe (z.B. „Lehrer“, wenn ich „schreiben“ sage). Dann kannst du das vorgelesene Wort noch mal selbst sagen. Oder es kann sein, dass du an das gleiche Wort denkst, das du schon mal gesagt hast. Dann kannst du es einfach noch mal sagen („Lehrer-Schule“, „schreiben-Schule“). Hast du das Spiel nun verstanden? Jetzt sage ich dir andere Wörter. „Schokolade, Zucker, Bonbon, lecker, Pudding, Eis, Kuchen“.

Ablenkaufgaben 1 und 2 (nach Lernphasen 1 und 2)

1) Jetzt zu etwas anderem. Ich zeige dir verschiedene Dinge und du sollst überlegen, ob sie in die Kategorien „Obst“ und/oder „rote Farbe“ passen. Ich sage „Apfel“ (Orange, Teufel, Bär). Gehört der Apfel zum Obst? Gehört der Apfel zu den Dingen, die rot sind?

2) Nun sollst du verschiedene Dinge den Kategorien „Blume“ und/oder „gelbe Farbe“ zuordnen. Ich sage „Sonnenblume“ (Rose, Stift, Feuerzeug). Gehört die Sonnenblume zu den Blumen? Gehört die Sonnenblume zu den Dingen, die gelb sind?

Ablenkaufgabe 3 (nach Wiedererkennen 1): Standardform des Progressiven Matrizen Tests (Raven, 1976)

Nun wollen wir uns etwas anderes anschauen. Hier habe ich ein Muster das unvollständig ist. Es gibt immer verschiedene Stücke, die in die Lücke passen würden, aber nur eines ist richtig. Finde bitte das passende Stück.

Testphase: Wiedererkennen 1

Ohne Generieren: Nun frage ich dich nach den Wörtern, die ich dir eben vorgelesen habe. Einige habe ich dir schon vorgelesen, an andere hast du vielleicht selbst gedacht als ich dir die Wörter vorgelesen habe und wieder andere sind ganz neu. Bei jedem Wort sollst du genau überlegen, ob es eben schon einmal vorkam und ob du daran gedacht hast. Alles klar? Dann lese ich dir jetzt die Wörter vor.

Ich sage nun „Mähne, Raubtier, Zoo, Tiger, gefährlich, Tatze, Fleisch“.

Habe ich „Zoo“ vorgelesen? (ja). Hast du an „Zoo“ gedacht? (nein).

Habe ich „singen“ vorgelesen? (nein). Hast du an „singen“ gedacht? (nein).

Habe ich „Tiger“ vorgelesen? (ja). Hast du an „Tiger“ gedacht? (ja).

Habe ich „Löwe“ vorgelesen? (nein). Hast du an „Löwe“ gedacht? (ja).

Testphase: Wiedererkennen 2

Mit Generieren: Nun frage ich dich nach den Wörtern, die ich dir eben vorgelesen habe. Einige habe ich dir schon vorgelesen, andere hast du vielleicht selbst gesagt und wieder andere sind ganz neu. Bei jedem Wort sollst du genau überlegen, ob es eben schon einmal vorkam und ob du es selbst gesagt hast. Alles klar? Dann lese ich dir jetzt die Wörter vor.

Ich sage nun „Zöpfe“ (klein, Puppe, Kleid, Kind, rosa, Barbie). Was fällt dir dazu ein?

Habe ich „klein“ vorgelesen? (ja). Hast du „klein“ gesagt? (nein).

Habe ich „Bier“ vorgelesen? (nein). Hast du „Bier“ gesagt? (nein).

Habe ich „Puppe“ vorgelesen? (ja). Hast du „Puppe“ gesagt? (ja).

Habe ich „Mädchen“ vorgelesen? (nein). Hast du „Mädchen“ gesagt? (ja).

Erwachsene

Lernphase 1

Ohne Generieren: Ich möchte mit dir ein Gedächtnisexperiment durchführen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor, die du dir gut einprägen sollst. Wenn ich mit dem Vorlesen aller Listen fertig bin sage ich dir wie es weitergeht.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Lernphase 2

Mit Generieren: Ich möchte mit dir ein Gedächtnisexperiment durchführen. Ich lese dir nacheinander mehrere Listen mit Wörtern vor, die du dir gut einprägen sollst. Zu jedem Wort, das ich dir vorlese, sollst du ein Wort sagen, das dir als erstes einfällt. Ich sage z.B. „Universität“, dann kannst du „Student“ sagen. Wenn ich „lesen“ sage, kannst du „schreiben“ sagen. Manchmal kann es sein, dass dir ein Wort einfällt, das ich schon vorgelesen habe (z.B. „Universität“, wenn ich „lesen“ sage). Dann kannst du das vorgelesene Wort noch mal selbst sagen. Oder es kann sein, dass du an das gleiche Wort denkst, das du schon mal gesagt hast. Dann kannst du es einfach noch mal sagen („Universität-Student“; „lesen-Student“). Hast du das Experiment nun verstanden? Jetzt sage ich dir andere Wörter. „Schokolade, Zucker, Bonbon, lecker, Pudding, Eis, Kuchen“.

Ablenkaufgaben 1 und 2 (nach Lernphasen 1 und 2)

1) Jetzt zu etwas anderem. Ich zeige dir verschiedene Dinge und du sollst überlegen wie diese aussehen und mir zwei Fragen dazu beantworten. Ich sage „Apfel“ (Orange, Teufel, Bär). Gehört der Apfel zum Obst? Gehört der Apfel zu den Dingen, die rot sind?

2) Ich sage „Sonnenblume“ (Rose, Stift, Feuerzeug). Gehört die Sonnenblume zu den Blumen? Gehört die Sonnenblume zu den Dingen, die gelb sind?

Ablenkaufgabe 3 (nach Wiedererkennen 1): Standardform des Progressiven Matrizen Tests (Raven, 1976)

Nun wollen wir uns etwas anderes anschauen. Hier habe ich ein Muster das unvollständig ist. Es gibt immer verschiedene Stücke, die in die Lücke passen würden, aber nur eines ist richtig. Finde bitte das passende Stück.

Testphase: Wiedererkennen 1

Ohne Generieren: Jetzt frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe. Einige habe ich dir schon am Anfang vorgelesen, an andere hast du vielleicht selbst gedacht als ich dir die Wörter vorgelesen habe und wieder andere sind ganz neu. Bei jedem Wort

sollst du genau überlegen, ob es eben schon einmal vorkam und ob du daran gedacht hast.

Alles klar? Dann lese ich dir jetzt die Wörter vor.

Ich sage nun „Mähne, Raubtier, Zoo, Tiger, gefährlich, Tatze, Fleisch“.

Habe ich „Zoo“ vorgelesen? (ja). Hast du an „Zoo“ gedacht? (nein).

Habe ich „singen“ vorgelesen? (nein). Hast du an „singen“ gedacht? (nein).

Habe ich „Tiger“ vorgelesen? (ja). Hast du an „Tiger“ gedacht? (ja).

Habe ich „Löwe“ vorgelesen? (nein). Hast du an „Löwe“ gedacht? (ja).

Testphase: Wiedererkennen 2

Mit Generieren: Jetzt frage ich dich nach den Wörtern, die ich dir gerade vorgelesen habe.

Einige habe ich dir schon am Anfang vorgelesen, andere hast du vielleicht selbst gesagt und wieder andere sind ganz neu. Bei jedem Wort sollst du genau überlegen, ob es eben schon einmal vorkam und ob du es selbst gesagt hast. Alles klar? Dann lese ich dir jetzt die Wörter vor.

Ich sage nun „Zöpfe“ (klein, Puppe, Kleid, Kind, rosa, Barbie). Was fällt dir dazu ein?

Habe ich „klein“ vorgelesen? (ja). Hast du „klein“ gesagt? (nein).

Habe ich „Bier“ vorgelesen? (nein). Hast du „Bier“ gesagt? (nein).

Habe ich „Puppe“ vorgelesen? (ja). Hast du „Puppe“ gesagt? (ja).

Habe ich „Mädchen“ vorgelesen? (nein). Hast du „Mädchen“ gesagt? (ja).

Anhang E: Eltern- und Schülerinformationsbriefe

E1 Elterninformationsbrief

Dipl.-Psych. Anette Hüsken
Fachbereich I - Psychologie
Universität Trier
54286 Trier

Trier, den 17.10.2005

Liebe Eltern,

im Rahmen eines Forschungsprojektes beschäftige ich mich an der Universität Trier mit der Lern- und Gedächtnisentwicklung bei Kindern und Erwachsenen. Dabei interessiert mich unter anderem wie sich Kindergarten-, Schulkinder und junge Erwachsene in ihren Lernleistungen unterscheiden.

Im Rahmen des Projektes möchte ich Kindergartenkinder (Erst-/Drittklässler) jeweils einzeln ca. 15 Minuten untersuchen. Die Kinder sollen mehrere Listen mit Wörtern lernen.

Meine Studie ist vom 14. bis zum 18. November in Ihrem Kindergarten (Ihrer Schule) geplant. Ich bin weniger daran interessiert wie das einzelne Kind antwortet als vielmehr daran wie sich Kinder dieser Altersstufe im Durchschnitt verhalten. Deshalb werden lediglich Geschlecht und Alter der teilnehmenden Kinder festgehalten. Die Ergebnisse der einzelnen Kinder bleiben anonym.

Bei der Durchführung meiner Untersuchung bin ich auf die Mitarbeit und Unterstützung vieler Kinder, Eltern, Erzieher/innen und Lehrer/innen angewiesen. Ich würde mich deshalb sehr freuen, wenn Sie damit einverstanden wären, dass Ihr Kind an dieser Untersuchung teilnimmt. Die Teilnahme ist natürlich freigestellt.

Ich möchte Sie bitten den folgenden Abschnitt auszufüllen und bei dem/der Erzieher/in (Lehrer/in) möglichst bald abzugeben.

In der Hoffnung auf Ihre Mitarbeit verbleibe ich
mit freundlichen Grüßen

(Anette Hüsken)

Bitte folgenden Abschnitt abtrennen und im Kindergarten (in der Schule) abgeben

Bitte ankreuzen:

JA, ich bin damit einverstanden, dass mein Sohn/meine Tochter an der Studie teilnimmt.

NEIN, ich bin mit der Teilnahme an der Studie nicht einverstanden.

.....
(Vor- und Nachname des Kindes)

.....
(Geburtsdatum des Kindes)

Trier, den

.....
(Unterschrift der Eltern)

E2 Schülerinformationsbrief

Dipl.-Psych. Anette Hüsken
Fachbereich I - Psychologie
Universität Trier
54286 Trier

Trier, den 18.11.2005

Liebe Schülerin,
lieber Schüler,

im Rahmen meiner Dissertation beschäftige ich mich an der Universität Trier mit der Lern- und Gedächtnisentwicklung bei Kindern und Erwachsenen. Dabei interessiert mich unter anderem wie sich Kindergarten-, Schulkinder und junge Erwachsene in ihren Lernleistungen unterscheiden.

An ihrer Schule möchte ich gerne meine Untersuchung mit den Schüler/innen der 12. Klassen jeweils einzeln ca. 20 Minuten durchführen. Es gilt mehrere Listen mit Wörtern zu lernen und diese wiederzugeben.

Die Studie findet voraussichtlich an fünf Unterrichtstagen in der nächsten Woche an Ihrer Schule statt.

Ich bin weniger daran interessiert wie die einzelne Person antwortet als vielmehr daran wie sich diese Altersstufe im Durchschnitt verhält. Deshalb werden lediglich Geschlecht und Alter der teilnehmenden Personen festgehalten und die einzelnen Ergebnisse bleiben anonym.

Bei der Durchführung meiner Untersuchung bin ich auf die Mitarbeit und Unterstützung vieler Kinder, Schüler/innen, Eltern, Erzieher/innen und Lehrer/innen angewiesen. Ich würde mich deshalb sehr freuen, wenn Sie damit einverstanden wären an dieser Untersuchung teilzunehmen.

Durch die Teilnahme an meiner Studie können Sie einen Einblick in die psychologische Forschung erhalten und einen Beitrag zu dieser leisten. Sobald ich die Ergebnisse der Untersuchung ausgewertet habe, werde ich Ihnen diese mitteilen.

In der Hoffnung auf Ihre Mitarbeit verbleibe ich
mit freundlichen Grüßen

(Anette Hüsken)

Bitte folgenden Abschnitt abtrennen und abgeben

Bitte ankreuzen:

JA, ich bin damit einverstanden, an der Studie teilzunehmen.

NEIN, ich bin mit der Teilnahme an der Studie nicht einverstanden.

.....
(Vor- und Nachname der Schülerin/des Schülers)

Erklärung zur Dissertation

Hiermit erkläre ich, Anette Jammers, geboren am 21. Februar 1980 in Trier, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet zu haben.

Diese Arbeit wurde bisher nicht veröffentlicht und auch nicht als Dissertation an einer anderen Hochschule eingereicht.

Trier, im Oktober 2007