

Institut für Psychologie und Pädagogik, Abteilung Entwicklungspsychologie

**Der Testeffekt im Altersvergleich:  
Auswirkungen von Testdurchgängen auf die Lernleistung  
abhängig vom Lernmaterial bei jüngeren und älteren  
Erwachsenen**

Philipp Handschuh  
aus Ellwangen

DISSERTATION

vorgelegt der Fakultät für Ingenieurwissenschaften, Informatik und  
Psychologie der Universität Ulm  
zur Erlangung des akademischen Grades  
eines Doktors der Philosophie (Dr. phil.)

2019

Amtierender Dekan: Prof. Dr. Maurits Ortmanns

Gutachter: Prof. Dr. Daniel Zimprich

Gutachter: Apl. Prof. Dr. Oliver Schilling

Gutachter: Apl. Prof. Dr. Markus Kiefer

Tag der Promotion: 12.12.2019

Die Inhalte der Seite wurden aus Gründen des Datenschutzes entfernt.

## ZUSAMMENFASSUNG

Tests während des Lernens dienen nicht nur zur Messung des momentanen Wissenstandes, sondern fördern darüber hinaus die Lernleistung. Doch inwieweit profitieren Lerner höherer Altersgruppen von diesem sogenannten Testeffekt? Bisher haben wenige Studien untersucht, ob Erwachsene im höheren Lebensalter, trotz altersbedingter kognitiver Veränderungen, durch erfolgreiches Abrufen lernen. In 4 Experimenten wird in der vorliegenden Arbeit untersucht, wie groß die Altersunterschiede im Testeffekt bei unterschiedlichen Lernmaterialien und Lernbedingungen sind. In den Experimenten 1 und 2 werden in einem Vergleich von Standardlernkurven, bestehend aus mehreren Lern-Test-Durchläufen mit „pure-study“-Lernkurven, mögliche Altersunterschiede in den Effekten von Tests untersucht. Hierbei setzt sich die „pure-study“-Lernbedingung aus mehreren Lernaufgaben mit unterschiedlicher Anzahl an Lerndurchgängen zusammen. Dabei werden in Experiment 1 die Lernkurven anhand von freiem Abruf von Wortlisten und in Experiment 2 anhand einer Paarassoziationsaufgabe untersucht. In Experiment 3 und 4 werden kurze Geschichten gelernt und durch freies Abrufen getestet. Dabei bearbeiteten die jüngeren und älteren Versuchspersonen neben einer Lernbedingung mit mehreren Lern-Test-Durchgängen eine Lernbedingung mit einem abschließenden Testdurchgang und einer Dropout-Bedingung, bei der erfolgreich erinnerte Lerninhalte in den folgenden Lerndurchgängen nicht mehr präsentiert wurden. In Experiment 3 folgte nach 2 Tagen und in Experiment 4 nach einer Woche ein verzögerter Abruf. In allen 4 Studien zeigten sich am Ende der Experimente Testeffekte. In Experiment 1 und 2 profitierten ältere Lerner im geringeren Umfang als jüngere Erwachsene vom Testeffekt. Zwar profitierten in Experiment 3 und 4 jüngere und ältere Lerner in vergleichbarem Maße vom Testeffekt beim letzten Testdurchgang des Experiments, jedoch nicht beim verzögerten Abruf, wobei sich keine Testeffekte zeigten. Ausnahme hierbei ist, dass bei älteren Lernern in Experiment 4 ein Testeffekt der Dropout-

Bedingung beim verzögerten Abruf vorlag. Die Ergebnisse zeigen, dass es Altersunterschiede im Testeffekt gibt, diese aber vom Lernmaterial abhängen.

## Inhaltsverzeichnis

1. Theorie .....	4
1.1. Lernen und Testen .....	4
1.1.1. Was ist Lernen? .....	4
1.1.2. Was ist Testen? .....	6
1.1.2.1. Wann sind Testeffekte beobachtbar? .....	10
1.1.2.2. Erklärungsansätze des Testeffekts .....	13
1.2. Lernen und Testen im Alter .....	23
1.2.1. Lernen im Alter .....	23
1.2.2. Testen im Alter .....	28
1.2.3. Individuelle Unterschiede beim Lernen und Testen .....	34
1.2.4. Vorteile von Mehrebenenmodellen .....	35
1.3. Hypothesen und Ziele der Arbeit .....	36
2. Studienblock 1 – Lernkurven mit und ohne den Einfluss des Testeffekts .....	37
2.1. Theoretische Einführung: Lernkurven .....	37
2.2. Lernmaterial – Wortlisten und Wortpaare .....	39
2.3. Experiment 1 – Wortlisten .....	40
2.3.1. Methode Experiment 1 .....	41
2.3.2. Ergebnisse Experiment 1 .....	49
2.3.3. Zusammenfassung der Ergebnisse aus Experiment 1 .....	61
2.4. Experiment 2 – Wortpaare .....	63
2.4.1. Methode Experiment 2 .....	63
2.4.2. Ergebnisse Experiment 2 .....	68
2.4.3. Zusammenfassung der Ergebnisse aus Experiment 2 .....	81
3. Studienblock 2 – Testeffekt bei Geschichten als Lernmaterial .....	83
3.1. Theoretische Einführung: Lernmaterial und Dropout .....	83
3.1.1. Lernmaterial – Geschichten .....	83
3.1.2. Das Dropout-Lerndesign .....	85
3.2. Experiment 3 - Geschichten .....	87
3.2.1. Methode Experiment 3 .....	88
3.2.2. Ergebnisse Experiment 3 .....	95
3.2.3. Zusammenfassung der Ergebnisse aus Experiment 3 .....	113
3.3. Experiment 4 – längere Geschichten .....	116
3.3.1. Methode Experiment 4 .....	117

3.3.2.	Ergebnisse Experiment 4.....	121
3.3.3.	Zusammenfassung der Ergebnisse aus Experiment 4 .....	142
4.	Diskussion.....	144
4.1.	Lernkurven beim freien und hinweisgesteuerten Abruf.....	145
4.1.1.	Ergebnisse Studienblock 1 .....	145
4.1.2.	Testeffekt beim freien Erinnern und hinweisgesteuerten Abruf.....	149
4.2.	Geschichten als Lernmaterial und Dropout-Design .....	152
4.2.1.	Ergebnisse Studienblock 2 .....	154
4.2.2.	Implikationen für Geschichten als Lernmaterial .....	159
4.3.	Überblick Testeffekt .....	161
4.4.	Ausblick.....	167
4.4.1.	Testeffekt und Lernereigenschaften .....	167
4.4.2.	Testeffekt und Eigenschaften des Lerndesigns .....	170
4.5.	Fazit .....	172
	LITERATURVERZEICHNIS.....	173



## Tabellenverzeichnis

Tabelle 1 Lernaufgaben Experiment 1 .....	43
Tabelle 2 Designs Experiment 1 .....	44
Tabelle 3 Experiment 1: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden .....	51
Tabelle 4 Experiment 1: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden.....	53
Tabelle 5 Experiment 1 Modelle 0 bis 2: Parameterschätzungen und Fit-Indizes .....	56
Tabelle 6 Experiment 1 Modelle 3 bis 6: Parameterschätzungen und Fit-Indizes .....	60
Tabelle 7 Lerndesign Experiment 2 .....	66
Tabelle 8 Experiment 2: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden	70
Tabelle 9 Experiment 2: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden.....	72
Tabelle 10 Experiment 2 Modelle 0 bis 2: Parameterschätzungen und Fit-Indizes .....	75
Tabelle 11 Experiment 2 Modelle 3 bis 6: Parameterschätzungen und Fit-Indizes .....	79
Tabelle 12 Lernbedingungen Experiment 3 .....	90
Tabelle 13 Versuchsdesigns Experiment 3 .....	90
Tabelle 14 Experiment 3: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden .....	98
Tabelle 15 Experiment 3: Lernleistung nach Geschichte und Design aufgeteilt .....	100
Tabelle 16 Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden.....	102
Tabelle 17 Experiment 3 Modelle 0 bis 2: Schätzungen der festen Effekten und Fit-Indizes .....	106
Tabelle 18 Experiment 3 Modelle 0 bis 2: Schätzungen der Residualvarianzen und zufälligen Effekte .....	107
Tabelle 19 Experiment 3 Modelle 3 und 4: Schätzungen der festen Effekten und Fit-Indizes .....	110
Tabelle 20 Experiment 3 Modelle 3 und 4: Schätzungen der Residualvarianzen und zufälligen Effekte .....	111
Tabelle 21 Experiment 3 Zusatzmodelle: Schätzungen der Residualvarianzen, zufälligen Effekte und Fit-Indizes .....	112
Tabelle 22 Lernbedingungen Experiment 4 .....	119

Tabelle 23 Experiment 4: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden .....	124
Tabelle 24 Experiment 4: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden.....	126
Tabelle 25 Experiment 4: Lernleistung nach Geschichte und Alter aufgeteilt .....	127
Tabelle 26 Experiment 4: Lernleistung in den Designs .....	128
Tabelle 27 Experiment 4 Modelle 1 bis 4: Schätzungen der festen Effekten und Fit-Indizes .....	132
Tabelle 28 Experiment 4 Modelle 1 bis 4: Schätzungen der Residualvarianzen und zufälligen Effekte .....	133
Tabelle 29 Experiment 4 Modelle 5 und 6: Schätzungen der festen Effekten und Fit-Indizes .....	138
Tabelle 30 Experiment 4 Modelle 5 und 6: Schätzungen der Residualvarianzen und zufälligen Effekte .....	139
Tabelle 31 Experiment 4 Zusatzmodelle: Schätzungen der Residualvarianzen, zufälligen Effekte und Fit-Indizes .....	141

**Abbildungsverzeichnis**

Abbildung 1. Ergebnisse von Roediger und Karpicke (2006b, S. 253).....	9
Abbildung 2. Experiment 1: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene .....	52
Abbildung 3. Experiment 1: Altersunterschiede in der Standardbedingung und in der „pure-study“-Bedingung sowie Differenz zwischen Standard- und „pure-study“-Bedingung für jüngere und ältere Erwachsene.....	52
Abbildung 4. Experiment 1: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene für Modell 5 .....	61
Abbildung 5. Experiment 2: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene .....	71
Abbildung 6. Experiment 2: Altersunterschiede in der Standardbedingung und in der „pure-study“-Bedingung sowie Differenzen zwischen Standard- und „pure-study“-Bedingung für jüngere und ältere Erwachsene.....	71
Abbildung 7. Experiment 2: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene für Modell 5 .....	80
Abbildung 8. Experiment 2: Mittlere Anzahl erinnerter Wortpaare der fünf Wortlisten beim verzögerten Abruf, aufgeteilt nach Altersgruppe. ....	81
Abbildung 9. Experiment 3: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene sowie die Altersunterschiede .....	99
Abbildung 10. Experiment 3: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene für Modell 3 .....	113
Abbildung 11. Experiment 4: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene sowie die Altersunterschiede .....	125
Abbildung 12. Experiment 4: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene für Modell 3 und Modell 4.....	134
Abbildung 13. Experiment 4: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene und ältere Erwachsene für Modell 5 .....	140

DER TESTEFFEKT IM ALTERSVERGLEICH:  
AUSWIRKUNGEN VON TESTDURCHGÄNGEN AUF DIE LERNLEISTUNG  
ABHÄNGIG VOM LERNMATERIAL BEI JÜNGEREN UND ÄLTEREN  
ERWACHSENEN

Die Mäeutik, der Begriff geht auf den griechischen Philosophen Sokrates zurück, stellt eine Lehrmethode dar, in der durch gezieltes Fragestellen dem Lernenden die Möglichkeit gegeben wird, Erkenntnisse aus seinen Antworten und aus sich selbst heraus hervorzubringen (Brockhaus, 2006). Hierbei wird die Abfrage nicht nur verwendet, um den Wissenstand zu messen, sondern auch um den Lernenden beim Erkenntnisgewinn zu unterstützen. Aus dem heutigen Forschungsstand würde man das didaktische Potenzial der Methode nicht nur in der selbstinitiierten Aufdeckung von Irrtümern sehen, sondern ebenfalls in dem Abrufprozess des Wissens (siehe Karpicke, 2017). Den Befund, dass der erfolgreiche Abruf beim Testen von Lerninhalten den erfolgreichen Abruf eben dieser Lerninhalte zu einem späteren Zeitpunkt fördert, bezeichnet man als Testeffekt (Roediger & Karpicke, 2006a, 2006b). Folglich messen Tests nicht nur den momentanen Wissensstand, sondern verändern diesen und tragen damit langfristig zum Lernerfolg bei. Das Abfragen von Wissen stellt somit eine wertvolle Methode zur Förderung des Lernens dar. Daher überrascht es nicht, dass die Forschung zum Testeffekt diesen unter verschiedenen Bedingungen, wie etwa Testmaterialien und Personengruppen untersucht (vgl. Adesope, Trevisan & Sundararajan, 2017; Karpicke, 2017; Rowland, 2014). Eine Eigenschaft von Lernenden, für die vergleichsweise wenig Studien vorliegen, ist das Alter.

Dabei ist es in der heutigen Gesellschaft ein bedeutsames Ziel, Lernen im Alter zu unterstützen und Bildung für jeden gleichermaßen zugänglich zu machen (Bundesministerium für Familie, Senioren, Frauen und Jugend, 2016). Aus lernpsychologischer Sicht liegt eine Besonderheit älterer Lerner in den altersbedingten

Unterschieden im Gedächtnis vor. In der bisherigen Gedächtnisforschung konnten altersbedingte Abnahmen im episodischen Gedächtnis weitestgehend konsistent gezeigt und repliziert werden (bspw. Verhaeghen, Marcoen, & Goossens, 1993). Hierbei stellt sich die Frage, ob ältere Lerner ebenfalls, trotz der Defizite im episodischen Gedächtnis, von den Effekten durch Testdurchgänge profitieren und wie sich der Lernerfolg im Vergleich zu jüngeren Lernern unterscheidet. Eines der Ziele der vorliegenden Arbeit ist es, dieser Fragestellung nachzugehen und Altersunterschiede im Testeffekt unter der Verwendung verschiedener Lerndesigns und Lernmaterialien zu untersuchen.

Erfolgreiches Lernen hängt unter anderem stark vom wiederholten Lernen und Überprüfung des Gelernten ab, wodurch ein Lernprozess sich über mehrere Zyklen des Einprägens und Abfragens erstrecken kann sowie über längere Zeiträume. Um einen solchen Lernprozess abzubilden, werden in der Gedächtnisforschung typischerweise Lernkurven verwendet. Jedoch spiegeln sich in diesen nicht nur die Effekte des erfolgreichen Einprägens, sondern auch Effekte des Abrufens wider. Um den Testeffekt zu bestimmen, ist es also sinnvoll, den Lernerfolg jeweils für das Lernen sowie das Lernen mit Testen zu vergleichen. Um eine Lernkurve ohne Einflüsse von Testdurchgängen zu erstellen, wird die Methode von Roediger und Smith (2012) verwendet, bei der anhand von mehreren Wortlisten, bei welchen jeweils nach einer unterschiedlichen Anzahl an aufeinanderfolgenden Lerndurchgängen *ein* Testdurchgang folgt, erstellt. Nachdem nicht nur das Lernen und Testen eine wichtige Rolle spielt, sondern ebenso das gelernte Material und die Art der Abfrage, werden Lernkurven für den freien Abruf von Wortlisten sowie für den hinweisgesteuerten Abruf von Wortpaaren durchgeführt.

Jedoch zählt das in den bisherigen Untersuchungen zu Lernkurven typischerweise verwendete Lernmaterial, wie Listen zusammenhangsloser Wörter oder Wortpaare, nicht zu den im Alltag häufig verwendeten Lernmaterialien. Doch auch bei bildungsrelevanten

Texten zeigen sich positive Effekte von Abrufversuchen, was den Testeffekt zu einer alltagsrelevanten Einflussgröße beim Lernen macht (bspw. Meyer & Logan, 2013; Roediger & Karpicke, 2006b). Aufgrund des Erzählcharakters einer Geschichte – welche in der vorliegenden Arbeit als ein Beispiel für Lerntexte verwendet wird – jedoch, entsteht eine sinnhafte, zusammenhängende Struktur des Textes, welche den Lerner darin unterstützen kann, sich Lerninhalte einzuprägen. Von solchen Strukturen können vor allem ältere Lerner profitieren, da sie beim Lernen und Abspeichern häufig Schwierigkeiten mit der subjektiven Organisation von Lerninhalten haben (bspw. Davis et al., 2003) und durch die Struktur einer Geschichte darin unterstützt werden (Craig & Bosman, 1992). Wünschenswert für den Lernerfolg ist nicht nur eine hohe Behaltensleistung, sondern darüber hinaus, dass diese in möglichst kurzer Zeit erreicht wird – allerdings sind diese Ziele nicht immer vereinbar (Kornell & Bjork, 2008). Eine Möglichkeit eines effizienteren Lernprozesses ist es, bereits erfolgreich abgerufene Inhalte aus dem Lerndurchgang auszuschließen. Sprich, man lernt als Konsequenz in den nächsten Lerndurchgängen nur diejenigen Inhalte, die man noch nicht weiß und versucht den gesamten Lernstoff abzurufen. Ein solches Vorgehen ist nicht nur zeitlich effizient, sondern bietet weiterhin die Vorteile des Testeffekts (Karpicke & Roediger, 2008). In der vorliegenden Arbeit wird untersucht, wie effektiv eine solche Methode bei jüngeren und älteren Personen beim Lernen von Geschichten ist und in welchem Umfang die Testdurchgänge den Lernerfolg in den beiden Altersgruppen unterstützen.

Bevor die vier Experimente und deren Ergebnisse vorgestellt werden, werden die Befunde bisheriger Forschung diskutiert. Nachdem eine Arbeitsdefinition für Lernen präsentiert wurde (Kapitel 1.1.1), werden Befunde zum Testeffekt vorgestellt und verschiedene theoretische Modelle gegenübergestellt (Kapitel 1.1.2.). Um Schlüsse auf die Wirksamkeit von Testeffekten in verschiedenen Altersgruppen zu treffen, werden Theorien

und Befunde zu Altersunterschieden in der Kognition und im Gedächtnis ausgeführt (Kapitel 1.2.) und mit den theoretischen Überlegungen und empirischen Befunden zum Testeffekt in Verbindung gebracht. Auf dieser Basis werden Hypothesen zur Wirksamkeit des Testeffekts in verschiedenen Altersgruppen abgeleitet (Kapitel 1.3.). Die vier vorgestellten Experimente lassen sich in zwei Studienblöcke unterteilen: 1) Experiment 1 und 2 modellieren Lernkurven mit und ohne den Einfluss des Testeffekts und 2) in Experiment 3 und 4 werden die Fragestellungen mit kohärentem Lernmaterial in Form von kurzen Geschichten getestet. Die Analysen werden mit gemischten Regressionsmodellen durchgeführt, wobei verschiedene Quellen von Varianz berücksichtigt werden (bspw. Unterschiede zwischen Personen in der Lernleistung; oder aufgrund von unterschiedlichen Messprozessen). Abschließend werden die Ergebnisse innerhalb (Kapitel 4.1. und 4.2.) und zwischen den Studienblöcken (Kapitel 4.3.) bezüglich der Altersunterschiede im Testeffekt besprochen, sowie ein Ausblick bezüglich der Hauptfrage der Arbeit gegeben: Wann gibt es Altersunterschiede im Testeffekt und wie kommen diese zustande?

## 1. Theorie

### *1.1. Lernen und Testen*

#### *1.1.1. Was ist Lernen?*

Lernen ist ein zentraler Bestandteil unseres alltäglichen Lebens. Den meisten Definitionen des „Lernens“ ist gemein, dass Lernen als Prozess betrachtet wird, der aufgrund von Erfahrungen relativ langfristige Veränderungen im Verhaltenspotential hervorruft (Koch & Stahl, 2017). „Lernen“ ist eng mit dem Gedächtnis verbunden, welches den Lernprozess erst ermöglicht. Dabei zählen zu den zentralen Funktionen des Gedächtnisses Enkodieren, Speichern (Retention) und Abrufen von Informationen, welche wiederum die zentralen Prozesse beim Lernen sind. Dabei gibt es unterschiedliche Theorien dazu, wie Informationen

in bestehende Strukturen des Langzeitgedächtnisses gespeichert, integriert und repräsentiert werden. Semantisches Wissen kann bspw. in semantischen Netzwerken gespeichert werden, welche das Wissen durch eine hierarchische Ordnung von Ober- und Unterbegriffen strukturieren (Collins & Quillian, 1969). Ein anderes Notationssystem für kategoriales Wissen im Langzeitgedächtnis (LZG) ist die Strukturierung durch Schemata (Bartlett, 1977). „Ein Schema ist eine kognitive Repräsentation einer Klasse von Objekten, Personen, Ereignissen oder Situationen. Es ist ein Wissensbündel, das Generalisierungen encodiert. Schemata beeinflussen sowohl Prozesse des Encodierens in das LZG als auch Prozesse des Abrufs.“ (Becker-Carus & Wendt, 2017b, S. 399). Dabei gehen Schemata über die Darstellung von Eigenschaften und Konzepten hinaus, wodurch Vorhersagen über ein bestimmtes Exemplar einer Kategorie getroffen werden (Anderson & Funke, 2007). Durch die Strukturierung, welche durch Schemata bereitgestellt wird, kann die subjektive Organisation des Lernprozesses (Tulving, 1962) und gleichzeitig der Abrufprozess der Information unterstützt werden. Dabei kann der Lerner bei einem Abrufprozess von Hinweisreizen auf Basis der Strukturierung profitieren (Craig & Bosman, 1992).

Lernen wird hier im Sinne von verbalem Lernen untersucht, welches das Einprägen von Symbolzuordnungen, wie Wörter, in Sequenzen oder paarweise durch Assoziationsbildung beinhaltet (Becker-Carus & Wendt, 2017a). In der Forschung wird verbales Lernen üblicherweise mit Paarassoziationsaufgaben, Wiedererkennungsaufgaben oder freiem Erinnern erfasst (siehe Kahana, 2014). Aufgrund der zeitlichen und räumlichen Gebundenheit des Lerninhalts an den Erhebungskontext wird Lernen als Prozess des episodischen Gedächtnisses untersucht. Im Folgenden werden die Lernmechanismen durch Akquisition (hier operationalisiert durch Lerndurchgänge) und Lernen durch Abruf untersucht (hier operationalisiert durch Testdurchgänge). Durch den Fokus auf das episodische Gedächtnis, verbales Lernen und Lernmaterial als Gegenstand des

Lernprozesses ist Lernerfolg quantifizierbar (richtig vs. falsch gelöstes Item) und bewertbar (mehr Inhalte richtig erinnern sowie längeres Behalten stellen eine bessere Lernleistung dar). Nachdem der Lernprozess kurz beschrieben wurde, wird im Folgenden auf den Testeffekt eingegangen.

### *1.1.2. Was ist Testen?*

Der Testeffekt („testing effect“) beschreibt den Befund, dass der erfolgreiche Abruf beim Testen von Lerninhalten den erfolgreichen Abruf eben dieser Lerninhalte zu einem späteren Zeitpunkt fördert (Roediger & Karpicke, 2006a, 2006b). Mehr noch, wer nach einem Lerndurchgang einen Test bearbeitet, erzielt häufig zu einem späteren Zeitpunkt eine bessere Abrufleistung als jemand, der einen zusätzlichen Lerndurchgang anstatt des Tests absolviert (Eisenkraemer, Jaeger & Stein, 2013; Karpicke, 2017; Roediger & Karpicke, 2006b; Rowland, 2014). Das bedeutet, Lernen findet nicht nur während des Einprägens, sondern auch während des Abrufs (abrufbasiertes Lernen) statt. Dieser Effekt lässt sich nach wenigen Minuten (bspw. Rowland & DeLosh, 2015; Roediger & Smith, 2012), aber auch nach mehreren Tagen beobachten (bspw. Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013; Karpicke & Roediger, 2008).

Der Testeffekt konnte für unterschiedliche Materialien wie z.B. Wortpaare (Carrier & Pashler, 1992), Bilder (Wheeler & Roediger, 1992) und Texte (Roediger & Karpicke, 2006b) gezeigt werden (siehe auch Rickard & Pan, 2018). Zudem wurden die Effekte des erfolgreichen Abrufs in unterschiedlichen Stichproben, z.B. bei Kindern (Agarwal, Roediger, McDaniel, & McDermott, 2010) oder Studierenden (Whiffen & Karpicke, 2017), und in unterschiedlichen Umgebungen wie im Labor (Karpicke & Roediger, 2008) oder der Schule (Leeming, 2002) repliziert. Darüber hinaus profitieren Lerner vom Effekt des Testens auch bei unterschiedlichen Formen von Gedächtnisleistungen, wie dem freien Abruf (z.B.

Karpicke & Roediger, 2007), dem Abruf mit Hinweisreizen (z.B. Carpenter, Pashler & Vul, 2006), sowie beim Wiedererkennen (z.B. Glover, 1989) (in Kapitel 2.2 und 3.1.1 wird genauer auf den Testeffekt mit verschiedenen Lernmaterialien eingegangen).

In „klassischen“ Experimenten zum Testeffekt wird die Lernleistung einer Gruppe mit mehreren Testdurchgängen mit einer Gruppe mit nur einem finalen Testdurchgang verglichen. Nach einem initialen Lerndurchgang folgt entweder ein weiterer Lerndurchgang oder ein Testdurchgang–wobei die Bedingung mit zwei Lerndurchgängen die Kontrollgruppe darstellt. Im Testdurchgang haben die Probanden die Möglichkeit, das gelernte Material abzurufen. Um die beiden Lernbedingungen zu vergleichen, findet abschließend ein finaler Test statt, der die Lernleistungen erfasst. Dieses Vorgehen kann durch die Veränderung der Anzahl der Lern- und Testdurchgänge, des Lernmaterials und durch weitere Merkmale der Operationalisierung (bspw. Feedback oder Abstände der Testdurchgänge) modifiziert werden (siehe Karpicke, 2017).

In ihrer wegweisenden Studie untersuchten Roediger und Karpicke (2006b) den Testeffekt beim Lernen bei pädagogisch relevantem Lernmaterial ohne Feedback. Hierfür lernten in einem ersten Experiment 120 Studierende (im Alter von 18 bis 24 Jahre) zwei Texte zu unterschiedlichen wissenschaftlichen Themen. Nach dem ersten sieben-minütigen Lerndurchgang folgte entweder ein weiterer Lerndurchgang (Kontrollbedingung) oder ein sieben-minütiger Testdurchgang, in dem das Material erinnert werden sollte (Experimentalbedingung). Dabei bearbeiten alle Probanden beide Bedingungen. Zwischen den Probanden variierte die Dauer bis zum finalen Testdurchgang (entweder fünf Minuten, zwei Tage oder eine Woche nach dem Experiment). Die Ergebnisse der Studie zeigen eine signifikante Interaktion der Lernbedingung mit dem finalen Abfrageintervall ( $F(2, 117) = 32.10, \eta_p^2 = .35$ ). Zwar wurden im Testdurchgang nach fünf Minuten durchschnittlich mehr Wörter in der Kontrollbedingung erinnert, als in der Lern-Testbedingung ( $t(39) = 3.22, d =$

0.52), allerdings zeigte sich in den späteren Abfragedurchgängen ein umgekehrtes Muster – nach zwei Tagen sowie nach einer Woche erinnerten sich die Probanden an mehr Wörter aus der Lern-Testbedingung (zwei Tage später:  $t(39) = 6.97$ ,  $d = 0.95$ ; eine Woche später:  $t(39) = 6.41$ ,  $d = 0.83$ ).

In einem zweiten Experiment konnte dieses Muster repliziert werden (Roediger & Karpicke, 2006b). Ziel dieses weiteren Experiments war es, die Ergebnisse des ersten Experimentes zu replizieren und darüber hinaus den Testeffekt anhand mehrerer Lern- und Testdurchgänge zu untersuchen. Hierfür bearbeiteten 180 Studierende (zwischen 18 und 24 Jahren) dieselben Geschichten wie in Experiment 1 in einer von drei Lernbedingungen. Die Teilnehmer lernten das Material entweder in 1) vier fünf-minütigen Lerndurchgängen (SSSS), 2) in drei Lerndurchgängen mit einem anschließenden Testdurchgang (SSST), 3) in einem Lerndurchgang und drei darauffolgenden Testdurchgängen (STTT). Die Probanden bearbeiteten entweder fünf Minuten oder eine Woche nach dem Lernexperiment einen finalen Testdurchgang. Hierbei zeigte sich wie im ersten Experiment eine signifikante Interaktion der Lernbedingung mit dem Abfrageintervall ( $F(2, 174) = 18.48$ ,  $\eta_p^2 = .18$ ) (siehe Abbildung 1). Diese Interaktion zeigt, dass wiederholtes Lernen vor allem die kurzzeitige Lernleistung unterstützt, wohingegen wiederholtes Testen den langfristigen Abruf unterstützt und dem Vergessen vorbeugt (Unterschied zwischen der SSSS- und der STTT-Bedingung eine Woche später:  $t(58) = 4.78$ ,  $d = 1.26$ ). Auch wenn nach einer Woche die Probanden mit drei Testdurchgängen im Durchschnitt die beste Lernleistung erzielten, unterschied sich die STTT- von der SSST-Bedingung nicht signifikant ( $t(58) = 1.21$ ,  $d = 0.31$ ). Jedoch wurde in der SSST-Bedingung anteilig mehr vergessen (Differenz der Lernleistung nach fünf Minuten und nach einer Woche, geteilt durch die Lernleistung nach fünf Minuten) im Vergleich zur STTT-Bedingung, bei der nach fünf Minuten 7% des gesamten Materials weniger erinnert wurde (relativer Vergessens-Anteil SSST 28%, STTT 14%). Ein weiterer/späterer

Testdurchgang wäre hier nötig, um zu überprüfen, ob die drei Testdurchgänge in der STTT stärker vor dem Vergessen schützen als die Kombination aus drei Lern- und einem Testdurchgang. Durch die beiden Experimente konnten die Autoren die Befundlage zu den positiven Effekten von Tests (ohne Feedback) auf die Lernleistung weiter stützen und für pädagogisch relevantes Material erweitern (siehe Roediger & Karpicke, 2018).

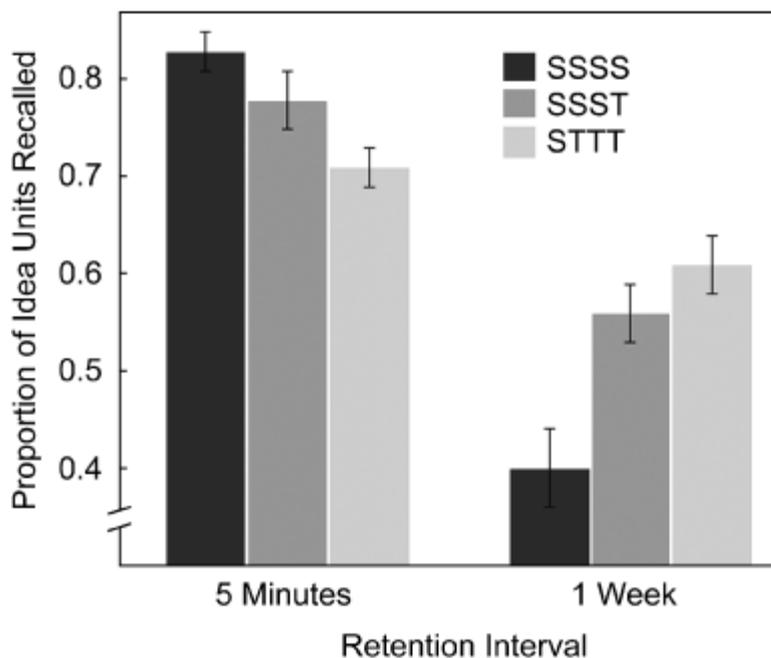


Abbildung 1. Ergebnisse von Roediger und Karpicke (2006b, S. 253). Durchschnittliche Anzahl korrekt erinnerter Einheiten in einem finalen Test nach fünf Minuten oder nach einer Woche. S = Lerndurchgang (Study); T = Testdurchgang (Test). Fehlerbalken stellen die Standardfehler der Mittelwerte dar. Aus “Test-enhanced learning: Taking memory tests improves long-term retention” von H. L. Roediger & J. D. Karpicke, 2006, *Psychological Science*, 17(3), S. 253. Copyright 2006 bei SAGE Publications. Wiedergabe mit Genehmigung.

### 1.1.2.1. Wann sind Testeffekte beobachtbar?

In Studien wird weitestgehend einheitlich berichtet, dass Testdurchgänge anstatt von Lerndurchgängen einen positiven Einfluss auf einen verzögerten Abruf nach Tagen bzw. Wochen haben (bspw. Dunlosky et al., 2013; Karpicke & Roediger, 2008). Hingegen zeigen sich bei Abrufversuchen nach wenigen Minuten gemischte Ergebnisse. So fanden Karpicke und Roediger (2006b) in ihren beiden Experimenten nach fünf Minuten höhere Abrufleistungen bei den Bedingungen mit wiederholten Lerndurchgängen im Gegensatz zu den Testbedingungen (mit einem Lerndurchgang). Gleichwohl können sich Testeffekte schon nach wenigen Sekunden zeigen, wenn die Abrufleistung im ersten Testdurchgang relativ hoch ist (Rowland & DeLosh, 2015). In Experiment 1 von Rowland und DeLosh (2015) wurden Wortlisten nach einem anfänglichen Lerndurchgang entweder ein zweites Mal gelernt oder in einem Testdurchgang ohne Feedback abgefragt. Im Testdurchgang wurden der erste oder die ersten beiden Buchstaben des Wortes als Hinweisreiz verwendet (durchschnittlich wurden in der Testbedingung nach dem ersten Testdurchgang 36% der Wörter erfolgreich erinnert). In einem abschließenden Testdurchgang wurden die Wörter nach einer, vier oder acht Minuten abgefragt. Hier zeigte sich kein signifikanter Haupteffekt der Testbedingung. In einem zweiten Experiment mit drei anfänglichen Lerndurchgängen (und damit auch mit einer höheren durchschnittlichen Testleistung von 53% bei der ersten Abfrage der Testbedingung) konnten die Ergebnisse des ersten Experiments repliziert werden. Wird nun durch die Erinnerungsleistung im ersten Testdurchgang der Testbedingung auf durchschnittlich 88% erhöht–durch die Darbietung von mehr Hinweisreizen–, zeigen sich schon nach 30, 90 und 240 Sekunden höhere Abrufleistungen bei der Testbedingung im Gegensatz zum wiederholten Lernen ( $\eta_p^2 = .35$ ). Die scheinbar widersprüchlichen Ergebnisse der drei Experimente–bezüglich des Beobachtungszeitpunkts des Testeffekts–können durch die unterschiedliche Ausgangsleistung im ersten

Testdurchgang erklärt werden (Rowland & DeLosh, 2015). Nachdem der erfolgreiche Abruf beim Testen von Lerninhalten den erfolgreichen Abruf eben dieser Lerninhalte fördert, zeigen sich größere Effekte von Tests, je mehr Wörter in einem ersten Test erinnert werden. Demnach hängt es von der Art der Lernaufgabe ab, ob die Erinnerungsleistung in der Testbedingung schon nach wenigen Minuten höher ist als in der Lernbedingung oder erst nach einem Retentionsintervall von einigen Tagen sichtbar wird.

Auch bei Experimenten mit mehreren Lern-Test-Durchgängen zeigen sich höhere Lernkurven in der Testbedingung als in der Lernbedingung (Roediger & Smith, 2012). Die Testbedingung kann hierbei aus mehreren hintereinander folgenden Lern-Test-Durchgängen (Standardlernkurven) bestehen und bspw. mit mehreren Lerndurchgängen ohne dazwischenliegende Testungen verglichen werden (Roediger & Smith, 2012). Hierbei zeigten sich höhere Lernraten in der Standardbedingung (mit Testeffekt) und demnach auch höhere Lernleistung am Ende der Experimente (Roediger & Smith, 2012). Karpicke und Roediger (2007) haben eine solche Standardlernkurve (zehn Lern-Test-Durchgänge) mit einer Lernkurve von fünf Durchgängen, jeder Durchgang bestand aus drei Lernphasen und einer Testphase (SSST), verglichen und fanden einen Vorteil der Standardbedingung bei einem verzögerten Abruf nach einer Woche. Im Gegensatz zu den Untersuchungsdesigns, in welchen ein Lern-Test-Durchgang einem doppelten Lerndurchgang gegenübergestellt werden, kumulieren sich die Testeffekte bei mehreren Lern-Test-Durchgängen. Das bedeutet, dass in aufeinanderfolgenden Lern-Test-Durchgängen durch die Testeffekte vorheriger Tests mehr Wörter erinnert werden, wobei der Lerner bei weiteren Abfrageversuchen wiederum von diesen Testeffekten profitieren kann. Besonders bei Designs, in welchen mehrere Lerndurchgänge durch Testdurchgänge ersetzt werden (STTT vs. SSSS, bspw. Roediger & Karpicke, 2006b), werden keine unmittelbar höheren Abrufleistungen gegenüber der Lernbedingung erwartet (bspw. nach 5 Minuten). In solchen

Designs ist die Erinnerungsleistung in einem ersten Testdurchgang nach einem Lerndurchgang üblicherweise geringer als die erwartete potentielle Erinnerungsleistung nach vier Lerndurchgängen in der Lernbedingung. Dementsprechend würde man vermuten, dass unmittelbar beobachtbare Testeffekte vor allem bei Designs auftreten, die in der Anzahl von Lerndurchgängen in der Lern- und in der Testbedingung vergleichbar sind oder gleich viele Lerndurchgänge aufweisen (bspw. Roediger & Smith, 2012).

Falls die Anzahl der Lerndurchgänge in der Testbedingung zu gering ist oder relativ wenig Wörter im ersten Testdurchgang erinnert werden, können zusätzliche bedingte Analysen unmittelbare Effekte aufdecken. Rowland und DeLosh (2015) betrachten in bedingten Analysen der finalen Erinnerungsleistung in der Testbedingung nur diejenigen Items, welche zuvor im ersten Testdurchgang erfolgreich erinnert wurden (und vergleichen diese mit allen erneut gelernten Items der Lernbedingung). So finden die Autoren in Experiment 1 und 2 höhere Lernleistungen, im finalen Test der Testbedingung bereits nach wenigen Minuten. Allerdings ergeben sich bei bedingten Analysen zusätzliche Schwierigkeiten. Einerseits könnten die erinnerten Items im ersten Testdurchgang selektiv zusammenhängen (bspw. besonders leichte Items des Lernmaterials) (Rowland & DeLosh, 2015). Andererseits spielt die anfängliche Lernleistung eine wesentliche Rolle für das Ausmaß des Testeffekts (Karpicke, 2017; Marsh, Agarwal & Roediger, 2009; Meyer & Logan, 2013), was durch ipsative Messungen schwierig abzubilden ist.

Zusammenfassend kann gesagt werden, dass der Testeffekt bei unterschiedlichem Material, Stichproben, Umgebungen und für unterschiedliche Formen der Gedächtnisleistung nach wenigen Minuten sowie Tage später auftreten kann. Die Effekte von Testdurchgängen können sich nicht nur nach Tagen zeigen, sondern bereits wenige Sekunden nach den Lern-Test-Durchgängen. Dies ist allerdings nur dann beobachtbar, wenn spezifische Eigenschaften des Lernexperiments zutreffen (u.a. Anzahl Lern-Test-

Durchgänge, Schwierigkeit der Abfrage). Im Allgemeinen kann man folgern, dass Testdurchgänge und Abfrageversuche sehr hilfreich für den Lernprozess sein können. Im Folgenden Kapitel werden verschiedene Erklärungsansätze des Testeffekts vorgestellt und anschließend gegenübergestellt, um einen theoretischen Rahmen für die vorgestellten Experimente zu definieren.

#### *1.1.2.2. Erklärungsansätze des Testeffekts*

Es gibt unterschiedliche theoretische Ansätze, die auf eine Erklärung des Testeffekts abzielen. Diese schließen sich nicht unbedingt gegenseitig aus. Die verschiedenen Erklärungsansätze basieren auf 1) der Modifizierung der Gedächtnisspur, 2) der transferangemessenen Verarbeitung, 3) dem elaborierenden Abruf, und 4) dem episodischen Kontext. Im Folgenden werden ausgewählte Theorien zur Erklärung des Testeffekts dargestellt und anschließend kritisch gewürdigt.

#### *Modifizierung der Gedächtnisspur*

Die frühesten Erklärungsansätze des Testeffekts fokussieren auf die Veränderung der Gedächtnisspur durch Abrufversuche. Demnach zeigen Tests nicht nur, welche Inhalte gelernt wurden, sondern modifizieren auch das Gedächtnis (Bjork, 1975). Hierbei wird angenommen, dass die gelernten Informationen im Gedächtnis repräsentiert und diese Repräsentationen verstärkt werden können. Diese Verstärkung kann beispielsweise durch wünschenswerte Schwierigkeiten operationalisiert werden („Desirable Difficulties“, Bjork 1994; Bjork & Bjork, 2011). Zu diesen wünschenswerten Schwierigkeiten gehört auch der Abrufversuch. Je schwieriger ein erfolgreicher Abruf, desto tiefer die Verarbeitung des Lerninhaltes (siehe auch Pyc & Rawson, 2009, Rowland, 2014).

Aus dem Ansatz der modifizierbaren Repräsentationen im Gedächtnis entwickelte sich die „neue Theorie der Nichtnutzung“ („New Theory of Disuse“, Bjork & Bjork, 1992). In dieser Theorie wird zwischen Speicherstärke („Storage Strength“) und Abrufstärke („Retrieval Strength“) unterschieden. Die Speicherstärke bezeichnet die Qualität des gelernten Inhalts und hat keinen direkten Effekt auf die Erinnerungsleistung. Wohingegen die Abrufstärke beschreibt, wie verfügbar ein Item im Gedächtnis zu einem bestimmten Zeitpunkt ist. Die zeitlich stabile Speicherstärke mindert den Rückgang der Abrufstärke über die Zeit (d.h. Speicherstärke reduziert Vergessen). Lernen und Abrufen von Informationen verstärkt die Speicherstärke sowie die Abrufstärke, allerdings ist der erfolgreiche Abruf wirksamer als ein Lerndurchgang. Die Zunahme in der Abrufstärke durch Lern- oder Testdurchgänge wird durch das Niveau der Speicher- und Abrufstärke beeinflusst. Hierbei verbessert sich die Abrufstärke durch einen erfolgreichen Abruf mit steigender Speicherstärke, verringert sich jedoch mit steigender Abrufstärke des Lerninhalts. Das bedeutet, dass der positive Einfluss des erfolgreichen Abrufs auf die Erinnerungsstärke größer wird, je schwieriger der Abrufversuch ist (geringere Abrufstärke) und je besser der Lerninhalt im Gedächtnis repräsentiert ist (höhere Speicherstärke; Bjork & Bjork, 1992; Storm, Friedman, Murayama & Bjork, 2014).

### *Bifurkationsmodell*

Ein weiteres Modell zur Erklärung des Testeffekts ist das Bifurkationsmodell, welches mit der „neuen Theorie der Nichtnutzung“ verwandt ist (Bifurcation Model, Kornell, Bjork, & Garcia, 2011). In diesem Modell wird angenommen, dass Testdurchgänge ohne Feedback besonders die Gedächtnisstärke von erfolgreich abgerufenen Items fördern, während nicht abgerufene Items durch den Testdurchgang nicht weiter verstärkt werden. Die Gedächtnisstärke stellt dabei ein Kontinuum dar, welches bei der Überschreitung eines

Grenzwerts dazu führt, dass ein Item erinnert wird. Durch diese Veränderung der Gedächtnisstärke werden die Items zunehmend in eine Gruppe mit stärkerer (nämlich die erfolgreich abgerufenen) und eine mit schwacher Gedächtnisstärke (nicht abgerufene Items) unterteilt. Im Gegensatz dazu, fördern Lerndurchgänge die Gedächtnisstärke aller Items (erfolgreich abgerufen und nicht abgerufen); dies jedoch in einem insgesamt geringeren Umfang als durch Testdurchgänge. Die dahinterliegende Annahme ist, dass 1) die Gedächtnisstärke kontinuierlich verbessert werden kann und 2) ein Item erfolgreich abgerufen wird, sobald die Gedächtnisstärke einen bestimmten Grenzwert überschreitet. Somit kann anhand des Bifurkationsmodells der Effekt erklärt werden, dass der Testungsvorteil, mit zeitlichem Abstand zum finalen Test, wächst. Durch den erfolgreichen Abruf verstärkt sich die Gedächtnisstärke der Items über den Grenzwert hinaus, wohingegen die Gedächtnisstärke von Items, die nur gelernt werden, im geringeren Maße über den spezifischen Grenzwert hinaus verstärkt werden. Wenn die Gedächtnisstärke im Laufe der Zeit abnimmt, fallen Items mit hoher Gedächtnisstärke erst später unter den Grenzwert–und werden dementsprechend später nicht mehr erinnert. Demnach profitieren besonders Lerner mit hoher Abrufleistung im ersten Testungsdurchgang vom Testeffekt (vgl. Marsh et al., 2009; Meyer & Logan, 2013; Rowland, 2014).

### *Transferangemessene Verarbeitung*

Mit dem Ansatz der transferangemessenen Verarbeitung (Transfer Appropriate Processing, Morris, Bransford & Franks, 1977), wurde ein weiterer Erklärungsansatz für die Lerneffekte von Testdurchgängen vorgestellt. Nach der transferangemessenen Verarbeitung wird die Lernleistung dann verbessert, wenn bei einem Abruf vergleichbare kognitive Prozesse wie beim Lernen benötigt werden und auftreten (Morris et al., 1977; Veltre, Cho & Neely, 2015). Der Ansatz der transferangemessenen Verarbeitung kann auf den Testeffekt

erweitert werden (Veltre, Cho & Neely, 2015). Demnach ist der Testeffekt umso größer, je ähnlicher die Prozesse beim initialen Abruf den Prozessen bei einem späteren finalen Test sind. Strenggenommen sollte ein Test dann den größten Einfluss auf die Lernleistung haben, wenn dieser identisch zu dem finalen Test ist–und demnach auch die gleichen kognitiven Prozesse erfordert (Karpicke, Lehman & Aue, 2014). Allerdings konnte in mehreren Studien der Ansatz der transferangemessenen Verarbeitung nicht unterstützt werden (vgl. Carpenter & DeLosh, 2006; für einen Überblick siehe Rowland, 2014). Beispielsweise zeigte sich, dass bei einer abschließenden Multiple-Choice Aufgabe die Items, welche zuvor mit Kurzantwortformaten abgefragt wurden (mit Feedback), besser abschnitten als Items, die zuvor im Multiple-Choice Format (mit Feedback) getestet wurden (Kang, McDermott & Roediger, 2007; McDaniel, Anderson, Derbish & Morrisette, 2007).

#### *Ansatz des elaborierenden Abrufs*

Im Ansatz des elaborierenden Abrufs (*Elaborative Retrieval Hypothesis*; Carpenter, 2009) wird angenommen, dass bei einem Abrufversuch zusätzlich weitere Informationen aktiviert werden, die mit dem Zielinhalt verwandt sind und den späteren Abruf erleichtern können. Demnach werden bei Abrufversuchen–im Gegensatz zu Lerndurchgängen–Gedächtnisinhalte aktiviert, die weitere verwandte Konzepte aktivieren können (vgl. Theorie zur ausbreitenden Aktivierung semantischer Prozesse von Collins & Loftus, 1975). Dies geschieht (bei einem erfolgreichen Abruf) solange, bis ein semantisches Netzwerk aktiviert wird, welches Verbindungen zum Zielinhalt hat (Carpenter, 2009). Durch diese vertiefte Informationsverarbeitung werden durch Hinweisreize (v.a. beim Paarassoziationslernen) semantisch verwandte Informationen aktiviert, die mit dem erfolgreichen Abruf verarbeitet werden. Diese neuen Informationen bilden daraufhin einen elaborierten Gedächtnispfad, der zukünftige Abrufversuche unterstützen kann. Hierbei gilt (wie schon bei den Ansätzen zur

Modifizierung der Gedächtnisstärke oder dem Bifurkationsmodell), dass mit zunehmender Schwierigkeit des erfolgreichen Abrufs ebenfalls die positiven Effekte des Testens zunehmen. Dieser Zusammenhang wird durch die vertiefte, breitere Aktivierung zusätzlicher Informationen erklärt. Diese zusätzlichen Informationen können bei einem späteren Abruf als zusätzliche Gedächtnispfade im semantischen Netzwerk dienen. In einer späteren Studie wurde der Ansatz des elaborierenden Abrufs durch die Hypothese der Effektivität von Mediatoren (*Mediator Effectiveness Hypothesis*; Pyc & Rawson, 2010) dahingehend erweitert, dass die beim Abrufversuch gebildeten Informationen genauer beschrieben werden. Diese neuen Informationen können demnach Mediatoren sein (bspw. ein Wort oder ein Konzept), welche einen Hinweisreiz mit der Zielinformation verbinden (Carpenter, 2011; Pyc & Rawson, 2010). Mithilfe von Paarassoziationsaufgaben konnte der Ansatz des elaborierenden Abrufs überprüft und mehrfach unterstützt werden (Carpenter, 2009; Carpenter, 2011; Carpenter & Yeung, 2017). Beispielsweise wurden bei den Experimenten von Carpenter (2011) Wortpaare, die schwach miteinander verwandt sind (*Mother – Child*)<sup>1</sup> zweimal gelernt bzw. einmal gelernt und einmal getestet (Carpenter, 2011). In einem finalen Test wurden die Zielwörter nicht mit dem ursprünglichen Hinweisreiz abgefragt, sondern durch ein neues Hinweiswort. Dieses neue Wort ist stark mit dem vorherigen Hinweiswort assoziiert (bspw. *Father*) und stellt einen semantischen Mediator dar. Die Abrufleistung fiel hierbei bei den Lernern, welche zuvor getestet worden waren, besser aus als bei jenen, die das Material zweimal gelernt hatten. Diese Befunde können mit der Aktivierung semantisch ähnlicher Informationen beim Abrufversuch erklärt werden. Allerdings zeigen sich keine Unterschiede im Testeffekt, wenn zu einem Wortpaar ein Mediator beim Abrufversuch erinnert wurde oder nicht (Cho, Neely, Brennan, Vitrano & Crocco, 2017). Darüber hinaus

---

<sup>1</sup> Typisch verwendetes Beispiel aus Carpenter, 2011. Ein anderes Beispiel ist das Wortpaar *Donor – Heart* und der Mediator *Blood*.

zeigen weitere Studien, dass semantische Elaboration alleine nicht die Effekte des Abrufversuchs erklären kann (Lehman & Karpicke, 2016; Lehman, Smith & Karpicke, 2014).

### *Ansatz des episodischen Kontextes*

Der Ansatz des episodischen Kontextes von Karpicke, Lehman und Aue (2014) (*Episodic Context Account*) ist ein weiterer Erklärungsansatz, der verstärkt auf etablierten Modellvorstellungen der Gedächtnisforschung aufbaut. Er basiert auf vier Grundannahmen. Beim Lernen werden 1) Attribute des zu lernenden Materials *und* Attribute des Kontextes der Lernsituation gespeichert. Dazu zählen z.B. situative, räumliche oder temporale Attribute eines Kontextes. Beim späteren Abruf versucht die Person, 2) sich in den Kontext des Lerndurchgangs zurückzusetzen, um die (erinnerten) Attribute dieses Kontextes zur Unterstützung für den Suchprozess nach dem erlernten Material im Gedächtnis zu verwenden. Wenn ein Item, 3) im Kontext A gelernt wurde und anschließend im Kontext B abgerufen wird, werden die mit dem Item assoziierten Kontextattribute so verändert, dass nunmehr Eigenschaften aus beiden Kontexten (A *und* B) im Gedächtnis repräsentiert sind. Schließlich unterstützen, 4) bei einem erneuten Abrufversuch zu einem späteren Zeitpunkt eben diese Kontextattribute (Kontext A und B) den Suchprozess und führen zu einer gesteigerten Gedächtnisleistung. Demnach werden bestimmte Attribute aus dem im Gedächtnis gespeicherten aktualisierten Kontext als zusätzliche Hinweisreize (aufgrund ihrer spezifischen Verbindung zum Zielinhalt) für spätere Abrufversuche verwendet (vgl. Raaijmakers & Shiffrin, 1981). Zwar können Kontextattribute auch beim wiederholten Lernen gespeichert und verknüpft werden. Jedoch findet beim Abruf episodischer Erinnerungen die Suche nach Informationen aus dem vorherigen Kontext (Kontext A)

bewusst statt, im Gegensatz zum eher zufälligen Abrufprozess vorheriger Kontexte (Kontext A) beim wiederholten Lernen (Karpicke, Lehman, & Aue, 2014).

Um die Kernannahmen des Ansatzes des episodischen Kontextes zu überprüfen, haben Whiffen und Karpicke (2017) in drei Experimenten untersucht, ob das Zurückerinnern an den Lernkontext die Lernleistung im Vergleich zu einem weiteren Lerndurchgang steigert. Nachdem die Versuchspersonen zwei Wortlisten gelernt hatten, folgte entweder ein weiterer Lerndurchgang mit den Wörtern aus beiden Listen oder eine Diskriminierungsaufgabe. In der Diskriminierungsaufgabe bestand die Aufgabe für den Probanden darin, zu entscheiden, ob das präsentierte Wort aus der ersten oder der zweiten Wortliste stammt. In beiden Bedingungen wurden alle Wörter gleich lang präsentiert; allerdings mussten sich die Personen in der Diskriminierungsbedingung in die vorherigen Lerndurchgänge zurückversetzen, um zu bestimmen, in welcher Liste das Zielwort präsentiert wurde. In allen drei Experimenten wurden in der Diskriminierungsbedingung mehr Wörter erinnert als in der Lernbedingung ( $d = 0.61$ ). Zusätzlich konnte gezeigt werden, dass im finalen Testdurchgang die Wörter anhand des episodischen Kontextes organisiert wurden (Wiedergabe der Wörter stärker nach Listenzugehörigkeit strukturiert), wenn die Probanden sich zuvor an den vorherigen Lernkontext zurückerkennen mussten (Diskriminierungsaufgabe) (siehe auch Lehman, Smith & Karpicke, 2014). Diese Ergebnisse zeigen, dass durch das Zurückversetzen in die Lernsituation während eines Abrufversuchs die Lernleistung verbessert werden kann. Somit wird zwar eine Grundannahme des Ansatzes des episodischen Kontextes gestützt, allerdings erklärt dies noch nicht, warum genau diese Vorgänge des Zurückversetzens in einen früheren Kontext für den Testeffekt verantwortlich sind.

Der Ansatz des episodischen Kontextes ermöglicht es, viele Phänomene des Testeffekts adäquat zu erklären und dahinterliegende Mechanismen zu beschreiben (für eine

ausführlichere Zusammenfassung siehe Karpicke, Lehman, & Aue, 2014). Jedoch merken Rawson, Vaughn und Carpenter (2015) an, dass der Ansatz des episodischen Kontextes den Testeffekt bei der Verwendung neuer Hinweisitems mit alten Zielinhalten nicht eindeutig erklären kann, aufgrund fehlenden vorherigen Kontextes.

### *Zusammenfassung und Würdigung der Theorien*

Die vorgestellten Theorien schließen sich nicht zwangsläufig gegenseitig aus und haben keinen rein exklusiven Anspruch. Stellenweise beleuchten die Ansätze spezifische Aspekte des Testeffekts aus verschiedenen Perspektiven–wie der Modifizierung der Gedächtnisspur, der transferangemessenen Verarbeitung, dem elaborierenden Abruf, oder dem episodischen Kontext–und mit unterschiedlicher Genauigkeit. Wie bereits erwähnt fördern Testdurchgänge den Lernerfolg. Aus entwicklungspsychologischer Perspektive stellt sich die Frage, ob Personen verschiedener Altersgruppe von den Testeffekten profitieren können und ob dieser von Prozessen der kognitiven Alterung beeinflusst wird. Jedoch wurden die meisten Studien zum Testeffekt bisher mit jüngeren Erwachsenen oder Schülern durchgeführt und das höhere Erwachsenenalter weitgehend ausgelassen. Dementsprechend wird auch in keiner der Theorien des Testeffekts der Aspekt der Altersveränderungen mit den Effekten des erfolgreichen Abrufs direkt in Zusammenhang gebracht. Gleichwohl bieten die Theorien Ansatzpunkte, um Auswirkungen altersbezogener Veränderungen auf den Testeffekt zu erschließen. Im Folgenden werden die Theorien kritisch gewürdigt, um im Anschluss daran altersspezifische Hypothesen zur Effektivität des Testeffekts im jüngeren und höheren Erwachsenenalter abzuleiten.

Den meisten Erklärungsansätzen des Testeffekts ist gemeinsam, dass sie besonders große Testeffekte bei schwierigem und erfolgreichem Abruf von Lerninhalten erwarten (abgesehen vom Ansatz der transferangemessenen Verarbeitung). Dieser Zusammenhang

wird jedoch unterschiedlich erklärt: durch wünschenswerte Schwierigkeiten des Abrufversuchs (Bjork, 1994), tiefere semantische Verarbeitung (Carpenter, 2009) oder durch das stärker ausgeprägte Zurückversetzen in den vorherigen Kontext und damit einhergehendes Erinnern der Kontextattribute (Karpicke, Lehman, & Aue, 2014). Die Effekte des erfolgreichen Abrufs schwieriger Lerninhalte werden in dem Ansatz der transferangemessenen Verarbeitung nicht spezifisch erklärt. Im anwendungsorientierten Kontext hilft der Ansatz der transferangemessenen Verarbeitung bei dem Verständnis, dass Tests unterschiedlich gut mit den Zielen eines Lerners zusammenpassen (bspw. hilft beim Erlernen einer Sprache–mit dem Ziel sich im Urlaub zu unterhalten–eine Testsituation, in der tatsächlich mit einem Einheimischen gesprochen wird, anstatt einer Multiple-Choice- oder Lückentextaufgabe).

In der „neuen Theorie der Nichtnutzung“ (Bjork & Bjork, 1992) werden v.a. die Effekte des erfolgreichen Abrufs beschrieben, jedoch weniger die dahinterliegenden Mechanismen. Ähnlich verhält es sich bei dem verwandten Bifurkationsmodell (Kornell et al., 2011), welches die Effekte des Testens vor allem als statistisches Modell beschreibt. Bei der „neuen Theorie der Nichtnutzung“ sowie beim Bifurkationsmodell wird beschrieben, dass beim erfolgreichen Abruf die Gedächtnisspur in größerem Umfang verstärkt wird als beim Lernen; dieser Einfluss hängt ebenfalls von der Schwierigkeit des Abrufs ab. Diese Befunde wurden vielfach in Studien gefunden und repliziert (Rowland, 2014). Daher sind weitere Theorien zur Erklärung dieser Muster beim Testeffekt notwendig, welche darüber hinaus Schlüsse auf Lerner mit verschiedenen Merkmalen (bspw. Alter) zulassen.

Eine solche Erklärung wird im Ansatz des elaborierenden Abrufs geboten (Carpenter, 2009; Carpenter, 2011). Demnach werden während eines erfolgreichen Abrufversuchs Mediatoren gebildet (Pyc & Rawson, 2010), die bei einem späteren Test den Abruf unterstützen können. Diese Mediatoren können Wörter, Sätze oder Konzepte sein, die einen

Hinweisreiz mit einem Zielinhalt verbinden. Diese breite Auslegung kann jedoch stark von den bekannten und auch aktiv verwendeten Strategien einer Person beeinflusst werden. Vor allem dann, wenn während eines Testdurchgangs bewertet wird, ob ein Mediator effektiv ist oder nicht (vgl. Pyc & Rawson, 2010). Zwar spielen bei Lernprozessen immer auch das Wissen und die Expertise von Lernstrategien eine zentrale Rolle, jedoch ist es hilfreich, zuerst die basalen Erklärungsmechanismen des Testeffekts zu beschreiben (wie wirken sich Tests ohne die Einflüsse verschiedener Lernstrategien aus), um diese anschließend in Zusammenhang mit Lernstrategien zu betrachten (letzteres wurde noch nicht ausreichend in der aktuellen Forschung zum Testeffekt beachtet).

Der Ansatz des episodischen Kontextes von Karpicke, Lehman und Aue (2014) ist ein Erklärungsansatz, der verstärkt auf etablierten Modellvorstellungen der Gedächtnisforschung aufbaut. Hier spielt das Zurückversetzen in vorherige Lern- und Testsituationen sowie das Erinnern passender Kontextattribute eine wichtige Rolle für den Testeffekt. Je unterschiedlicher Lern- und Testkontexte sind, desto effektiver ist das Zurückversetzen in den Lernkontext bezogen auf die Effektivität des Testeffekts. Nach dieser Annahme wären in Laborexperimenten, welche innerhalb eines Settings mehrere standardisierte Lern-Test-Durchgänge beinhalten, die Auswirkungen des Testeffekts eingeschränkt. Dennoch zeigen sich bei wiederholten Lern- und Testdurchgängen große Effektstärken (bspw. Karpicke & Roediger, 2006b; Roediger & Smith, 2012). Whiffen und Karpicke (2017) haben ebenfalls gezeigt, dass sich in Laborexperimenten das Zurückversetzen in die Lernsituationen positiv auf den Lernerfolg sowie auf die interne Organisation des Lernmaterials auswirkt. Der Ansatz des episodischen Kontextes beschreibt viele Phänomene, welche mit dem Testeffekt zusammenhängen (wie bspw. den Zusammenhang mit schwierigem und zeitlich verzögerten Abruf; der Auswirkung initialer Lernleistung auf den Testeffekt) und ist dabei so spezifisch, dass sich dieser mit Theorien

des Alterns verbinden lässt (siehe Kapitel 1.2.2). Daher wird im Folgenden der Ansatz des episodischen Kontextes als Erklärung für den Testeffekt angenommen.

## *1.2. Lernen und Testen im Alter*

### *1.2.1. Lernen im Alter*

In der Forschung konnten altersbedingte Abnahmen im episodischen Gedächtnis weitestgehend konsistent gezeigt und repliziert werden (für einen Überblick siehe Park & Festini, 2017; Verhaeghen et al., 1993). Dabei gibt es verschiedene Erklärungsansätze zur kognitiven Alterung (siehe bspw. Park & Festini, 2017). Eine der etablierteren Theorien, welche den Ansatz des episodischen Kontextes (Karpicke et al., 2014) um den Aspekt des Alters ergänzt, ist die Assoziationsdefizithypothese (Naveh-Benjamin, 2000). Im Folgenden wird diese Hypothese eingeführt und anschließend weitere Theorien zu kognitiven Altersunterschieden im episodischen Gedächtnis präsentiert.

### *Assoziations-Defizit-Hypothese*

Naveh-Benjamin (2000) sieht in den im Alter zunehmenden Schwierigkeiten Gedächtnisinhalte assoziativ zu verknüpfen eine mögliche Erklärung der Abnahme der Gedächtnisleistung (Naveh-Benjamin, 2000; Naveh-Benjamin & Mayr, 2018). In der sogenannten Assoziations-Defizit-Hypothese (ADH) (Naveh-Benjamin, 2000) wird angenommen, dass im höheren Alter nicht nur die Attribute eines Lernkontextes (z.B. semantischer, räumlich-dinglicher sowie temporaler Kontext) beim Einprägen weniger erfolgreich miteinander verknüpft werden als im jungen Alter, sondern auch, dass diese Attribute bei einem Abrufversuch weniger zugänglich sind. In einer Metaanalyse konnten Old und Naveh-Benjamin (2008)–v.a. bei intentionalem Abruf–für die Quelle, den Kontext, die zeitliche Reihenfolge, die Orientierung im Raum und Paarverbindungen als Attribute von

Gedächtnisinhalten altersbedingte, assoziative Defizite zeigen. Demnach zeigen ältere Personen höhere Defizite im Gedächtnis, wenn sie sich assoziative Informationen einprägen sollen (bspw. Verknüpfung von Lernitems oder Kontexten). Im Gegensatz dazu, sind die Beeinträchtigungen für nicht zu verknüpfende Informationen geringer—bspw. nur ein Wort eines Wortpaares (Old & Naveh-Benjamin, 2008).

Aus den Annahmen der ADH lassen sich spezifische Hypothesen bezüglich der Altersunterschiede in der Lernleistung bei verschiedenen Lernaufgaben ableiten (siehe Naveh-Benjamin, 2000). Aufgaben mit Abrufhinweisen (bspw. Paarassoziationsaufgaben) sind aufgrund der Enkodierung und des Abrufs spezifischer Assoziationen von nicht verwandten Wortpaaren besonders sensitiv gegenüber Altersunterschieden. Im Gegensatz zu Wiedererkennungsaufgaben, werden in Aufgaben zum freien Erinnern (bspw. Lernen und Abrufen einer Wortliste) größere Altersunterschiede erwartet, da letztere indirekt assoziationspezifische Operationen für die Enkodierung und den Abruf benötigen (bspw. Hinweisreize für den Abruf eines Wortes generieren oder mehrere Worte während des Lernens miteinander verknüpfen).

Die vielseitige Anwendbarkeit der ADH konnte in vielen Studien mehrfach bestätigt werden (für einen Überblick siehe Old & Naveh-Benjamin, 2008; Naveh-Benjamin & Mayr, 2018) und hat sich somit als Hypothese zu kognitiven Altersunterschieden im episodischen Gedächtnis für assoziative Informationen etabliert. Jedoch spezifiziert die ADH weder die Ursache dieser Altersunterschiede, noch geht sie genauer auf die Mikroebene der Wirkmechanismen ein—bspw. durch die Verknüpfung mit altersassoziierten kognitiven Fähigkeiten (bspw. Verarbeitungsgeschwindigkeit, Salthouse, 1996). Auch wenn die ADH bisher vor allem 1), in querschnittlichen Studien untersucht wurde und daher bzgl. Annahmen über Altersunterschiede—und nicht Altersveränderungen—getestet wurde, sowie 2) den Fokus vor allem auf Enkodierung und Abruf, aber weniger auf Behalten gelegt wurde,

lassen sich aus ihr spezifische Hypothesen kognitiven Alterns und Lernens ableiten (Naveh-Benjamin & Mayr, 2018).

### *Weitere Theorien*

Neben der ADH gibt es noch weitere Theorien zur kognitiven Entwicklung im Alter. Im Folgenden wird auf die Inhibitions-Defizit-Theorie (Hasher & Zacks, 1988), die Hypothese der umweltbezogenen Unterstützung (Craik, 1986) und die Theorie der Verarbeitungsgeschwindigkeit (Salthouse, 1996) eingegangen. Während erstere postuliert, dass im Alter irrelevante Informationen das Arbeitsgedächtnis belasten, vermutet die zweite Altersunterschiede aufgrund altersbezogener Beeinträchtigungen bei selbstinitiierten Abrufprozessen, wohingegen letztere annimmt, dass durch die begrenzte Verarbeitungsgeschwindigkeit weniger Informationen gleichzeitig verarbeitet werden können.

In der Inhibitions-Defizit-Theorie von Hasher und Zacks (1988) wird angenommen, dass Altersunterschiede im episodischen Gedächtnis auf altersbedingte Defizite in der kognitiven Inhibition irrelevanter Informationen zurückzuführen sind. Dabei belastet die beeinträchtigende Inhibition das episodische Gedächtnis dadurch, dass 1) irrelevante Information ins Arbeitsgedächtnis gelangen (und daher weniger Platz für relevante Informationen besteht), 2) irrelevante Informationen schlechter aus dem Arbeitsgedächtnis gelöscht werden und 3) unpassende Reaktionen seltener unterdrückt werden können (Lustig, Hasher, & Zacks, 2007). Beispielsweise kann durch den zunehmenden Stellenwert persönlicher Werte und Erfahrungen im Alter sowie die Tendenz, Gedanken in Zusammenhang mit einem größeren Informationsnetzwerk zu bringen, dazu beitragen, dass die Inhibition im Alter beeinträchtigt wird und somit weniger Informationen in das Arbeitsgedächtnis gelangen (Hasher & Zacks, 1988).

Laut dem Ansatz des episodischen Kontextes werden Abrufversuche durch das Zurückversetzen in den Zeitpunkt des Lernens unterstützt. In der Hypothese der umweltbezogenen Unterstützung (*Environmental Support Hypothesis* (ESH); Craik, 1986) wird dieses aktive Zurückversetzen als selbstinitiierte Aktivität gesehen, welche umso wichtiger ist, je weniger Hinweisreize für den Erinnerungsprozess zur Verfügung stehen. Jedoch sind eben diese selbstinitiierten Aktivitäten beim aktiven Erinnern (Abruf- und Wiederherstellungsprozesse) im Alter schwieriger auszuführen. Demnach wird in der ESH erwartet, dass sich Altersunterschiede vor allem bei Erinnerungsversuchen zeigen, welche durch wenige Hinweisreize unterstützt werden (bspw. freier Abruf), als wenn viele Hinweisreize gegeben sind (bspw. Wiedererkennungsaufgaben). Selbst wenn ältere Lerner wesentlich von dem Prozess des Zurückversetzens profitieren, können sie sich an weniger Inhalte des vorherigen Kontextes erinnern (Craik & Schloerscheidt, 2011). Die ADH und ESH postulieren gleichermaßen, dass im Alter Enkodierung und Abruf von Kontextinformationen beeinträchtigt sind. Jedoch treffen die beiden Hypothesen unterschiedliche Annahmen bezüglich der zu erwartenden Altersunterschiede bei Paarassoziations- und freien Erinnerungsaufgaben. In der ESH wird vorhergesagt, dass die Altersunterschiede umso größer werden, je weniger Kontextinformationen beim Abruf gegeben sind. Demnach sollten die größten Altersunterschiede bei freien Erinnerungsaufgaben zu finden sein, gefolgt von Aufgaben mit Paarassoziationsaufgaben mit Hinweisreizen. Die geringsten Altersunterschiede sollten sich bei Wiedererkennungsaufgaben finden lassen. Wohingegen in der ADH angenommen wird, dass die Altersunterschiede vor allem bei assoziativen Informationen auftreten (Naveh-Benjamin, 2000). Demnach erwartet die ADH die größten Altersunterschiede bei Paarassoziationsaufgaben mit Hinweisreizen, gefolgt von freien Abrufaufgaben. Die geringsten Altersunterschiede werden laut ADH bei Wiedererkennungsaufgaben erwartet.

Demnach werden bei freiem Abruf von Wortlisten sowie beim hinweisgesteuerten Abruf von Wortpaaren Unterschiede in der Lernleistung zwischen jüngeren und älteren Erwachsenen erwartet.

In der Theorie der Verarbeitungsgeschwindigkeit von Salthouse (1996) wird postuliert, dass die Altersunterschiede im episodischen Gedächtnis auf kognitive Veränderungen in der Verarbeitungsgeschwindigkeit zurückgehen. Salthouse (1996) schlägt zwei Mechanismen vor, welche die Auswirkungen der Verarbeitungsgeschwindigkeit auf die kognitive Leistung erklären; den Mechanismus der begrenzten Zeit (*Limited Time Mechanism*) und den Mechanismus der Gleichzeitigkeit (*Simultaneity Mechanism*). Der Mechanismus der begrenzten Zeit hebt hervor, dass in einem begrenzten Zeitintervall ältere Personen weniger Informationen verarbeiten als jüngere. Dadurch können für den Lernerfolg relevante Operationen nicht erfolgreich ausgeführt werden. Dies ist beispielsweise besonders relevant, wenn Wörter/Wortpaare in der Lernphase nur für eine kurze Zeit präsentiert werden. Die geringere Lernleistung älterer Lerner kann teilweise darauf zurückgeführt werden, dass in der Lernphase weniger zusätzliche Informationen zu den Wörtern/Wortpaaren enkodiert wurde. Der Mechanismus der begrenzten Zeit bezieht sich daher auf ein externes Zeitlimit, wohingegen der Mechanismus der Gleichzeitigkeit sich auf ein internes Zeitlimit bezieht. Letzterer beschreibt den Zusammenhang gleichzeitig verfügbarer Informationen und der Verarbeitungsgeschwindigkeit. Demnach sind bei höherer Verarbeitungsgeschwindigkeit mehr Informationen gleichzeitig verfügbar, welche für eine tiefere Verarbeitung von Informationen notwendig sind. Für ältere Lerner kann dies bedeuten, dass für gewisse Lernstrategien weniger Informationen verwendet werden als bei jüngeren und dadurch nur auf oberflächlicher Ebene elaboriert wird (bspw. Assoziationen zwischen Wörtern eines Wortpaars finden und sich diese bildlich vorstellen). Die Befunde zur Theorie der Verarbeitungsgeschwindigkeit basieren jedoch vor allem auf

querschnittlichen Daten. In längsschnittlichen Analysen fallen die altersassoziierten Veränderungen der Verarbeitungsgeschwindigkeit geringer aus (Hultsch, Hertzog, Small, McDonald-Miszczak & Dixon, 1992; Sliwinski & Buschke, 1999; Zimprich, 2002; Zimprich & Martin, 2002).

Den vorgestellten Theorien zur kognitiven Entwicklung im Alter ist gemein, dass Altersunterschiede beim Lernen auf altersbedingte Veränderungen in basalen Prozessen des episodischen Gedächtnisses zurückzuführen sind, wie Beeinträchtigungen bei der Abschirmung von irrelevanten Informationen, Schwierigkeiten bei selbstinitiierten Abrufprozessen sowie abnehmender Informationsverarbeitung und damit einhergehender Defizite bei der Enkodierung. Die vorgestellten Theorien der kognitiven Entwicklung im Alter sind direkt auf Lernprozesse anwendbar und können so Unterschiede in der Lernleistung zwischen Altersgruppen aufklären. Die drei Theorien liefern wertvolle Ansätze, um der Frage nachzugehen, ob sich der Einfluss des Testeffekts im Alter verändert. Dieser Zusammenhang von Alter und Wirksamkeit des Testeffekts, wird im Folgenden theoretisch und empirisch aufgearbeitet.

### *1.2.2. Testen im Alter*

#### *Ausgangsleistung*

Da der Testeffekt auf dem erfolgreichen Abruf basiert, spielt eben diese Abrufleistung eine entscheidende Rolle für Lernerfolg durch den Testeffekt. Der Effekt ist umso größer, je mehr Informationen während des Testdurchgangs erfolgreich abgerufen wird („*level of retrieval success*“, siehe Karpicke, 2017; Marsh et al., 2009; Meyer & Logan, 2013)–da mit steigender Anzahl erfolgreich abgerufener Items mehr Abrufprozesse von den oben beschriebenen Prozessen profitieren (siehe Kapitel 1.1.2.2.). Jedoch ist bei älteren Personen diese Abrufleistung durch die Defizite im episodischen Gedächtnis geringer,

wodurch ältere Individuen weniger von Tests profitieren. Doch selbst bei einem erfolgreichen Abruf ist davon auszugehen, dass die Attribute des Abrufkontextes weniger gut im Gedächtnis repräsentiert und daher für die Suche nach Gedächtnisinhalten nicht mehr im gleichen Maß verfügbar sind, wie bei jüngeren Erwachsenen. Daher wird im Folgenden anhand von etablierten Theorien zum kognitiven Altern sowie zum Testeffekt dargestellt, wie sich die Effekte des erfolgreichen Abrufs zwischen Altersgruppen unterscheiden könnten.

### *Theorien der kognitiven Alterung im Zusammenhang mit dem Ansatz des episodischen Kontextes*

Der Testeffekt wird (wie in Kapitel 1.1.2.2. beschrieben) im Rahmen des Ansatzes des episodischen Kontextes interpretiert. In diesem Ansatz wird davon ausgegangen, dass das Sich-Zurückversetzen in den Kontext der Lern- sowie der Testphase einen wesentlichen Einfluss auf die spätere Abrufleistung hat. Dies ist deckungsgleich mit der Annahme, dass 1) Enkodierung in einem Kontext stattfindet, 2) die Kontexteigenschaften zusätzlich im Gedächtnis abgespeichert werden, um 3) später als Hinweisreize zu fungieren (Becker-Carus & Wendt, 2017b). Die ADH nimmt jedoch an, dass im Alter Defizite beim Abspeichern und Abrufen dieser Kontextattribute auftreten. Allerdings sind bei der Erklärung des Testeffekts eben diese Kontextattribute von zentraler Bedeutung. Wenn folglich altersbedingt weniger Attribute des Gedächtnisinhalts selbst und vor allem des Kontextes–dessen Kombination aus Lern- und Abrufkontext als eine „Anreicherung“ von Attributen zu sehen ist–bei einem Abruf zur Verfügung stehen, wird dadurch die Effektivität des Testeffekts eingeschränkt. Jedoch finden sich nur wenige Studien, welche die ADH im Vergleich von freien Abrufformaten mit hinweisgesteuerten Formaten testen (Naveh-Benjamin, 2000; für einen Überblick siehe Old & Naveh-Benjamin, 2008). In ihrer Metaanalyse konnten Old und

Naveh-Benjamin (2008) die ADH für Wiedererkennungsaufgaben unterstützen, jedoch nicht die Interaktion von Altersgruppe und Art der Information (alleinstehende oder assoziative Informationen) bei Aufgaben mit freiem Abruf oder Abruf mit Hinweisreizen. Dieser Befund erscheint insofern überraschend, da durch die erhöhte Schwierigkeit des freien Abrufformats gelernte Hinweisreize (assoziationsbezogene, selbstgenerierte Informationen) eine wesentlich größere Rolle spielen sollten als bei leichteren Abrufformaten (bspw. Wiedererkennungsaufgaben). Old und Naveh-Benjamin (2008) interpretieren diesen Befund 1) als Resultat der ohnehin starken altersspezifischen Defizite, welche bei alleinstehenden Informationen so groß sind, dass diese durch assoziative Erinnerungsaufgaben nicht größer werden können. Und 2) wird bei Wiedererkennungsaufgaben Vertrautheit und bei Erinnerungsaufgaben der bewusste Abruf einer Episode und seiner Details getestet (Jacoby, Jones & Dolan, 1998), wobei letztere durch Alternsprozesse beeinflusst wird. Jedoch ist zu beachten, dass die Metaanalyse von Old und Naveh-Benjamin (2008) den Fokus vor allem auf die objektiv erfassbaren Kontexteigenschaften legt, bspw. zeitlicher Kontext. Bei einer weiter gefassten Definition der ADH können neben Assoziationen zwischen objektiv beobachtbarem Kontext und Lerngegenstand auch Verbindungen zwischen Lerngegenstand und internen kognitiven Prozessen des Lerners als Assoziation interpretiert werden (vgl. Naveh-Benjamin, 2000). Bei dieser Definition erscheint es nachvollziehbar, dass v.a. beim Lernen von Wortpaaren Assoziationseinheiten mitgelernt und enkodiert werden, die bei einem späteren Abruf als Hinweisreize fungieren können. Dies ist nicht nur bei stark assoziierten Wörtern der Fall, sondern–viel mehr noch–bei schwach assoziierten Wortpaaren (für die Verwendung von Strategien bei Paarassoziationsaufgaben siehe Pyc & Rawson, 2010; Dunlosky & Hertzog, 1998). Es ist daher möglich, dass altersbedingte, assoziative Defizite den Testeffekt bei Aufgaben mit intentionalem Abrufformat beeinflussen können.

Die ESH thematisiert den gleichen basalen Erinnerungsprozess, wie der Ansatz des episodischen Kontextes. Bei beiden wird das Zurückversetzen in die vorherige Lernsituation als Kernkonzept für die Erklärung des Testeffektes dargestellt. Jedoch wird in der ESH angenommen, dass vor allem bei selbstinitiierten Abrufprozessen dieses Zurückversetzen durch altersspezifische Defizite beeinträchtigt ist. Demnach würde man auch im Rahmen der ESH in Verbindung mit dem Ansatz des episodischen Kontextes geringere Testeffekte erwarten, im Vergleich zu jüngeren Lernern.

Auch im Rahmen der Inhibitionsdefizithypothese und der Theorie der Verarbeitungsgeschwindigkeit können entsprechende Annahme bezüglich des Lernerfolgs im Alter durch das Testen formuliert werden. Wenn ältere Lerner aufgrund geringerer Verarbeitungsgeschwindigkeit weniger Lerninhalte abrufen, können weniger Abrufprozesse von dem Testeffekt beeinflusst werden. Gemäß der Inhibitionsdefizithypothese werden durch Beeinträchtigungen bei Enkodierungsprozessen, 1) bei Abrufprozessen durchschnittlich weniger korrekt erinnerte Items erwartet und es wird angenommen, dass 2) das Zurückversetzen in die vorherige Lernsituation aufgrund weniger abgespeicherter Kontextinformationen im Vergleich zu jüngeren Personen in einem geringeren Testeffekt resultiert. Dies führt dazu, dass ältere Personen nicht nur seltener von den Effekten des erfolgreichen Abrufs profitieren, sondern im geringeren Umfang vom Testeffekt profitieren als jüngere Erwachsene; d.h. dass bei älteren Probanden die erhöhte Abruftrate einer Testbedingung früher auf das Abrufniveau einer Lernbedingung ohne Testungen fallen sollte als bei jüngeren.

#### *Aktuelle Befundlage zu Altersunterschieden im Testeffekt*

Es liegen vergleichsweise wenige Studien vor, welche das Ausmaß des Testeffekts in verschiedenen Altersgruppen betrachten. Beispielsweise untersuchten Meyer und Logan

(2013) den Testeffekt bei jüngeren und älteren Erwachsenen. In ihrer Studie wurden gelernte Textabschnitte besser erinnert, wenn vor dem finalen Abruftest mit Hinweisreizen eine Wiedererkennungsaufgabe, anstatt eines erneuten Lerndurchgangs durchgeführt wurde. Ältere und jüngere Probanden profitierten in dieser Studie im Durchschnitt gleichermaßen vom Testeffekt. Coane (2013) untersuchte Altersunterschiede in den Effekten des Testens anhand einer Paarassoziationsaufgabe (mit korrigierendem Feedback). Beim verzögerten Abruf nach zwei Tagen ergab sich keine Interaktion aus Altersgruppe und Lernbedingung, was einen vergleichbar starken Testeffekt in beiden Altersgruppen nahelegt. Bishara und Jacoby (2008) untersuchten die Effekte des Abrufs anhand einer Paarassoziationsaufgabe, bei der nur Teile des Zielworts bei der Abfrage zu sehen waren (mit anschließendem korrigierendem Feedback). Hierbei profitierten ältere und jüngere Teilnehmer im ähnlichen Maß von zwei Testdurchgängen im Vergleich zu zwei wiederholten Lerndurchgängen. In einer Studie von Pastötter und Bäuml (2019) erinnerten 40, 50, 60 und 70-jährige Probanden bei einem freien Abrufdurchgang am gleichen Tag mehr Wörter in der Lernbedingung (zweimal Lernen), im Gegensatz zur Testbedingung (einmal Lernen und einmal Testen). Dieses Ergebnis ist kohärent mit der Annahme, dass der Testeffekt erst nach einem verzögerten Abruf von mehreren Tagen auftritt. Jedoch zeigten sich hier ebenfalls keine Interaktionen von Altersgruppe und Lernbedingung.

Dagegen zeigen die Ergebnisse aus einer Aufgabe zum Lernen von Gesichter-Namen-Assoziationen ein anderes Bild (Tse, Balota & Roediger, 2010). Hier profitierten Ältere bei einem finalen Abruf nach 1.5 Stunden nur dann vom Testeffekt, wenn während des Testens Feedback gegeben wurde. Die Autoren interpretieren dieses Ergebnis vor dem Hintergrund einer geringeren Abrufleistung bei älteren Lernern während des Testens („*level of retrieval success*“, s.o.), die ohne Feedback einen geringen Lernzuwachs zeigten.

Selbst wenn diese Studienergebnisse überwiegend keine Interaktion von Alter und Testbedingung zeigen, in dem Sinne, dass ältere geringer von den Effekten des Testens profitieren als jüngere Personen, so bieten die Studien jedoch theoretische Überlegungen und Begründungen für einen solchen Altersunterschied. So sehen Tse et al. (2010) das vergleichsweise schlechtere episodische Gedächtnis als mögliche Ursache für einen geringeren (oder gar negativen) Effekt des Testens, wohingegen Meyer und Logan (2013) zusätzlich die geringere Lernleistung Älterer als Erklärungsansatz annehmen. Coane (2013) führen neben selteneren, spontanen Elaborationsprozesse als Begründung (vgl. Craik, 1986; Craik & Rabinowitz, 1985), ebenso altersbedingte Assoziationsdefizite an (vgl. Naveh-Benjamin, 2000). Dennoch finden sich in den Studien mehr Belege für einen altersunabhängigen Testeffekt als für einen altersabhängigen.

Woher kann diese Diskrepanz zwischen Theorie und Empirie kommen? Einerseits sind die bisherigen Studien heterogen bezüglich des Lernmaterials, der Stichprobe und dem Feedback. Jedoch ist die Rolle von Feedback kritisch zu sehen, da es eher auf der Seite eines (erneuten) Einprägens verortet werden kann und insofern die „reinen“ Effekte des Testens verfälschen könnte (siehe Karpicke, 2017). Andererseits wird in den vorgestellten Studien in der Testbedingung nur eine Abfrage und in der Lernbedingung jeweils ein doppelter Lerndurchgang verwendet (nicht jedoch bei Bishara & Jacoby, 2008 sowie bei Tse et al., 2010). Durch den einmaligen Effekt des Abrufens könnte—ähnlich wie bei den unmittelbaren Effekten des Testens (siehe obiges Kapitel 1.1.2.1.)—der Altersunterschied zu gering sein, um beobachtbar zu sein. Erst, wenn sich die Testeffekte kumulieren und somit mehrere Kontexte für ein erfolgreich erinnertes Item abgespeichert werden, könnte sich ein Altersunterschied schon unmittelbar während der Lern-Test-Durchgänge zeigen. Durch mehrere Testdurchgänge erhöht sich auch die Variabilität in den Ausmaßen des Testeffekts. Somit würde man größere Unterschiede im Ausmaß des Testeffekts zwischen Lernern (wenn

Personen unterschiedlich stark von den Testeffekten profitieren, siehe folgendes Kapitel) und innerhalb von Lernern erwarten (bspw. werden nach vier Lern-Test-Durchgängen größere Testeffekte als nach einem Lern-Test-Durchgang erwartet).

### *1.2.3. Individuelle Unterschiede beim Lernen und Testen*

Selbst wenn sich die altersbedingte Abnahme im episodischen Gedächtnis als robuster Befund in der Literatur finden lässt (bspw. Verhaeghen et al., 1993), so zeigen jedoch die Studien ebenso, dass diese Abnahme, bezogen auf Lernerfolg, nicht bei jeder Person im gleichen Maße erfolgt (MacDonald, Stigsdotter-Neely, Derwinger & Bäckman, 2006; Kurtz & Zimprich, 2014b; Zimprich, Rast & Martin, 2008). Dabei können, trotz einer generellen Abnahme der kognitiven Fähigkeiten, individuelle Unterschiede im Alter zunehmen (Ardila, 2007). Die Betrachtung der Abweichungen von den mittleren Verläufen ist zentral für das Verständnis kognitiven Alterns (bspw. Hofer & Sliwinski, 2001; Zimprich, 2002). Dadurch stellt sich die Frage, wieso Lerner–v.a. ältere Personen–in ihrer verbalen Lernfähigkeit variieren.

Individuelle Unterschiede lassen sich anhand von Lernkurven in der Lernrate, in der Ausgangsleistung sowie dem potentiellen Lernmaximum finden–und treten bei jüngeren und älteren Altersgruppen auf (Rast & Zimprich, 2009; Zimprich et al., 2008). Diese individuellen Unterschiede können bspw. von Unterschieden in den jeweiligen individuellen Ausgangsleistungen sowie der Interaktion von Person mit ihrer Umwelt stammen (Zimprich & Kurtz, 2015). Bspw. können kognitive Fähigkeiten Unterschiede im episodischen Gedächtnis und damit folglich in der Lernleistung aufklären. Dazu zählt ebenfalls die Verarbeitungsgeschwindigkeit (bspw. Kurtz, Mogle, Sliwinski & Hofer, 2013; Salthouse, 1994). Gemäß der Verarbeitungsgeschwindigkeitstheorie kann eine höhere Verarbeitungsgeschwindigkeit über die Mechanismen der begrenzten Zeit und der

Gleichzeitigkeit den Lernprozess unterstützen und somit Unterschiede innerhalb und zwischen Altersgruppen aufklären (siehe Kapitel 1.2.1).

Beim Testeffekt treten ebenfalls individuelle Unterschiede auf (bspw. Karpicke, Blunt, & Smith, 2016). Gemäß dem Ansatz des episodischen Kontextes basiert der Testeffekt auf dem Speichern und Abrufen von Kontextinformationen. Nachdem diese basalen Lernprozesse sich individuell unterscheiden und von kognitiven Fähigkeiten abhängig sind, liegt der Schluss nahe, dass Personen auch unterschiedlich von den Effekten des erfolgreichen Abrufs profitieren. Zusätzlich sind die Effekte des Testens nach längeren Zeiträumen zu beobachten, wodurch das Vergessen—hier vor allem von Kontextattributen und Assoziationen—eine entscheidende Rolle spielt. Auch beim Vergessen zeigen sich individuelle Unterschiede (bspw. MacDonald et al., 2006; Zimprich & Kurtz, 2013), wodurch die Annahme gestützt wird, dass Personen unterschiedlich von den Effekten des Testens profitieren.

#### *1.2.4. Vorteile von Mehrebenenmodellen*

Um die individuellen Unterschiede zwischen Personen zu berücksichtigen, werden in der vorliegenden Arbeit die Daten anhand von Mehrebenenmodellen ausgewertet. Diese Modelle haben unter anderem den Vorteil, dass sie nicht derart restriktiven Annahmen unterliegen, wie die Methode der Varianzanalyse mit Messwiederholung (bspw. ANOVA). Dadurch lassen sich verschiedene Quellen zufälliger Varianz modellieren. Diggle, Heagerty, Liang und Zeger (2013) unterscheiden drei qualitativ unterschiedliche Quellen zufälliger Varianz bei Analysen mit Messwiederholungen. Neben den 1) zufälligen Effekten und der 2) seriellen Korrelation können darüber hinaus 3) Fehler durch verschiedene Messprozesse hervorgerufen werden. Zufällige Effekte sind dabei typischerweise Unterschiede zwischen Personen, bspw. in der anfänglichen Lernleistung oder in der Lernrate bei mehreren Lern-

Test-Durchgängen (diese werden in allen Experimenten der Arbeit berücksichtigt). Zusätzlich können Messungen miteinander korrelieren, wobei die Korrelation typischerweise mit der Zeit geringer wird (bspw. Autokorrelation erster Ordnung in Experiment 2). Die dritte mögliche Varianzquelle bezieht sich auf die verwendeten Messprozesse. Da in der Arbeit zwei unterschiedliche Prozesse miteinander verglichen werden (Lernen mit Testeffekt vs. Lernen ohne den Einfluss des Testeffektes), wird diese Varianzquelle in allen Experimenten mitberücksichtigt. Darüber hinaus soll durch die Ergebnisse die Lernleistung in Lernkurven vorhergesagt werden und Verläufe dieser Lernkurven genauer beschrieben werden. Daher werden hierfür gemischte Regressionen modelliert. Nachdem die theoretischen Überlegungen zu Altersunterschieden im Testeffekt vorgestellt wurden, werden im Folgenden Kapitel Fragestellungen und Hypothesen formuliert.

### *1.3. Hypothesen und Ziele der Arbeit*

Eines der Hauptziele der Arbeit ist es, Altersunterschiede in der Wirksamkeit des Testeffekts zu untersuchen. Im Einklang mit etablierten Theorien des kognitiven Alterns mit dem Ansatz des episodischen Kontextes werden Altersunterschiede in der Wirksamkeit des Testeffekts erwartet. Und zwar in dem Sinne, dass zwar ältere Lerner auch von den Effekten des erfolgreichen Abrufs profitieren, aber im geringeren Umfang als jüngere Lerner. Dieser Effekt wird vor allem dann deutlich, wenn das Lernmaterial in mehreren Lern-Test-Durchgängen bearbeitet wird. Jedoch können die Altersunterschiede im Testeffekt von dem verwendeten Material abhängen (siehe Kapitel 2.2. und 3.1.1.). Hierbei stellt sich die Frage, inwieweit der Testeffekt bei jüngeren und älteren Erwachsenen über verschiedene Lernmaterialien generalisierbar ist und wie konsistent Altersunterschiede im Testeffekt sind. Um diese Fragestellungen zu beantworten, wurden vier Experimente durchgeführt, welche sich in zwei Studienblöcke unterteilen lassen: 1) Experiment 1 und 2 modellieren

Lernkurven mit und ohne die Effekte von Testdurchgängen und 2) in Experiment 3 und 4 werden die Fragestellungen mit kohärentem Lernmaterial in Form von kurzen Geschichten getestet.

## 2. Studienblock 1 – Lernkurven mit und ohne den Einfluss des Testeffekts

### 2.1. Theoretische Einführung: Lernkurven

Zu den klassischen Befunden der psychologischen Gedächtnisforschung zählt die sog. Lernkurve. Untersucht man die Gedächtnisleistung in Form eines freien Abrufs über mehrere Lern- und Abrufdurchgänge, zeigt sich als Konsequenz typischerweise eine negativ beschleunigte Leistungssteigerung (Ritter & Schooler, 2001). Vor dem Hintergrund des Testeffekts stellt sich allerdings die Frage, wie viel der Leistungssteigerung in solchen mehrfachen Lern-Test-Durchgängen auf das wiederholte Einprägen des zu lernenden Materials und wie viel der Leistungssteigerung auf das (wiederholte) Testen des Materials zurückzuführen ist. Der Testeffekt belegt eindrucksvoll, dass die Messung der Gedächtnisleistung die Gedächtnisleistung selbst steigert–über das wiederholte Einprägen hinaus. Insofern ist davon auszugehen, dass die klassische Lernkurve nicht nur das erfolgreiche Einprägen von zu erlernendem Material widerspiegelt, sondern ebenso die Effekte des Abrufens in den vermeintlich nur der Messung der Gedächtnisleistung dienenden Testdurchgängen.

Um Lernkurven ohne den Einfluss des Testeffekts zu ermitteln, schlugen Roediger und Smith (2012) die sog. „pure-study“-Lernkurve vor. Diese Lernkurve wird anhand von mehreren Wortlisten erstellt, welchen jeweils nach einer unterschiedlichen Anzahl an aufeinanderfolgenden Lerndurchgängen *ein* Testdurchgang folgt (siehe Experiment 1, Tabelle 1 zur Veranschaulichung). Durch den Vergleich dieser „pure-study“-Lernkurve mit einer Standardlernkurve, in der sich Lern- und Testdurchgänge abwechseln, lässt sich der

Effekt des Testens auf die Gedächtnisleistung untersuchen. Roediger und Smith (2012) konnten zeigen, dass sich die „pure-study“- und die Standardlernkurve in mehrerer Hinsicht unterscheiden. Zum einen ist die Gedächtnisleistung, wie sie in der Standardlernkurve zum Ausdruck kommt, in allen Durchgängen höher als in der „pure-study“-Lernkurve (außer in dem ersten Durchgang)<sup>2</sup>. Hier zeigt sich erwartungsgemäß, dass Testen eine Rolle für den Lernerfolg spielt und die Standardlernkurve somit sowohl Effekte des Einprägens als auch des Testens enthält. Darüber hinaus jedoch, war die Steigung der beiden Lernkurven unterschiedlich, da sich in der Standardlernkurve ein über die Durchgänge hinweg größerer Lernzuwachs ergab als in der „pure-study“-Lernkurve, der Abstand zwischen den beiden Lernkurven also immer größer wurde (Roediger & Smith, 2012; Interaktion von Durchgang und Lernbedingung beim 1) freien Erinnern:  $\eta_p^2 = .41$  und 2) Wortpaarassoziationsaufgaben:  $\eta_p^2 = .38$ ).

Eine Besonderheit von Lernkurven liegt in der dahinterliegenden Datenstruktur, welche sich in mehrere Ebenen untergliedern lässt, nämlich innerhalb und zwischen Personen. Dabei können die Daten einerseits von den personenspezifischen Mittelwerten (bspw. Unterschiede in der Ausgangsleistung zwischen Personen) und andererseits von zeitabhängigen individuellen Unterschieden abhängen (bspw. Unterschiede in der Lernrate zwischen Personen), denn Individuen unterscheiden sich beim Lernen in ihrer Ausgangsleistung sowie in ihrer Lernrate voneinander (Rast & Zimprich, 2009). Zusätzlich können individuelle Unterschiede bei der Wirksamkeit des Testeffekts auftreten—bspw. können sich demnach Personen darin unterscheiden, wie gut sie sich Kontextattribute merken und diese für den Suchprozess verwenden.

Altersunterschiede zeigen sich jedoch nicht nur bei einmaliger Testung der Gedächtnisleistung, sondern ebenfalls in Lernkurven, d.h., bei wiederholten Lern- und

---

<sup>2</sup>) Da vor dem ersten Lerndurchgang noch nicht getestet wurde, kann hier auch noch kein Testeffekt wirken.

Testdurchgängen. Dabei können älteren Personen eine niedrigere und flachere Lernkurve aufzeigen als jüngere (siehe Davis et al., 2003; Rast & Zimprich, 2009). Eine naheliegende und für die mögliche Erklärung von Altersunterschieden in der Gedächtnisleistung relevante Frage lautet folglich: Welcher Anteil des Lernerfolgs in Lernkurven ist durch Altersunterschiede im Einprägen von zu erlernendem Material und welcher Anteil des Lernerfolgs ist durch Altersunterschiede im Testeffekt, also auf Unterschiede im Lernen während des Abrufs, bedingt? Bislang wurden diese Effekte in der Untersuchung von Altersunterschieden in Lernkurven nicht separiert. Wie im Kapitel 1.2.2. „Testen im Alter“ genauer hergeleitet, werden aufgrund von altersbedingten Beeinträchtigungen im assoziativen Gedächtnis, mitunter Altersunterschiede im Testeffekt erwartet und zwar im dem Sinne, dass ältere Probanden im Durchschnitt weniger Items korrekt erinnern sowie eine geringere Lernrate (eine flachere Lernkurve) als jüngere Lerner aufweisen.

## 2.2. *Lernmaterial – Wortlisten und Wortpaare*

Zwei typische Verfahren um verbales Lernen zu messen, sind Aufgaben zum freien Erinnern oder Aufgaben mit Hinweisreizen. Bei den Aufgaben zum freien Erinnern werden vor allem Wortlisten in einem Lerndurchgang gelernt und zu einem späteren Zeitpunkt während eines Testdurchgangs abgefragt. Dabei können die Wörter in beliebiger Reihenfolge genannt werden. Durch die Besonderheit des freien Abrufs können mit diesem Aufgabentyp besonders die Effekte des Erinnerns bezogen auf die Reihenfolge der genannten Wörter sowie die dahinterliegenden Mechanismus des Suchverhaltens nach Lerninhalten im Gedächtnis untersucht werden (für einen genaueren Überblick siehe Kahana, 2014). Bei hinweisgesteuerten Abfragemechanismen–bspw. beim Lernen und Erinnern von Wortpaaren–wird die Reihenfolge der Abfrage (meist) randomisiert durch die Präsentation des Hinweiswortes vorgegeben. Dadurch können Effekte der

Reihenfolge/internen Organisation von Items sowie Vorteile von im Gedächtnis kurzfristig gespeicherten Wortpaaren kontrolliert werden (vgl. *Recency Effect*, Raaijmakers & Shiffrin, 1981).

In der Assoziations-Defizit-Hypothese werden aufgrund von altersbedingten Defiziten beim Enkodieren und Abrufen spezifischer Assoziationen größere Altersunterschiede bei Assoziationsaufgaben im Gegensatz zu Aufgaben mit freiem Erinnerungsformat erwartet (Naveh-Benjamin, 2000; siehe Kapitel 1.1.2.2.). Bezüglich der Wirksamkeit des Testeffekts werden größere Effekte bei Aufgaben zum freien Erinnern erwartet, da bei diesen Aufgaben eigene Hinweisreize generiert werden müssen–im Gegensatz zu Paarassoziationsaufgaben, die den Abruf mit Hinweisreizen unterstützen (Bregman & Wiener, 1970; Karpicke et al., 2014). Sobald eine geringe Anzahl an Hinweisreizen durch die Umwelt vorgegeben wird, wird das Zurückversetzen in vorherige Lern- und/oder Test-Kontexte umso wichtiger (nach dem Ansatz des episodischen Kontextes). Jedoch eignen sich beide Aufgabentypen für die Effekte des erfolgreichen Abrufs.

### 2.3. Experiment 1 – Wortlisten

#### *Ziele von Experiment 1*

Mit Experiment 1 werden zwei Ziele verfolgt. Zum einen sollte untersucht werden, welcher Anteil der Gedächtnisleistung in einer Lernkurve auf das Einprägen und welcher Anteil auf die Effekte des erfolgreichen Abrufs (Testeffekt) zurückgeführt werden kann. Dazu wird die Standardlernkurve mit einer „pure-study“-Lernkurve verglichen. Das erste Ziel ist als ein Versuch der Replikation der Ergebnisse von Roediger und Smith (2012) zu sehen. Das zweite, zentrale Ziel bestand in der Untersuchung von Altersunterschieden in der Gegenüberstellung der Standardlernkurve und der „pure-study“-Lernkurve, um somit

mögliche Altersunterschiede in der Wirksamkeit des Testeffekts zu ermitteln. Zu erwarten war, dass sich die Lernkurven der jüngeren und älteren Teilnehmer unterscheiden im Sinne von niedrigeren und flacheren Lernkurven der älteren Teilnehmer. Trotz dieser Altersunterschiede sollten beide Altersgruppen in der Standardbedingung vom Testeffekt profitieren, im Gegensatz zur „pure-study“-Bedingung ohne wiederholtes Testen. Zusätzlich wurde erwartet, dass ältere Lerner in einem geringeren Ausmaß vom Testeffekt profitieren– d.h., der Unterschied im Lernerfolg zwischen der Standard- und der „pure-study“-Lernkurve sollte bei den älteren Probanden mit jedem Durchgang weniger stark zunehmen, als bei den jüngeren Teilnehmern.

### 2.3.1. Methode Experiment 1

#### *Stichprobe*

Insgesamt nahmen 40 jüngere (62.5% weiblich) und 40 ältere Personen (67.5% weiblich) an der Studie teil (Gesamtanzahl:  $N = 80$ ; 65.0% weiblich).<sup>3</sup> Die Probanden der jüngeren Teilstichprobe waren im Mittel 24.20 Jahre alt ( $SD = 3.66$ , 18-30 Jahre), wohingegen die älteren Teilnehmer durchschnittlich 69.68 Jahre alt waren ( $SD = 7.94$ , 60-80 Jahre). Die subjektive Gesundheit der Probanden wurde mittels einer Likert-Skala von 1 = „sehr gut“ bis 5 = „mangelhaft“ erfasst und betrug im Mittel 2.38 ( $SD = 0.93$ ). Dabei gaben die jüngeren Teilnehmer im Durchschnitt eine bessere subjektive Gesundheit an ( $M = 2.08$ ,  $SD = 0.80$ ) als ältere ( $M = 2.68$ ,  $SD = 0.97$ ) ( $t(78) = -3.02$ ,  $p < .05$ , Cohen's  $d = -0.68$ ). Um ihren Schulabschluss anzugeben, hatten die Probanden vier Optionen zur Verfügung: "kein Abschluss", "Volks-/Hauptschule", "Realschule", "Gymnasium". In der jüngeren Stichprobe gaben 32 (80%) Abitur als höchsten erreichten Schulabschluss an, während bei den älteren

---

<sup>3</sup> ) Die Geschlechterverteilung zwischen den Altersgruppen war nicht signifikant verschieden,  $\chi^2(1, 79) = .22$ ,  $p = .64$ , Cohen's  $w = 0.10$ .

Probanden 17 (42.5%) Volks-/Hauptschulabschluss und 17 (42.5%) Abitur angaben. Alle 40 jüngeren Probanden gaben als Familienstand „ledig“, wohingegen bei den älteren Teilnehmern „verheiratet“ die häufigste Angabe war (65%).

### *Messinstrumente*

#### *Freies Erinnern*

Für diese Aufgabe wurden 108 zweisilbige deutsche Nomen aus der „Berlin Affective Word List Reloaded“ (BAWL-R; Vö et al. 2009) ausgewählt. Die Wörter weisen im Mittel eine neutrale Emotionalität von 0.39 auf (auf einer Skala von -3 bis 3—sehr negativ bis sehr positiv), eine mittlere Erregung von 2.18 (auf einer Skala von 1 bis 5—beruhigend bis aufregend) sowie eine mittlere Bildhaftigkeit von 2.67 (auf einer Skala von 1 bis 7—niedrige Bildhaftigkeit bis hohe Bildhaftigkeit). Alle Wörter kommen mit einer durchschnittlichen Häufigkeit von 80.60 pro Million relativ häufig vor. Die Wörter wurden zufällig auf vier Listen mit je 27 Wörtern aufgeteilt (Liste A, B, C und D); die Listen waren für alle Probanden die gleichen. Die Präsentationsreihenfolge der Wörter wurde für jeden Probanden sowie für jeden Lerndurchgang randomisiert.

#### *Lerndesign*

Jeder Proband durchlief vier Lernaufgaben (eine Aufgabe für die Standardbedingung und drei Aufgaben zur „pure-study“-Bedingung). Die Standardbedingung setzt sich aus vier direkt aufeinanderfolgenden Lern-Test-Durchgängen zusammen. In der „pure-study“-Bedingung hingegen lernten die Teilnehmer die Wortlisten unterschiedlich oft hintereinander (ohne Testen), wobei die Lerndurchgänge durch Füllaufgaben getrennt waren (Verarbeitungsgeschwindigkeitsaufgaben; Dauer: je Aufgabe ungefähr 90 Sekunden). Nach dem jeweils letzten Lerndurchgang in den drei Aufgaben der „pure-study“-Bedingung folgte unmittelbar im Anschluss ein Testdurchgang (siehe Tabelle 1). Um die Reihenfolge der vier

Lernaufgaben auszubalancieren, wurden acht verschiedenen Experimentdesigns erstellt (siehe Tabelle 2). Jeweils fünf jüngere und fünf ältere Probanden bearbeiteten je ein Design. Hierbei ist die Reihenfolge der vier Wortlisten (A, B, C und D) in jedem Design dieselbe, unabhängig von der Lernaufgabe. Somit wurde in jedem Design die Wortliste A in der ersten Aufgabe gelernt und abgefragt (in der zweiten die Wortliste B usw.). Durch dieses Vorgehen werden alle Wortlisten gleich oft mit den vier Lernaufgaben kombiniert. Der Punktwert einer Person eines Testdurchgangs ist die Prozentzahl korrekt genannter Wörter der abgefragten Liste (mögliche Punktezahl: 0–100).

Tabelle 1

*Lernaufgaben Experiment 1*

Bedingung	Durchgang 1	Durchgang 2	Durchgang 3	Durchgang 4
Standard	S T	S T	S T	S T
“Pure-Study” 2	S –	S T		
“Pure-Study” 3	S –	S –	S T	
“Pure-Study” 4	S –	S –	S –	S T

*Anmerkungen.* S = Lerndurchgang; T = Testdurchgang; – = Verarbeitungsgeschwindigkeitsaufgabe

Tabelle 2

*Designs Experiment 1*

	1. Aufgabe		2. Aufgabe		3. Aufgabe		4. Aufgabe	
	Wortliste A		Wortliste B		Wortliste C		Wortliste D	
Design 1	ST	F	PS 3	F	PS 4	F	PS 2	
Design 2	PS 2	F	ST	F	PS 3	F	PS 4	
Design 3	PS 3	F	PS 4	F	PS 2	F	ST	
Design 4	PS 4	F	PS 2	F	ST	F	PS 3	
Design 5	PS 2	F	PS 4	F	PS 3	F	ST	
Design 6	PS 4	F	PS 3	F	ST	F	PS 2	
Design 7	ST	F	PS 2	F	PS 4	F	PS 3	
Design 8	PS 3	F	ST	F	PS 2	F	PS 4	

*Anmerkungen.* ST = Standardbedingung; PS 2 = „Pure-study“-Bedingung mit zwei Lerndurchgängen; PS 3 = „Pure-study“-Bedingung mit drei Lerndurchgängen; PS 4 = „Pure-study“-Bedingung mit vier Lerndurchgängen; F = Füllaufgabe

*Verarbeitungsgeschwindigkeit*

Um die Verarbeitungsgeschwindigkeit zu messen, wurden drei Aufgaben zur Verarbeitungsgeschwindigkeit durchgeführt: Identische Bilder, einfache Reaktionsaufgabe, Buchstaben-Zahlen-Test.

*Identische Bilder*

In dieser Aufgabe wurden 60 Items des „Identical Pictures Test“ aus dem „Educational Testing Service“ präsentiert (ETS; Ekstrom, French, Harman & Dermen, 1976). Jedes Item besteht aus einer Reihe aus sechs Bildern: das Zielbild wurde auf der linken Seite dargestellt, die anderen fünf Bilder nummeriert daneben. Die Teilnehmer sollten möglichst schnell die Nummer des Bildes auf der Tastatur eingeben, welches identisch zum

Zielbild war. Die Probanden bearbeiteten während des Experiments zwei unterschiedliche Identische-Bilder-Tests für jeweils 60 Sekunden. Die erreichte Punktzahl ist die Anzahl der korrekt beantworteten Items aus beiden Tests (mögliche Reichweite: 0–120). Die Test-Retest Reliabilität entspricht bei Männern .87 und bei Frauen .81 (Ekstrom et al., 1976).

### *Einfache Reaktionsaufgabe*

Zu Beginn der Aufgabe wurde eines von zwölf Bildern als Zielbild präsentiert. Anschließend werden die zwölf Bilder einzeln und in zufälliger Reihenfolge präsentiert. Die Versuchspersonen wurden instruiert, „q“ auf der Tastatur zu drücken sobald das Zielbild zu sehen war, andernfalls „p“. Im Durchschnitt erscheint das Zielbild im Verhältnis von 1:2. Die Probanden bearbeiteten während des Experiments die Aufgabe zweimal für je 60 Sekunden—das Zielbild wurde jeweils zufällig ausgewählt. Die erreichte Punktzahl ist die Anzahl der korrekt beantworteten Items aus beiden Tests.

### *Buchstaben-Zahlen-Matrizen-test*

Der Test besteht aus 75 Items. Jedes Item beinhaltet eine Kodiertabelle, in der fünf verschiedene Buchstaben mit den Zahlen von eins bis fünf gepaart dargestellt werden. Unter der Tabelle wird ein einzelner Buchstabe mit einem Fragezeichen gezeigt. Die Teilnehmer wurden instruiert, die Zahl zu drücken, welche dem dementsprechenden Buchstaben zugeordnet war. Um Gedächtniseinflüsse zu reduzieren, wurde für jedes Item eine andere Kodiertabelle und ein neuer Hinweisbuchstabe präsentiert (Piccinin & Rabbitt, 1999). Die Probanden bearbeiteten während des Experiments zwei Buchstaben-Zahlen-Matrizen-tests, jeden für jeweils 60 Sekunden. Die erreichte Punktzahl ist die Anzahl der korrekt beantworteten Items aus beiden Tests (mögliche Punktzahl: 0–150). Die Test-Retest

Reliabilität in einem sehr ähnlich aufgebauten „Letter-Digit-Coding-Task“ betrug bei älteren Probanden .88 (Houx et al., 2002).

### *Allgemeine Verarbeitungsgeschwindigkeit*

Um einen zusammengefassten Messwert für Verarbeitungsgeschwindigkeit zu erhalten, wurden die z-standardisierten Werte der Verarbeitungsgeschwindigkeitsaufgaben addiert und gemittelt ( $M = 0.00$ ,  $SD = 0.98$ ). Die zusammengefasste Variable weist eine hohe interne Konsistenz auf (Cronbachs Alpha: .97).

### *Prozedur*

Nachdem die Probanden eine Einverständniserklärung unterzeichnet haben, bearbeiteten sie die Aufgabe in Einzeltestungen an einem Laptop (13 Zoll Bildschirm). Nach den Angaben zu demographischen Variablen (Alter, Geschlecht, Familienstand, Bildungshintergrund, sowie subjektive Gesundheit), lernten die Versuchspersonen die vier Wortlisten unter den unterschiedlichen Lernbedingungen.<sup>4</sup> Die Wörter wurden während des Lerndurchgangs einzeln und in schwarzen Buchstaben auf weißem Hintergrund für je drei Sekunden dargeboten (Schriftart: Arial; Schriftgröße: 9% des 13 Zoll Bildschirms). Im Testdurchgang hatten die Probanden 90 Sekunden Zeit, alle Wörter des Lerndurchgangs, an die sie sich erinnerten, auf ein Blatt Papier zu schreiben. Nach dem Testdurchgang legte die Versuchsleiterin das Blatt beiseite, sodass die Teilnehmer ihre Antworten nicht mehr sehen konnten. Die Füllaufgaben in den Aufgaben der „pure-study“-Bedingung bestanden jeweils aus einer 60-sekündigen Verarbeitungsgeschwindigkeitsaufgabe und einer ca. 30-sekündigen Instruktion (s.o.). Die Reihenfolge der Füllaufgabe ist dabei innerhalb eines Designs

---

<sup>4</sup>) Jeweils vor und nach dem Experiment wurde die aktuelle Motivation sowie das emotionale Erleben abgefragt.

konstant. Die vier Lernaufgaben sind durch nicht-kognitive Aufgaben getrennt.<sup>5</sup> Das gesamte Experiment dauerte ungefähr 70 Minuten.

### *Datenanalyse*

Für die Lernkurve der „pure-study“-Bedingung liegt keine Testung nach einem Lerndurchgang vor (sondern nur Testwerte für zwei, drei und vier Lerndurchgänge). Um jeweils für die Standard- sowie für die „pure-study“-Bedingung Lernkurven bestehend aus vier Messwerten zu modellieren, wurden die Testwerte der ersten Testung der Standardbedingung aufgeteilt (d.h. nach dem ersten Lerndurchgang). Hierfür wurde die Wortliste in zwei Hälften aufgeteilt, wobei eine Hälfte die Erinnerungswerte für den ersten Testdurchgang der Standardbedingung und die andere Hälfte die Lernleistung für den ersten Wert der „pure-study“-Bedingung darstellt (dementsprechend beziehen sich die Prozentwerte auf die dementsprechende Teilliste, sodass die Werte mit denen der anderen Testdurchgänge vergleichbar sind). Dadurch, dass der erste Testdurchgang der Standardbedingung nicht von Testeffekten vorheriger Testdurchgänge beeinflusst wird, ist dieser Wert gleichermaßen für die „pure-study“-Lernbedingung verwendbar (einmal Lernen und einmal Testen). Durch die vier Messzeitpunkte in den beiden Lernbedingungen ist es möglich quadratische Verläufe zu schätzen (anstatt eines saturierten Modells).

Die Daten wurden mittels mehrerer gemischter Regressionen modelliert. Es sei  $y_{it1}$  die Abrufleistung von Person  $i$  ( $i = 1, \dots, N$ ) im Durchgang  $t$  ( $t = 1, \dots, 4$ ) in der Standardbedingung. Entsprechend bezeichne  $y_{it2}$  die Abrufleistung in der „pure-study“-Bedingung. Im Folgendem wird das Modell zu Modell 5 beschrieben (siehe Tabelle 6).<sup>6</sup>

---

<sup>5</sup> Die Teilnehmer beantworteten Fragen zu ausgewählten Facetten der Big Five Faktoren Extraversion und Neurotizismus.

<sup>6</sup> Die Unterschiede der anderen Modelle zu dem präsentierten Modell werden im Ergebnisteil genauer beschrieben (siehe Kapitel 2.3.2.).

Aufgrund des non-linearen Verlaufs von Gedächtnisleistungen über die Durchgänge wurden die Effekte des Durchgangs logarithmisch modelliert (d.h.  $\ln(t) = 0, 0.69, 1.10, 1.39$ ). Um die Abrufleistung unter den beiden Lernbedingungen gleichzeitig zu modellieren, wurden die Variablen  $y_{it1}$  und  $y_{it2}$  zu einem Vektor  $\mathbf{y}_{it}$  zusammengefasst (vgl. MacCallum, Kim, Malarkey & Kiecolt-Glaser, 1997):

$$\mathbf{y}_{it} = \begin{pmatrix} y_{it1} \\ y_{it2} \end{pmatrix}. \quad (1)$$

Die Effekte der Bedingung, die sich in Form einer Dummyvariable  $x_{1i}$  ( $0 = \text{Standard}$ ,  $1 = \text{„pure-study“}$ ) modellieren lassen, sowie des Durchgangs, der Altersgruppe  $x_{2i}$  ( $0 = \text{jung}$ ,  $1 = \text{älter}$ ) und deren Interaktionen lassen sich dann beschreiben als

$$\begin{aligned} \mathbf{y}_{it} &= \begin{pmatrix} \beta_0 + b_{0i} + \alpha_0 x_{2i} + (\delta_0 + d_{0i} + \alpha_1 x_{2i}) x_{1i} \\ \beta_0 + b_{0i} + \alpha_0 x_{2i} + (\delta_0 + d_{0i} + \alpha_1 x_{2i}) x_{1i} \end{pmatrix} + \\ &+ \begin{pmatrix} \beta_1 + b_{1i} + \alpha_2 x_{2i} + (\delta_1 + d_{1i} + \delta_2 x_{2i}) x_{1i} \\ \beta_1 + b_{1i} + \alpha_2 x_{2i} + (\delta_1 + d_{1i} + \delta_2 x_{2i}) x_{1i} \end{pmatrix} \ln(t) + \begin{pmatrix} u_{ST} \\ u_{PS} \end{pmatrix} + \begin{pmatrix} e_{it1} \\ e_{it2} \end{pmatrix} \\ &= \begin{pmatrix} \beta_0 + b_{0i} + \alpha_0 x_{2i} \\ \beta_0 + b_{0i} + \alpha_0 x_{2i} + \delta_0 + d_{0i} + \alpha_1 x_{2i} \end{pmatrix} + \begin{pmatrix} \beta_1 + b_{1i} + \alpha_2 x_{2i} \\ \beta_1 + b_{1i} + \alpha_2 x_{2i} + \delta_1 + d_{1i} + \delta_2 x_{2i} \end{pmatrix} \ln(t) + \begin{pmatrix} u_{ST} \\ u_{PS} \end{pmatrix} + \begin{pmatrix} e_{it1} \\ e_{it2} \end{pmatrix} \quad (2) \end{aligned}$$

wobei  $\beta_0$  den festen Ausgangswert (d.h. die durchschnittliche Abrufleistung im ersten Durchgang),  $b_{0i}$  die zufällige individuelle Abweichung vom festen Ausgangswert in Person  $i$ ,  $\delta_0$  den festen Effekt der Bedingung und  $d_{0i}$  die zufällige individuelle Abweichung vom festen Effekt der Bedingung bezeichne. Darüber hinaus bezeichne  $\beta_1$  den festen Effekt des Durchgangs (d.h. die durchschnittliche Veränderung der Abrufleistung über die Durchgänge) und  $b_{1i}$  die zufällige individuelle Abweichung vom Effekt des Durchgangs in Person  $i$ . Die Altersunterschiede werden durch den festen Effekt der Altersgruppe  $\alpha_0$ , der Interaktion von Altersgruppe und Bedingung  $\alpha_1$ , der Interaktion von Altersgruppe mit den logarithmierten

Durchgängen  $\alpha_2$ , sowie der dreifach Interaktion  $\delta_2$  von Bedingung, Durchgang und Altersgruppe modelliert. Die Fehleranteile der Messmethode durch die Standard- und ohne „pure-study“-Lernkurve werden durch  $u_{ST}$  und  $u_{PS}$  dargestellt. Dazu kommen individuelle Fehleranteile, von denen angenommen wurde, dass sie mit konstanter Varianz normalverteilt sind, d.h.

$$e_{it1}, e_{it2} \sim N(0, \sigma_e^2). \quad (3)$$

Darüber hinaus wurde angenommen, dass die zufälligen Effekte des Ausgangsniveaus, der Bedingung und des Lernzuwachs multivariat normalverteilt sind mit jeweils Mittelwert 0 und Kovarianz-Matrix

$$\begin{pmatrix} b_{0i} \\ d_{0i} \\ b_{1i} \end{pmatrix} \sim MVN(\mathbf{0}, \mathbf{G}), \quad \text{mit } \mathbf{G} = \begin{pmatrix} \sigma_{b_0}^2 & & \\ \sigma_{b_0 d_0} & \sigma_{d_0}^2 & \\ \sigma_{b_0 b_1} & \sigma_{d_0 b_1} & \sigma_{b_1}^2 \end{pmatrix}. \quad (4)$$

Die Residualvarianzen der Standardbedingung  $\sigma_{ST}^2$  sowie der „pure-study“-Bedingung  $\sigma_{PS}^2$  und deren Kovarianz  $\sigma_{STPS}$  werden geschätzt.

Die Berechnungen wurden mit SAS Version (9.4) (SAS Institute, 2016) durchgeführt. Die gemischte Regression wurde mit der Prozedur MIXED und der „Maximum Likelihood“-Schätzmethode modelliert. Die Varianzaufklärung auf den beiden Ebenen des Modells werden mit  $R^2$  (Recchia, 2010) und Effektstärken mit Cohen's  $d$  angegeben—die Angaben dienen dabei vor allem der Veranschaulichung.

### 2.3.2. Ergebnisse Experiment 1

#### Deskriptive Statistiken

Tabelle 3 zeigt die deskriptiven Statistiken der Lernleistung aller Testdurchgänge. Wie erwartet, steigt die Lernleistung in der Standardbedingung sowie in der „pure-study“-

Bedingung mit jedem weiteren Durchgang bei jüngeren und älteren Erwachsenen an. Hierbei ist die Lernleistung negativ beschleunigt; das bedeutet, dass sich die Lernrate über die Durchgänge hinweg reduziert. Dieses Muster ist konsistent mit der Annahme einer abflachenden Lernkurve. Es zeigen sich Unterschiede in der Lernleistung zwischen den Altersgruppen ( $F(8, 71) = 14.13, p < .05$ ). Zusätzlich sind Altersunterschiede in der Standardbedingung größer als in der „pure-study“-Bedingung, was andeutet, dass neben Altersunterschieden beim Lernen, ebenfalls Altersunterschiede in der Effektivität von Testdurchgängen gibt (siehe Abbildung 2 und 3). Ausnahme hierbei stellen die Altersunterschiede in der ersten Testung der Standard- und der „pure-study“-Bedingung dar. Nachdem diese Werte aus demselben Testdurchgang stammen (siehe Kapitel 2.3.1.), können die verschiedenen Altersunterschiede in den beiden Bedingungen als zufällig interpretiert werden. Die Erinnerungsleistung im ersten Testdurchgang der Standard- und „pure-study“-Bedingung unterscheidet sich nicht statistisch signifikant ( $t(79) = -0.19, p = .85, d = -0.02$ ).<sup>7</sup> In Abbildung 3 sind die Differenzen der Standard- zur „pure-study“-Lernkurve für die beiden Altersstichproben abgebildet; bzw. der Vorteil der Standardbedingung gegenüber der „pure-study“-Bedingung. Hierbei fällt auf, dass vor allem von Durchgang 3 auf 4 sich die Lernleistung in der Standardbedingung gegenüber der „pure-study“-Bedingung zunimmt. Tabelle 4 zeigt die Interkorrelationen aller Testdurchgänge. Der vorherige Testdurchgang korreliert durchgängig höher (in der Standard- sowie in der „pure-study“-Bedingung) mit dem darauffolgenden Durchgang als der Testdurchgang vor zwei oder drei Durchgängen. Dieses Muster ist konsistent mit der Annahme, dass zeitlich näherliegende Testungen höher zusammenhängen, als Testungen die zeitlich weiter auseinanderliegen.

---

<sup>7</sup> Die Erinnerungsleistung im ersten Testdurchgang der Standard- und „pure-study“-Bedingung unterscheidet sich nicht signifikant bei der jüngeren Teilstichprobe ( $t(39) = -0.86, p = .40, d = -0.14$ ) sowie bei der älteren Stichprobe ( $t(39) = 0.81, p = .42, d = 0.13$ ).

Tabelle 3

*Experiment 1: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

Variablen		Jung (N = 40)	Alt (N = 40)	Gesamt (N = 80)	Diff(Jung – Alt)	Cohens d
Standard 1	<i>M</i>	29.26	17.04	23.15	$t(78) = 3.92^*$	.88
	<i>SD</i>	14.91	11.65	15.16		
Standard 2	<i>M</i>	48.52	26.02	37.27	$t(78) = 9.59^*$	2.14
	<i>SD</i>	11.09	9.87	15.39		
Standard 3	<i>M</i>	57.87	33.15	45.51	$t(78) = 8.98^*$	2.01
	<i>SD</i>	14.05	10.27	17.44		
Standard 4	<i>M</i>	66.02	39.35	52.69	$t(78) = 9.20^*$	2.06
	<i>SD</i>	13.68	12.20	18.60		
„pure-study“ 1	<i>M</i>	31.85	15.19	23.52	$t(78) = 6.69^*$	1.50
	<i>SD</i>	13.02	8.87	13.88		
„pure-study“ 2	<i>M</i>	40.83	25.09	32.96	$t(78) = 6.21^*$	1.39
	<i>SD</i>	11.45	11.21	13.76		
„pure-study“ 3	<i>M</i>	48.89	29.63	39.26	$t(78) = 7.56^*$	1.69
	<i>SD</i>	13.40	8.96	14.90		
„pure-study“ 4	<i>M</i>	49.26	32.04	40.65	$t(78) = 5.84^*$	1.31
	<i>SD</i>	14.39	11.88	15.71		

*Anmerkungen.* Standard 1–4 entsprechen der Lernleistung der Testdurchgänge in der Standardbedingung; „pure-study“ 1–4 entsprechen der Lernleistung der Testdurchgänge in den „pure-study“-Bedingungen.

\* $p < .05$

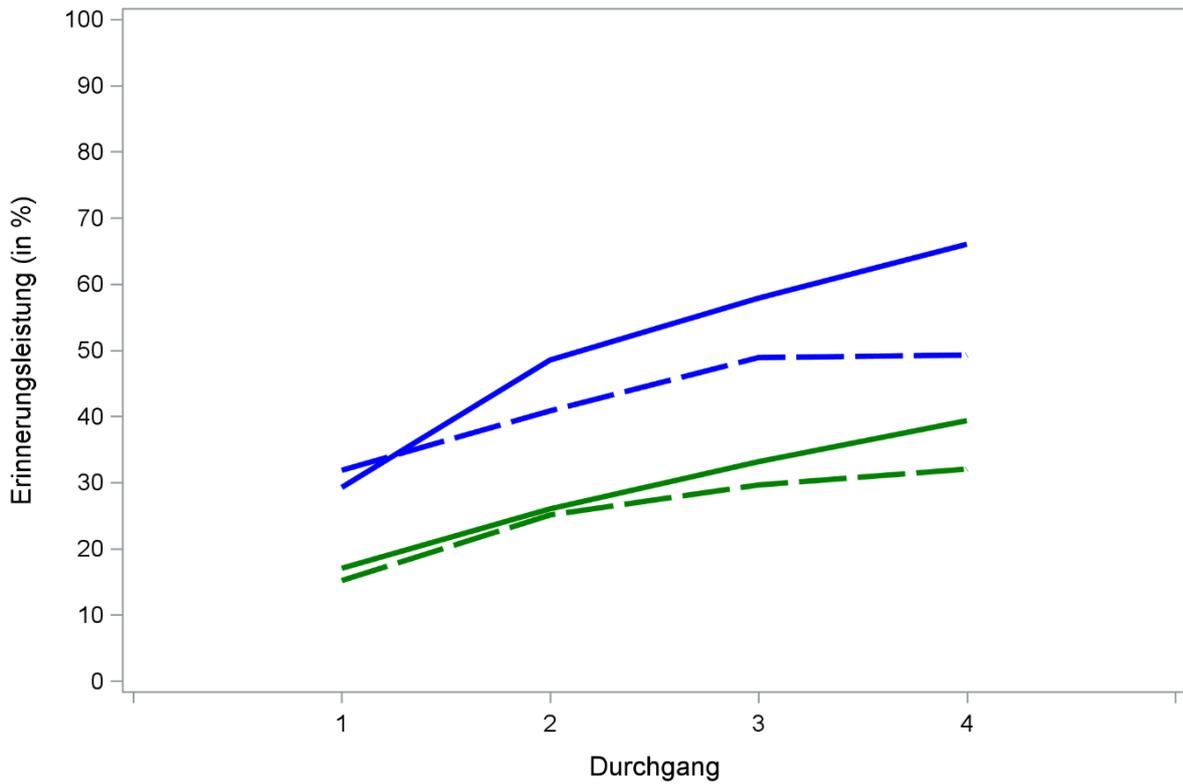


Abbildung 2. Experiment 1: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün). Die durchgehenden Linien repräsentieren die Standardlernbedingung und die gestrichelten Linien die „pure-study“-Bedingung.

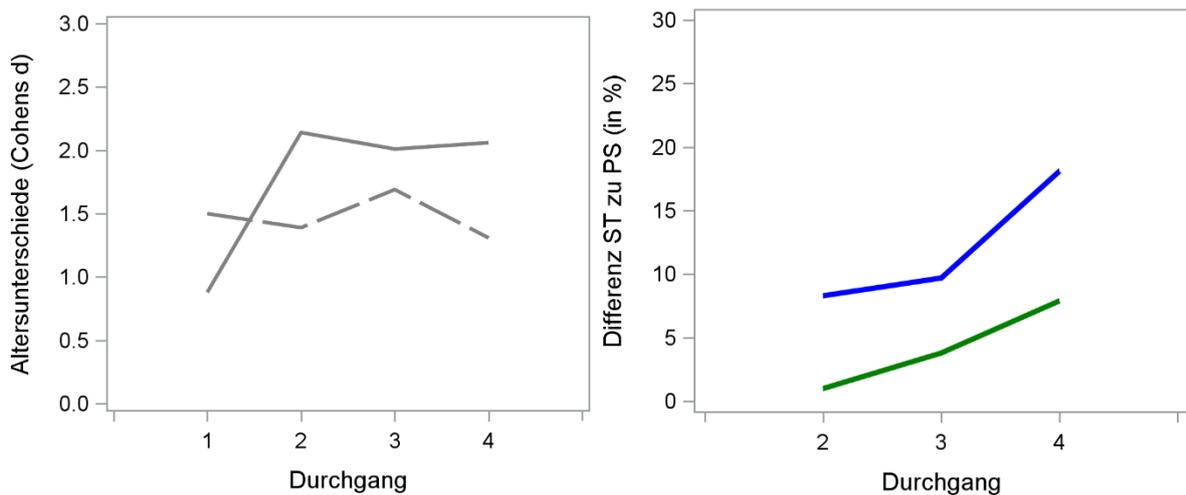


Abbildung 3. Experiment 1: Links die Altersunterschiede in der Standardbedingung (durchgehende Linie) und in der „pure-study“-Bedingung (gestrichelte Linie). Rechts die Differenz zwischen Standard- und „pure-study“-Bedingung (ST – PS) für jüngere Erwachsene (blau) und ältere Erwachsene (grün).

Tabelle 4

*Experiment 1: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

	1.	2.	3.	4.	5.	6.	7.	8.
1. Standard 1		.62*	.50*	.46*	.13	.28*	.20*	.37*
2. Standard 2	.62*		.84*	.81*	.46*	.47*	.41*	.60*
3. Standard 3	.45*	.81*		.88*	.34*	.63*	.55*	.61*
4. Standard 4	.50*	.77*	.79*		.41*	.61*	.50*	.61*
5. „Pure-Study“ 1	.02	.37*	.36*	.26		.19	.40*	.43*
6. „Pure-Study“ 2	.46*	.77*	.61*	.68*	.25		.63*	.67*
7. „Pure-Study“ 3	.49*	.57*	.54*	.55*	.22	.65*		.72*
8. „Pure-Study“ 4	.33*	.51*	.63*	.62*	.25	.59*	.65*	

*Anmerkungen.* Oberhalb der Diagonale werden die Korrelationen für die jüngere Stichprobe und unterhalb der Diagonale die Korrelationen der älteren Stichprobe dargestellt.

\* $p < .05$

### *Mehrebenenmodelle*

In mehreren gemischten Regressionen wird die Lernleistung (prozentualer Anteil korrekt erinnelter Wörter) der beiden Lernbedingungen in Abhängigkeit vom Durchgang modelliert (unterschiedliche Modellierungen des Effekts des Durchgangs: linear, quadratisch oder logarithmisch). Hierbei wird der Ausgangswert, der Effekt des Durchgangs, sowie der Effekt der Lernbedingung als zufällige Effekte mit in das Modell aufgenommen, d.h., die Ausprägungen dieser drei Variablen können sich zwischen den Personen unterscheiden. Damit wird ebenfalls die statistische Abhängigkeit der längsschnittlichen Daten berücksichtigt. Lernbedingung und Altersgruppe werden als Dummy-Variablen kodiert. Tabelle 5 und 6 zeigen die Parameterschätzungen der festen Effekte (Haupteffekte und Interaktionen), sowie der zufälligen Effekte (Varianzen, Kovarianzen und Korrelationen).

Die Interklassenkorrelation beträgt in einem „*Unconditional Model*“  $ICC = .46$ . Die Leistungsvarianz geht dabei zu 46% auf Unterschiede zwischen Personen zurück. Somit zeigt sich bei diesem Experiment, dass bei den Analysen die Abhängigkeit der Daten beachtet werden sollte, und zwar in dem Sinne, dass Effekte zwischen und innerhalb von Personen betrachtet werden. In einem ersten Modell (Modell 0) werden die Lernleistungen innerhalb einer Person in Abhängigkeit des Durchgangs sowie der Lernbedingung modelliert ( $t = 0, 1, 2, 3$ ) (siehe Tabelle 5). Die Lernrate wird als der lineare Effekt des Durchgangs für die Standardbedingung geschätzt. Die Interaktion des Durchgangs mit der Lernbedingung gibt hier den Unterschied der Lernrate in der „pure-study“-Bedingung an. Tabelle 5 zeigt die dazugehörigen festen sowie zufälligen Effekte. Im Durchschnitt erinnerten sich Personen im ersten Testdurchgang der Standardbedingung an 25.13% der Wörter (entspricht ungefähr sieben Wörtern) und lernten mit jedem weiteren Lern-Test-Durchgang 9.69% dazu (entspricht ungefähr drei Wörtern). Zwar erinnerten sich die Lerner im ersten Durchgang der „pure-study“-Bedingung an gleich viele Wörter wie in der Standardbedingung, wiesen hingegen eine geringere Lernrate in der „pure-study“-Bedingung auf ( $\delta_1 = 3.92\%$ , d.h. es werden pro Durchgang durchschnittlich ein Wort weniger dazu gelernt als in der Standardbedingung). Alle drei Varianzschätzungen der zufälligen Effekte sind signifikant von Null verschieden. Dies bedeutet, dass es zwischen Personen bedeutsame Unterschiede im Ausgangswert (Gedächtnisleistung im ersten Durchgang in der Standardbedingung), in den Effekten des (linearen) Durchgangs, sowie in den Effekten der Lernbedingung gab.

Um einen nicht-linearen Verlauf abbilden zu können, wird in Modell 1 zusätzlich der Effekt des quadrierten Durchgangs ( $t^2 = 0, 1, 4, 9$ ) geschätzt. Die festen Effekte für den Ausgangswert, den Durchgang sowie deren Interaktion bleiben statistisch signifikant. Ebenfalls wird der negative Effekt des quadrierten Durchgangs signifikant, welcher eine

abflachende Lernkurve beschreibt. Der Modellfit (hier AIC) verbesserte sich durch die Hinzunahme eines quadratischen festen Effektes um 34.5 Punkte.

Durch die Modellierung eines logarithmischen Effekts des Durchgangs ( $\ln(t) = 0, 0.69, 1.10, 1.39$ )–anstatt eines linearen und quadratischen Verlaufs–, verbessert sich der Modellfit um weitere 5.5 Punkte (siehe Modell 2, Tabelle 5). Die Effekte des logarithmierten Verlaufs sind wiederum statistisch signifikant. Dies bedeutet, dass bspw. im zweiten Durchgang unter der Standardbedingung 14.54% mehr Wörter erinnert wurden (entspricht ungefähr vier Wörtern) als im ersten Durchgang. Weiterhin bleibt der Unterschied in der Lernrate zwischen der Standard- und „pure-study“-Bedingung bestehen; d.h. im zweiten Testdurchgang der Standardbedingung erinnerten sich die Probanden an 5.65% mehr Wörter (entspricht ungefähr ein bis zwei Wörtern) als in der „pure-study“-Bedingung. Hierbei korreliert der individuelle Ausgangswert mit der Lernrate zu  $r = .37$  ( $p < .05$ ) und mit der Effektivität der Lernbedingung zu  $-.74$  ( $p < -.05$ ). Dies zeigt, dass Lerner, welche nach dem ersten Durchgang mehr Wörter korrekt erinnerten, mit jedem weiteren Durchgang mehr Wörter dazu lernten sowie ebenfalls mehr Wörter in der Standardbedingung lernten (im Gegensatz zur „pure-study“-Bedingung). Es zeigt sich kein signifikanter Zusammenhang zwischen der individuellen Effektivität der Durchgänge und der Effektivität der Lernbedingung. Der logarithmische Verlauf der Lernleistung pro Durchgang beschreibt die Daten besser, als der lineare oder der quadratische Verlauf und wird daher in den weiterführenden Modellen 3–6 verwendet.

Tabelle 5

*Experiment 1 Modelle 0 bis 2: Parameterschätzungen und Fit-Indizes*

Parameter	Modell 0	Modell 1	Modell 2
	Linear	Quadratisch	Logarithmisch
<b>Feste Effekte</b>			
$\beta_0$ : Ausgangswert	25.13*	23.25*	22.91*
$\beta_1$ : Durchgang	9.69*	15.31*	
$\beta_2$ : Durchgang <sup>2</sup>		-1.88*	
$\beta_3$ : ln(Durchgang)			21.07*
$\delta_0$ : Lernbedingung (0 = Standard, 1 = „pure-study“)	0.32	.32	0.95
$\delta_1$ : Lernbedingung*Durchgang	-3.92*	-3.92*	
$\delta_1$ : Lernbedingung*ln(Durchgang)			-8.19*
<b>Zufällige Effekte</b>			
$\sigma_e^2$ : Var(Residuen)	64.46*	58.84*	57.24*
$\sigma_{b_0}^2$ : Var(Ausgangswert)	155.93*	158.60*	154.38*
$\sigma_{b_1}^2$ : Var(Durchgang)	4.76*	5.32*	29.95* <sup>a</sup>
$\sigma_{d_0}^2$ : Var(Lernbedingung)	36.55*	39.36*	40.16*
Korr. Ausgangswert x Durchgang	.55*	.49*	.37* <sup>a</sup>
Korr. Ausgangswert x Lernbedingung	-.74*	-.73*	-.74*
Korr. Lernbedingung x Durchgang	.12	.11	.10 <sup>a</sup>
<b>Informationskriterium</b>			
-2 LL	4794.0	4757.5	4754.0
AIC	4816.0	4781.5	4776.0

*Anmerkungen.* -2LL =  $-2 \times \text{Log-Likelihood}$ , AIC = Akaike Information Criterion.

<sup>a</sup> Durchgang hier jeweils als ln(Durchgang)

\* $p < .05$

In einem nächsten Schritt wird das Modell 2 so erweitert, dass für die Standardbedingung sowie für die „pure-study“-Bedingung unterschiedliche Varianzen geschätzt werden (Modell 3, Tabelle 6). Das bedeutet, dass die Residuen in den beiden Lernkurven unterschiedlich weit um den durchschnittlichen Verlauf streuen können. Hierbei

ist die Varianz in der „pure-study“-Bedingung größer als in der Standardbedingung, was auf höhere Abweichungen der geschätzten von den gemessenen Werten hinweist. Die Zusammenhänge der zufälligen Effekte bleiben für die Korrelation des Ausgangswerts mit der Lernbedingung, sowie der Lernbedingung mit dem logarithmierten Effekt des Durchgangs ähnlich wie im Modell 2; allerdings sinkt die Korrelation zwischen Ausgangswert und Durchgang im Gegensatz dazu und wird statistisch nicht mehr signifikant. Der AIC dieses Modells verbessert sich um 12.5 Punkte, wenn die beiden Residuen der Lernbedingungen im Modell korrelieren können (Modell 4). Die negative Korrelation von  $r = -.27$  ( $p < .05$ ) zeigt einen schwachen bis mittelstarken Zusammenhang der Residuen der beiden Lernkurven. Demnach zeigten Personen, deren wahrer Wert in der Standardbedingung nah an dem geschätzten Wert lag, tendenziell höhere Residuen in der „pure-study“-Lernkurve.

In Modell 5 wird zusätzlich die Altersgruppe berücksichtigt (0 = junge; 1 = alt) (siehe Abbildung 4). Der Ausgangswert  $\beta_0$  beschreibt, dass ein jüngerer Erwachsener im ersten Durchgang im Mittel 29.53% (ungefähr acht Wörter) korrekt erinnert ( $t(78) = 16.35, p < .05, \beta_0 = 29.53, d = 3.70$ ). Die Analyse zeigt einen signifikanten Haupteffekt der Durchgänge ( $t(78) = 19.97, p < .05, \beta_1 = 26.29, d = 4.52$ ). Dementsprechend sagt das Modell vorher, dass jüngere Probanden mit jedem weiteren Durchgang in der Standardbedingung ihre Lernleistung erhöhen; d.h. ein jüngerer Erwachsener berichtet im zweiten Durchgang der Standardbedingung im Mittel fünf Wörter (18%) mehr als im ersten Durchgang. Zusätzlich zeigten ältere Probanden eine durchschnittlich niedrigere Lernleistung in der Standardbedingung als jüngere ( $t(398) = -5.18, p < .05, \alpha_0 = -13.24, d = -0.52$ ). D.h. ältere Erwachsene erinnerten sich bei der ersten Abfrage in der Standardbedingung im Mittel an vier Wörter (16.29%). Es zeigt sich kein signifikanter Haupteffekt der Lernbedingung in Bezug auf den Ausgangswert ( $t(78) = 1.34, p = .18, \delta_0 = 2.50, d = 0.30$ ). Allerdings

interagiert die Lernbedingung mit den Durchgängen ( $t(398) = -7.49, p < .05, \delta_1 = -12.85, d = -0.75$ ) und zwar in dem Sinne, dass bei jüngeren und älteren Lernern in der Standardbedingung mehr Wörter mit jedem weiteren Durchgang gelernt wurden als in der „pure-study“-Bedingung. Die Interaktion von Altersgruppe mit Lernbedingung ist nicht signifikant ( $t(398) = -1.18, p = .24, \alpha_1 = -3.09, d = -0.12$ ), hingegen ergibt sich ein signifikanter Interaktionseffekt von Altersgruppe mit der Anzahl der Durchgänge ( $t(398) = -5.60, p < .05, \alpha_2 = -10.43, d = -0.56$ ). Somit war die Lernrate älterer Probanden in der Standardbedingung im Mittel geringer als die Lernrate jüngerer Teilnehmer. Allerdings unterschieden sich die Altersgruppen nicht in der Lernrate in der „pure-study“-Bedingung<sup>8</sup>. Auch die dreifach-Interaktion zwischen Lernbedingung, Durchgang und Altersgruppe ist signifikant ( $t(398) = 3.84, p < .05, \delta_2 = 9.31, d = 0.39$ ). Diese zeigt, dass der Unterschied im Lernzuwachs zwischen der Standard- und der „pure-study“-Lernkurve in der älteren Stichprobe geringer ausfällt als in der jüngeren.

Ebenfalls in Modell 5 sind alle drei Varianzschätzungen der zufälligen Effekte signifikant von Null verschieden. Dies bedeutet, dass es zwischen Personen bedeutsame Unterschiede im Ausgangswert (Gedächtnisleistung im ersten Durchgang in der Standardbedingung), in den Effekten des (logarithmierten) Durchgangs, sowie in den Effekten der Lernbedingung gibt. Hierbei korrelieren der individuelle Ausgangswert mit der Effektivität der Lernbedingung zu  $r = -.77 (p < .05)$ , was zeigt, dass Lerner, die nach dem ersten Durchgang mehr Wörter korrekt erinnern, mehr Wörter in der Standardbedingung lernen (im Gegensatz zur „pure-study“-Bedingung), als Individuen mit niedrigeren Ausgangswerten. Es zeigen sich weder signifikante lineare Zusammenhänge zwischen der Effektivität der Durchgänge mit dem Ausgangswert, noch mit der Lernbedingung. Das

---

<sup>8</sup> Verwendet man diesem Modell die „pure-study“-Bedingung als Referenzgruppe, so ergibt sich für die Interaktion des Durchgangs mit der Altersgruppe  $t(398) = -0.51, p = .61, \alpha_{2\text{„pure-study“}} = -1.12, d = -0.05$ .

beschriebene Modell klärt 56.4% der Varianz innerhalb und 51.5% der Varianz zwischen den Personen auf.

In einem letzten Modell werden zusätzlich die Effekte der Verarbeitungsgeschwindigkeit mit aufgenommen—am Mittelwert zentriert. Zwischen den Altersgruppen gibt es große Unterschiede in der Verarbeitungsgeschwindigkeit ( $t(78) = 5.63$ ,  $p < .05$ , Differenz = 1.05,  $d = 1.26$ ). Es zeigen sich keine statistisch signifikanten Zusammenhänge zwischen der Lernleistung und der Verarbeitungsgeschwindigkeit; weder im Ausgangswert noch im Durchgang, der Altersgruppe oder in deren Interaktion. Daher bleiben die Schätzungen für die festen und zufälligen Effekte ähnlich zu Modell 5. Selbst wenn Modell 6 0.6% mehr innerhalb und 1.2% mehr Varianz zwischen Personen aufklärt, verschlechtert sich das Informationskriterium AIC verschlechtert von Modell 5 auf 6 um 5.2 Punkte. Damit beschreibt Modell 5, gemäß dem AIC, die Daten am besten.

Tabelle 6

*Experiment 1 Modelle 3 bis 6: Parameterschätzungen und Fit-Indizes*

Parameter	Modell 3 Heterogene V.	Modell 4 +Korrelation	Modell 5 + Altersgruppe	Modell 6 + Speed
<b>Feste Effekte</b>				
$\beta_0$ : Ausgangswert	22.91*	22.91*	29.53*	28.20*
$\beta_1$ : ln(Durchgang)	21.07*	21.07*	26.29*	26.44*
$\delta_0$ : Lernbedingung (0 = Standard, 1 = "pure-study")	0.95	0.95	2.50	2.50
$\delta_1$ : Lernbed.*ln(Durchgang)	-8.19*	-8.19*	-12.85*	-12.85*
$\alpha_0$ : Altersgruppe			-13.24*	-11.73*
$\alpha_1$ : Altersgruppe*Lernbed.			-3.09	-3.09
$\alpha_2$ : Altersgruppe*ln(t)			-10.43*	-10.27*
$\delta_2$ : Altersgruppe*ln(t) *			9.31*	9.31*
Lernbedingung				
Speed				2.53
Speed*ln(t)				-0.29
Speed*Altersgruppe				-2.19
Speed*ln(t)*Altersgruppe				0.88
<b>Zufällige Effekte</b>				
$\sigma_{eST}^2$ : Var(Residuen) – Standard	43.03*	39.41*	37.72*	37.45*
$\sigma_{ePS}^2$ : Var(Residuen) – „pure- study“	69.27*	68.28*	66.02*	66.36*
Korr. $\sigma_{eST} \times ePS$		-.27*	-.24*	-.24*
$\sigma_{b_0}^2$ : Var(Ausgangswert)	158.04*	163.61*	99.12*	97.34*
$\sigma_{b_1}^2$ : Var(ln(Durchgang))	39.05*	45.69*	34.51*	34.55*
$\sigma_{d_0}^2$ : Var(Lernbedingung)	40.70*	34.89*	32.22*	32.21*
Korr. Ausgangswert x ln(Durchgang)	.27	.18	-.20	-.20
Korr. Ausgangswert x Lernbedingung	-.78*	-.79*	-.77*	-.78*
Korr. Lernbedingung x ln(Durchgang)	.11	.11	.35	.36
<b>Informationskriterium</b>				
-2 LL	4745.5	4731.0	4654.5	4651.6
AIC	4769.5	4757.0	4688.5	4693.6

Anmerkungen. -2LL =  $-2 \times \text{Log-Likelihood}$ , AIC = Akaike Information Criterion.

\* $p < .05$

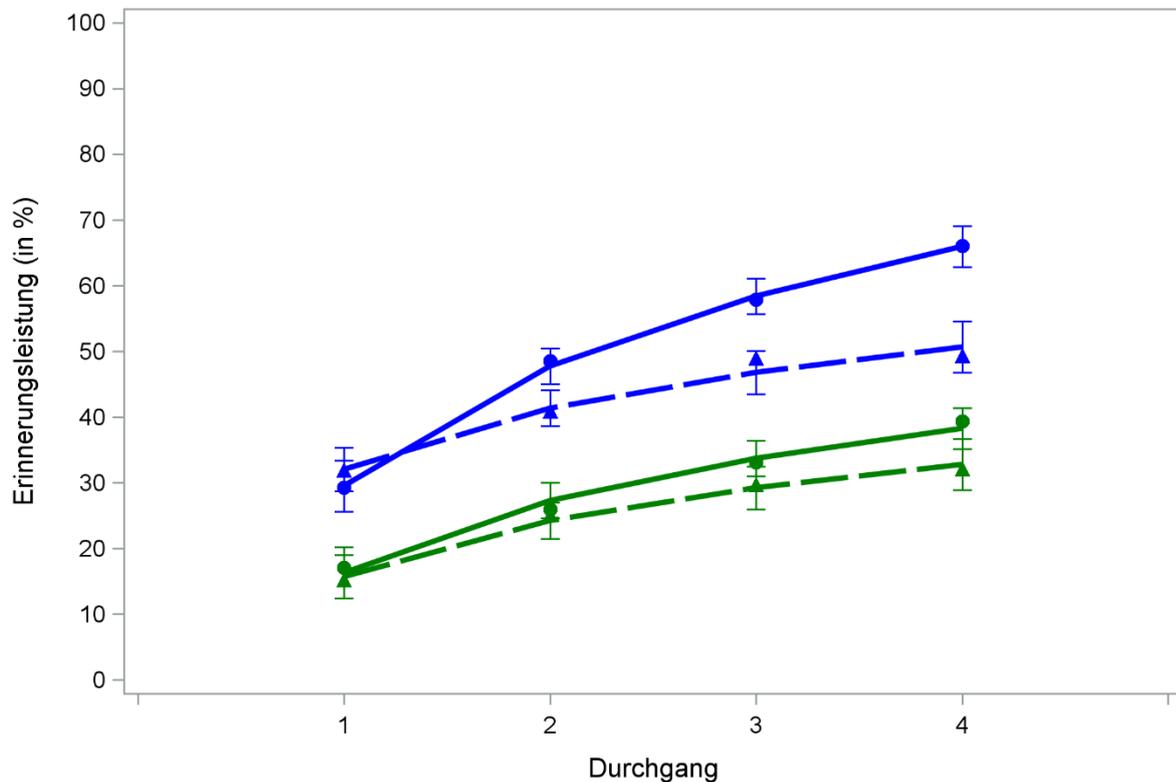


Abbildung 4. Experiment 1: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) für Modell 5. Die durchgehenden Linien repräsentieren die Standardlernbedingung und die gestrichelten Linien die „pure-study“-Bedingung. Die Kreise zeigen die beobachteten Werte der Standardbedingung, die Dreiecke die beobachteten Werte der „pure-study“-Lernbedingung in den beiden Altersgruppen. Die Fehlerbalken stellen die Standardfehler dar.

### 2.3.3. Zusammenfassung der Ergebnisse aus Experiment 1

Aus den Ergebnissen geht hervor, dass die Probanden von wiederholten Testdurchgängen profitierten, und zwar schon ab dem ersten Testdurchgang. Damit replizieren die Ergebnisse des Experiments die Befunde von Roediger und Smith (2012) und zeigen, dass in der Standardbedingung mehr Wörter gelernt werden als in der „pure-study“-Bedingung. Die ältere Stichprobe zeigt dabei eine geringere Lernrate und profitiert weniger von den Effekten zusätzlicher Tests. Dieser altersspezifische Unterschied im Testeffekt ist von kleiner bis mittlerer Stärke (Interaktion aus Bedingung, Durchgang und Altersgruppe:  $d = 0.39$ ). Ältere erinnern sich am Anfang (nach dem ersten Lerndurchgang) an weniger

Wörter als Jüngere, lernen weniger mit jedem weiteren Lerndurchgang dazu und profitieren im geringeren Umfang von dem Testeffekt.

Dabei zeigen sich individuelle Unterschiede im Ausgangswert, der Lernrate und in der Lernbedingung. Dies legt den Schluss nahe, dass nicht alle Lerner im gleichen Umfang von den Testeffekten profitieren. Hierbei korreliert der individuelle Ausgangswert hoch mit der Effektivität der Lernbedingung ( $r = -.77$ ), was zeigt, dass Lerner, die nach dem ersten Durchgang mehr Wörter korrekt erinnern, folglich mehr Wörter in der Standardbedingung lernen (im Gegensatz zur „pure-study“-Bedingung). Demnach hängt die Wirksamkeit des Testeffekts von der individuellen Ausgangsleistung ab. Dieses Ergebnis stützt das Kernkonzept des Testeffekts: je mehr Wörter erinnert werden, desto mehr erfolgreiche Abrufprozesse werden durch den Testeffekt beeinflusst (vgl. Karpicke, 2017; Marsh et al., 2009; Meyer & Logan, 2013). Die Residualvarianzen für die „pure-study“-Lernkurve sind höher als die Residualvarianzen der Standardlernkurve. Diese Unterschiede in der Fehlervarianz der beiden Lernkurven können durch die Art der Erhebung zustande kommen (die Standardbedingung besteht aus einer Aufgabe, wohingegen die „pure-study“-Bedingung aus den Werten mehrerer Lernaufgaben zusammengesetzt wird).

Allerdings klärt die Verarbeitungsgeschwindigkeit keinen statistisch signifikanten Anteil individueller Unterschiede auf. Die Ergebnisse sind unerwartet, da ursprünglich erwartet wurde, dass die Verarbeitungsgeschwindigkeit individuelle Unterschiede in der Lernrate aufklärt (bspw. Kurtz et al., 2013). Aufgrund der hohen Altersunterschiede in der Verarbeitungsgeschwindigkeit könnten die Effekte der Verarbeitungsgeschwindigkeit stark mit den Effekten der Altersgruppe zusammenhängen (Modell 6 ohne Kontrolle von Altersgruppeneffekten:  $\beta_{VG}: t(399) = 3.94, p < .05, \beta_{VG} = 4.25, d = .39; \beta_{VG*ln(t)}: t(399) = 2.39, p < .05, \beta_{VG*ln(t)} = 2.08, d = .24$ ). Die Rolle der Verarbeitungsgeschwindigkeit für die Lernleistung wird in den weiteren Experimenten der Arbeit weiter untersucht.

Als Limitation der Studie ist anzuführen, dass für die Lernkurve der „pure-study“-Bedingung der Wert für die Lernleistung des ersten Testdurchgangs nicht separat erhoben wurde. Zwar konnte dies durch die Aufteilung der Ergebnisse des ersten Testdurchgangs der Standardbedingung kompensiert werden, führt jedoch einerseits zu untypischen Korrelationsmustern und andererseits zu höheren Standardabweichungen und damit zu höheren Standardfehlern der Bedingung (siehe Tabelle 4). Dennoch zeigt das Experiment, dass es statistisch bedeutsame Altersunterschiede im Testeffekt gibt.

#### *2.4. Experiment 2 – Wortpaare*

In einem zweiten Experiment sollen Altersunterschiede im Testeffekt für eine andere Form von Lernmaterial getestet werden, nämlich für Wortpaaraufgaben. Bezüglich des Testeffekts und der Altersunterschiede werden die gleichen Annahmen wie für Experiment 1 getroffen. Aufgrund der spezifischeren Assoziationsaufgabe sollten Ältere bei Wortpaaren in geringem Ausmaß von den zusätzlichen Testdurchgängen profitieren (siehe Überlegungen Kapitel 1.2.2.). Zusätzlich wird ein anderes Design verwendet. In Experiment 1 wurden die Werte für die Standard- und die „pure-study“-Lernkurve aus verschiedenen Aufgaben erstellt. In Experiment 2 wurden alle Testwerte innerhalb einer Aufgabe erhoben. Dadurch ist ebenfalls das Lernmaterial in allen Lernbedingungen gleich schwer (da in den Lerndurchgängen immer dieselben Wortpaare gelernt werden) und es benötigt keine Experimentdesigns, um die Reihenfolge verschiedener Lernaufgaben auszubalancieren.

##### *2.4.1. Methode Experiment 2*

###### *Stichprobe*

Am zweiten Experiment nahmen 53 jüngere und 55 ältere Erwachsene (insgesamt N = 108) teil. Ein älterer Erwachsener brach das Experiment frühzeitig ab und wurde daher aus

den Berechnungen ausgeschlossen. Damit gehen die Daten von 107 Probanden in die Berechnungen ein. Die jüngeren Teilnehmer der Stichproben waren im Mittel 22.58 Jahre alt ( $SD = 3.11$ ; 18-30 Jahre), wohingegen die Probanden im höheren Erwachsenenalter durchschnittlich 68.85 Jahre alt waren ( $SD = 5.80$ ; 60-83 Jahre). Von den 107 Probanden waren 54 (50.5%) weiblich (jüngere Erwachsene: 53% Frauen; ältere Erwachsene: 48%).<sup>9</sup> Um ihren Schulabschluss anzugeben, hatten die Teilnehmer sieben Optionen zur Verfügung: „keinen weiterführenden Abschluss“, „Haupt-/Volksschulabschluss“, „Mittlere Reife“, „Berufsausbildung“, „Fachabitur“, „Abitur“, „(Fach-) Hochschulabschluss“. In der jüngeren Stichprobe gaben 29 (54.7%) Abitur und 13 (24.5%) (Fach-) Hochschulabschluss als höchsten erreichten Schulabschluss an. Wohingegen die beiden häufigsten Antworten bei den älteren Erwachsenen, Berufsausbildung (18 Teilnehmer, 33.3%) sowie (Fach-) Hochschulabschluss (15 Teilnehmer, 27.8%) waren. Die meisten jüngeren Probanden waren ledig ( $N = 49$ , 92,5%) und Studenten ( $N = 36$ , 67.9%), wohingegen bei meisten älteren Teilnehmer verheiratet ( $N = 46$ , 85.2%) und im Ruhestand waren ( $N = 40$ , 74.1%).

### *Messinstrumente*

#### *Paar-Assoziations-Lernen*

Um die Lernleistung zu messen, wurde eine Paar-Assoziations-Aufgabe verwendet. Hierfür wurden 100 zweisilbige deutsche Nomen aus der “Berlin Affective Word List Reloaded” (BAWL-R; Vö et al. 2009) ausgewählt. Dabei fungierten 50 Wörter als Hinweiswörter und die anderen 50 als Zielwörter (zufällige Aufteilung). Die Wörter wiesen im Mittel eine neutrale Emotionalität von .10 auf (auf einer Skala von -3 bis 3 – sehr negativ bis sehr positiv), sowie eine geringe bis mittlere Bildhaftigkeit von 3.28 (auf einer Skala von

---

<sup>9</sup> Die Geschlechterverteilung zwischen den Altersgruppen war nicht signifikant verschieden,  $\chi^2(1, 106) = .23$ ,  $p = .63$ , Cohen’s  $w = 0.09$ .

1 bis 7–niedrige Bildhaftigkeit bis hohe Bildhaftigkeit). Zusätzlich wiesen die Wörter eine mittlere Erregung von 2.18 auf (auf einer Skala von 1 bis 5–beruhigend bis aufregend). Alle Wörter kamen mit einer durchschnittlichen Häufigkeit von 91.79 pro Million relativ häufig vor. Die 50 Wortpaare wurden für jeden Probanden zufällig auf fünf Listen mit je 10 Wortpaaren verteilt.

### *Lerndesign*

Das Lernexperiment bestand aus fünf aufeinanderfolgenden Lern-Test-Durchgängen. Während jeder Lernphase wurden alle 50 Wortpaare in schwarzen Buchstaben (Schriftart: Arial; Schriftgröße: 10% des Laptopbildschirms) für je 3 Sekunden präsentiert. In dem anschließenden Testdurchgang wurden hintereinander die Hinweisreize der zehn Wortpaare aus der ersten Liste präsentiert. Die Versuchsperson wurde aufgefordert, die jeweiligen Zielwörter zu nennen. Im zweiten Testdurchgang wurden die zehn Wortpaare der ersten Liste plus die 10 Wortpaare der zweiten Wortliste abgefragt. Dementsprechend wurde im dritten Testdurchgang 30 Wortpaare abgefragt (Liste 1 bis 3), im vierten 40 Wortpaare und im fünften Durchgang alle 50 Wortpaare aus den fünf Wortlisten (siehe Tabelle 7). Die Präsentationsreihenfolge der Wortpaare wurde in Lerndurchgang randomisiert. Auch die Hinweiswörter im Testdurchgang wurden in jedem Durchgang in zufälliger Reihenfolge präsentiert. Jede Versuchsperson hat die komplette Lernaufgabe (alle fünf Wortlisten) bearbeitet.

Tabelle 7

*Lerndesign Experiment 2*

	Durchgang 1	Durchgang 2	Durchgang 3	Durchgang 4	Durchgang 5
Wortliste 1	S T	S T	S T	S T	ST
Wortliste 2	S	S T	S T	S T	ST
Wortliste 3	S	S	S T	S T	ST
Wortliste 4	S	S	S	S T	ST
Wortliste 5	S	S	S	S	ST

*Anmerkungen.* S = Lerndurchgang; T = Testdurchgang.

*Zahlen vergleichen*

Die Aufgabe war es, so schnell wie möglich Zahlenreihen miteinander zu vergleichen, welche am Bildschirm präsentiert werden, wobei eine Zahlenreihe dabei drei bis 13 Zahlen lang sein kann (Ekstrom et al., 1976). Die Teilnehmer hatten 90 Sekunden Zeit die Aufgabe zu bearbeiten. Die Punktzahl ergab sich aus der Anzahl korrekt wiedergegebener Zahlenreihen (mögliche Werte: 0–60). Für die Berechnungen wurden die Werte der Aufgabe z-standardisiert. Bei älteren Erwachsenen zeigt sich ein Cronbachs Alpha von .79 (Rast, 2011).

*Prozedur*

Nachdem die Probanden eine Einverständniserklärung unterzeichnet haben, bearbeiteten sie die Aufgabe in Einzeltestungen an einem Laptop. Anfangs wurden demographische Variablen abgefragt (Alter, Geschlecht, Familienstand, Bildungshintergrund, sowie die berufliche Tätigkeit)<sup>10</sup>. Anschließend folgte die Paar-

<sup>10</sup> Zusätzlich wurden auf einer 10-Punkte Skala gefragt „Wie fit fühlen Sie sich gerade?“, „Wie häufig benutzen Sie einen Computer?“ und „Wie groß ist Ihre Motivation für die folgenden Aufgaben?“.

Assoziations-Aufgabe (wie oben beschrieben). Es folgten ein Zahlen-Rückwärts-Test sowie der Zahlenvergleichstest, um die Zeit bis zu einem weiteren Abrufversuch zu überbrücken (Dauer ungefähr 9 Minuten). Nachdem nochmal alle 50 Wortpaare abgefragt wurden, konnten die Teilnehmer angeben, ob Sie eine Lernstrategie verwendet haben. Das Experiment dauerte insgesamt ca. 60 Minuten.

### *Datenanalyse*

Die Werte des fünften und des verzögerten Testdurchgangs fehlt für eine ältere Person. Daher wurden bei dieser Person nur die Werte der vorherigen Testdurchgänge und die Leistung im Verarbeitungsgeschwindigkeitstest in die Berechnungen mit aufgenommen. Für die Analyse der Daten wurde eine gemischte Regression verwendet. Hierbei wurde ein ähnliches Modell wie in Experiment 1 verwendet mit dem Unterschied, dass der Einfluss des Durchgangs linear (und nicht logarithmisch) geschätzt wurde (Beschreibungen beziehen sich auf Modell 5 in Tabelle 11).

$$\mathbf{y}_{it} = \begin{pmatrix} \beta_0 + b_{0i} + \alpha_0 x_{2i} \\ \beta_0 + b_{0i} + \alpha_0 x_{2i} + \delta_0 + d_{0i} + \alpha_1 x_{2i} \end{pmatrix} + \begin{pmatrix} \beta_1 + b_{1i} + \alpha_2 x_{2i} \\ \beta_1 + b_{1i} + \alpha_2 x_{2i} + \delta_1 + d_{0i} + \delta_2 x_{2i} \end{pmatrix} t + \begin{pmatrix} u_{ST} \\ u_{PS} \end{pmatrix} + \begin{pmatrix} e_{it1} \\ e_{it2} \end{pmatrix} \quad (5)$$

Die Effekte der Bedingung, sowie die Effekte der Altersgruppe wurden in Form von Dummyvariablen modelliert ( $x_{1i}$ : 0 = Standard, 1 = „pure-study“;  $x_{2i}$ : 0 = jung, 1 = älter). Die Parameter werden wie in Experiment 1 interpretiert (siehe Kapitel 2.3.1.). Im Gegensatz zum ersten Experiment gibt es hier fünf Durchgänge. Es wird Autoregression erster Ordnung zwischen den Fehlern angenommen. Zusätzlich wird angenommen, dass die individuellen Fehler mit konstanter Varianz normalverteilt sind sowie, dass die Kovarianz-Matrix der zufälligen Effekte des Ausgangsniveaus, der Bedingung und des Lernzuwachs wie in Experiment 1 aufgebaut ist (siehe Formel 4).

Die Berechnungen wurden mit SAS Version (9.4) (SAS Institute, 2016) durchgeführt. Die gemischte Regression wurde mit der Prozedur MIXED und der „Maximum Likelihood“-Schätzmethode modelliert. Die Varianzaufklärung auf den beiden Ebenen des Modells werden mit  $R^2$  (Recchia, 2010) und Effektstärken mit Cohen's  $d$  angegeben—die Angaben dienen dabei vor allem der Veranschaulichung.

#### 2.4.2. Ergebnisse Experiment 2

##### *Deskriptive Statistiken*

Tabelle 8 zeigt die deskriptiven Statistiken der Lernleistung aller Testdurchgänge. Wie erwartet, steigt die Lernleistung in der Standardbedingung sowie in der „pure-study“-Bedingung mit jedem weiteren Durchgang bei jüngeren und älteren Erwachsenen an. Die Lernleistung jüngerer Probanden unterscheidet sich signifikant von der älterer ( $F(9, 96) = 6.54, p < .05$ ). Hierbei werden die Altersunterschiede in der Standardbedingung mit jedem weiteren Lern-Test-Durchgang größer, wohingegen in der „pure-study“-Bedingung die Altersunterschiede relativ konstant bleiben (siehe Abbildung 6). Dass es sich hier um eine relativ schwere Lernaufgabe handelt, ist daran zu erkennen, dass in keiner der Altersgruppen bzw. Lernbedingungen im Durchschnitt mehr als 50% Wortpaare der Liste erinnert wurden (mögliche Lernleistung: 0 bis 10 Wortpaare). In Abbildung 5 sind die beobachteten Mittelwerte der Standard- und der „pure-study“-Bedingung für die beiden Altersgruppen abgetragen. Diese zeigen allerdings nicht die erwarteten negativ beschleunigten Lernkurven, sondern eher lineare Verläufe (v.a. bei der älteren Altersgruppe; siehe Abbildung 5).

Diese Form der Lernkurve kann mit der Art und Weise der Erhebung zusammenhängen, da in den ersten vier Durchgängen mehr Wortpaare gelernt wurden als abgefragt. So haben sich die Probanden im ersten Testdurchgang im Mittel an 0.92 ( $SD = 1.41$ ) Wortpaare erinnert, allerdings erinnerten sich Probanden im gesamten

(Erinnerungsleistung aller abgefragten Wortlisten) zweiten Testdurchgang an 3.38 ( $SD = 3.51$ ), im dritten an 7.54 ( $SD = 6.45$ ), im vierten an 12.50 ( $SD = 9.50$ ) und im letzten Testdurchgang an 18.94 ( $SD = 12.91$ ) erinnert. Demnach vergrößern sich die Abstände der mittleren Lernleistung in jeden weiteren Testdurchgang (wenn die korrekt erinnerten Wortpaare aus jeweils allen abgefragten Listen berücksichtigt werden). Nachdem in der den Testdurchgängen der Standard- und der „pure-study“-Bedingung immer nur die Anzahl korrekt erinnelter Wortpaare einer Liste eine Rolle spielen, könnte sich daraus die eher untypische lineare Lernkurvenform ergeben.

Tabelle 8

*Experiment 2: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

Variablen		Jung (N = 53)	Alt (N = 54)	Gesamt (N = 107)	Diff(Jung – Alt)	Cohens d
Standard 1	<i>M</i>	1.21	0.63	0.92	$t(105) = 2.16^*$	0.42
	<i>SD</i>	1.46	1.31	1.41		
Standard 2	<i>M</i>	2.55	1.19	1.86	$t(105) = 3.78^*$	0.73
	<i>SD</i>	2.04	1.67	1.98		
Standard 3	<i>M</i>	3.83	1.69	2.75	$t(105) = 5.45^*$	1.05
	<i>SD</i>	2.17	1.89	2.29		
Standard 4	<i>M</i>	4.81	2.20	3.50	$t(105) = 5.78^*$	1.12
	<i>SD</i>	2.37	2.29	2.66		
Standard 5	<i>M</i>	5.81	2.57	4.19	$t(104) = 6.79^*$	1.32
	<i>SD</i>	2.29	2.62	2.94		
„pure-study“ 2	<i>M</i>	2.32	0.74	1.52	$t(105) = 5.08^*$	0.98
	<i>SD</i>	1.90	1.26	1.79		
„pure-study“ 3	<i>M</i>	3.21	1.13	2.16	$t(105) = 5.61^*$	1.09
	<i>SD</i>	2.32	1.41	2.17		
„pure-study“ 4	<i>M</i>	3.66	1.65	2.64	$t(105) = 4.90^*$	0.95
	<i>SD</i>	2.22	2.03	2.34		
„pure-study“ 5	<i>M</i>	4.49	1.89	3.19	$t(104) = 5.61^*$	1.09
	<i>SD</i>	2.58	2.19	2.72		

*Anmerkungen.* Standard 1 = erster Testdurchgang der Standardbedingung (wird zudem als erster Testdurchgang der „pure-study“-Lernkurve verwendet).

\* $p < .05$

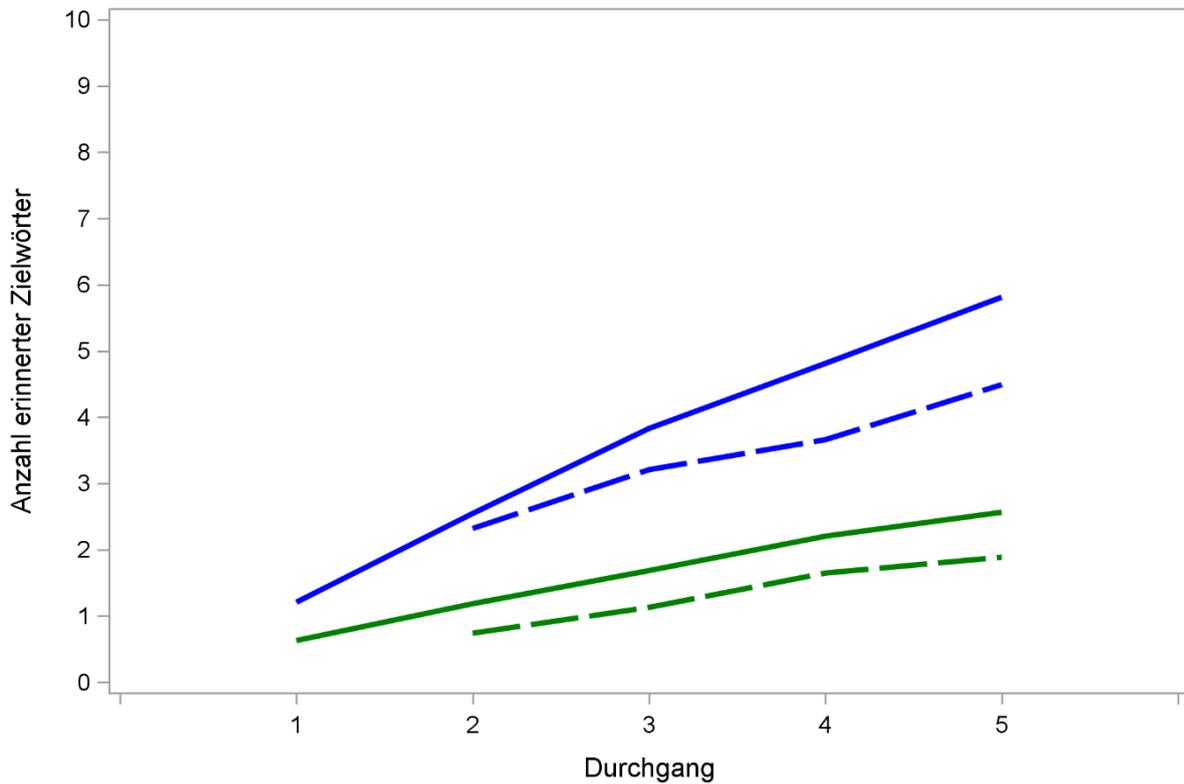


Abbildung 5. Experiment 2: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün). Die durchgehenden Linien repräsentieren die Standardlernbedingung und die gestrichelten Linien die „pure-study“-Bedingung.

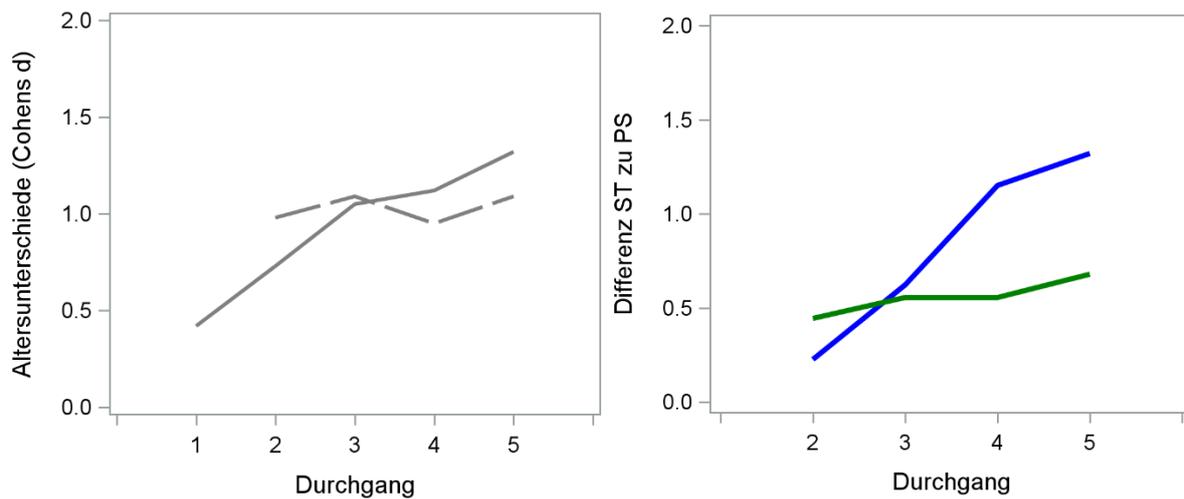


Abbildung 6. Experiment 2: Links die Altersunterschiede in der Standardbedingung (durchgehende Linie) und in der „pure-study“-Bedingung (gestrichelte Linie). Rechts die Differenz zwischen Standard- und „pure-study“-Bedingung (ST – PS) für jüngere Erwachsene (blau) und ältere Erwachsene (grün).

Tabelle 9 zeigt die Interkorrelationen aller Testdurchgänge aus der Standard- sowie aus der „pure-study“-Bedingung. Der vorherige Testdurchgang korreliert fast durchgängig höher (in der Standard- sowie in der „pure-study“-Bedingung) mit dem darauffolgenden Durchgang als der Testdurchgang vor zwei, drei oder vier Durchgängen. Einzige Ausnahme stellt hierbei die Korrelation der Lernleistung im ersten Durchgang der Standardbedingung mit der Lernleistung in „pure-study“ 3-Aufgabe bei der älteren Stichprobe dar ( $r_{st1ps3} = .68$ ;  $r_{ps2ps3} = .64$ ). Dieses Muster ist konsistent mit der Annahme, dass zeitlich näherliegende Testungen höher zusammenhängen als Testungen, die zeitlich weiter auseinanderliegen.

Tabelle 9

*Experiment 2: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Standard 1		.72*	.65*	.62*	.50*	.60*	.59*	.47*	.38*
2. Standard 2	.82*		.78*	.72*	.64*	.75*	.62*	.59*	.54*
3. Standard 3	.72*	.85*		.88*	.77*	.74*	.70*	.69*	.61*
4. Standard 4	.55*	.79*	.85*		.81*	.71*	.69*	.66*	.62*
5. Standard 5	.60*	.82*	.85*	.86*		.61*	.61*	.70*	.67*
6. „pure-study“ 2	.59*	.60*	.58*	.64*	.57*		.64*	.61*	.55*
7. „pure-study“ 3	.68*	.72*	.76*	.69*	.70*	.64*		.73*	.69*
8. „pure-study“ 4	.65*	.69*	.73*	.76*	.69*	.72*	.77*		.69*
9. „pure-study“ 5	.64*	.71*	.65*	.69*	.73*	.63*	.78*	.79*	

*Anmerkungen.* Oberhalb der Diagonale werden die Korrelationen für die jüngere Stichprobe und unterhalb der Diagonale die Korrelationen der älteren Stichprobe dargestellt.

\* $p < .05$

*Mehrebenenmodelle*

Die Interklassenkorrelation beträgt in einem „*Unconditional Model*“  $ICC = .58$ . Somit zeigt sich bei diesem Experiment, dass bei den Analysen die Abhängigkeit der Daten beachtet werden sollte, und zwar in dem Sinne, dass Effekte zwischen und innerhalb von Personen betrachtet werden. In einem ersten Modell (Modell 0) wird die Lernleistungen innerhalb einer Person in Abhängigkeit des Durchgangs sowie der Lernbedingung modelliert ( $t = 0, 1, 2, 3, 4$ ) (siehe Tabelle 10). Die Lernrate wird als der lineare Effekt des Durchgangs für die Standardbedingung geschätzt. Im Durchschnitt erinnerten sich Personen im ersten Testdurchgang der Standardbedingung an ein Wort und lernen mit jedem weiteren Durchgang 0.81 Wortpaare dazu. Die Lernrate war in der Standardbedingung im Mittel 0.27 Wortpaare höher als in der „pure-study“-Bedingung. Die Varianzschätzungen für den Ausgangswert, den Durchgang sowie für den Effekt der Lernbedingung sind signifikant von Null verschieden, was statistisch bedeutsame Unterschiede zwischen Personen andeutet. Wobei die Ausprägung des individuellen Ausgangswert mit der Lernrate und der Effektivität der Lernbedingung korrelieren.

Um einen nichtlinearen Verlauf abbilden zu können, wird in Modell 1 zusätzlich der Effekt des quadrierten Durchgangs ( $t^2 = 0, 1, 4, 9, 16$ ) geschätzt. Die festen Effekte für den Ausgangswert, den Durchgang sowie der Interaktion von Lernbedingung mit Effekt des Durchgangs bleiben, im Vergleich zu Modell 0, statistisch signifikant. Ebenfalls ist der negative Effekt des quadrierten Durchgangs signifikant, welcher eine abflachende Lernkurve beschreibt. Der Modellfit des AIC verbesserte sich durch die Hinzunahme eines quadratischen festen Effektes um 2.2 Punkte. Die zufälligen Effekte zeigen, dass sich Personen in ihren Ausgangswert, der Lernrate und in der Effektivität der Lernbedingung statistisch signifikant unterscheiden. Die Korrelation der individuellen Ausgangswerte mit dem Effekt des Durchgangs (Lernrate) zeigt, dass Lerner, die nach dem ersten

Lerndurchgang mehr Wortpaare erinnern konnten, tendenziell mit jedem weiteren Lerndurchgang mehr Wortpaare dazugelernt haben. Zusätzlich profitieren Individuen mit höherem Ausgangswert von der Standardbedingung. Dieses Muster ist konsistent mit der Annahme, dass der Testeffekt beim erfolgreichen Abruf auftritt.

In Modell 2 wird der Effekt des Durchgangs logarithmisch modelliert ( $\ln(t) = 0, 0.69, 1.10, 1.39, 1.61$ ). Allerdings steigt, im Gegensatz zum quadratischen Modell, der AIC um 15.9 Punkte. Diese Veränderung der Modellpassung entspricht nicht dem erwarteten Verlauf der Lernkurve (13.7 Punkte höher als das lineare Modell) und könnte mit der Art der Erhebung zusammenhängen.

Tabelle 10

*Experiment 2 Modelle 0 bis 2: Parameterschätzungen und Fit-Indizes*

Parameter	Modell 0	Modell 1	Modell 2
	Linear	Quadratisch	Logarithmisch
<b>Feste Effekte</b>			
$\beta_0$ : Ausgangswert	1.01*	0.92*	0.72*
$\beta_1$ : Durchgang	0.81*	1.00*	
$\beta_2$ : Durchgang <sup>2</sup>		-0.05*	
$\beta_3$ : ln(Durchgang)			2.00*
$\delta_0$ : Lernbedingung (0 = Standard, 1 = “pure-study”)	-0.01	-0.13	-0.45*
$\delta_1$ : Lernbedingung*Durchgang	-0.27*	-0.23*	
$\delta_1$ : Lernbedingung*ln(Durchgang)			-0.24
<b>Zufällige Effekte</b>			
$\sigma_e^2$ : Var(Residuen)	0.98*	0.97*	0.98*
$\sigma_{b_0}^2$ : Var(Ausgangswert)	2.03*	2.03*	1.99*
$\sigma_{b_1}^2$ : Var(Durchgang)	0.22*	0.22*	1.39 <sup>aa</sup>
$\sigma_{d_0}^2$ : Var(Lernbedingung)	0.67*	0.61*	0.43*
Korr. Ausgangswert x Durchgang	.41*	.41*	.23 <sup>a</sup>
Korr. Ausgangswert x Lernbedingung	-.45*	-.45*	-.58*
Korr. Lernbedingung x Durchgang	-.28	-.28	-.18 <sup>a</sup>
<b>Informationskriterium</b>			
-2 LL	3271.5	3267.3	3285.2
AIC	3293.5	3291.3	3307.2

*Anmerkungen.* -2LL =  $-2 \times \text{Log-Likelihood}$ , AIC = Akaike Information Criterion.

<sup>a</sup> Durchgang hier jeweils als ln(Durchgang)

\* $p < .05$

In den weiteren Modellen wurde der Effekt des Durchgangs als linearer Parameter geschätzt, da dieses Vorgehen am sparsamsten ist und Modell 0 nur gering schlechtere AIC-Werte aufwies als das Modell mit zusätzlichem quadratischen Effekt (BIC Modell 0: 3322.9; BIC Modell 1: 3323.4). In einem nächsten Schritt wird das Modell 0 so erweitert, dass für

die Standardbedingung sowie für die „pure-study“-Bedingung unterschiedliche Varianzen geschätzt wurden (Tabelle 11, Modell 3) und somit die Residuen in den beiden Lernkurven unterschiedlich weit um den durchschnittlichen Verlauf streuen können. Hierbei ist die Varianz in der „pure-study“-Bedingung größer als in der Standardbedingung, was auf höhere Abweichungen der geschätzten von den gemessenen Werten hinweist. Die Zusammenhänge der Varianzen zufälliger Effekte sowie deren Interaktionen bleiben weitgehend ähnlich zu Modell 0, mit der Ausnahme, dass die individuelle Abweichung von der Effektivität der Lernbedingung statistisch signifikant negativ mit der individuellen Abweichung der Lernrate korreliert. D.h. Personen die mehr von der Standardbedingung profitierten, zeigen zudem eine höhere Lernrate.

In Modell 4 wird neben der Korrelation der Residualvarianzen der beiden Lernbedingungen zudem die Autokorrelation erster Ordnung geschätzt, welche statistische Signifikanz aufwies. Demnach korrelieren benachbarte Residuen. Dadurch verbessert sich der Modellfit, gemäß dem AIC, um 11.4. Punkte. Die Residuen der beiden Lernkurven korrelieren nicht statistisch signifikant miteinander.

In Modell 5 wird zusätzlich die Altersgruppe berücksichtigt (0 = junge; 1 = alt) (siehe Abbildung 7). Der Ausgangswert  $\beta_0$  beschreibt, dass ein jüngerer Erwachsener im ersten Durchgang im Mittel 1.33 Wortpaare korrekt erinnerte ( $t(105) = 6.50, p < .05, \beta_0 = 1.33, d = 1.27$ ). Zwar hat die Lernbedingung keinen statistisch relevanten Haupteffekt ( $t(105) = 1.17, p = .25, \delta_0 = 0.30, d = 0.23$ ), jedoch in der Interaktion von Lernbedingung und Durchgang zeigt sich, dass jüngere Probanden in der „pure-study“-Bedingung mit jedem Lerndurchgang im Mittel 0.43 Wortpaare weniger dazulernten als in der Standardbedingung ( $t(638) = -5.06, p < .05, \delta_1 = -0.43, d = -0.40$ ). Ältere Lerner erinnerten sich durchschnittlich an weniger Wortpaare im ersten Testdurchgang ( $t(638) = -2.26, p < .05, \alpha_0 = -0.65, d = -0.18$ ), und lernten weniger Wortpaare mit jedem weiteren Lern-Test-Durchgang als jüngere Lerner

$(t(638) = -7.14, p < .05, \alpha_2 = -0.67, d = -0.57)$ .<sup>11</sup> Die Interaktion von Altersgruppe mit Lernbedingung ( $t(638) = -1.69, p = .09, \alpha_1 = -0.60, d = -0.13$ ) ist nicht signifikant. Auch die dreifach-Interaktion zwischen Lernbedingung, Durchgang und Altersgruppe wurde signifikant ( $t(638) = 2.81, p < .05, \delta_2 = 0.34, d = 0.22$ ). Diese zeigt, dass der Unterschied im Lernzuwachs zwischen der Standard- und der „pure-study“-Lernkurve in der älteren Stichprobe geringer ausfällt als in der jüngeren. Bei der älteren Stichprobe unterscheidet sich die Lernrate in der Standardbedingung nicht statistisch signifikant von der Lernrate der „pure-study“-Bedingung (werden in Modell 5 die älteren Erwachsenen als Referenzgruppe modelliert ergibt sich für die Interaktion des Durchgangs und der Bedingung: ( $t(638) = -1.11, p = .27, \delta_{1\text{alt}} = -0.09, d = -0.09$ ). Auch in Modell 5 sind die Residualvarianzen der beiden Lernbedingungen, die Autokorrelation erster Ordnung sowie die zufälligen Varianzen der Ausgangsleistung, des Effekts des Durchgangs und der Lernbedingung signifikant. Das beschriebene Modell klärt 34% der Varianz innerhalb und 29% der Varianz zwischen den Personen auf.

In einem letzten Modell (Modell 6) werden zusätzlich die Effekte der Verarbeitungsgeschwindigkeit mit aufgenommen. Es zeigt sich eine Interaktion aus Verarbeitungsgeschwindigkeit und Durchgang welche andeutet, dass die Lernrate positiv mit der Verarbeitungsgeschwindigkeit zusammenhängt.<sup>12</sup> Durch die Hinzunahme der Verarbeitungsgeschwindigkeit–welche sich stark zwischen den beiden Altersgruppen unterscheidet ( $t(105) = 6.63, p < .05, \text{Differenz} = 1.08, d = 1.28$ )–verringert sich der Haupteffekt der Altersgruppe und wird dadurch nicht mehr statistisch signifikant. Das

---

<sup>11</sup> Verwendet man diesem Modell die „pure-study“-Bedingung als Referenzgruppe, so ergibt sich für die Interaktion des Durchgangs mit der Altersgruppe:  $t(638) = -2.57, p < .05, \alpha_{2\text{„pure-study“}} = -0.33, d = -0.20$ .

<sup>12</sup> Schätzt man zusätzlich die Interaktion von Verarbeitungsgeschwindigkeit mit Altersgruppe und die dreifach-Interaktion von Verarbeitungsgeschwindigkeit, Altersgruppe und Durchgang (wie in Experiment 1), werden keine der Verarbeitungsgeschwindigkeit statistisch signifikant.

Modell verbessert sich, gemäß dem AIC, um 10.9 Punkte. Modell 6 klärt 40% der Varianz innerhalb und 38% der Varianz zwischen den Personen auf.

Tabelle 11

*Experiment 2 Modelle 3 bis 6: Parameterschätzungen und Fit-Indizes*

Parameter	Modell 3 Heterogene V.	Modell 4 +Autokorr.	Modell 5 + Altersgruppe	Modell 6 + Speed
<b>Feste Effekte</b>				
$\beta_0$ : Ausgangswert	1.01*	1.00*	1.33*	1.17*
$\beta_1$ : Durchgang (= t)	0.81*	0.81*	1.15*	1.06*
$\delta_0$ : Lernbedingung (0 = Standard, 1 = "pure-study")	0.01	-0.01	0.30	0.30
$\delta_1$ : Lernbedingung*t	-0.27*	-0.26*	-0.43*	-0.43*
$\alpha_0$ : Altersgruppe			-0.65*	-0.33
$\alpha_1$ : Altersgruppe*Bedingung			-0.60	-0.60
$\alpha_2$ : Altersgruppe*t			-0.67*	-0.49*
$\delta_2$ : Altersgruppe*t*Bedingung			0.34*	0.34*
Speed				0.30
Speed*t				0.17*
<b>Zufällige Effekte</b>				
$\sigma_{eST}^2$ : Var(Residuen) – Standard	0.78*	0.99*	0.94*	0.93*
$\sigma_{ePS}^2$ : Var(Residuen) – „pure- study“	1.22*	1.53*	1.47*	1.46*
Korr. $\sigma_{eST} \times ePS$		.11	.11	.11
AR(1) Durchgang		0.27*	0.24*	0.23*
$\sigma_{b_0}^2$ : Var(Ausgangswert)	2.01*	1.63*	1.52*	1.41*
$\sigma_{b_1}^2$ : Var(t)	0.24*	0.21*	0.13*	0.10*
$\sigma_{d_0}^2$ : Var(Lernbedingung)	0.68*	0.36	0.42*	0.43*
Korr. Ausgangswert x t	.38*	.58*	.47*	.39*
Korr. Ausgangswert x Lernbedingung	-.44*	-.54*	-.50*	-.42*
Korr. Lernbedingung x t	-.34*	-.47*	-.45*	-.30
<b>Informationskriterium</b>				
-2 LL	3257.7	3242.3	3191.2	3176.3
AIC	3281.7	3270.3	3227.2	3216.3

Anmerkungen. -2LL =  $-2 \times \text{Log-Likelihood}$ , AIC = Akaike Information Criterion.

\* $p < .05$

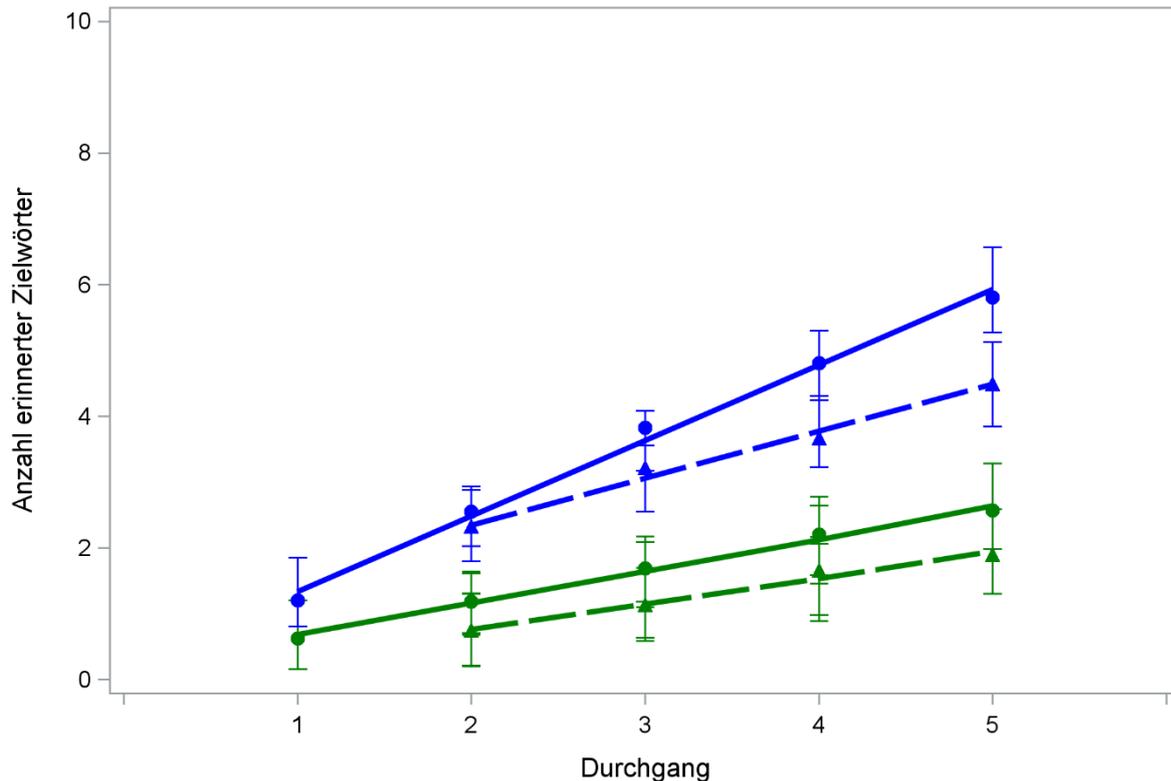


Abbildung 7. Experiment 2: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) für Modell 5. Die durchgehenden Linien repräsentieren die Standardlernbedingung und die gestrichelten Linien die „pure-study“-Bedingung. Die Kreise zeigen die beobachteten Werte der Standardbedingung, die Dreiecke die beobachteten Werte der „pure-study“-Lernbedingung in den beiden Altersgruppen. Die Fehlerbalken stellen die Standardfehler dar.

Abbildung 8 zeigt die mittlere Anzahl erinnerter Wortpaare beim verzögerten Abruf, nach Wortliste aufgeteilt (Wortliste 1 entspricht fünfmal lernen und fünfmal lernen; Wortliste 2 entspricht fünfmal lernen und viermal testen, usw.). Hier zeigt sich in beiden Altersgruppen unerwarteterweise, dass tendenziell mehr Wörter aus Listen erinnert wurden, die seltener abgefragt wurden. In einer 5 (Wortlisten) x 2 (Altersgruppen) ANOVA gibt es zwar signifikante Unterschiede in der Anzahl erinnerter Wörter zwischen den Altersgruppen ( $F(1, 520) = 208.90; \eta^2 = .29$ ), nicht aber Wortlisten ( $F(4, 520) = 2.29; \eta^2 = .02$ ). In post-hoc t-Testanalysen unterscheiden sich Wortliste 1 (fünfmal lernen und fünfmal abfragen) und

Wortliste 5 (fünfmal lernen und einmal abfragen) bei jüngeren ( $t(52) = -4.06, p < .05, d = -0.56$ ) und bei älteren Teilnehmern ( $t(52) = -2.32, p < .05, d = -0.32$ ).

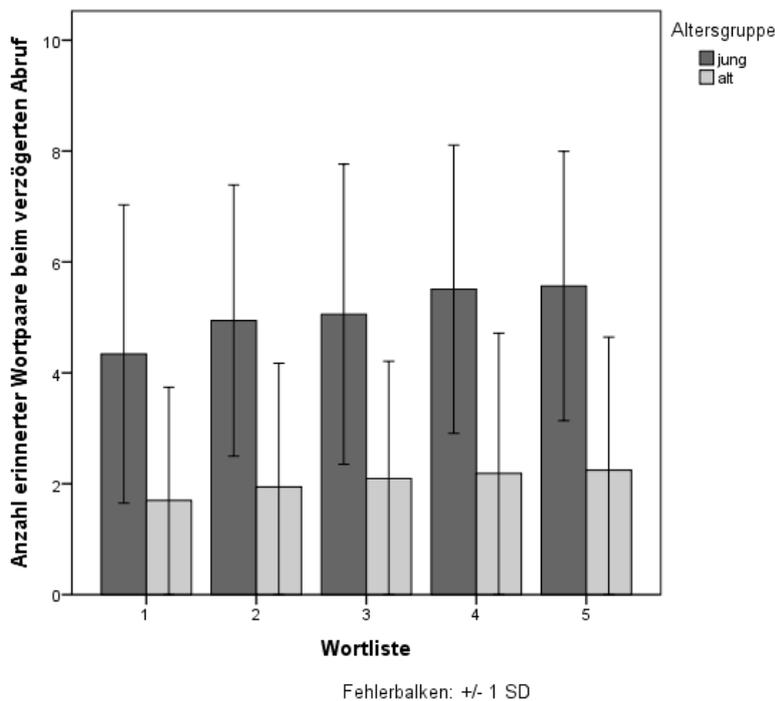


Abbildung 8. Experiment 2: Mittlere Anzahl erinnerter Wortpaare der fünf Wortlisten beim verzögerten Abruf, aufgeteilt nach Altersgruppe.

### 2.4.3. Zusammenfassung der Ergebnisse aus Experiment 2

Experiment 2 repliziert die Ergebnisse aus Experiment 1 in dem Sinne, dass es einen Altersunterschied in der Wirksamkeit des Testeffekts gibt. Jüngere Probanden lernten in der Standardbedingung mit jedem Lern-Test-Durchgang eine höhere Anzahl an Wörtern als in der „pure-study“-Bedingung und zeigten darüber hinaus eine allgemein höhere Lernrate als Ältere. Jedoch profitierten ältere Lerner nicht von den Effekten des Testens. Hierfür kommen zwei Erklärungsansätze in Frage: 1) die Art der Aufgabe und 2) die Schwierigkeit der Aufgabe. Zu 1) sei gesagt, dass bei Aufgaben mit Abrufhinweisreizen gemäß der ADH besonders große Altersunterschiede bei der Enkodierung und dem Abrufen erwartet werden

(Naveh-Benjamin, 2000), welche sich dementsprechend auf den Testeffekt auswirken (siehe Überlegungen in Kapitel 1.2.2.). Dadurch können die Testeffekte bei Älteren geringer ausfallen und dementsprechend schwieriger statistisch nachzuweisen sein. Zusätzlich kann 2) die Schwierigkeit der Aufgabe eine zentrale Rolle spielen. Nachdem der Testeffekt auf den Lernerfolg durch erfolgreiches Abrufen zurückgeht, könnte die hohe Schwierigkeit der Aufgabe für ältere Erwachsene einen Einfluss auf das Ergebnis haben. Bspw. erinnerten sich ältere Lerner im fünften Durchgang der Standardbedingung–der Testdurchgang mit der durchschnittlich höchsten Erinnerungsleistung–an zwei bis drei Wortpaare von 10 möglichen (im Vergleich dazu erinnerten sich jüngere Lerner an ungefähr sechs Wortpaare, siehe Tabelle 8). Des Weiteren erinnerten sich ältere Probanden im ersten Durchgang durchschnittlich nicht einmal an ein Wortpaar (vgl. Einfluss Schwierigkeit bei Tse et al., 2010). Der Zusammenhang der anfänglichen Lernleistung und des Testeffekts wird zudem in der hohen Korrelation des Ausgangswert mit der Lernbedingung deutlich ( $r = -.50$ ).

Außerdem hängt der individuelle Ausgangswert stark mit der Effektivität der Durchgänge zusammen. Dies deutet an, dass je mehr Wortpaare eine Person im ersten Lerndurchgang erfolgreich gelernt hat, desto mehr Wortpaare wird diese Person mit jedem weiteren Durchgang zusätzlich lernen. Dieser Zusammenhang könnte darauf zurückzuführen sein, dass die kognitiven Eigenschaften, welche mit hohem, anfänglichem Lernerfolg zusammenhängen, ebenfalls mit der Lernrate zusammenhängen (bspw. Verarbeitungsgeschwindigkeit, vgl. Rast & Zimprich, 2009; Zimprich, Rast & Martin, 2008). Bspw. hängt die in dem vorliegenden Experiment erfasste Verarbeitungsgeschwindigkeit positiv mit der Lernrate zusammen (siehe Tabelle 11, Modell 6). Es zeigen sich darüber hinaus statistisch nicht signifikante Tendenzen für einen schwachen Zusammenhang der Verarbeitungsgeschwindigkeit mit der anfänglichen Lernleistung ( $t(638) = 1.90, p = .06, \delta_{VG} = .30, d = .15$ ); dieser Zusammenhang sollte in

zukünftiger Forschung weiter berücksichtigt werden, um die Zusammenhänge genauer aufzuklären. Je höher die Verarbeitungsgeschwindigkeit, desto mehr Wortpaare werden mit jedem Lerndurchgang zusätzlich gelernt, was damit die individuellen Unterschiede zwischen Personen aufklärt.

Somit konnten die Hypothesen zum Testeffekt und zu den Altersunterschieden größtenteils bestätigt werden. Aufgrund der hohen Schwierigkeit der Aufgabe, fällt für die ältere Altersgruppe der Testeffekt aus (bzw. hat nur einen kleinen statistisch nicht signifikanten Einfluss). Zusammenfassend zum Design kann man folgern, dass eine Standard- sowie eine „pure-study“-Lernkurve gleichzeitig durch eine Lernaufgabe erhoben werden kann, jedoch unter den Einschränkungen 1) einer höheren Schwierigkeit der ansonsten einfacheren Aufgaben (bspw. zehn Wortpaare erinnern vs. zufällige zehn Wortpaare aus 50 möglichen erinnern) und 2) des verwendbaren Materials–dies funktioniert jedoch ausschließlich mithilfe von hinweisreizgesteuertem Material.

### 3. Studienblock 2 – Testeffekt bei Geschichten als Lernmaterial

#### 3.1. Theoretische Einführung: Lernmaterial und Dropout

##### 3.1.1. Lernmaterial – Geschichten

In Experiment 1 und 2 wurde der Testeffekt anhand von Wortlisten und Wortpaaren genauer untersucht. Da diese–meist unzusammenhängenden Wörter–keine alltagsnahe Lernsituation abbilden, wird in den folgenden Experimenten Lernmaterial verwendet, das inhaltlich zusammenhängender, also kohärenter ist. Hierfür werden Geschichten aus dem WMS-R (Härtling et al., 2000) und WMS-IV (Petermann & Lepach, 2012) verwendet. Vorteile kohärenter Geschichten liegen 1) in der einfacheren Verständlichkeit des Materials, v.a. dann, wenn der Rahmen der Geschichte mit den Erwartungen des Lesers übereinstimmt, 2) darüber hinaus darin, dass der Rahmen der Geschichte eigene Abrufhinweisreize bietet,

(bspw. gibt es einen Akteur oder Ort in der Geschichte? Was ist die Handlung? Usw.) und 3) in den Vernetzungen innerhalb der Geschichte (und damit einhergehenden Wiederholungen, wie bspw., dass der Akteur der Geschichte wiederholt genannt wird) (Bower, 1976). Je nach Kontext der zu lernenden Geschichte können Vorwissen (bspw. McNamara & Kintsch, 1996) oder vorhandene Schemata die Enkodierung sowie Abrufprozesse unterstützen (*Schematic Support*, Craik & Bosman, 1992). Dabei können Schemata ebenso effektiv sein wie umweltbezogene Unterstützungen. Die Vorteile einer Schemaaktivierung können vor allem bei kohärentem Material auftreten, im Gegensatz zu inhaltlich zusammenhangslosen Wortlisten oder Wortpaaren.

Für die Untersuchung des Testeffekts von besonderem Interesse ist die Eigenschaft von kohärenten Geschichten, dass die einzelnen Propositionen der Geschichten durch ihre zusammenhängende Struktur als Hinweisreize und materialbezogene Kontextattribute dienen können (durch den Rahmen der Geschichte und die Vernetzungen innerhalb dieser). Bspw. konnten Roediger und Karpicke (2006b) langfristige Testeffekte bei pädagogisch relevantem Lernmaterial zeigen (nach zwei Tagen und nach einer Woche).

Nach der Hypothese der Unterstützung durch Schemata (*Schematic Support*, Craik & Bosman, 1992; Peterson, Schmidt, & Naveh-Benjamin, 2017) profitieren vor allem ältere Lerner von kohärentem Lernmaterial. Ältere können hierbei auf detaillierte Schemata (bspw. Erinnern der Lebensmittelpreise; Castel, 2005) und Vorwissen zurückgreifen (Umanath & Marsh, 2014). In einer Metaanalyse von 194 Studien zu Altersunterschieden beim Lernen und Erinnern von Texten konnte gezeigt werden, dass Ältere durchschnittlich schlechter bei diesem Lernmaterial abschneiden als jüngere (Johnson, 2003). Jedoch zeigen sich keine altersbezogenen Unterschiede bezogen auf wiederholte Testungen im Gegensatz zu einmaligen Testungen. Auch bei Meyer und Logan (2013) lagen keine Altersunterschiede in den Effekten zusätzlicher Testdurchgänge vor–beim Lernen wissenschaftlicher Artikel.

Aufgrund der hohen Interkonnektivität logisch kohärenter Texte (wie es bei dem WMS-R und WMS-IV der Fall ist), kann es für den Testeffekt weniger relevant sein, dass neue Assoziationen der Propositionen zwischen den bereits bestehenden Assoziationen gebildet werden. Wobei vor allem beim Bilden von Assoziationen zwischen Einheiten Altersunterschiede auftreten (Naveh-Benjamin, 2000), welche sich auf Unterschiede in der Wirksamkeit des Testeffekts auswirken können. Durch ein fundierteres Schema über den Aufbau einer Geschichte und über den Ablauf von Alltagsgeschichten bei Älteren können altersbedingte Defizite beim Verknüpfen und Erinnern von Kontextattributen kompensiert werden. Demnach werden hier keine Altersunterschiede im Testeffekt erwartet, wohl aber generelle Altersunterschiede bei der Lernleistung.

### *3.1.2. Das Dropout-Lerndesign*

Eine typische Lernstrategie bei Studenten ist das Lernen mit Karteikarten (Kornell & Bjork, 2008). Bei dieser Strategie ist es möglich, bereits gelernte und erfolgreich abgerufene Inhalte aus dem Lern-Test-Zyklus herauszunehmen, wodurch sich die Lernzeit verkürzt und die Aufmerksamkeit auf die noch nicht gelernten Inhalte gelenkt werden kann. Ein solches Design ist eine Variante eines Dropout-Designs. Bei Dropout-Designs werden Items, sobald diese ein bestimmtes Kriterium erfüllt haben—bspw. einmal korrekt erinnert—aus dem Lern- und/oder dem Testdurchgang genommen. Dadurch bilden sich für eine Lernbedingung mit einer festen Anzahl an Lern-Test-Durchgängen neben der Standardvariante (es wird jedes Mal alles gelernt und alles abgefragt) drei mögliche Dropout-Bedingungen: 1) ein Item wird aus dem Lerndurchgang genommen und weiterhin getestet, 2) es wird nur aus dem Testdurchgang genommen, verbleibt jedoch in den Lerndurchgängen oder 3) es wird aus dem Lern- sowie Testdurchgang genommen (äquivalent zu der oben beschriebenen Lernstrategie des Karteikartenlernens). Die Konsequenz eines Dropout-Designs ist, dass die

Lern-Test-Durchgänge mit jedem Durchgang kürzer werden (siehe. Kornell & Bjork, 2008; Pyc & Rawson, 2011).

Jedoch sind die Dropout-Designs unterschiedlich effizient für den Lernerfolg. In ihrer Studie haben Karpicke und Roediger (2008) die vier oben genannten Designs in Bezug auf die Effekte von Testdurchgängen getestet. Hierbei erinnerten Probanden sich in den Bedingungen mit kompletten Testdurchgängen (sprich der Standardbedingung und der Bedingung mit Dropout im Lerndurchgang) an mehr Wortpaare als in den Bedingungen mit Dropout in den Testdurchgängen. Darüber hinaus erinnerten die Probanden sich in der Standardbedingung an ähnlich viele Wortpaare, wie in der Bedingung mit Dropout im Lerndurchgang. Dadurch wird abermals die besondere Rolle des Testeffekts deutlich: Wiederholtes Abfragen unterstützt langfristige Lerneffekte wirksamer als wiederholte Lerndurchgänge.

Bezüglich der Altersunterschiede in der Lernleistung bei Dropout-Designs gibt es bisher wenig Literatur. Kurtz und Zimprich (2014a) fanden in einer älteren Stichprobe ( $M = 69.3$ ,  $SD = 6.4$ ) bei der Bearbeitung eines doppelten Dropout-Designs—d.h. erfolgreich abgerufene Inhalte werden aus dem Lern- und Testdurchgang genommen—interindividuelle Unterschiede, welche durch kognitive Fähigkeiten, wie der Verarbeitungsgeschwindigkeit und verbales Wissens, teilweise erklärbar sind. Demnach gibt es unter älteren Erwachsenen starke Unterschiede in der Effektivität des Dropout-Designs bezüglich der Lernleistung. Zusätzlich hat die individuelle Lerngeschwindigkeit einen positiven Einfluss auf die Erinnerungsleistung bei einem finalen Abruf aller gelernter Wörter. D.h., ältere Lerner mit einer durchschnittlich höheren Lernrate vergessen weniger Wörter als Lerner mit einer niedrigeren Lernrate.

In den folgenden Studien wird ein Dropout-Design verwendet, bei dem nach erfolgreicher Testung das dementsprechende Item aus dem Lerndurchgang genommen wird

und weiterhin im Testdurchgang abgefragt wird (im Weiteren wird dies als Dropout-Bedingung bezeichnet). Dropout-Designs werden vor allem mit Lernkarten, bzw. mit eigenständigen Informationseinheiten (Wortpaaren oder spezifischen Fragen) kombiniert. Um zu überprüfen, ob diese zeitlich sparsamere Variante ebenso für andere Lernmaterialien eine effiziente Strategie darstellt, wird diese im Zusammenhang mit Geschichten Lernen untersucht. Bezüglich des Testeffekts wird erwartet, dass in der Standard- sowie der Dropout-Bedingung ähnlich viele Wörter gelernt werden und Lerner im gleichen Maße von den Effekten des Testens profitieren.

### *3.2. Experiment 3 – Geschichten*

Im dritten Experiment werden Geschichten als Lernmaterial verwendet, und zwar unter den folgenden Bedingungen: Standard-, Dropout- und „pure-study“-Bedingung. Es wird erwartet, dass sich die Effekte des Testens im gleichen Maße in der Standardbedingung sowie der Dropout-Bedingung zeigen, wobei beide Bedingungen eine höhere Lernleistung am Ende des Experiments, als in der „pure-study“-Bedingung zeigen sollten. Generell wird erwartet, dass ältere Erwachsene sich durchschnittlich an weniger Texteinheiten erinnern, jedoch aufgrund der Eigenschaften des Lernmaterials in ähnlichem Umfang von den Testeffekten profitieren, wie jüngere Lerner–dies wird für den unmittelbaren Testeffekt während des Experiments sowie für den Testeffekt beim verzögerten Abruf erwartet. Weiterhin wird angenommen, dass beim verzögerten Abruf durchschnittlich weniger Texteinheiten in der Standard- und Dropout-Bedingung vergessen werden und gleichzeitig mehr Texteinheiten erinnert werden.

### 3.2.1. Methode Experiment 3

#### Stichprobe

Insgesamt nahmen 52 jüngere Probanden zwischen 18 und 30 Jahren ( $M = 22.79$ ,  $SD = 3.56$ ) und 51 ältere Erwachsene zwischen 59 und 80 Jahren an der Studie teil ( $M = 65.29$ ,  $SD = 5.49$ ). An der Studie nahmen 55 Frauen (53.4%) teil (jüngere Erwachsene: 59.6% Frauen; ältere Erwachsene 47.1% Frauen).<sup>13</sup> Die subjektive Gesundheit der Probanden wurde mittels einer Likert-Skala von 1 = „sehr gut“ bis 5 = „mangelhaft“ erfasst und betrug im Mittel 2.08 ( $SD = 0.78$ ). Dabei gaben die jüngeren Teilnehmer im Mittel eine bessere subjektive Gesundheit an ( $M = 1.87$ ,  $SD = 0.66$ ), als ältere ( $M = 2.29$ ,  $SD = 0.83$ ) ( $t(101) = -2.90$ ,  $p < .05$ , Cohen's  $d = -.57$ ). Um ihren Schulabschluss anzugeben, hatten die Probanden fünf Optionen zur Verfügung: "kein Abschluss", "Volks-/Hauptschule", "Realschule", "Abitur", „Hochschulabschluss“. In der jüngeren Stichprobe gaben 39 (75.0%) Abitur und 8 (15.4%) einen Hochschulabschluss als höchsten erreichten Schulabschluss an, während bei der älteren Probanden 19 (37.3%) einen Hochschulabschluss und 14 (27.5%) einen Volks-/Hauptschulabschluss angaben. Die meisten jüngeren Probanden gaben an ledig ( $N = 50$ , 96.2%) und Schüler/Studenten ( $N = 34$ , 65.4%), wohingegen bei meisten älteren Teilnehmer verheiratet ( $N = 31$ , 60.8%) und im Ruhestand ( $N = 22$ , 43.1%) oder vollzeitbeschäftigt waren ( $N = 20$ , 39.2%).

#### Messinstrumente

##### Geschichten

Als verbale Lernaufgabe wurden die Geschichten B aus dem Subtest für logisches Gedächtnis der deutschen Version des WMS-R (Härting, et al., 2000) (72 Wörter lang)

---

<sup>13</sup> Die Geschlechterverteilung zwischen den Altersgruppen war nicht signifikant verschieden,  $\chi^2(1, 102) = 1.63$ ,  $p = .20$ , Cohen's  $w = 0.26$ .

sowie die Geschichten B und C aus dem Subtest für logisches Gedächtnis der deutschen Version des WMS-IV (Petermann & Lepach, 2012) (69 und 87 Wörter lang). Im Lern-Durchgang lasen die Probanden die präsentierte Geschichte laut vor. Im Testdurchgang wurden die Teilnehmer aufgefordert, sich an so viele Details wie möglich in einer beliebigen Reihenfolge zu erinnern—mit der Empfehlung die Geschichte von Anfang an zu erzählen. Dabei können in jeder Geschichte insgesamt 25 Propositionen richtig erinnert werden. Der mögliche zu erreichende Punktwert geht von 0 bis 100 (wobei 100 einem Prozentwert von 100% korrekt genannt entspricht). Die Reliabilität beträgt für die Aufgabe des WMS-R .79 (Härting, et al., 2000) und für die Geschichten aus dem WMS-IV .87 und .89 (Petermann & Lepach, 2012).

### *Lerndesign*

Jeder Proband bearbeitete drei Lernbedingungen: eine Standard-, eine Dropout- und eine „pure-study“-Lernbedingung. Die Standardbedingung setzt sich aus drei Lern-Test-Durchgängen zusammen. In der „Dropout“-Bedingung wurde ebenfalls dreimal gelernt und abgefragt, jedoch wurden in den Lerndurchgängen nur die Teile der Geschichte präsentiert, welche zuvor nicht korrekt erinnert wurden (im Testdurchgang sollte die Teilnehmer sich wieder an die gesamte Geschichte erinnert werden). Hierfür wurden beim Lerndurchgang die vorher korrekt erinnerten Propositionen weiß überdeckt waren (gleiche Farbe wie der Hintergrund). In der reinen Lernbedingung lernten die Probanden die Geschichte fünfmal bevor ein finaler Testdurchgang folgte. Bei den drei Aufgaben folgte jeweils nach einem Lerndurchgang eine 60-sekündige Füllaufgabe<sup>14</sup> (siehe Tabelle 12). Die Versuchspersonen wurden zufällig auf eines von sechs möglichen Designs aufgeteilt. Dabei wurde bei den

---

<sup>14</sup> Die Füllaufgaben waren verschiedene Suchbilder, Labyrinth und Zahlenbilder welche jeweils für 60 Sekunden bearbeitet bzw. fortgeführt werden konnte.

Designs darauf geachtet, dass jede Geschichte bzw. Lernbedingung gleich oft an jeder Position vorkommt (da es drei Lernbedingungen gibt, gibt es zudem drei mögliche Positionen). Dafür wurden die Versuchsdesigns nach griechisch-lateinisch Quadrat erstellt (siehe Tabelle 13).

Tabelle 12

*Lernbedingungen Experiment 3*

Bedingung	Durchgang 1		Durchgang 2		Durchgang 3		Durchgang 4		Durchgang 5		Finaler Test	2 Tage später
	1	2	3	4	5	6	7	8	9			
Standard	S	F	T	S	F	T	S	F	T	S	F	T
„Dropout“	S	F	T	S <sub>n</sub>	F	T	S <sub>n</sub>	F	T	S <sub>n</sub>	F	T
„pure-study“	S	F	S	F	S	F	S	F	S	F	T	T

*Anmerkungen.* S = Lerndurchgang; T = Testdurchgang; S<sub>n</sub> = Lerndurchgang, bei dem nur vorher nicht korrekt erinnerte Propositionen gezeigt werden; F = 60-sekündige Füllaufgabe.

Tabelle 13

*Versuchsdesigns Experiment 3*

	Durchgang 1	Durchgang 2	Durchgang 3
Design 1	G2 – ST	G3 – S <sub>n</sub> T	G1 – SS
Design 2	G3 – SS	G1 – ST	G2 – S <sub>n</sub> T
Design 3	G1 – S <sub>n</sub> T	G2 – SS	G3 – ST
Design 4	G1 – SS	G3 – S <sub>n</sub> T	G2 – ST
Design 5	G2 – S <sub>n</sub> T	G1 – ST	G3 – SS
Design 6	G3 – ST	G2 – SS	G1 – S <sub>n</sub> T

*Anmerkungen.* ST = Standardbedingung; S<sub>n</sub>T = „Dropout“-Bedingung; SS = reine Lernbedingung; G1 = Geschichte B aus dem WMS-IV; G2 = Geschichte C aus dem WMS-IV; G3 = Geschichte B aus dem WMS-R.

### *Verarbeitungsgeschwindigkeit*

Zur Bestimmung der Verarbeitungsgeschwindigkeit wurden ein Testwert aus den zwei Messwerten zweier einfacher Reaktionszeitaufgaben (einfache Reaktionszeitaufgabe aus Experiment 1). Zu Beginn der Aufgabe wurde eines von zwölf Bildern als Zielbild präsentiert (in beiden Aufgaben wurde das Bild jeweils zufällig ausgewählt). Anschließend wurden die zwölf Bilder einzeln und in zufälliger Reihenfolge präsentiert. Die Versuchspersonen wurden instruiert, „q“ auf der Tastatur zu drücken sobald das Zielbild zu sehen war, andernfalls „p“. Im Durchschnitt erscheint das Zielbild im Verhältnis von 1:2. Die Probanden bearbeiteten während des Experiments die Aufgabe zweimal für je 90 Sekunden–das Zielbild wurde jeweils zufällig ausgewählt. Die erreichte Punktzahl der beiden Aufgaben wurde gemittelt (Cronbachs Alpha: .88). Der Wert für Verarbeitungsgeschwindigkeit ist die z-standardisierte erreichte Anzahl der korrekt beantworteten Items aus beiden Tests.

### *Verzögerter Abruf*

An einem zweiten Termin zwei Tage später, wurden die Versuchspersonen nochmal gebeten die Geschichten wiederzugeben. Hierbei wurde für jede Geschichte vier Hinweisreize gegeben: Vor- und Zuname der Hauptperson, Ort sowie ein Hinweis auf die Handlung. Die Probanden konnten dementsprechend 21 weitere Propositionen richtig benennen (aus Gründen der Vergleichbarkeit von Haupt- und Nachtest wurde als Einheit der Lernleistung die Prozentwerte 0-100 verwendet–wobei 100 bedeutet, dass eine Person alle 21 Propositionen im verzögerten Abruf genannt hat). Die Reihenfolge der Abfrage der Geschichten wurde dabei für jeden Teilnehmer randomisiert.

### *Prozedur*

Die Erhebung fand an zwei Terminen statt. Am ersten Termin (Experiment) haben die Probanden die Einverständniserklärung unterzeichnet und die Aufgaben in Einzeltestungen an einem Laptop bearbeitet. Anfangs wurden demographische Variablen abgefragt (Alter, Geschlecht, Familienstand, Schulabschluss, subjektive Gesundheit sowie die berufliche Tätigkeit).<sup>15</sup> Nachdem die Probanden den Fragebogen zur aktuellen Motivation (FAM; Rheinberg, Vollmeyer & Burns, 2001) beantwortet haben folgte die verbale Lernaufgabe. Vor jeder Lernbedingung gab es eine kurze Instruktion bzgl. der Lernbedingung und wie oft gelernt und abgefragt wird. Die Geschichten wurden in schwarzen Buchstaben auf weißen Hintergrund dargeboten (Schriftart: Arial; Schriftgröße: 5.5% des Bildschirms). Zwischen den drei Lernbedingungen bearbeiteten die Versuchspersonen für 90 Sekunden eine einfache Reaktionsaufgabe. Zum Schluss des Haupttermins wurden die Probanden gefragt, ob und welche Strategie sie verwendet haben, um sich die Geschichten zu merken. Dieser erste Termin dauerte ungefähr 45 Minuten. Eine weitere Testung wurde zwei Tage später telefonisch durchgeführt. Die Teilnehmer wurden nochmal nach Alter und Geschlecht gefragt, bevor der verzögerte Abruf begann. Dieser zweite Teil des Experiments dauerte ungefähr 10 Minuten.

### *Datenanalyse*

Die Daten wurden mittels mehrerer gemischter Regressionen modelliert. Es sei 1)  $y_{iPS}$  die Abrufleistung von Person  $i$  ( $i = 1, \dots, N$ ) in der „pure-study“-Bedingung, 2)  $y_{iST}$  die Abrufleistung im Durchgang  $t$  ( $t = 1, 2, 3$ ) in der Standardbedingung und 3)  $y_{iDO}$  die Abrufleistung im Durchgang  $t$  der Dropout-Bedingung. Entsprechend bezeichne  $y_{iPS2}$ ,  $y_{iST2}$

---

<sup>15</sup> Zusätzlich wurde angegeben mit wie vielen Personen die Versuchsperson gemeinsam in einem Haushalt lebt, wie oft sie regelmäßig Sport betreibt oder an Weiterbildungen teilnimmt.

und  $y_{itDO2}$  die Abrufleistung im verzögerten Abruf zwei Tage später der „pure-study“-, der Standard- bzw. der Dropout-Bedingung. Aufgrund des non-linearen Verlaufs von Gedächtnisleistungen über die Durchgänge wurden die logarithmierten Durchgänge als Prädiktorvariable verwendet (d.h.  $\ln(t) = \ln(1) - \ln(3)$ ,  $\ln(2) - \ln(3)$ ,  $\ln(3) - \ln(3) = -1.10$ ,  $-0.41$ ,  $0$ ). Dabei wurde der letzte Messzeitpunkt des Lernexperiments als Referenzwert „0“ verwendet (der dritte Testdurchgang der Standard- und Dropout-Bedingung, bzw. der Testdurchgang der „pure-study“-Bedingung). Um die Abrufleistung unter den beiden Lernbedingungen gleichzeitig zu modellieren, wurden die Variablen  $y_{itPS}$ ,  $y_{itST}$ ,  $y_{itDO}$  und  $y_{itPS2}$ ,  $y_{itST2}$ ,  $y_{itDO2}$  zu einem Vektor  $\mathbf{y}_{it}$  zusammengefasst (vgl. MacCallum et al., 1997). Die Effekte der Altersgruppe, die sich in Form einer Dummyvariable  $x_i$  ( $0 = \text{jung}$ ,  $1 = \text{älter}$ ) modellieren lassen, sowie der Bedingungen (ebenfalls Dummy-codiert), des Durchlaufs und deren Interaktionen lassen sich dann beschreiben als

$$\mathbf{y}_{it} = \begin{pmatrix} y_{itPS} \\ y_{itST} \\ y_{itDO} \\ y_{itPS2} \\ y_{itST2} \\ y_{itDO2} \end{pmatrix} = \begin{pmatrix} \beta_{0PS} + \alpha_{0PS}x_i \\ \beta_{0PS} + \alpha_{0PS}x_i + \beta_{0ST} + b_{STi} + \alpha_{0ST}x_i \\ \beta_{0PS} + \alpha_{0PS}x_i + \beta_{0ST} + \alpha_{0ST}x_i + \beta_{0DO} + b_{DOi} + \alpha_{0DO}x_i \\ \beta_{0PS2} + \alpha_{0PS2}x_i \\ \beta_{0PS2} + \alpha_{0PS2}x_i + \beta_{0ST2} + \alpha_{0ST2}x_i \\ \beta_{0PS2} + \alpha_{0PS2}x_i + \beta_{0ST2} + \alpha_{0ST2}x_i + \beta_{0DO2} + \alpha_{0DO2}x_i \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_{1ST} + \alpha_{1ST}x_i + b_{ii} \\ \beta_{1ST} + \alpha_{1DO}x_i + \beta_{1DO} + b_{ii} \\ 0 \\ 0 \\ 0 \end{pmatrix} \ln(t) + \begin{pmatrix} u_{PS} \\ u_{Test} \\ u_{Test} \\ u_{PS} \\ u_{Test} \\ u_{Test} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \\ e_{i6} \end{pmatrix} \quad (6)$$

wobei  $\beta_{0PS}$  die durchschnittliche Abrufleistung im Testdurchgang der „pure-study“-Bedingung,  $\beta_{0ST}$  den Unterschied zum dritten Durchgang der Standardbedingung und  $\beta_{0DO}$  den Unterschied des dritten Durchgangs der Dropout-Bedingung zur Standardbedingung

darstellen. Es bezeichne  $b_{PSi}$  die zufällige individuelle Abweichung der „pure-study“-Bedingung,  $b_{STi}$  die zufällige individuelle Abweichung vom festen Effekt der Standardbedingung in Person  $i$  und  $b_{DOi}$  die zufällige individuelle Abweichung vom festen Effekt der Dropout-Bedingung. Dabei stellen  $\alpha_{0PS}$ ,  $\alpha_{0ST}$  sowie  $\alpha_{0DO}$  die Altersunterschiede zu den von  $\beta_{0PS}$ ,  $\beta_{0ST}$  und  $\beta_{0DO}$  beschriebenen Unterschieden dar. Darüber hinaus bezeichne  $\beta_{1ST}$  den festen Effekt des logarithmierten Durchgangs der Standardbedingung (d.h. die durchschnittliche Veränderung der Abrufleistung über die Durchgänge). Dementsprechend stellt  $\beta_{1DO}$  den festen Effekt des logarithmierten Durchgangs in der Dropout-Bedingung dar.  $b_{ti}$  bezeichne die zufällige individuelle Abweichung vom Effekt des Durchgangs in Person  $i$  (Standard- und Dropout-Bedingung gemeinsam). Die Altersunterschiede im Lernzuwachs werden durch die Interaktion von Altersgruppe mit den logarithmierten Durchgängen der Standardbedingung  $\alpha_{1ST}$  und Dropout-Bedingung  $\alpha_{1DO}$  modelliert. Die Vergessensraten des jeweils letzten Testdurchgangs einer Lernbedingung während des Experiments zu dem verzögerten Abruf zwei Tage später werden durch  $\beta_{0PS2}$ ,  $\beta_{0ST2}$  und  $\beta_{0DO2}$  dargestellt. Dabei stellt (wie schon bei den festen Effekten  $\beta_{0PS}$ ,  $\beta_{0ST}$  und  $\beta_{0PS}$ )  $\beta_{0ST2}$  den Unterschied in zu  $\beta_{0PS2}$  in der Vergessensleistung dar; und  $\beta_{0DO2}$  dementsprechend den Unterschied von der Dropout- zur Standardbedingung dar.  $\alpha_{0PS2}$ ,  $\alpha_{0ST2}$  und  $\alpha_{0DO2}$  bezeichnen die dazugehörigen Altersunterschiede in der Differenz vom letzten Testdurchgang einer Lernbedingung zum dazugehörigen verzögerten Erinnerungsdurchgang. Die Fehleranteile der Messmethode mit Testdurchgängen (Standard- und Dropout-Bedingung) und ohne Testdurchgänge („pure-study“-Bedingung) werden durch  $u_{Test}$  und  $u_{PS}$  dargestellt. Dazu kommen individuelle Fehleranteile, von denen angenommen wurde, dass sie mit konstanter Varianz normalverteilt sind, d.h.

$$e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5}, e_{i6} \sim N(0, \sigma_e^2). \quad (7)$$

Darüber hinaus wurde angenommen, dass die zufälligen Effekte der „pure-study“-bedingung, der Standard- und der Dropout-Bedingung, sowie die zufälligen Effekte Lernzuwachses multivariat normalverteilt sind mit jeweils Mittelwert 0 und Kovarianz-Matrix

$$\begin{pmatrix} b_{STi} \\ b_{DOi} \\ b_{PSi} \\ b_{ti} \end{pmatrix} \sim MVN(\mathbf{0}, \mathbf{G}), \quad \text{mit } \mathbf{G} = \begin{pmatrix} \sigma_{b_{ST}}^2 & & & \\ \sigma_{b_{ST}b_{DO}} & \sigma_{b_{DO}}^2 & & \\ \sigma_{b_{ST}b_{PS}} & \sigma_{b_{DO}b_{PS}} & \sigma_{b_{PS}}^2 & \\ \sigma_{b_{ST}b_t} & \sigma_{b_{DO}b_t} & \sigma_{b_{PS}b_t} & \sigma_{b_t}^2 \end{pmatrix}. \quad (8)$$

In Modellen, welche zuerst ohne den verzögerten Abruf modelliert wurden (siehe Tabelle 18), werden die zufällige Effekte für die Standard- und Dropout-Bedingung sowie für die deren jeweiligen Lernraten getrennt geschätzt. In allen Modellen wurden die Residualvarianzen der Bedingung ohne Testen  $\sigma_{e_{PS}}^2$  sowie der beiden Bedingungen mit Testdurchgängen  $\sigma_{e_{Test}}^2$  (Standard- und Dropout-Bedingung gemeinsam) getrennt geschätzt.

Die Berechnungen wurden mit SAS Version (9.4) (SAS Institute, 2016) durchgeführt. Die gemischte Regression wurde mit der Prozedur MIXED und der „Maximum Likelihood“-Schätzmethode modelliert. Die Varianzaufklärung auf den beiden Ebenen des Modells werden mit  $R^2$  (Recchia, 2010) und Effektstärken mit Cohen’s  $d$  angegeben—die Angaben dienen dabei vor allem der Veranschaulichung.

### 3.2.2. Ergebnisse Experiment 3

In den Daten von Experiment 3 treten mehrere fehlende Werte auf: Aufgrund von technischen Problemen während der Erhebung fehlen bei zwei jüngeren Personen die Werte der Dropout-Bedingung und bei einer weiteren jüngeren Versuchsperson die Werte der Standardbedingung (die Messungen während des Experiments sowie zwei Tage später). Bei

einer älteren Versuchsperson wurde der zweite Dropout-Durchgang übersprungen (dementsprechend werden folgende Testdurchgänge in der Dropout-Bedingung sowie in der Messung zwei Tage später nicht berücksichtigt). Aufgrund einer Verwechslung zweier Geschichten im verspäteten Abfragedurchgang nach zwei Tagen fehlen bei einer weiteren älteren Person die Werte für die Dropout- und „pure-study“-Bedingung des verzögerten Abrufs. Eine ältere Versuchsperson hat während der drei Testdurchgänge in der Dropout-Bedingung sich an 19, dann vier und abschließend wieder an 19 Einheiten erinnert. Aufgrund dieses außergewöhnlichen Verlaufs wurden die Testwerte für den zweiten und dritten Dropout-Durchgang aus den Analysen ausgeschlossen (der verzögerte Abruf nach zwei Tagen wurde jedoch berücksichtigt). Aufgrund von technischen Problemen bei der Erhebung, fehlen bei acht Personen die Werte der Verarbeitungsgeschwindigkeitsaufgabe.

In Tabelle 14 sind die beobachteten Mittelwerte der einzelnen Testdurchgänge für jüngere sowie ältere Probanden dargestellt (siehe Abbildung 9). Die Lernleistung der beiden Altersgruppen steigt in der Standard- sowie in der Dropout-Bedingung erwartungsgemäß mit jedem weiteren Lern-Test-Durchgang an. Der Lernzuwachs vom ersten auf den zweiten Testdurchgang ist durchschnittlich größer als der Lernzuwachs vom zweiten auf den dritten Test; dies ist konform mit der Theorie einer abflachenden Lernkurve. Die Altersunterschiede werden mit jedem weiteren Lern-Test-Durchgang größer. Obgleich die älteren weniger von jedem weiteren Lerndurchgang profitieren als die jüngeren Probanden, so unterscheiden sich die beiden Altersgruppen im ersten Testdurchgang der Standard- sowie der Dropout-Bedingung nicht signifikant voneinander. Bei einem Abruf zwei Tage später erinnern sich die Probanden an die Geschichten der Standard- und Dropout-Bedingung im Mittel ähnlich gut, wohingegen durchschnittlich eine Informationseinheit weniger der Geschichte der „Pure-Study“-Bedingung erinnert wurde. Bezüglich der Altersunterschiede beim verzögerten Abruf sind diese in der Dropout-Bedingung am größten, dann in der Standard- und am

geringsten in der „Pure-Study“-Bedingung. Die Altersunterschiede in den beiden Testungen der „Pure-Study“-Bedingung sind geringer als die Altersunterschiede in den Designs mit Testdurchgängen. Die (vergleichsweisen) hohen Altersunterschiede in den Testdurchgängen der Dropout-Bedingung könnten zudem auf die spezifische Art des Lernens dieser Bedingung zurückgeführt werden. Das Lernen der noch nicht gelernten Einheiten ist zwar eine typische Strategie beim Vokabellernen, jedoch eher außergewöhnlich bei zusammenhängenden Geschichten. Ältere Probanden könnten daher Schwierigkeiten bei der flexiblen Adaption dieses Designs haben, was wieder einen Einfluss auf die Altersunterschiede mit sich bringt.

Tabelle 14

*Experiment 3: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

Variablen				Jung	Alt	Gesamt	Diff(Jung – Alt)	Cohens d
Standard 1	$N_{jung}$	51	$M$	63.14	57.96	60.55	$t(100) =$	0.31
	$N_{alt}$	51	$SD$	13.65	19.26	16.81	1.57	
Standard 2	$N_{jung}$	51	$M$	84.55	76.63	80.59	$t(100) =$	0.56
	$N_{alt}$	51	$SD$	12.83	15.50	14.70	2.81*	
Standard 3	$N_{jung}$	51	$M$	92.08	83.14	87.61	$t(100) =$	0.78
	$N_{alt}$	51	$SD$	7.86	14.25	12.30	3.92*	
Dropout 1	$N_{jung}$	50	$M$	64.48	59.45	61.94	$t(99) =$	0.30
	$N_{alt}$	51	$SD$	15.26	17.69	16.64	1.53	
Dropout 2	$N_{jung}$	50	$M$	82.08	70.78	76.58	$t(97) =$	0.78
	$N_{alt}$	49	$SD$	14.07	14.93	15.51	3.88*	
Dropout 3	$N_{jung}$	50	$M$	91.36	78.69	85.09	$t(97) =$	0.99
	$N_{alt}$	49	$SD$	9.30	15.52	14.20	4.94*	
„pure-study“ 3	$N_{jung}$	52	$M$	86.14	78.51	82.37	$t(101) =$	0.70
	$N_{alt}$	51	$SD$	10.19	11.65	11.54	3.55*	
Standard 2T	$N_{jung}$	51	$M$	72.92	62.46	67.69	$t(100) =$	0.62
	$N_{alt}$	51	$SD$	13.18	19.86	17.57	3.13*	
Dropout 2T	$N_{jung}$	50	$M$	73.24	60.62	67.06	$t(96) =$	0.71
	$N_{alt}$	48	$SD$	13.84	21.29	18.88	3.49*	
„pure-study“ 2T	$N_{jung}$	52	$M$	66.03	58.67	62.42	$t(100) =$	0.41
	$N_{alt}$	51	$SD$	15.12	20.10	18.03	2.10*	

*Anmerkungen.* Lernleistung als Prozentangabe korrekt erinnerter Textpropositionen. Standard 1-3 entsprechen den Testdurchgängen der Standardbedingung nach 1 bis 3 Lerndurchgängen; Dropout 1-3 entsprechen den Testdurchgängen der Dropout-Bedingung nach 1 bis 3 Lerndurchgängen; „Pure-Study“ 3 entspricht der Erinnerungsleistung nach fünfmaligem Lernen; Standard/Dropout/„Pure-Study“ 2T = Standard/Dropout/„Pure-Study“ Abfragedurchgang zwei Tage nach der Testung.

\* $p < .05$

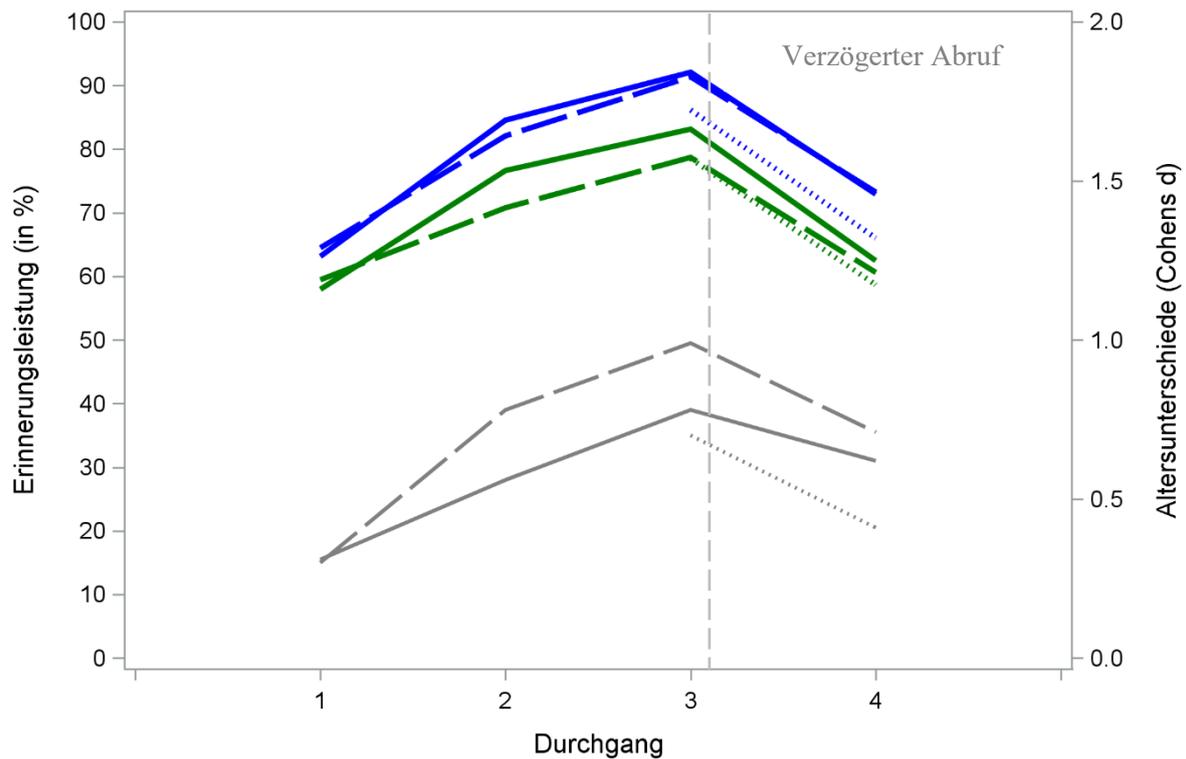


Abbildung 9. Experiment 3: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) sowie die Altersunterschiede (grau – rechte y-Achse). Die durchgehenden Linien repräsentieren die Standardlernbedingung, die gestrichelten Linien die Dropout-Bedingung und die gepunktete Linie die „pure-study“-Bedingung.

Die Mittelwerte der Lernleistung innerhalb der drei Geschichten unterscheidet sich signifikant ( $F(2, 1007) = 36.66, p < .05$ ) (siehe Tabelle 15). Hierbei schneiden Personen bei der Geschichte B aus dem WMS-IV am besten (Geschichte mit den wenigsten Wörtern) und bei der Geschichte C aus dem WMS-IV am schlechtesten ab (Geschichte mit den meisten Wörtern). Die durchschnittlichen Erinnerungswerte über alle Personen und Messungen hinweg unterscheiden sich über die sechs Designs ( $F(5, 1004) = 2.64, p < .05$ ). Das leichteste Design (Design 1) unterscheidet von Design (Design 2) mit einer Differenz von 7.16 Prozentpunkten).

Tabelle 15

*Experiment 3: Lernleistung nach Geschichte und Design aufgeteilt*

	Geschichte			Design						
	G1	G2	G3	1	2	3	4	5	6	Alle
<i>N</i>	342	332	336	166	177	156	157	174	180	1010
<i>M</i>	79.63	68.12	71.62	69.44	76.60	73.62	72.44	73.24	73.47	73.18
<i>SD</i>	15.67	19.07	18.88	20.44	16.46	17.21	20.09	18.90	17.58	18.55

*Anmerkungen.* G1 = Geschichte B aus dem WMS-IV; G2 = Geschichte C aus dem WMS-IV; G3 = Geschichte B aus dem WMS-R.

Tabelle 16 zeigt die Interkorrelationen aller Testdurchgänge sowie den Tests nach zwei Tagen. Innerhalb der Standard- und Dropout-Bedingung korrelieren aufeinanderfolgende Testdurchgänge höher als mit dem vorletzten Testdurchgang (sprich dritter mit ersten Testdurchgang). In den beiden Altersgruppen korrelieren die Testdurchgänge der drei Bedingungen (Standard, Dropout und „pure-study“) mit den jeweiligen Werten zwei Tage später. Das bedeutet, wer schon innerhalb einer Bedingung eine hohe Abrufleistung erzielte, erinnerte durchschnittlich mehr Teile der Geschichte bei dem verzögerten Abruf nach zwei Tagen. Ausnahme hierbei stellt die Dropout-Bedingung bei den jüngeren Probanden dar, bei der sich keine signifikante Korrelation des verzögerten Abrufs mit den drei Testwerten innerhalb des Experiments zeigt. Diese etwas überraschende Korrelation zeigt nochmal den besonderen Stellenwert der Dropout-Bedingung gegenüber der Standard- oder „pure-study“-Bedingung. Die jüngeren Probanden könnten demnach unterschiedlich gut mit diesem besonderen Format der Erhebung zurechtkommen. Je weniger man im vorherigen Testdurchgang der Dropout-Bedingung erinnert, desto mehr gleicht dieser der Standardbedingung; wohingegen je mehr man in einem Test erfolgreich erinnert hat, desto weniger Informationen enthält der folgende Lerndurchgang (bis hin zu garkeiner Textproposition) und gleicht einer Bedingung mit wiederholter Testung ohne

Lerndurchgang. Zwischen diesen beiden möglichen Extremen (STSTST vs. STTT) können unterschiedlich Aufgabentypen entstehen, welche vielleicht bei einem mittleren Erinnerungsniveau von 50% maximal ungewohnt sind (bspw., wenn nur noch eine nicht zusammenhängende Hälfte der Geschichte präsentiert wird). Bei älteren zeigen sich hingegen hohe Zusammenhänge beim verzögerten Abruf der Dropout-Bedingung mit den drei Testwerten dieser Bedingung. Im Gegensatz zur jüngeren Vergleichsgruppe ist hier zu erwarten, dass die Dropout-Bedingung einheitlicher zu Schwierigkeiten führt (bspw. höhere Altersunterschiede in der Dropout-Bedingung im Gegensatz zur Standardbedingung in Tabelle 14).

Generell ergeben sich bei den älteren Teilnehmern höhere Korrelationen zwischen den Testdurchgängen der verschiedenen Bedingungen als bei den jüngeren Probanden. Dies wird besonders deutlich, wenn man die Korrelationen der Testdurchgänge der Standardbedingung mit denen der Dropout-Bedingung betrachtet. Dies kann einerseits an den Deckeneffekten bei jüngeren Teilnehmer und der damit einhergehenden eingeschränkten Varianz liegen (bspw. sind die Standardabweichungen bei den jüngeren Probanden durchweggehend niedriger als bei den älteren, siehe Tabelle 14); oder andererseits kann dies daran liegen, dass die Lernleistung bei älteren aufgrund Dedifferenzierung kognitiver Ressourcen im Alter stärker zusammenhängen können (Zimprich & Martin, 2010).

Tabelle 16

*Experiment 3: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Standard 1		.62*	.56*	.02	-.18	-.14	-.14	.47*	.04	-.05
2. Standard 2	.80*		.72*	-.05	-.02	-.17	-.18	.34*	.05	-.24
3. Standard 3	.77*	.86*		-.13	-.13	-.16	-.28*	.47*	.10	-.29*
4. Dropout 1	.48*	.46*	.35*		.77*	.36*	.22	-.11	.23	.31*
5. Dropout 2	.41*	.40*	.43*	.79*		.56*	.34*	-.32*	.23	.20
6. Dropout 3	.42*	.36*	.39*	.75*	.77*		.30*	-.19	.23	.19
7. „pure-study“	.33*	.31*	.23	.23	.14	.25		-.13	.17	.58*
8. Standard 2T	.64*	.66*	.59*	.41*	.31*	.32*	.14		.31*	.09
9. Dropout 2T	.39*	.36*	.35*	.60*	.47*	.65*	-.05	.48*		.28*
10. „pure-study“ 2T	.38*	.24	.20	.21	-.02	.15	.52*	.39*	.33*	

*Anmerkungen.* Oberhalb der Diagonale werden die Korrelationen für die jüngere Stichprobe und unterhalb der Diagonale die Korrelationen der älteren Stichprobe dargestellt.

\* $p < .05$

### *Mehrebenenmodelle*

Die Leistungsvarianz aller Abfragedurchgängen während des Experiments und beim verzögerten Abrufversuch gehen zu 23% auf die Unterschiede zwischen Personen zurück (ICC = .23,  $p < .01$ ). Somit zeigt sich zudem bei diesem Experiment, dass bei den Analysen die Abhängigkeit der Daten beachtet werden sollte, und zwar in dem Sinne, dass Effekte zwischen und innerhalb von Personen betrachtet werden. In einem ersten Schritt wird Lernrate–und damit die Effekte der Lern-Test-Durchgängen–genauer betrachtet (siehe

Tabelle 17 und 18), um in einem zweiten Schritt die Effekte der drei Lernbedingungen auf die unmittelbare und die langfristige Lernleistung zu betrachten.

In einem ersten Modell (Modell 0) wurden die Lernleistungen innerhalb einer Person in Abhängigkeit der Lernbedingung sowie des Durchgangs modelliert ( $t = -2, -1, 0$ ) (siehe Tabelle 17). Hierbei stellt die Dummyvariable  $\beta_{0ST}$  den Unterschied der Standardbedingung im dritten Abfragedurchgang zur ersten Abfrage in der „pure-study“-Bedingung dar; und  $\beta_{0DO}$  beschreibt den Unterschied der Dropout-Bedingung zur Standardbedingung im jeweils dritten Durchgang. Als zufällige Effekte wurden die Effekte der Standard- und Dropout-Bedingung sowie deren jeweiligen Lernraten geschätzt. Die Lernrate wurde als linearer Effekt des Durchgangs einmal für die Standardbedingung und einmal als Unterschied der Lernrate in der Dropout-Bedingung im Gegensatz zur Standardbedingung modelliert. In dem Modell wird vorhergesagt, dass im Durchschnitt in der Standardbedingung am meisten gelernt wurde, danach in der Dropout-Bedingung. Jedoch zeigt das lineare Modell Schwierigkeiten beim Schätzen der Korrelationen der zufälligen Effekte (siehe Tabelle 18). Dies könnte an der schlechten Passung des Modells an den nichtlinearen Lernverlauf liegen (vgl. Tabelle 17).

Um einen nichtlinearen Verlauf abbilden zu können, wurde in Modell 1 der logarithmierte Effekt des Durchgangs ( $\ln(t) = (-1.10, -0.41, 0)$ ) geschätzt. Der Modellfit, gemäß dem AIC, verbesserte sich durch den logarithmisch modellierten Effekt um 51.4 Punkte. Das Modell zeigt, dass im letzten Testdurchgang Personen in der Standardbedingung einen größeren Lernerfolg als in der Dropout- oder „pure-study“-Bedingung erzielten.<sup>16</sup>

Die Varianzen für die „pure-study“-Bedingung und Testbedingungen (Standard- und Dropout-Bedingung) wurden separat geschätzt (siehe Tabelle 18). Das bedeutet, dass die

---

<sup>16</sup> Schätzt man anstatt der Differenz der Dropout-Bedingung zur Standardbedingung (DO) die Differenz der Dropout-Bedingung zur „pure-study“-Bedingung, so zeigt sich, dass sich die beiden Bedingungen nicht signifikant voneinander unterscheiden  $t(100) = 1.46, p = .15, \beta_{0DO} = 2.63$ .

Residuen für die Bedingungen mit oder ohne Testdurchgänge unterschiedlich weit um ihre durchschnittlichen Werte streuen können. Hierbei ist die Varianz in der „pure-study“-Bedingung größer als die Residualvarianz der Testbedingungen, was auf höhere Abweichungen der geschätzten von den gemessenen Werten hinweist (diese können jedoch auf unterschiedliche Modellierung der zufälligen Effekte zurückgeführt werden, da für die „pure-study“-Bedingung kein zufälliger Effekt geschätzt wird). Die Varianzen der zufälligen Effekte sind alle im statistischen Maße bedeutsam, wobei die Varianzen bezüglich der Dropout-Bedingung jeweils größer sind als die der Standardbedingung (Varianz der Lernbedingung und für deren Lernrate). Dabei korrelieren die individuellen Werte der Standardbedingung mit denen der Dropout-Bedingung in dem Maße, dass Personen deren wahrer Wert in der Standardbedingung tendenziell über dem geschätzten Wert lag, tendenziell höhere Residuen in der Dropout-Bedingung zeigten. Ein ähnliches Bild zeigt sich auch in Zusammenhang der beiden Lernraten (der Standard- und der Dropout-Bedingung). Lerner, welche in der Standardbedingung tendenziell schneller als der Durchschnitt lernten, taten dies wahrscheinlicher in der Dropout-Bedingung.

In einem nächsten Schritt wurde die Altersgruppe berücksichtigt, wodurch sich die Passung des Modells, gemäß dem AIC, von 39.0 Punkten verbesserte (siehe Tabelle 17, Modell 2; 0 = jung; 1 = alt). Der Wert  $\beta_{0PS}$  beschreibt, dass ein jüngerer Erwachsener sich durchschnittlich an 22 Texteinheiten nach fünfmaligem Lernen erinnert; sprich in der „pure-study“-Bedingung ( $t(302) = 57.37, p < .05, \beta_{0PS} = 86.15, d = 6.60$ ). In der Dropout- sowie in der Standardbedingung wurden im dritten Testdurchgang im Mittel mehr Texteinheiten von den jüngeren Probanden erinnert als in der „pure-study“-Bedingung (ein bis zwei Einheiten mehr), diese unterschieden sich jedoch nicht signifikant voneinander. Zwar wussten ältere Versuchspersonen in der „pure-study“-Bedingung im Mittel zwei Informationseinheiten weniger als jüngere Lerner ( $t(302) = -3.58, p < .05, \beta_{Alter} = -7.64, d = -0.41$ ), jedoch

unterschied sich dieser Wert nicht signifikant in der Standard- oder in der Dropout-Bedingung.<sup>17</sup> Somit zeigten Ältere in den drei Bedingungen zwar eine geringere Lernleistung, welche aber statistisch unabhängig von der Art der Lernbedingung ist (sprich mit oder ohne Testen). Auch wenn sich die beiden Altersgruppen nicht in der Lernrate der Standardbedingung unterscheiden, so gibt es Altersunterschiede in der Lernrate der Dropout-Bedingung ( $t(302) = -3.13, p < .05, \alpha_{2DO} = -7.38, d = -0.36$ )<sup>18</sup>.

Die individuellen Abweichungen der geschätzten Werte in der Standardbedingung hängen statistisch bedeutsam positiv mit denen der Dropout-Bedingung und negativ mit den Abweichungen in der Lernrate der Standardbedingung zusammen. Dabei zeigt letztere, dass Lerner mit durchschnittlich höheren Werten in der Standardbedingung eine geringere Lernrate zeigten. Diese scheinbar unerwartete Korrelation könnte aufgrund von Deckeneffekten zustande kommen. Je mehr Personen in einem ersten Testdurchgang wussten, desto weniger konnten sie mit jedem weiteren Lerndurchgang dazulernen.

---

<sup>17</sup> Schätzt man für die Interaktion mit der Altersgruppe anstatt der Differenz der Dropout-Bedingung zur Standardbedingung die Differenz der Dropout-Bedingung zur „pure-study“-Bedingung, so zeigt sich, dass sich die beiden Bedingungen nicht signifikant voneinander unterschieden ( $t(302) = -1.76, p = .08, \alpha_{0DO} = -5.76, d = -0.20$ ).

<sup>18</sup> Hierfür wurde Modell 2 wiederum so modelliert, dass nicht die Differenzen der Lernrate zur Standardbedingung, sondern unabhängig von dieser geschätzt wurden.

Tabelle 17

*Experiment 3 Modelle 0 bis 2: Schätzungen der festen Effekten und Fit-Indizes*

Parameter	Modell 0	Modell 1	Modell 2
	Linear <sup>a</sup>	Logarithmisch	+ Altersgruppe
<b>Feste Effekte</b>			
$\beta_{0PS}$ : „pure-study“ (PS)	82.37*	82.37*	86.15*
$\beta_{0ST}$ : Standard (ST)	7.47*	6.52*	7.27*
$\beta_{0DO}$ : Dropout (DO)	-3.82*	-3.89*	-1.79
$\beta_{1PS}$ : ST*ln(t)	13.54*	25.11*	26.85*
$\beta_{1DO}$ : DO*ln(t)	-1.97*	-4.08*	-2.19
$\alpha_{0PS}$ : Altersgruppe*PS			-7.64*
$\alpha_{0ST}$ : Altersgruppe*ST			-1.50
$\alpha_{0DO}$ : Altersgruppe*DO			-4.26
$\alpha_{1ST}$ : Altersgruppe*ST*ln(t)			-3.50
$\alpha_{2DO}$ : Altersgruppe*DO*ln(t)			-3.88
<b>Informationskriterium</b>			
-2 LL	5472.6	5421.2	5372.2
AIC	5506.6	5455.2	5416.2

*Anmerkungen:*  $-2LL = -2 \times \text{Log-Likelihood}$ , AIC = Akaike Information Criterion.

<sup>a</sup> hier wurde statt ln(t) der lineare Durchgang t (-2, -1, 0) verwendet.

\* $p < .05$

Tabelle 18

*Experiment 3 Modelle 0 bis 2: Schätzungen der Residualvarianzen und zufälligen Effekte*

Parameter	Modell 0	Modell 1	Modell 2
	Linear <sup>a</sup>	Logarithmisch	+ Altersgruppe
<b>Residualvarianzen</b>			
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“	131.86*	131.86*	117.26*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO	63.17*	47.34*	47.34*
<b>Zufällige Effekte</b>			
$\sigma_{b_0}^2$ : Var(ST)	103.89*	123.73*	102.87*
$\sigma_{b_1}^2$ : Var(DO)	149.64*	163.81*	118.93*
$\sigma_{d_0}^2$ : Var(ST*ln(t))	3.32	44.08*	41.02*
$\sigma_{d_0}^2$ : Var(DO*ln(t))	16.23*	75.70*	62.45*
Korr. ST x DO	.53*	.46*	.33*
Korr. ST x ST*ln(t)	-1.00*	-.35	-.52*
Korr. ST x DO*ln(t)	.19	.14	-.03
Korr. DO x ST*ln(t)	.09	.04	-.10
Korr. DO x DO*ln(t)	-.04	.07	-.20
Korr. ST*ln(t) x DO*ln(t)	1.00*	.56*	.52*

<sup>a</sup> hier wurde statt ln(t) der lineare Durchgang t (-2, -1, 0) verwendet.

\* $p < .05$

In Modell 3 wird der vorhergesagte Bereich der Abrufleistung zudem auf die verzögerten Testdurchgänge zwei Tage später erweitert (siehe Tabelle 19 und 20) (siehe Abbildung 10). Hierfür werden zusätzliche Prädiktoren für die Differenzen zwischen den drei Lernbedingungen sowie deren Altersunterschiede mit aufgenommen (vgl. Formel 3.1). Es zeigt sich, dass jüngere Probanden beim verzögerten Abruf der „pure-study“-Bedingung sich im Durchschnitt an vier Texteinheiten weniger erinnern als im Testdurchgang während des Experiments ( $t(593) = -9.82, p < .05, \beta_{0PS2} = -20.13, d = -0.81$ ).<sup>19</sup> Darüber hinaus gab

<sup>19</sup> Wird die ältere Altersgruppe als Referenzgruppe verwendet zeigt sich ein statistisch signifikanter Unterschied in der Lernrate der Dropout-Bedingung im Gegensatz zur Standardbedingung ( $t(593) = -2.58, p < .05, \beta_{1DOalt} = -6.16, d = -0.21$ ).

es keine Altersunterschiede in der Vergessensrate in den drei Lernbedingungen. Demnach haben jüngere und ältere Lerner ähnlich viele Einheiten der Geschichten der einzelnen Lernbedingungen im Laufe von zwei Tagen vergessen. Aufgrund der Hinzunahme einer weiteren Testung, kann in den Modellen ebenso die zufällige Varianz der „pure-study“-Bedingung geschätzt werden, welche im Modell 3 signifikant ist und zusätzlich mit den Abweichungen der Lernleistung in der Dropout-Bedingung positiv korreliert. So zeigen Personen, welche vom geschätzten Effekt der Dropout-Bedingung abweichen (also von einem Haupteffekt der Dropout-Bedingung), höhere Abweichungen in der Standard- und „pure-study“-Bedingung.

Berücksichtigt man mögliche Effekte des Designs<sup>20</sup> –modelliert als zufälliger Effekt zwischen den Designvarianten; *cross-classified-random-effect* –so verändern sich zwar nicht die festen Effekte im Wesentlichen, aber die zufälligen Effekte und deren Korrelationen (siehe Tabelle 21). Durch diese Modellierung wird das Design als mögliche Varianzquelle berücksichtigt, wodurch sich die Fehlervarianz reduziert. Der Modellfit verbessert sich, gemäß dem AIC, um 18.3 Punkte.

Dadurch, dass jüngere Probanden durchschnittliche mehr Textpropositionen erinnern als ältere, könnten v.a. bei jüngeren Erwachsenen vermehrt Deckeneffekte auftreten. Um dies zu berücksichtigen, werden die Residualvarianzen für Modell 3 in einem weiteren Schritt für die beiden Altersgruppen aufgeteilt (jeweils eine geschätzte Residualvarianz für die Lernbedingung mit und ohne Testbedingung für die beiden Altersgruppen; siehe Tabelle 20). Die größte Varianz zeigt sich bei der „pure-study“-Lernbedingung der älteren

---

<sup>20</sup> Die Variable für den zufälligen Effekt des Designs besitzt 18 Stufen, wovon jeweils eine Stufe eine verwendete Kombination aus der verwendeten Geschichte, deren Position im Experiment und Lernbedingung darstellt (= 6 Designs mit je drei Lernbedingungen/Geschichten/Positionen, siehe Tabelle 13).

Erwachsenen. Demnach fällt bei diesem Modell vor allem die Fehlervarianz bei dieser Gruppe am höchsten aus.<sup>21</sup>

In einem weiteren Modell wurden zusätzlich die Effekte der Verarbeitungsgeschwindigkeit mit aufgenommen (Modell 4), wodurch sich der AIC um 4.5 Punkte im Vergleich zu Modell 3 verbesserte und gleichzeitig 1,8% mehr Varianz innerhalb und 3,7% mehr Varianz zwischen Individuen aufklärt.<sup>22</sup> Insgesamt gibt es zwischen den beiden Altersgruppen sehr große Unterschiede in der Verarbeitungsgeschwindigkeit ( $t(93) = 6.18, p < .05$ , Differenz = 1.07,  $d = 1.27$ ). Die durchschnittliche Leistung bei der verbalen Lernaufgabe (unabhängig von der Lernbedingung) hängt im statistisch signifikanten Maße mit der Verarbeitungsgeschwindigkeit zusammen ( $t(545) = 2.60, p < .05$ ,  $\beta_{VG} = 2.85, d = 0.22$ ).<sup>23</sup> Das bedeutet, Personen welche eine Standardabweichung über der durchschnittlichen Verarbeitungsgeschwindigkeit liegen, erinnern im Mittel knapp eine Worteinheit mehr. Dies zeigt sich im schwindenden Einfluss der Altersgruppe auf die allgemeine Lernleistung (von Modell 3 auf Modell 4). Jedoch hat die Verarbeitungsgeschwindigkeit keinen direkten Einfluss auf die Lernrate in den beiden Testbedingungen ( $t(545) = -1.11, p = .27$ ,  $\beta_{VG*\ln(t)} = -1.21, d = -0.10$ ).

---

<sup>21</sup> Berücksichtigt man die vier Residualvarianzen in Modell 1 verringert sich die  $-2$  LL um einen Punkt und das AIC erhöht sich um 3.1 Punkte; für Modell 2 verringert sich die  $-2$  LL um 0.9 Punkte und das AIC erhöht sich um drei Punkte. Für Modell 4 zeigt sich ein ähnliches Bild wie für Modell 3 ( $-2$  LL = 7303.4; AIC = 7367.4).

<sup>22</sup> In Modell 4 wurden aufgrund von fehlenden Werten der Verarbeitungsgeschwindigkeit weniger Daten in das Modell mit aufgenommen (insgesamt 930 Beobachtungen). Für den Vergleich der Varianzaufklärung und der Modellgüte wurde daher auch Modell 3 930 Beobachtungen verwendet.

<sup>23</sup> Werden die Parameter der Verarbeitungsgeschwindigkeit für die Altersgruppen getrennt geschätzt, wird kein Prädiktor statistisch signifikant (durch die Hinzunahme der Prädiktoren verschlechtert sich Modell 4, gemäß dem AIC, um 3.1 Punkte).

Tabelle 19

*Experiment 3 Modelle 3 bis 5: Schätzungen der festen Effekten und Fit-Indizes*

Parameter	Modell 3 + verzögert	Modell 4 <sup>b</sup> + VG
<b>Feste Effekte</b>		
$\beta_{0PS}$ : „pure-study“ (PS)	86.15*	84.73*
$\beta_{0ST}$ : Standard (ST)	7.26*	7.21*
$\beta_{0DO}$ : Dropout (DO)	-1.91	-1.81
$\beta_{1ST}$ : ST*ln(t)	26.85*	28.76*
$\beta_{1DO}$ : DO*ln(t)	-2.17	-1.77
$\alpha_{0PS}$ : Altersgruppe*PS	-7.64*	-4.97
$\alpha_{0ST}$ : Altersgruppe*ST	-1.49	-1.28
$\alpha_{0DO}$ : Altersgruppe*DO	-4.25	-4.63
$\alpha_{1ST}$ : Altersgruppe*ST*ln(t)	-3.50	-5.20
$\alpha_{1DO}$ : Altersgruppe*DO*ln(t)	-3.99	-4.39
$\beta_{0PS2}$ : PS*2Tage	-20.13*	-20.67*
$\beta_{0ST2}$ : ST*2Tage	-0.32	-1.02
$\beta_{0DO2}$ : DO*2Tage	2.07	2.09
$\alpha_{0PS2}$ : Altersgruppe*PS*2Tage	0.08	0.74
$\alpha_{0ST2}$ : Altersgruppe*ST*2Tage	-1.44	-1.38
$\alpha_{0DO2}$ : Altersgruppe*DO*2Tage	1.60	1.49
VG		2.85*
VG*ln(t)		-1.21
<b>Informationskriterium</b>		
-2 LL	7954.2	7310.9 <sup>b</sup>
AIC	8010.2	7370.9 <sup>b</sup>
R <sup>2</sup> Level 1	32.3%	34.6%
R <sup>2</sup> Level 2	11.6%	15.3%

<sup>b</sup> Aufgrund von insgesamt 80 fehlenden VG-Werten gehen in diese Berechnung 930 statt 1010 Beobachtungen ein (dadurch sind die Fit-Indizes nicht mehr direkt mit den vorherigen Modellen vergleichbar).

\* $p < .05$

Tabelle 20

*Experiment 3 Modelle 3 und 4: Schätzungen der Residualvarianzen und zufälligen Effekte*

Parameter	Modell 3 + verzögert	Modell 4 <sup>b</sup> + VG
<b>Residualvarianzen</b>		
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“	109.29*	109.52*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO	88.40*	85.26*
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“-jung		
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“-alt		
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO-jung		
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO-alt		
<b>Zufällige Effekte</b>		
$\sigma_{b_0}^2$ : Var(ST)	127.89*	137.24*
$\sigma_{b_1}^2$ : Var(DO)	126.76*	136.10*
$\sigma_{d_0}^2$ : Var(PS)	103.31*	97.91*
$\sigma_{d_0}^2$ : Var(ln(t))	24.64*	22.63*
Korr. ST x DO	.36*	.33*
Korr. ST x PS	.21	.15
Korr. ST x ln(t)	-.11	-.01
Korr. DO x PS	.39*	.34*
Korr. DO x ln(t)	-.18	-.04
Korr. PS x ln(t)	-.22	-.14

<sup>b</sup> Aufgrund von insgesamt 80 fehlenden VG-Werten gehen in diese Berechnung 930 statt 1010 Beobachtungen ein.

Tabelle 21

*Experiment 3 Zusatzmodelle: Schätzungen der Residualvarianzen, zufälligen Effekte und Fit-Indizes*

Parameter	Modell 3 mit Design-Effekt	Modell 3 mit Residualvarianzen für Testungsart und Altersgruppe
<b>Residualvarianzen</b>		
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“	85.78*	
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO	71.70*	
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“-jung		73.88*
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“-alt		151.26*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO-jung		80.89*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO-alt		96.05*
$\sigma_{Design}^2$ : Var (Design)	0.28*	
<b>Zufällige Effekte</b>		
$\sigma_{b_0}^2$ : Var(ST)	125.38*	126.66*
$\sigma_{b_1}^2$ : Var(DO)	109.11*	124.54*
$\sigma_{d_0}^2$ : Var(PS)	102.97*	99.44*
$\sigma_{d_0}^2$ : Var(ln(t))	28.32*	26.00*
Korr. ST x DO	.48*	.35*
Korr. ST x PS	.17	.18
Korr. ST x ln(t)	-.11	-.10
Korr. DO x PS	.44*	.42*
Korr. DO x ln(t)	-.17	-.19
Korr. PS x ln(t)	-.29	-.21
<b>Informationskriterium</b>		
-2 LL	7933.9	7944.9
AIC	7991.9	8004.9

\* $p < .05$

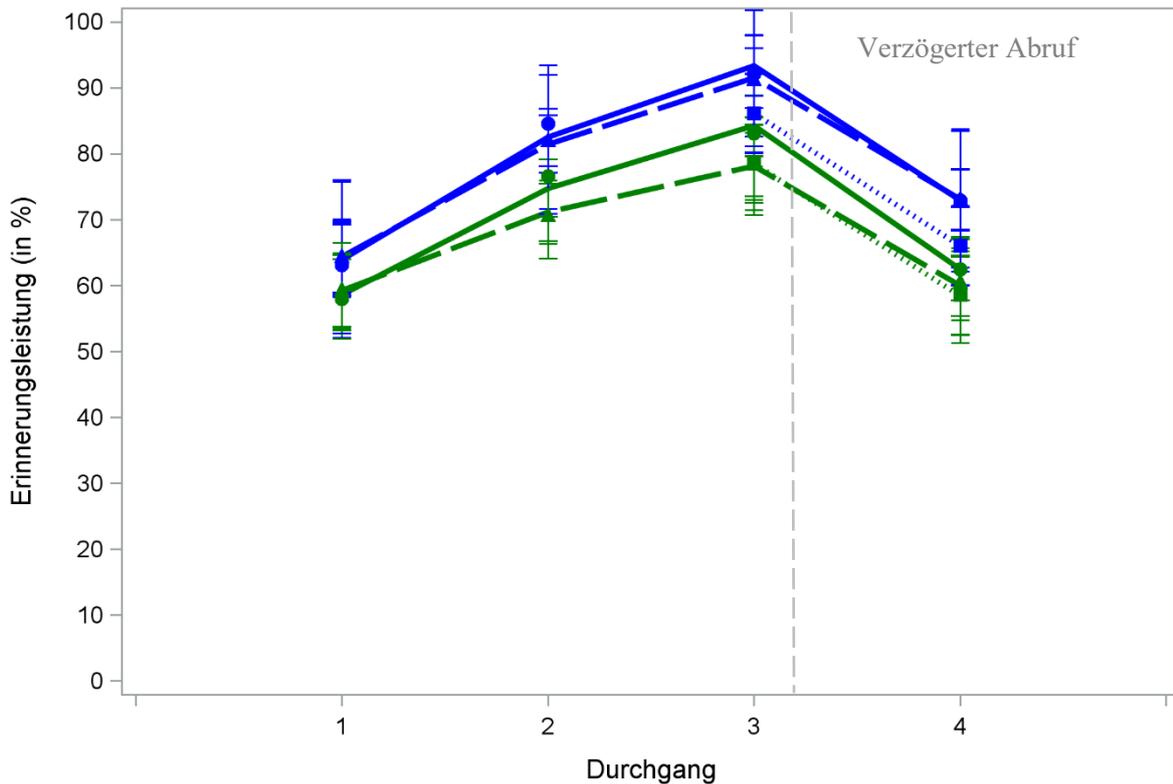


Abbildung 10. Experiment 3: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) für Modell 3. Die durchgehenden Linien repräsentieren die Standardlernbedingung, die gestrichelten Linien die Dropout-Bedingung und die gepunktete Linie die „pure-study“-Bedingung. Die Kreise zeigen die beobachteten Werte der Standardbedingung, die Dreiecke die beobachteten Werte Dropout-Bedingung dar und die Quadrate die beobachteten Werte der „pure-study“-Lernbedingung in den beiden Altersgruppen. Die Fehlerbalken stellen die Standardfehler dar.

### 3.2.3. Zusammenfassung der Ergebnisse aus Experiment 3

Die festen Effekte in den Modellen beschreiben, dass jüngere Lerner während des Experiments in der Standard- und Dropout-Bedingung im Durchschnitt einen höheren Lernerfolg erzielt haben als in der „pure-study“-Bedingung. Hier zeigen sich unmittelbare Effekte des Testens. In der älteren Gruppe ist zwar die Lernleistung höher in der Standardbedingung als in der „pure-study“-Bedingung, jedoch unterscheidet die Dropout-Bedingung sich hinsichtlich des Lernerfolgs bei Älteren nicht von der „pure-study“-

Bedingung. Der Testeffekt zeigt sich bei den älteren Lernern nur in der Standardbedingung. Eine weitere Sonderrolle kommt der Dropout-Bedingung darüber hinaus durch die Lernrate zu. Dadurch, dass in der Standard- sowie in der Dropout-Bedingung jeweils die gesamte Geschichte abgefragt wurde, lernten Probanden mit jedem Testdurchgang ähnlich viele, neue Propositionen in beiden Bedingungen. Allerdings lernen ältere Individuen in der Dropout-Bedingung ungefähr ein bis zwei Propositionen im ersten, und ungefähr eine Proposition im zweiten Testdurchgang weniger als jüngere Erwachsene. Dieses Ergebnis hebt die besondere Rolle der Dropout-Bedingung hervor und zeigt, dass ältere Erwachsene größere Probleme mit dieser Art des Lernens und Testens haben als jüngere. Dies kann vor allem mit dem Material der Geschichte zusammenhängen. Auf die Besonderheiten von kohärenten Texten als Lernmaterial wird in der Diskussion (Kapitel 4.2.) genauer eingegangen.

Personen erinnern beim verzögerten Abruf der Standardbedingung mehr Wörter, als in der „pure-study“-Bedingung mit den Geschichten als Lernmaterial (die Dropout-Bedingung unterscheidet sich nicht signifikant von den anderen beiden Bedingungen). Jedoch vergessen die Probanden im Durchschnitt bei allen drei Bedingungen ähnlich viele Propositionen (vgl. Experiment 3 bei Roediger & Smith, 2012). Dies spiegelt jedoch nicht den typischen Testeffekt wider (beim freien Abruf von Texten: Roediger & Karpicke, 2006b). Durch die Effekte des wiederholten, erfolgreichen Abrufs (bspw. zusätzliche Kontexthinweise) sollten die gelernten bzw. abgerufenen Einheiten langfristiger gespeichert werden. Das Abrufintervall ist im Vergleich zu anderen Studien, welche mehrere Lern-Test-Durchgänge verwenden, mit zwei Tagen relativ kurz (vgl. Karpicke & Roediger, 2007). Das Material könnte durch das fünfmalige Lernen in der „pure-study“-Bedingung so gemeistert werden, dass es in den ersten Tagen relativ stabil gegen Vergessen ist (vgl. Bifurkationsmodell, Kornell et al., 2011). Daher könnten etwaige Testeffekte in Bezug auf das Vergessen sich erst Tage später auswirken.

Die Verarbeitungsgeschwindigkeit hängt mit der Leistung beim letzten Testdurchgang des Experiments zusammen, und zwar in dem Sinne, dass Personen sich mit höheren Werten in der Verarbeitungsgeschwindigkeit allgemein an mehr Textpropositionen erinnern. Der Effekt der Verarbeitungsgeschwindigkeit mindert den Unterschied der beiden Altersgruppen bezüglich der Lernleistung im letzten Durchgang des Experiments. Um zu überprüfen, ob der Einfluss der Verarbeitungsgeschwindigkeit nicht nur in den Altersunterschieden in dieser kognitiven Ressource besteht, wurde Modell 4 wiederholt mit einem gruppencentrierten Verarbeitungsgeschwindigkeitsprädiktor (bezüglich Altersgruppe) berechnet. Hierbei bleibt der signifikante Einfluss der Verarbeitungsgeschwindigkeit bestehen ( $t(545) = 2.60, p < .05, \beta_{VG} = 2.85, d = 0.22$ ). Zusätzlich zeigt sich, wie schon in Modell 3, ein negativer Einfluss der Altersgruppe auf die Testleistung in der „pure-study“-Bedingung ( $t(545) = -2.72, p < .05, \alpha_{PS} = -8.04, d = -0.23$ ). Dieses Ergebnis unterstützt die Annahme, dass die Verarbeitungsgeschwindigkeit als basale mentale Ressource einen Einfluss auf die Lernleistung hat.

Die Fehlervarianzen der Lernbedingung mit Testen unterscheiden sich von denen der Lernbedingung ohne Testdurchgänge in dem Sinne, dass in der „pure-study“-Bedingung mehr Fehlervarianz vorliegt. Werden die Residualvarianzen zusätzlich für die beiden Altersgruppen getrennt betrachtet, zeigt sich die größte Fehlervarianz in der „pure-study“-Bedingung bei älteren Erwachsenen (siehe Tabelle 21); demnach beschreibt das Modell die Verläufe der anderen Gruppen besser. Dies könnte auf größere Unterschiede zwischen Lernern im Umgang mit der Lernbedingung deuten. Es wird vermutet, dass bspw. unterschiedliche Lernstrategien der Probanden hierauf einen Einfluss haben. Jedoch liegt die Analyse verwendeter Lernstrategien der Probanden nicht im Fokus dieser Arbeit.

Durch die Berücksichtigung des Designs als Varianzquelle wird die Fehlervarianz, welche auf die Kombination aus Lernbedingung, Geschichte und deren Position im Text

geschätzt. Einerseits kann der unterschiedliche Schwierigkeitsgrad der Geschichten die Lernleistung beeinflusst haben (siehe Tabelle 15) und andererseits Übungseffekte beim Lernen im Laufe des Experiments. In Modell 3 beeinflussen die Effekte des Designs vor allem die Residualvarianz, wohingegen die statistische Signifikanz und Werte der festen Effekte weitestgehend unverändert bleiben. Demnach liegen in dem Experiment zusätzliche Störfaktoren der Designs vor, welche jedoch die Interpretierbarkeit der Lern- sowie Testeffekte nicht beeinträchtigen.

Eine Limitation der Studie besteht in eher zu geringen Aufgabenschwierigkeit. Dadurch, dass die Geschichten jeweils dreimal gelernt und abgefragt werden (bzw. fünfmal gelernt und einmal abgefragt werden), entstehen schnell Deckeneffekte. Selbst wenn alle drei Geschichten sich in ihrer Schwierigkeit unterscheiden, so sticht vor allem Geschichte 1 als besonders leicht hervor (Geschichte 1 vs. 2: zwei bis drei Texteinheiten mehr erinnert,  $d = .66$ ; Geschichte 1 zu 3: ca. zwei Texteinheiten,  $d = .46$ ). Dieses Muster zeigt sich bei beiden Altersgruppen, jedoch bei der älteren Stichprobe im Besonderen (Geschichte 1 zu 2:  $d = .87$ ; Geschichte 1 zu 3:  $d = .59$ ). Trotz der Randomisierung können diese Unterschiede die Varianzen der zufälligen Effekte sowie die Fehlervarianzen verzerren. Jedoch ist andererseits mit drei Lern-Test-Durchgängen die Aussagekraft der Lernkurven eingeschränkt – bspw. zu wenig Messzeitpunkte, um einen quadratischen Verlauf zu schätzen. Durch zusätzliche Lern-Test-Durchgänge können validere Aussagen über die Lernrate in den Altersgruppen getroffen werden. Die Bedeutung der Ergebnisse für das verwendete Material wird in der Diskussion (Kapitel 4.2) diskutiert.

### *3.3. Experiment 4 – längere Geschichten*

Im Vergleich zu Experiment 3 werden im folgenden Experiment schwerere Aufgaben verwendet, mehr Test-Durchgänge durchgeführt (dadurch wird ein quadratischer Einfluss

der Lern-Test-Durchgänge modellierbar) und der verzögerte Abruf auf eine Woche nach dem Experiment verschoben. Es wird erwartet, dass sich ein Testeffekt im unmittelbaren sowie im verzögerten Abruf nach einer Woche ergibt—d.h. ein Vorteil der Standard- und Dropout-Bedingung gegenüber der „pure-study“-Bedingung—und dass sich dieser Effekt bei jüngeren und älteren Erwachsenen gleichermaßen finden lässt. Aufgrund der sinnhaften Struktur von Geschichten als Lernmaterial, werden keine Altersunterschiede erwartet (siehe Kapitel 3.1.1.: Theorie Studienblock 2), was in Einklang mit den Ergebnissen aus Studie 3 ist.

### 3.3.1. Methode Experiment 4

#### Stichprobe

Am vierten Experiment nahmen 50 jüngere und 52 ältere Erwachsene (insgesamt  $N = 102$ ) teil. Die Teilnehmer der jüngeren Stichprobe waren im Mittel 24.32 Jahre alt ( $SD = 3.67$ ; 18-30 Jahre), wohingegen die Probanden im höheren Erwachsenenalter durchschnittlich 71.88 Jahre alt waren ( $SD = 5.12$ ; 65-81 Jahre). Von den 103 Probanden waren 62 (60%) weiblich (jüngere Erwachsene: 58% Frauen; ältere Erwachsene: 62%).<sup>24</sup> Die subjektive Gesundheit der Probanden wurde mittels einer Likert-Skala von 1 = „sehr gut“ bis 5 = „mangelhaft“ erfasst und betrug im Mittel 2.15 ( $SD = 0.81$ ). Dabei gaben die jüngeren Teilnehmer im Mittel eine bessere subjektive Gesundheit an ( $M = 1.78$ ,  $SD = 0.71$ ) als ältere ( $M = 2.50$ ,  $SD = 0.75$ ) ( $t(100) = -4.97$ ,  $p < .05$ , Cohen's  $d = -0.98$ ). Um ihren Schulabschluss anzugeben, hatten die Teilnehmer fünf Optionen zur Verfügung: „kein Abschluss“, „Volks-/Hauptschule“, „Realschule“, „Abitur“ und „Hochschulabschluss“. In der jüngeren Stichprobe gaben 24 (48%) Abitur und 19 (38%) Hochschulabschluss als höchsten

---

<sup>24</sup> Die Geschlechterverteilung zwischen den Altersgruppen war nicht signifikant verschieden,  $\chi^2(1, 101) = .13$ ,  $p = .72$ , Cohen's  $w = 0.07$ .

erreichten Schulabschluss an. Wohingegen die beiden häufigsten Antworten bei den älteren Erwachsenen, Volks-/Hauptschulabschluss (18 Teilnehmer, 35%) sowie Realschulabschluss (16 Teilnehmer, 31%) waren. Durchschnittlich verbrachte die jüngere Stichprobe ( $M = 16.64$ ,  $SD = 3.29$ ) drei Jahre mehr in Bildungseinrichtungen als die ältere Stichprobe ( $M = 13.37$ ,  $SD = 3.83$ ) ( $t(100) = 4.62$ ,  $p < .05$ , Cohen's  $d = 0.91$ ). Die meisten jüngeren Probanden waren ledig ( $N = 46$ , 92%) und Schüler/Studenten ( $N = 27$ , 54%) oder vollzeitbeschäftigt ( $N = 18$ , 36%), wohingegen die meisten älteren Teilnehmer verheiratet ( $N = 36$ , 69%) oder verwitwet ( $N = 9$ , 17%) und Rentner waren ( $N = 50$ , 96%).

### *Geschichten*

Als verbale Lernaufgabe wurden dieselben Geschichten wie aus Experiment 3 verwendet. Diese wurden jeweils um sieben zusätzliche Propositionen erweitert. Dabei handelt es sich vor allem um spezifizierende und ergänzende Propositionen, welche die Handlung der Geschichten nicht verändert (bspw.: in der Originalversion der Geschichte transportiert ein LKW Eier; in der erweiterten Version der Geschichte sind es Eier und Gurken). Somit wurde Geschichte G1 (Geschichte B aus dem WMS-IV) von 69 auf 85 Wörter; Geschichte G2 (Geschichte C aus dem WMS-IV) von 87 auf 100 und G3 (Geschichte B aus dem WMS-R) von 72 auf 89 Wörter erweitert. Die Teilnehmer konnten in jedem Testdurchgang bis zu 32 Propositionen korrekt erinnern. Lern- sowie Testdurchgang wurden wie in Experiment 3 durchgeführt.

### *Lerndesign*

Die Standard-, Dropout- sowie „pure-study“-Lernbedingungen wurden wie in Experiment 3 durchgeführt, mit zwei Ausnahmen. Erstens wurden in diesem Experiment vier statt drei Lern-Test-Durchgänge durchgeführt und zweitens wurde in der „pure-study“-

Bedingung die Geschichte insgesamt vier Mal gelernt bis der finale Testdurchgang folgte. Hierbei folgte auf jeden Lerndurchgang jeweils eine 60-sekündige Füllaufgabe<sup>25</sup> (siehe Tabelle 22). Die Versuchspersonen wurden zufällig auf eines von sechs möglichen Designs aufgeteilt (nach griechisch-lateinischen Quadrat erstellt) (siehe Experiment 3, Tabelle 13).

Tabelle 22

*Lernbedingungen Experiment 4*

Bedingung	Durchgang 1		Durchgang 2		Durchgang 3		Durchgang 4		Finaler Test	2 Tage später
	S	F	T S	F	T S	F	T S	F		
Standard	S	F	T S	F	T S	F	T S	F	T	T
Dropout	S	F	T S <sub>n</sub>	F	T S <sub>n</sub>	F	T S <sub>n</sub>	F	T	T
„pure-study“	S	F	S	F	S	F	S	F	T	T

*Anmerkungen.* S = Lerndurchgang; T = Testdurchgang; S<sub>n</sub> = Dropout-Bedingung; F = 60-sekündige Füllaufgabe.

*Messinstrumente**Buchstaben-Zahlen-Matrizen-test*

Der Buchstaben-Zahlen-Matrizen-test ist der gleiche wie in Experiment 1 (siehe Kapitel 2.3.1.). Dieser Test wurde von den Probanden zweimal für jeweils 60 Sekunden bearbeitet. Die erreichte Punktzahl ist die Anzahl der korrekt beantworteten Items aus beiden Tests (mögliche Punktzahl: 0–150).

<sup>25</sup> Die Füllaufgaben waren einfache kurze Aufgaben wie bspw. ein Fahrrad zeichnen, Kopfrechnen, „Streichholzrätsel“, Ziffernblatt einer Uhr zeichnen, Labyrinth lösen, Zahlen-verbinden.

*Verzögerter Abruf*

Der verzögerte Abruf eine Woche später wurde wie in Experiment 3 durchgeführt (außer dass nun insgesamt 28 Propositionen pro Geschichte erinnert werden konnten). Als Hinweisreize dienten pro Geschichte dieselben vier Propositionen wie in Experiment 3.

*Prozedur*

Die Erhebung fand an zwei Terminen statt. Am ersten Termin (Hauptuntersuchung) unterzeichneten die Probanden die Einverständniserklärung und bearbeiteten die Aufgaben in Einzeltestungen an einem Laptop. Anfangs wurden demographische Variablen abgefragt (Alter, Geschlecht, Familienstand, Schulabschluss, Bildungsdauer in Jahren, derzeitige Tätigkeit sowie subjektive Gesundheit). Anschließend wurde die Gedächtnisaufgabe „Bilder erinnern“ bearbeitet (NAI; Oswald & Fleischmann, 1999). Darauf folgte die verbale Lernaufgabe, wobei vor jeder Lernbedingung eine kurze Instruktion präsentiert wurde (bzgl. der Lernbedingung, wie oft gelernt und abgefragt wird). Die Geschichten wurden in schwarzen Buchstaben auf weißen Hintergrund dargeboten (Schriftart: Arial; Schriftgröße: 5.2% des Bildschirms). Zwischen den drei Lernbedingungen bearbeiteten die Versuchspersonen jeweils für 90 Sekunden den Buchstaben-Zahlen-Test (ca. 30 Sekunden Instruktion und 60 Sekunden Aufgabe). Jede Versuchsperson bearbeitete alle drei Lernbedingungen. Abschließend bearbeiteten die Probanden für bis zu sieben Minuten den Mehrfachwahlwortschatztest (MWT-B, Lehrl, 1999). Zum Schluss des Haupttermins wurden die Probanden gefragt, ob und gegebenenfalls welche Strategie sie verwendet haben, um sich die Geschichten zu merken. Dieser erste Termin dauerte ungefähr eine Stunde. Eine weitere Testung wurde eine Woche später telefonisch durchgeführt. Die Teilnehmer wurden nochmal nach Alter und Geschlecht gefragt, bevor der verzögerte Abruf begann hat. Dieser zweite Teil des Experiments dauerte ungefähr 10 Minuten.

*Datenanalyse*

Für die Analyse der Daten wurden mehrere gemischte Regressionen modelliert. Hierbei wurde ein ähnliches Modell wie in Experiment 3 verwendet (die Modellbeschreibung bezieht sich auf Modell 5 in Tabelle 29 und 30). Im Gegensatz zu Experiment 3 gibt es hier vier Durchgänge (d.h. der logarithmierte Durchgang  $\ln(t)$  kann dementsprechend die Ausprägungen  $-1.39$ ,  $-0.69$ ,  $-0.29$  und  $0$  annehmen). Es wurde angenommen, dass die individuellen Fehler mit konstanter Varianz normalverteilt sind und dass die Kovarianz-Matrix der zufälligen Effekte der drei Bedingungen und des Lernzuwachs wie in Experiment 3 aufgebaut ist (siehe Formel 3.4). Die Residualvarianzen der Bedingung ohne Testen  $\sigma_{ePS}^2$  sowie der beiden Bedingungen mit Testdurchgängen  $\sigma_{eTest}^2$  (Standard- und Dropout-Bedingung gemeinsam) wurden getrennt geschätzt.

Die Berechnungen wurden mit SAS Version (9.4) (SAS Institute, 2016) durchgeführt. Die gemischte Regression wurde mit der Prozedur MIXED und der „Maximum Likelihood“-Schätzmethode modelliert. Die Varianzaufklärung auf den beiden Ebenen des Modells werden mit  $R^2$  (Recchia, 2010) und Effektstärken mit Cohen's  $d$  angegeben—die Angaben dienen dabei vor allem der Veranschaulichung.

*3.3.2. Ergebnisse Experiment 4**Datenbereinigung*

Aufgrund von Problemen bei der Durchführung fehlt bei einer älteren Person die Leistung zum ersten Testzeitpunkt der Standardbedingung. Eine andere ältere Person war zum verzögerten Abruf nicht mehr erreichbar, wodurch deren Werte für diese Abfragedurchgänge fehlen. Die Abrufleistung beim verzögerten Erinnern einer jüngeren Person war so auffällig niedrig, dass diese in fehlende Werte umgewandelt wurden (die Geschichten der Standard- und „pure-study“-Bedingung wurden komplett nicht erinnert).

Aufgrund des längeren Abrufintervalls konnten mehrere ältere Personen bei manchen Geschichten keine weiteren Textpropositionen abrufen. Zwei ältere haben die Geschichte aus der Standardbedingung nicht mehr erinnert. Zehn ältere Personen konnten keine weiteren Propositionen der „pure-study“-Geschichte nennen, wovon eine Person zusätzlich keine Texteinheiten der Dropout-Geschichte nennen konnte. Dementsprechend haben diese Personen sich an 0-Einheiten erinnert. Da hier der Unterschied zum letzten Testdurchgang des Experiments–wie auch bei teilweise erinnerten Geschichten–Vergessen darstellt, werden die nicht-erinnerten Geschichten in die Berechnungen mit aufgenommen. Allerdings weisen eben diese Fälle Besonderheiten gegenüber Geschichten auf, welche nur teilweise vergessen wurden. Bei Geschichten die komplett vergessen wurden, hängt der Umfang des vergessenen Materials vom letzten Testdurchgang im Experiment ab: wenn bspw. eine Person im vierten Testdurchgang der „pure-study“-Bedingung 21 Einheiten erinnert hat, würde sie dementsprechend 21 Propositionen vergessen haben, wohingegen eine mit 10 Propositionen nur 10 Einheiten vergessen hätte. Daher wird bei der Ergebnispräsentation an spezifischer Stelle zusätzlich auf die Ergebnisse eines Modells verwiesen, bei dem die vergessenen Geschichten dieser Personen als fehlende Werte behandelt werden (siehe Ergebnisbeschreibung von Modell 5).

### *Deskriptive Statistiken*

In Tabelle 23 sind die beobachteten Mittelwerte der einzelnen Testdurchgänge für jüngere sowie ältere Probanden dargestellt (siehe Abbildung 11). Die Lernleistung in den beiden Altersgruppen steigt in der Standard- sowie in der Dropout-Bedingung erwartungsgemäß mit jedem weiteren Lern-Test-Durchgang an. Wie schon in Experiment 3 wird der Lernzuwachs mit jedem weiteren Lern-Test-Durchgang in der Standard- und Dropout-Bedingung durchschnittlich geringer. Es zeigen sich bei jeder Testung

Altersunterschiede, wobei die durchschnittlichen Lernkurven der jüngeren und älteren Erwachsenen in den ersten drei Messungen fast parallel verlaufen. Dies gilt jedoch nicht für die die Altersunterschiede im dritten und vierten Durchgang der Standardbedingung. Dies könnte daran liegen, dass die jüngere Gruppe im dritten Testdurchgang schon Deckeneffekte zeigen (bei den jüngeren Versuchspersonen erinnern sich elf Personen an alle, und acht Personen an 31 Propositionen—bei der älteren Gruppe sind es jeweils nur zwei Probanden). Die Altersunterschiede im letzten Testdurchgang des Experiments sind bei der Dropout-Bedingung am größten, in der Standard- sowie in der „pure-study“-Bedingung sind diese ähnlich geringer. Hier zeigt sich wie in Experiment 3, dass am Ende des Experiments ältere Probanden mehr Schwierigkeiten mit der Dropout-Bedingung haben. Jedoch zeigt sich ein anderes Muster im verzögerten Abruf nach einer Woche. Hier liegen die geringsten Altersunterschiede im Testdurchgang der Dropout-Geschichte vor (auf diesen Unterschied wird in den Analysen zu den Vergessensraten sowie in der Diskussion nochmal Bezug genommen). Allgemein erinnerten sich die Probanden beim Abruf nach einer Woche durchschnittlich an die Geschichten der Standard- und Dropout-Bedingung ähnlich gut, wohingegen im Mittel vier Informationseinheit weniger bei der Geschichte der „Pure-Study“-Bedingung erinnert wurden.

Tabelle 23

*Experiment 4: Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

Variablen				Jung	Alt	Gesamt	Diff(Jung – Alt)	Cohens d
Standard 1	$N_{jung}$	50	$M$	55.13	43.63	49.32	$t(99) =$	0.71
	$N_{alt}$	51	$SD$	15.19	17.05	17.08	3.58*	
Standard 2	$N_{jung}$	50	$M$	79.56	64.60	71.94	$t(100) =$	0.97
	$N_{alt}$	52	$SD$	13.16	17.46	17.16	4.87*	
Standard 3	$N_{jung}$	50	$M$	90.81	74.04	82.26	$t(100) =$	1.34
	$N_{alt}$	52	$SD$	7.92	15.74	15.05	6.76*	
Standard 4	$N_{jung}$	50	$M$	94.94	81.43	88.05	$t(100) =$	1.21
	$N_{alt}$	52	$SD$	5.78	14.54	13.00	6.12*	
Dropout 1	$N_{jung}$	50	$M$	56.25	42.55	49.26	$t(100) =$	0.84
	$N_{alt}$	52	$SD$	14.40	18.10	17.70	4.22*	
Dropout 2	$N_{jung}$	50	$M$	75.06	58.89	66.82	$t(100) =$	1.02
	$N_{alt}$	52	$SD$	12.43	18.49	17.71	5.16*	
Dropout 3	$N_{jung}$	50	$M$	86.81	70.13	78.31	$t(100) =$	1.06
	$N_{alt}$	52	$SD$	12.99	18.06	17.80	5.34*	
Dropout 4	$N_{jung}$	50	$M$	94.44	77.64	85.88	$t(100) =$	1.31
	$N_{alt}$	52	$SD$	6.95	16.54	15.26	6.64*	
„pure-study“	$N_{jung}$	50	$M$	83.56	65.87	74.54	$t(100) =$	1.19
	$N_{alt}$	52	$SD$	13.18	16.41	17.30	5.99*	
Standard 1W	$N_{jung}$	49	$M$	57.65	39.34	48.31	$t(98) =$	1.06
	$N_{alt}$	51	$SD$	14.19	19.70	19.45	5.32*	
Dropout 1W	$N_{jung}$	49	$M$	53.83	42.03	47.81	$t(98) =$	0.69
	$N_{alt}$	51	$SD$	13.61	19.82	17.98	3.46*	
„pure-study“ 1W	$N_{jung}$	49	$M$	42.67	23.35	32.81	$t(98) =$	1.04
	$N_{alt}$	51	$SD$	15.59	20.93	20.82	5.22*	

*Anmerkungen.* Standard 1-4 entsprechen den Testdurchgängen der Standardbedingung nach 1 bis 4 Lerndurchgängen; Dropout 1-4 entsprechen den Testdurchgängen der Dropout-Bedingung nach 1 bis 4 Lerndurchgängen; „pure-study“ entspricht der Erinnerungsleistung nach fünfmaligem Lernen; Standard/Dropout/„pure-study“ 1W = Standard/Dropout/„pure-study“ Abfragedurchgang eine Woche nach der Testung.

\* $p < .05$

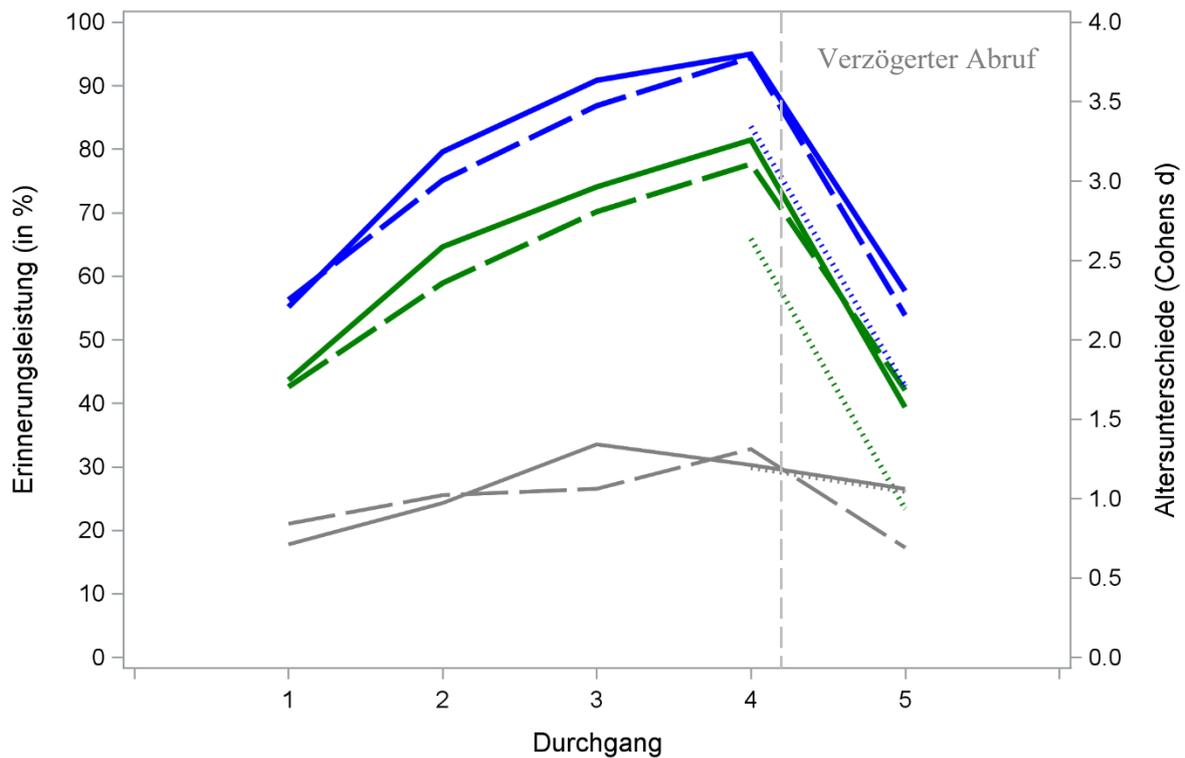


Abbildung 11. Experiment 4: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) sowie die Altersunterschiede (grau – rechte y-Achse). Die durchgehenden Linien repräsentieren die Standardlernbedingung, die gestrichelten Linien die Dropout-Bedingung und die gepunktete Linie die „pure-study“-Bedingung.

Tabelle 24 zeigt die Interkorrelationen aller Testdurchgänge sowie den Tests nach einer Woche. Innerhalb der Standard- und Dropout-Bedingung korrelieren aufeinanderfolgende Testdurchgänge höher als mit den vorangegangenen Tests. Bei den jüngeren Erwachsenen korreliert der letzte Durchgang der Standardbedingung gering (aber nicht signifikant) mit den Werten des verzögerten Abrufs, wohingegen die letzten Testwerte der Dropout-Bedingung und der „pure-study“-Bedingung jeweils mittel bis hoch mit den jeweiligen Testwerten des verzögerten Erinnerns korrelieren. Durch die hohen Abrufleistungen im dritten und vierten Durchgang der Standardbedingung bei den jüngeren Probanden könnte die daraus folgende eingeschränkte Varianz Ursache für die fehlenden statistisch signifikanten Korrelationen sein. Bei den älteren zeigen sich durchwegs hohe

Korrelationen der Testwerte im Experiment mit denen des Abrufs nach einer Woche. Wie in Experiment 3 zeigen sich hier bei den älteren Teilnehmern durchschnittlich höhere Korrelationen zwischen den Testdurchgängen der verschiedenen Bedingungen als bei den jüngeren Probanden.

Tabelle 24

*Experiment 4: Korrelationstabelle der Lernleistung aller Testdurchgänge jüngerer und älterer Probanden*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Standard 1		.62*	.52*	.42*	.35*	.36*	.34*	.25	.46*	.33*	.30*	.27
2. Standard 2	.83*		.78*	.66*	.31*	.39*	.39*	.29*	.48*	.27	.37*	.22
3. Standard 3	.78*	.91*		.73*	.13	.18	.13	-.01	.36*	.27	.26	.22
4. Standard 4	.61*	.81*	.88*		.12	.32*	.24	-.00	.43*	.25	.28*	.24
5. Dropout 1	.66*	.56*	.51*	.47*		.79*	.63*	.55*	.54*	.22	.49*	.48*
6. Dropout 2	.67*	.59*	.58*	.56*	.88*		.79*	.64*	.57*	.29*	.49*	.42*
7. Dropout 3	.61*	.57*	.56*	.59*	.86*	.90*		.80*	.66*	.29*	.55*	.45*
8. Dropout 4	.62*	.58*	.53*	.59*	.79*	.84*	.90*		.56*	.29*	.55*	.41*
9. „pure-study“	.60*	.68*	.69*	.69*	.58*	.57*	.69*	.58*		.40*	.57*	.61*
10. Standard 1W	.65*	.67*	.66*	.63*	.40*	.45*	.44*	.47*	.48*		.57*	.61*
11. Dropout 1W	.51*	.47*	.43*	.43*	.57*	.61*	.66*	.63*	.46*	.68*		.63*
12. „pure-study“ 1W	.37*	.39*	.43*	.33*	.34*	.43*	.41*	.33*	.59*	.54*	.67*	

*Anmerkungen.* Oberhalb der Diagonale werden die Korrelationen für die jüngere Stichprobe und unterhalb der Diagonale die Korrelationen der älteren Stichprobe dargestellt. 1W = Abruf nach einer Woche.

In Tabelle 25 ist die mittlere Erinnerungsleistung für die drei verwendeten Geschichten dargestellt. Selbst wenn sich Geschichte 1 signifikant von Geschichte 2 und 3 unterscheidet, so kommt dieser Unterschied vor allem durch die Unterschiede in der älteren Stichprobe zustande (in der jüngeren gibt es keine signifikanten Unterschiede zwischen den drei Lerndesigns). Die durchschnittlichen Erinnerungswerte über alle Personen und Messungen hinweg unterscheiden sich nicht statistisch signifikant über die sechs Designs ( $F(5, 1211) = 2.10, p = .06$ ) (siehe Tabelle 26).<sup>26</sup>

Tabelle 25

*Experiment 4: Lernleistung nach Geschichte und Alter aufgeteilt*

	Geschichte			Geschichte			Geschichte			Alle
	Jung			Alt			Gesamt			
	G1	G2	G3	G1	G2	G3	G1	G2	G3	
<i>N</i>	198	201	198	211	205	204	409	406	402	1217
<i>M</i>	74.26	71.16	72.60	63.49	51.30	56.28	68.70	61.13	64.32	64.73
<i>SD</i>	22.40	21.06	21.14	24.68	23.81	24.64	24.18	24.57	24.37	24.55

*Anmerkungen.* G1 = Geschichte B aus dem WMS-IV; G2 = Geschichte C aus dem WMS-IV; G3 = Geschichte B aus dem WMS-R.

<sup>26</sup> Bei der jüngeren Stichprobe unterscheidet sich Design 2 ( $M_2 = 68.28, SD_2 = 23.10$ ) von Design 4 und 5 ( $M_4 = 75.65, SD_4 = 21.40; M_5 = 75.64, SD_5 = 21.74$ ). Jedoch unterscheidet sich für die ältere Teilstichprobe Design 6 ( $M_6 = 66.32, SD_6 = 25.30$ ) signifikant von Design 1, 2, 3, 4 und 5 ( $M_1 = 52.34, SD_1 = 22.09; M_2 = 56.28, SD_2 = 23.64; M_3 = 54.21, SD_3 = 24.47; M_4 = 56.41, SD_4 = 24.77; M_5 = 56.25, SD_5 = 26.89$ ).

Tabelle 26

*Experiment 4: Lernleistung in den Designs*

	Design						Alle
	1	2	3	4	5	6	
<i>N</i>	204	213	200	192	204	204	1217
<i>M</i>	63.01	61.52	63.48	66.03	66.51	68.01	64.73
<i>SD</i>	22.67	24.10	25.29	25.02	26.11	23.67	24.55

*Mehrebenenmodelle*

Die Leistungsvarianz aller Abfragedurchgänge während des Experiments und beim verzögerten Abrufversuch gehen zu 25% auf die Unterschiede zwischen Personen zurück ( $ICC = .25, p < .01$ ). Somit zeigt sich bei diesem Experiment, dass bei den Analysen die Abhängigkeit der Daten beachtet werden sollte, und zwar in dem Sinne, dass Effekte zwischen und innerhalb von Personen betrachtet werden. In einem ersten Schritt wird die Lernrate–und damit die Effekte der Lern-Test-Durchgängen–genauer betrachtet (siehe Tabelle 27 und 28), um in einem zweiten Schritt die Effekte der drei Lernbedingungen auf die unmittelbare und die langfristige Lernleistung zu betrachten. Für die Beschreibung der Lernrate wird der Effekt des Lern-Test-Durchgangs einmal als logarithmischer Effekt und einmal als quadratischer Effekte modelliert (Daten zunächst ohne den verzögerten Abruf). Aufgrund der methodischen Unterschiede der „pure-study“-Bedingung und den beiden Lernbedingungen (Standard- und Dropout-Bedingung gemeinsam) mit mehreren Testdurchgängen werden unterschiedliche Fehlervarianzen für diese geschätzt. Als zufällige Effekte werden je die Effekte der Standard- und die der Dropout-Bedingung, sowie der Einfluss des Lern-Test-Durchgangs innerhalb der Standard- und der Dropout-Bedingung modelliert (getrennt für die beiden Lernbedingungen). Durch die Aufteilung der zufälligen

Effekte nach Lernbedingung wird berücksichtigt, dass die Standard- und die Dropout-Bedingung aufgrund ihrer spezifischen Besonderheiten bei der Enkodierung des Lernmaterials zudem in der Lernleistung unterscheiden können.

Damit die „pure-study“-Bedingung gleichzeitig mit den anderen beiden Lernbedingungen modelliert und die Effekte weiterhin intuitiv interpretiert werden können, nimmt der Prädiktor Durchgang  $t$  für den vierten Abfragedurchgang in der Standard- und Dropout-Bedingung sowie für den Testdurchgang der „pure-study“ den Wert 0 an (dementsprechend nimmt  $t$  für den ersten Lern-Test-Durchgang den Wert  $-3$ , für den zweiten den Wert  $-2$  und für den dritten den Wert  $-1$  an). Somit stellen die festen Effekte der drei Lernbedingungen jeweils die durchschnittliche Lernleistung am Ende des Experiments dar (vor dem verzögerten Abruf).

In Modell 1 wird für die Lernrate ein logarithmischer Effekt modelliert ( $\ln(t) = -1.39, -0.69, -0.29, 0$ ) (siehe Tabelle 27). In der Standard- sowie in der Dropout-Bedingung rufen die Probanden im Durchschnitt mit jedem Lern-Test-Durchgang mehr Textpropositionen ab, wobei sich die Lernraten der beiden Lernbedingungen nicht statistisch signifikant unterscheiden. Zusätzlich zeigen sich interindividuelle Unterschiede in der Effektivität der Lern-Test-Durchgänge. Dabei ist die Varianz in der Standardbedingung größer als in der Dropout-Bedingung, was auf höhere Abweichungen der individuellen Lernraten vom durchschnittlichen Effekt deutet. Die Korrelation der beiden zufälligen Effekte der Lernraten zeigt, dass Lerner, welche in der Standardbedingung schneller als der Durchschnitt lernten, dies wahrscheinlicher in der Dropout-Bedingung taten. Die Passung des Modells beträgt, gemäß dem AIC, 6842.6 Punkte und ist damit 242.5 Punkte besser als ein Modell mit einem linear geschätzten Effekt des Durchgangs, was an dem typischerweise abflachenden Verlauf der Lernkurven liegt (wie schon in Experiment 1 und 3 gezeigt).

Wird Modell 1 um die Effekte der Altersgruppe erweitert, verbessert sich der Modellfit, gemäß dem AIC, um 66.9 Punkte (Modell 3, siehe Tabelle 27 und Abbildung 11). Ältere Lerner erinnerten sich im Testdurchgang der „pure-study“-Bedingung durchschnittlich an fünf bis sechs Texteinheiten weniger als die jüngere Vergleichsstichprobe ( $t(507) = -6.05, p < .05, \alpha_{0PS} = -17.70, d = -0.54$ ). Im Gegensatz zu den jüngeren Probanden unterscheiden sich die Lernleistungen älterer Probanden im letzten Testdurchgang der Standardbedingung von der Dropout-Bedingung ( $t(507) = -2.62, p < .05, \alpha_{0DO-alt} = -4.69, d = -0.23$ ).<sup>27</sup> Jedoch unterscheiden sich in beiden Altersgruppen nicht die Lernraten der beiden Lernbedingungen. Demnach lernen beide Altersgruppen in den Lernbedingungen ähnlich viele Textpropositionen mit jedem Lern-Test-Durchgang dazu. Durch die Berücksichtigung der Altersgruppe verringert sich, im Gegensatz zu Modell 1, vor allem die Varianz interindividueller Unterschiede der Standard- und Dropout-Bedingung (da sich die Altersgruppen in den Erinnerungsleistungen der verschiedenen Testdurchgänge stark unterscheiden, siehe Tabelle 28).

Als Alternative für den logarithmischen Verlauf wird in Modell 2 ein linearer gemeinsam mit einem quadratischen Verlaufsprädiktor modelliert ( $t^2 = -9, -8, -5, 0$ ), wodurch sich der das AIC-Gütekriterium gegenüber dem Modell 1 um 2.7 Punkte verringert (siehe Tabelle 27). Im Vergleich zu Modell 1 unterscheiden sich die Lernraten der Dropout-Bedingung von denen der Standardbedingung im linearen sowie im quadratischen Effekt signifikant (ST\*t:  $t(506) = 25.77, d = 2.29$ ; DO\*t:  $t(506) = -4.27, d = -0.38$ ; ST\*t<sup>2</sup>:  $t(506) = -14.48, d = 1.29$ ; DO\*t<sup>2</sup>:  $t(506) = 4.18, d = 0.37$ ). Dabei flacht der Verlauf in der Dropout-Bedingung nicht so schnell ab wie in der Standardbedingung (der quadratische Effekt der Dropout-Bedingung ist von Null statistisch signifikant verschieden). Im Gegensatz zu

---

<sup>27</sup> Bei dieser Art der Kodierung wird der Unterschied der Standard- und Dropout-Bedingung ersichtlich, wenn die höhere Altersgruppe als Referenzgruppe verwendet wird.

Modell 1 verringern sich die Varianzen des zufälligen Effekts des Durchgangs für die Standard- und Dropout-Bedingung, welche aus Gründen der Vergleichbarkeit und Sparsamkeit nur für den linearen Effekt geschätzt werden.<sup>28</sup>

Ergänzt man nun Modell 2 um die Effekte der Altersgruppe so verbessert sich der Modellfit, gemäß dem AIC, um 73.0 Punkte (Modell4, siehe Tabelle 27, siehe Abbildung 11). Der Altersunterschied in der Abrufleistung der „pure-study“-Bedingung ist vergleichbar mit Modell 3. Konsistent zu Modell 2 zeigen sich hier bei beiden Altersgruppen unterschiedliche Lernraten für die Standard- und Dropout-Bedingung (höhere Altersgruppe: DO\*t:  $t(503) = -2.22$ ,  $\alpha_{1DO-alt} = -4.11$ ;  $p < .05$ ,  $d = -0.20$ ; DO\*t<sup>2</sup>:  $t(503) = 2.05$ ,  $\alpha_{2DO-alt} = 1.17$ ;  $p < .05$ ,  $d = 0.10$ ). Hierbei unterscheiden sich die Lernraten der Standardbedingung (linearer und quadratischer Effekt) der jüngeren und älteren Individuen. In Abbildung 11 werden Modell 3 und Modell 4 gegenübergestellt, wobei vor allem bei der Standardlernkurve der jüngeren Erwachsenen der Unterschied im quadratischen Verlauf gegenüber der Dropout-Lernkurve erkennbar ist.

Zwar weisen die Modelle mit quadratisch modellierter Lernrate gemäß dem AIC jeweils einen etwas besseren Fit auf als die dementsprechenden Modelle mit logarithmischem Effekt, jedoch sind aufgrund der zusätzlichen Parameter die Werte für das BIC der Modelle etwas größer (siehe Tabelle 27). Im Folgenden werden die Lernbedingungen für alle Abrufdurchgänge (inkl. verzögerter Abruf) mit dem logarithmisch modellierten Effekt durchgeführt, da dieses Modell 1) sparsamer ist und 2) sich im Gegensatz zum quadratischen Effekt mit weiteren Verläufen theoretisch sinnvolle Vorhersagen treffen lassen (da es im quadratischen Modell möglich ist, mit weiteren Lern-Test-Durchgängen Wörter zu vergessen anstatt dazuzulernen).

---

<sup>28</sup> Der AIC von Modell 2 verbessert sich nochmal um 35.0 Punkte, wenn zufällige Effekte für jeweils den quadratischen Effekt der Standard- und für die Dropout-Bedingung modelliert werden ( $\sigma_{ST*t^2}^2 = 3.49$ ,  $p < .05$ ;  $\sigma_{DO*t^2}^2 = 2.68$ ,  $p < .05$ ). Der BIC verringert sich hierbei um 6.0 Punkte.

Tabelle 27

*Experiment 4 Modelle 1 bis 4: Schätzungen der festen Effekten und Fit-Indizes*

Parameter	Modell 1 <sup>a</sup>	Modell 2	Modell 3 <sup>a</sup>	Modell 4
	Logarithmisch	Quadratisch	Logarithmisch Altersgruppe	Quadratisch Altersgruppe
<b>Feste Effekte</b>				
$\beta_{0PS}$ : „pure-study“ (PS)	74.54*	74.54*	83.56*	83.56*
$\beta_{0ST}$ : Standard (ST)	15.08*	13.12*	13.98*	11.07*
$\beta_{0DO}$ : Dropout (DO)	-3.88*	-1.89	-3.03	-0.34
$\beta_{1PS}$ : ST*t	28.28*	25.32*	29.46*	28.30*
$\beta_{1DO}$ : DO*t	-1.79	-5.69*	-1.80	-7.28*
$\beta_{2ST}$ : ST*t <sup>2</sup>		-4.22*		-5.08*
$\beta_{2DO}$ : DO*t <sup>2</sup>		1.72*		2.28*
$\alpha_{0PS}$ : Altersgruppe*PS			-17.70*	-17.70*
$\alpha_{0ST}$ : Altersgruppe*ST			2.16	4.02
$\alpha_{0DO}$ : Altersgruppe*DO			-1.66	-3.04
$\alpha_{1ST}$ : Altersgruppe*ST*t			-2.33	-5.91*
$\alpha_{1DO}$ : Altersgruppe*DO*t			0.03	3.16
$\alpha_{2ST}$ : Altersgruppe*ST*t <sup>2</sup>				1.70*
$\alpha_{2DO}$ : Altersgruppe*DO*t <sup>2</sup>				-1.11
<b>Informationskriterium</b>				
-2 LL	6842.6	6835.9	6765.7	6748.9
AIC	6876.6	6873.9	6809.7	6800.9
BIC	6921.2	6923.7	6867.4	6869.1
R <sup>2</sup> Level 1	17.89%	18.04%	37.61%	37.76%
R <sup>2</sup> Level 2	10.07%	10.11%	35.72%	35.74%

*Anmerkungen.* -2LL =  $-2 \times \text{Log-Likelihood}$ , AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion.

<sup>a</sup> hier wurde statt dem linearen Durchgang  $t(-3, -2, -1, 0)$  der logarithmierte Durchgang  $\ln(t)$   $(-1.39, -0.69, -0.29, 0)$  verwendet.

\* $p < .05$

Tabelle 28

*Experiment 4 Modelle 1 bis 4: Schätzungen der Residualvarianzen und zufälligen Effekte*

Parameter	Modell 1 <sup>a</sup>	Modell 2	Modell 3 <sup>a</sup>	Modell 4
	Logarithmisch	Quadratisch	Logarithmisch Altersgruppe	Quadratisch Altersgruppe
<b>Residualvarianzen</b>				
$\sigma_{e_{PS}}^2$ : Var(Residuen) „pure-study“	296.30*	296.30*	218.03*	218.03*
$\sigma_{e_{Test}}^2$ : Var(Residuen) ST & DO	34.11*	34.44*	34.11*	33.65*
<b>Zufällige Effekte</b>				
$\sigma_{b_0}^2$ : Var(ST)	182.13*	173.09*	121.85*	114.62*
$\sigma_{b_1}^2$ : Var(DO)	241.23*	234.18*	167.39*	159.93*
$\sigma_{d_0}^2$ : Var(ST*t)	66.55*	13.27*	65.23*	13.26*
$\sigma_{d_0}^2$ : Var(DO*t)	37.64*	7.86*	36.33*	7.78*
Korr. ST x DO	.70*	.71*	.57*	.56*
Korr. ST x ST*t	.11	.08	.03	.03
Korr. ST x DO*t	.16	.14	.06	.05
Korr. DO x ST*t	-.11	-.12	-.23	-.23
Korr. DO x DO*t	.06	.03	-.05	-.07
Korr. ST*t x DO*t	.46*	.48*	.44*	.46*

<sup>a</sup> hier wurde statt dem linearen Durchgang t (-3, -2, -1, 0) der logarithmierte Durchgang ln(t) (-1.39, -0.69, -0.29, 0) verwendet.

\* $p < .05$

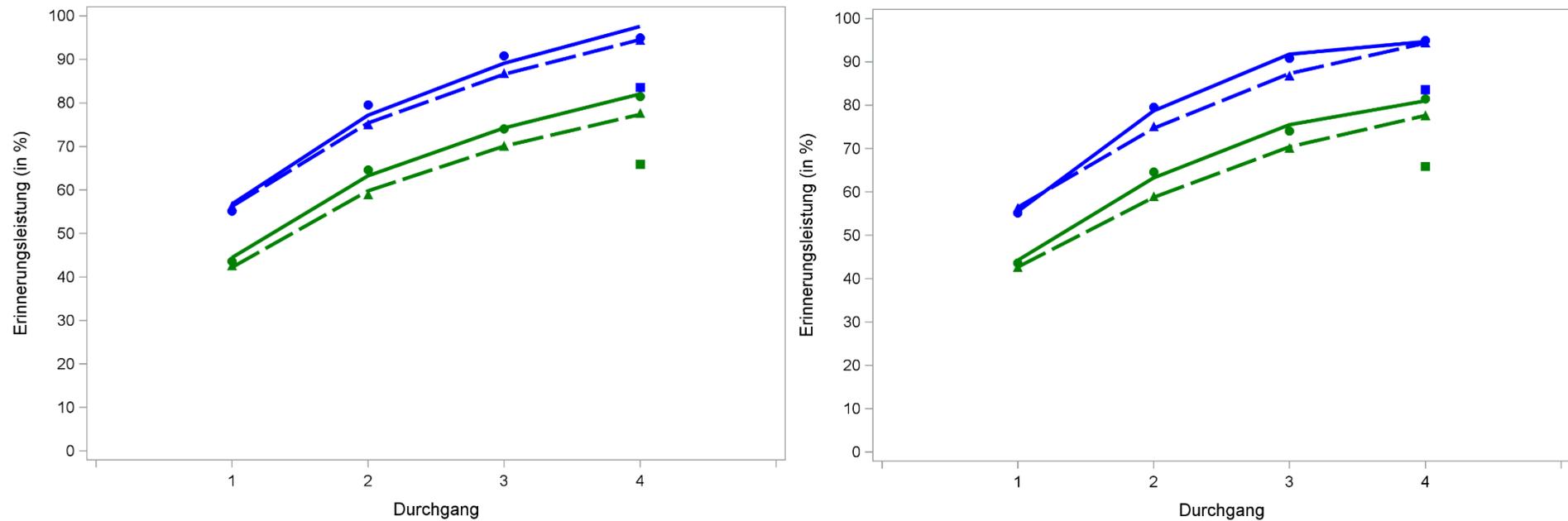


Abbildung 12. Experiment 4: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) für Modell 3 (links) und Modell 4 (rechts). Die durchgehenden Linien repräsentieren die Standardlernbedingung, die gestrichelten Linien die Dropout-Bedingung. Die Kreise zeigen die beobachteten Werte der Standardbedingung, die Dreiecke die beobachteten Werte Dropout-Bedingung dar und die Quadrate die beobachteten Werte der „pure-study“-Lernbedingung in den beiden Altersgruppen.

In den Modellen 5 und 6 wird der vorhergesagte Bereich der Abrufleistung auf die verzögerten Testdurchgänge nach einer Woche erweitert (siehe Tabelle 29 und 30) (siehe Abbildung 13). Wie in Studie 3 werden hierfür zusätzliche dummy-codierte Prädiktoren für den verzögerten Abruf und für Altersunterschiede aufgenommen (vgl. Formel 3.1 im Abschnitt Methoden der Studie 3). Modell 5 klärt 58.22% der Varianz innerhalb von Personen und 28.29% der Varianz zwischen Personen auf. Im Durchschnitt vergessen jüngere Erwachsene 41% des Lernmaterials im Laufe einer Woche ( $t(801) = -18.94, p < .05, \beta_{0PS2} = -40.99, d = -1.34$ ). Jedoch zeigen sich darüber hinaus keine Altersunterschiede in den Vergessensraten der drei Lernbedingungen bei der jüngeren Vergleichsstichprobe. Allerdings liegt ein statistisch signifikanter Altersunterschied beim verzögerten Abruf der Dropout-Bedingung vor. Ältere Erwachsene vergessen in der Dropout-Bedingung durchschnittlich zwei bis drei Textpropositionen weniger als jüngere ( $t(801) = 2.89, p < .05, \alpha_{DO2} = 8.32, d = 0.20$ ). Die individuellen Abweichungen von den geschätzten Werten der drei Lernbedingungen korrelieren hoch untereinander. D.h. Personen, die hoch vom geschätzten Wert einer Lernbedingung abweichen, werden wahrscheinlicher zudem höhere Abweichungen in den beiden anderen Bedingungen zeigen.<sup>29</sup> Berücksichtigt mögliche Effekte des Designs<sup>30</sup>—modelliert als zufälliger Effekt—so verändern sich zwar nicht die festen Effekte im Wesentlichen, aber die zufälligen Effekte und deren Korrelationen (siehe Tabelle 31). Der Modellfit verbessert sich, gemäß dem AIC, um 53.6 Punkte. Demnach ist das Design eine Quelle der Fehlervarianz, welche durch den zufälligen Effekt jedoch berücksichtigt wird.

---

<sup>29</sup> Kodiert man beim verzögerten Erinnern die nicht-erinnerten Geschichten als Missings, so verändern sich die festen und zufällige Effekte geringfügig, nicht aber die statistische Signifikanz der geschätzten Effekte.

<sup>30</sup> Die Variable für den zufälligen Effekt des Designs besitzt 18 Stufen, wovon jeweils eine Stufe eine verwendete Kombination aus der verwendeten Geschichte, deren Position im Experiment und Lernbedingung darstellt (= 6 Designs mit je drei Lernbedingungen/Geschichten/Positionen, siehe Tabelle 3.2).

Dadurch, dass jüngere Probanden durchschnittliche mehr Textpropositionen erinnern als ältere, könnten v.a. bei der jüngeren Erwachsenen vermehrt Deckeneffekte auftreten. Um dies zu berücksichtigen werden die Residualvarianzen für Modell 5 in einem weiteren Schritt für die beiden Altersgruppen modelliert (jeweils eine geschätzte Residualvarianz für die Lernbedingung mit und ohne Testbedingung für die beiden Altersgruppen; siehe Tabelle 31). Die Varianz der älteren Erwachsenen bei der „pure-study“-Lernbedingung ist die größte der vier Residualvarianzen. Demnach fällt bei diesem Modell vor allem die Fehlervarianz bei dieser Gruppe am höchsten aus.

In einem weiteren Modell wurde zusätzlich die Effekte der Verarbeitungsgeschwindigkeit (VG) mit aufgenommen (Modell 6), wodurch sich der AIC um 28.4 Punkte im Vergleich zu Modell 5 verbesserte und gleichzeitig 4.25% mehr Varianz innerhalb und 12.03% mehr Varianz zwischen Individuen aufklärt. Insgesamt gibt es zwischen den beiden Altersgruppen sehr große Unterschiede in der Verarbeitungsgeschwindigkeit ( $t(100) = 12.41, p < .05$ , Differenz = 1.55,  $d = 2.46$ ). Die durchschnittliche Leistung bei der verbalen Lernaufgabe–unabhängig von der Lernbedingung–hängt bei der älteren Stichprobe im statistisch signifikanten Maße von der Verarbeitungsgeschwindigkeit ab ( $t(798) = 3.24, p < .05$ ,  $\beta_{\text{Speed}} = 10.91, d = 0.23$ ). Das bedeutet, ältere Personen, welche eine Standardabweichung über der durchschnittlichen Verarbeitungsgeschwindigkeit liegen, erinnern im Mittel drei Wörteinheiten mehr. Dies zeigt sich zudem im schwindenden Einfluss der Altersgruppe auf die allgemeine Lernleistung (von Modell 5 auf Modell 6). Zusätzlich hängt die Verarbeitungsgeschwindigkeit bei der älteren Stichprobe mit dem verzögerten Abruf zusammen ab ( $t(798) = -7.47, p < .05$ ,  $\beta_{\text{Speed}} = -3.61, d = 0.26$ ).<sup>31</sup> Ältere Erwachsene mit hoher Verarbeitungsgeschwindigkeit erinnern zwar im

---

<sup>31</sup> Die Interaktion von Verarbeitungsgeschwindigkeit und Altersgruppe sowie die Interaktion von Verarbeitungsgeschwindigkeit, Altersgruppe und verzögerten Abruf bleiben jeweils statistisch signifikant von Null verschieden, wenn man die ältere Altersgruppe als Referenzgruppe verwendet.

letzten Durchgang des Experiments mehr Texteinheiten, vergessen jedoch auch nach einer Woche mehr als die jüngeren.

Tabelle 29

*Experiment 4 Modelle 5 und 6: Schätzungen der festen Effekten und Fit-Indizes*

Parameter	Modell 5 + verzögert	Modell 6 + VG
<b>Feste Effekte</b>		
$\beta_{0PS}$ : „pure-study“ (PS)	83.56*	82.67*
$\beta_{0ST}$ : Standard (ST)	13.98*	13.98*
$\beta_{0DO}$ : Dropout (DO)	-3.03	-3.03
$\beta_{1ST}$ : ST*ln(t)	29.46*	30.49*
$\beta_{1DO}$ : DO*ln(t)	-1.80	-1.80
$\alpha_{0PS}$ : Altersgruppe*PS	-17.70*	-7.62
$\alpha_{0ST}$ : Altersgruppe*ST	2.15	2.15
$\alpha_{0DO}$ : Altersgruppe*DO	-1.65	-1.65
$\alpha_{1ST}$ : Altersgruppe*ln(t)*ST	-2.36	-3.25
$\alpha_{1DO}$ : Altersgruppe*ln(t)*DO	0.07	0.07
$\beta_{0PS2}$ : PS*1Woche	-40.99*	-42.59*
$\beta_{0ST2}$ : ST*1Woche	1.08	1.08
$\beta_{0DO2}$ : DO*1Woche	-0.88	-0.87
$\alpha_{0PS2}$ : Altersgruppe*PS*1Woche	-1.92	-4.42
$\alpha_{0ST2}$ : Altersgruppe*ST*1Woche	-1.29	-1.30
$\alpha_{0DO2}$ : Altersgruppe*DO*1Woche	8.32*	8.32*
VG		1.13
VG*Altersgruppe		10.91*
VG*ln(t)		-1.31
VG*ln(t)*Altersgruppe		1.50
VG*1Woche		1.98
VG*1Woche*Altersgruppe		-7.47*
<b>Informationskriterium</b>		
-2 LL	9269.6	9229.2
AIC	9325.6	9297.2
BIC	9399.1	9386.4
R <sup>2</sup> Level 1	58.22%	62.47%
R <sup>2</sup> Level 2	28.29%	40.32%

\* $p < .05$

Tabelle 30

*Experiment 4 Modelle 5 und 6: Schätzungen der Residualvarianzen und zufälligen Effekte*

<b>Residualvarianzen</b>	Modell 5 mit verzögertem Abruf	Modell 6 + Speed
<b>Within</b>		
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“	115.52*	109.99*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO	65.90*	64.76*
<b>Zufällige Effekte</b>		
$\sigma_{b_0}^2$ : Var(ST)	131.01*	102.48*
$\sigma_{b_1}^2$ : Var(DO)	156.85*	143.44*
$\sigma_{d_0}^2$ : Var(PS)	161.66*	147.02*
$\sigma_{d_0}^2$ : Var(ln(t))	30.47*	30.37*
Korr. ST x DO	.66*	.61*
Korr. ST x PS	.76*	.71*
Korr. ST x ln(t)	.05	.12
Korr. DO x PS	.81*	.78*
Korr. DO x ln(t)	-.10	-.06
Korr. PS x ln(t)	.00	.04

\* $p < .05$

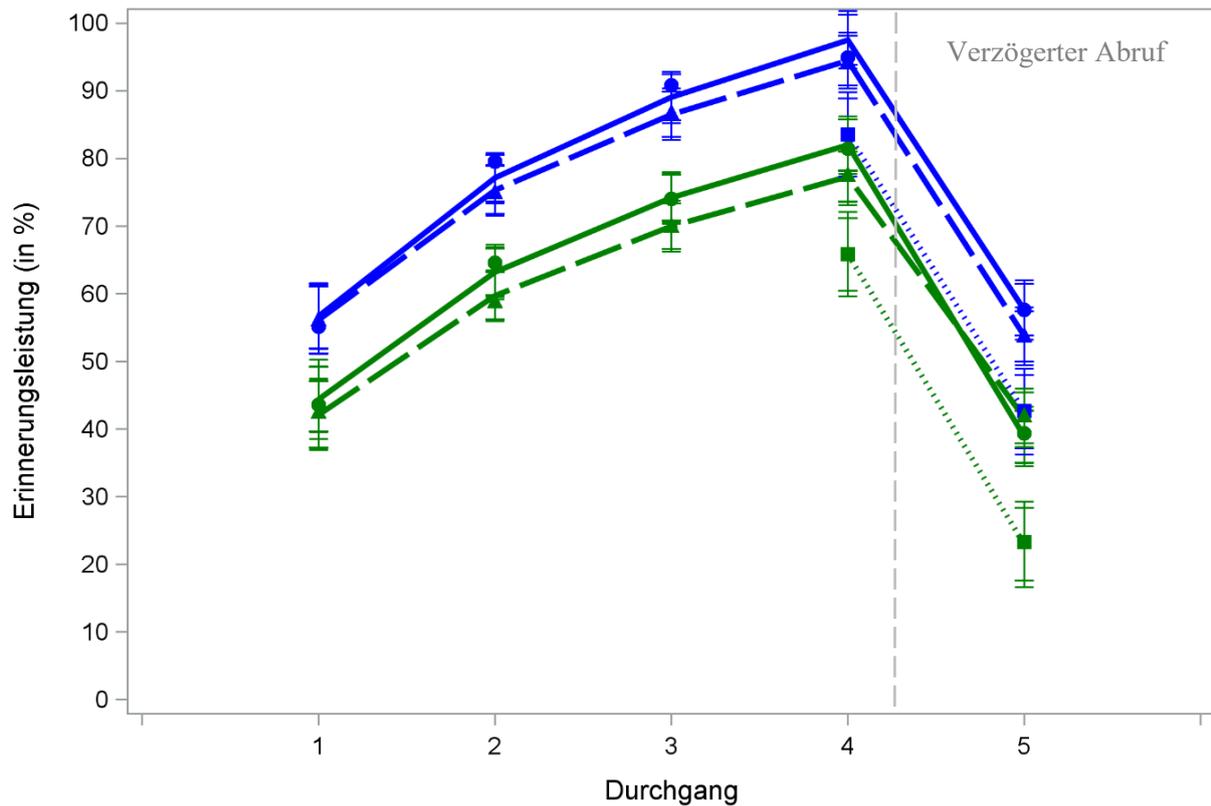


Abbildung 13. Experiment 4: Mittlere Anzahl erinnerter Texteinheiten für jüngere Erwachsene (blau) und ältere Erwachsene (grün) für Modell 5. Die durchgehenden Linien repräsentieren die Standardlernbedingung, die gestrichelten Linien die Dropout-Bedingung und die gepunktete Linie die „pure-study“-Bedingung. Die Kreise zeigen die beobachteten Werte der Standardbedingung, die Dreiecke die beobachteten Werte Dropout-Bedingung dar und die Quadrate die beobachteten Werte der „pure-study“-Lernbedingung in den beiden Altersgruppen. Die Fehlerbalken stellen die Standardfehler dar.

Tabelle 31

*Experiment 4 Zusatzmodelle: Schätzungen der Residualvarianzen, zufälligen Effekte und Fit-Indizes*

Parameter	Modell 5 mit Design-Effekt	Modell 5 mit Residualvarianzen für Testungsart und Altersgruppe
<b>Residualvarianzen</b>		
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“	99.06*	
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO	51.53*	
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“-jung		77.55*
$\sigma_{ePS}^2$ : Var(Residuen) – „pure-study“-alt		163.82*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO-jung		62.92*
$\sigma_{eTest}^2$ : Var(Residuen) – ST & DO-alt		68.98*
$\sigma_{Design}^2$ : Var (Design)	0.30*	
<b>Zufällige Effekte</b>		
$\sigma_{b_0}^2$ : Var(ST)	124.06*	129.90*
$\sigma_{b_1}^2$ : Var(DO)	140.37*	155.82*
$\sigma_{d_0}^2$ : Var(PS)	156.10*	163.59*
$\sigma_{d_0}^2$ : Var(ln(t))	30.62*	30.43*
Korr. ST x DO	.78*	.66*
Korr. ST x PS	.73*	.78*
Korr. ST x ln(t)	.03	.04
Korr. DO x PS	.89*	.85*
Korr. DO x ln(t)	-.14	-.10
Korr. PS x ln(t)	-.09	-.02
<b>Informationskriterium</b>		
-2 LL	9214.0	9260.2
AIC	9272.0	9320.2
BIC	9214.0	9399.0

\* $p < .05$

### 3.3.3. Zusammenfassung der Ergebnisse aus Experiment 4

In allen Modellen zeigt sich ein unmittelbarer Testeffekt in dem Sinne, dass in den Lernbedingungen mit mehreren Testdurchgängen ungefähr vier Texteinheiten mehr gelernt wurden als in der „pure-study“-Bedingung. Obgleich ältere Erwachsene durchschnittlich weniger Textpropositionen lernten, so profitierten beide Altersgruppen in vergleichbarem Umfang von den Effekten des Testens–bezogen auf den unmittelbaren Effekt am Ende des Experiments. Bezüglich dem verzögerten Abruf nach einer Woche erinnerten die Probanden deutlich weniger Texteinheiten als im letzten Testdurchgang während des Experiments (ungefähr elf bis zwölf Einheiten weniger). Obgleich bei beiden Altersgruppen im Durchschnitt die meisten Wörter in den Testbedingungen (Standard- und Dropout-Bedingung) erinnert wurden, so zeigte sich kein signifikanter Vorteil der Standardbedingung gegenüber der „pure-study“-Bedingung in der Vergessensrate (jüngere Stichprobe:  $d = 0.03$ ; für die ältere Stichprobe:  $d = -0.01$ ). Demnach haben Ältere und Jüngere ähnlich viele Propositionen in der Bedingung mit wiederholten Lern-Test-Durchgängen, wie in der Bedingung ohne dazwischenliegende Testungen vergessen. Eine Sonderrolle nimmt die Dropout-Bedingung ein. Hier zeigt sich in der älteren Altersgruppe ein direkter Testeffekt in den absoluten Werten des Vergessens. Ältere vergessen im Laufe einer Woche ungefähr zwei Texteinheiten weniger in der Dropout-Bedingung als jüngere Erwachsene. Dieses Ergebnis ist–bezogen auf die Literatur zum Testeffekt–untypisch, da der Effekt nicht in der Standardbedingung und nur bei älteren Lernern zu beobachten ist. Wie schon in Experiment 3 beschrieben, kommt der Dropout-Bedingung aufgrund ihres im Alltag untypischen Abfrageformats eine gewisse Sonderrolle zu. Diese Lernbedingung erfordert (aufgrund des ungewohnten Formats) besonders viel Aufmerksamkeit, was bei älteren Lernern aufgrund geringerer Aufmerksamkeitsressourcen schlechter kompensiert werden könnte ( Craik & Byrd, 1982). Durch die Einzigartigkeit und Schwierigkeit des Durchgangs könnten Effekte der

wünschenswerten Schwierigkeit auftreten (siehe Kapitel 1.1.2.1.), welcher den Testeffekt in dieser Konstellation besonders fördert (Bjork 1994; Bjork & Bjork, 2011).

Wie in Experiment 3, liegt die größte Fehlervarianz der „pure-study“-Bedingung bei den älteren Erwachsenen vor (siehe Tabelle 31). Diese höhere Fehlervarianz könnte auf größere Unterschiede zwischen Lernern im Umgang mit der Lernbedingung und dem unterschiedlichen Bildungsniveau der Altersgruppen zurückgehen (siehe auch Kapitel 4.2.1.).

Ältere Erwachsene mit hoher Leistungsfähigkeit in der Verarbeitungsgeschwindigkeit erinnern zwar im letzten Durchgang des Experiments mehr Texteinheiten, vergessen jedoch nach einer Woche mehr Inhalte als die jüngeren (jedoch durchschnittlich nicht mehr, als sie dazugelernt hatten, siehe Tabelle 30). Aufgrund der großen Altersunterschiede könnte es sein, dass eine geringere Verarbeitungsgeschwindigkeit nur bis zu einem gewissen Grenzwert einen Einfluss auf die Lernleistung hat, und daher bei den jüngeren anhand der vorliegenden Daten nicht sichtbar ist (beide Altersgruppen unterscheiden sich stark in der Verarbeitungsgeschwindigkeit: Differenz = 1.55,  $d = 2.46$ ;  $M_{VGjung} = 0.79$ ;  $SD_{VGjung} = .59$ ;  $M_{VGalt} = -0.76$ ;  $SD_{VGalt} = .66$ ).

Schließlich klärt darüber hinaus die Lerndesignvariable Fehlervarianz auf (entspricht der Kombination aus Lernbedingung, Geschichte und deren Position im Experiment). Hierbei könnte der unterschiedliche Schwierigkeitsgrad der Geschichten einerseits die Lernleistung beeinflusst haben (siehe Tabelle 25) und andererseits Übungseffekte beim Lernen im Laufe des Experiments.

#### 4. Diskussion

Eines der Hauptziele der Arbeit ist es, Altersunterschiede in der Wirksamkeit des Testeffekts zu untersuchen. Die grundlegenden Annahmen dabei sind, dass nicht nur Prozesse beim Lernen, sondern zudem beim Abrufen von altersassoziierten kognitiven Veränderungen betroffen sind. Hierbei wird, im Einklang mit dem Ansatz des episodischen Kontextes zur Erklärung des Testeffekts (Karpicke et al., 2014), erwartet, dass durch altersbedingte Defizite beim Enkodieren und Abrufen von Assoziationen, sowie durch Defizite beim Erinnern von Kontextattributen, ältere Erwachsene weniger von Testdurchgängen profitieren als jüngere. Bisherige Untersuchungen zeigten aufgrund unterschiedlicher Operationalisierungen (bspw. mit oder ohne Feedback; nur einem Testdurchgang) uneindeutige Ergebnisse bezüglich Altersunterschiede im Testeffekt (bspw. Tse et al., 2010). In der vorliegenden Arbeit wird unter der Verwendung von Lernkurven und Designs mit unterschiedlicher Anzahl an Lern- und Test-Durchgängen gezeigt, dass sich 1) bei unmittelbarer Abfrage mit unterschiedlichem Lernmaterial–Wortlisten mit freiem Abruf, Wortpaarassoziationsaufgaben und kleinen Geschichten–Testeffekte zeigen und 2) abhängig vom Material Altersunterschiede im Testeffekt auftreten können. Letztere können so extrem sein, dass sich bei Wortpaarassoziationsaufgaben keine Testeffekte für ältere Lerner ergeben. Jedoch zeigten sich für Geschichten–welche als Lernmaterial kohärenter sind als standardisierte Wortlisten und Wortpaare–bei einem freien Abrufformat im Test zwar unmittelbare Testeffekte, jedoch konnten Testungen nicht das Vergessen verhindern. Die altersspezifischen Unterschiede im Testeffekt wurden im Rahmen dieser Arbeit mit vier Experimenten untersucht, welche sich in zwei Studienblöcke unterteilen: 1) Experiment 1 und 2 modellieren Lernkurven mit und ohne Effekte von Testdurchgängen und 2) in Experiment 3 und 4 werden die Fragestellungen mit kohärentem Lernmaterial in Form von kurzen Geschichten getestet. In der Diskussion werden zuerst die Ergebnisse, Implikationen und Limitationen der beiden Studienblöcke präsentiert, um anschließend die Ergebnisse im theoretischen Rahmen zu betrachten. Abschließend wird

Ausblick auf offene Themen bezüglich Eigenschaften des Lernenden und des Lerndesigns für effiziente Testeffekte eingegangen sowie ein Fazit präsentiert.

#### *4.1. Lernkurven beim freien und hinweisgesteuerten Abruf*

Das übergreifende Ziel dieser Arbeit ist es zu untersuchen, ob der Testeffekt sich zwischen jüngeren und älteren Erwachsenen unterscheidet. Um die Effekte von Testdurchgängen genauer zu verstehen, werden diese in mehreren Lern-Test-Durchgängen untersucht. In Experiment 1 und 2 werden hierfür die Lernkurven mit und ohne die Effekte des Testens verglichen. Um Lernkurven ohne den Einfluss des Testeffekts zu ermitteln, wird das Lerndesign von Roediger und Smith (2012) verwendet und aus mehreren Lernaufgaben mit unterschiedlich vielen Lerndurchgängen eine „pure-study“-Lernkurve erstellt. Diese wird anschließend mit der Standardlernkurve (mehrere Lern-Test-Durchgänge) verglichen, wobei die Unterschiede der beiden Lernkurven die Testeffekte repräsentieren. Vor dem Hintergrund der Lernkurve stellt sich die Frage, wie viel der Leistungssteigerung in solchen mehrfachen Lern-Test-Durchgängen auf das wiederholte Einprägen des zu lernenden Materials und wie viel der Leistungssteigerung auf das (wiederholte) Testen des Materials zurückzuführen ist. Aufgrund des im Alter abnehmenden episodischen Gedächtnisses und der Schwierigkeit, Assoziationen zu enkodieren und abzurufen (Naveh-Benjamin, 2000), werden größere Testeffekte in der Lernleistung bei jüngeren Lernern erwartet als bei älteren.

##### *4.1.1. Ergebnisse Studienblock 1*

Im Folgenden werden zunächst die zentralen Ergebnisse bezüglich der Lernkurven jüngerer und älterer Erwachsener genauer betrachtet, bevor anschließend die Unterschiede bezüglich des verwendeten Lernmaterials diskutiert werden. In Übereinstimmung mit Roediger und Smith (2012) zeigten jüngere Probanden in Experiment 1 und 2 einen mittleren bis starken Testeffekt, d.h. sie lernten mit jedem weiteren Lern-Test-Durchgang mehr Inhalte

aus dem Material, als in der Bedingung ohne wiederholtes Testen (d.h. im Vergleich zur Lernrate, hervorgerufen durch jeden weiteren Lerndurchgang). Hier zeigen sich die unmittelbaren Vorteile von Testdurchgängen auf den Lernerfolg, wobei jeder weitere Testdurchgang auch zur Verbesserung der Lernleistung beiträgt. Einhergehend mit den theoretischen Überlegungen zu den abnehmenden Testeffekten im Alter, aufgrund des Defizits Assoziationen zu bilden und des abnehmenden episodischen Gedächtnis–wodurch Enkodierung und Abrufprozesse der Kontextattribute erschwert werden–finden sich in beiden Experimenten Unterschiede im Testeffekt zwischen den Altersgruppen. Jedoch wurden diese Unterschiede in der Empirie bisher kaum gefunden (vgl. Coane, 2013; Meyer & Logan, 2013). Durch die Verwendung mehrerer Lern-Test-Durchgänge, ist es in den hier vorgestellten Experimenten möglich, dass sich die Testeffekte aufsummieren und dadurch mehr Wörter durch die Testdurchgänge erinnert werden. Dadurch können Lerner von mehreren Lern- und Abrufkontexten profitieren. Jedoch zeigten ältere Probanden im Durchschnitt keine Testeffekte im Experiment 2. Wie in Kapitel 2.4.3. ausgeführt, kann der ausbleibende Testeffekt aufgrund des starken assoziativen Charakters der Wortpaaraufgabe–wird gemäß der Assoziations-Defizit-Hypothese erwartet–sowie der hohen Schwierigkeit der Aufgabe zustande kommen. Aufgrund der niedrigen Abruftrate können nur wenige erfolgreiche Abrufprozesse von den Effekten des Testens profitieren, was mit dem Erklärungsansatz erfolgreich erinnelter Items übereinstimmt.

Für die Interpretation von Standardlernkurven bedeuten die Ergebnisse aus den Experimenten 1 und 2, dass (Standard-)Lernkurven nicht nur die Effekte des Enkodierens, sondern ebenfalls die Effekte von Testdurchgängen gemessen werden. Dies Effekte spielt eine wesentliche Rolle für die Lernleistung. In Experiment 1 unterscheidet sich in der „pure-study“-Lernkurve die Lernrate älterer Erwachsener nicht von der Lernrate der jüngeren Vergleichsstichprobe. Obgleich sich ältere im Durchschnitt an weniger Wörter erinnerten, lernten sie mit jedem Lerndurchgang ähnlich viele Wörter, wie die jüngere

Vergleichsstichprobe. In Experiment 2 zeigt sich dieses Muster nicht, jedoch zeigt auch hier der Testeffekt einen wesentlichen Einfluss auf die Lernrate jüngerer Erwachsener. Dies deutet an, dass Alterseffekte in der Lernrate in der Standardlernbedingung z.T. auch auf Altersunterschiede bei Effekten des Testens und nicht nur auf Enkodierungsprozesse zurückgehen können. Daraus ergibt sich die Empfehlung für zukünftige Forschung zu Altersunterschieden beim verbalen Lernen, die Effekte des Lernens/Enkodierens getrennt von den Effekten des Testens zu betrachten. Dies ermöglicht es, Aussagen über die Herkunft von Altersunterschieden beim Lernen zu treffen und darauf aufbauend gezielt Lernstrategien zu entwickeln. Diese Strategieempfehlungen können darüber hinaus altersspezifische Empfehlungen beinhalten; bspw.: Wenn große Altersunterschiede in den Testeffekten erwartet werden, könnte diesen durch zusätzliche Darbietung von Feedback entgegengewirkt werden (vgl. Tse et al., 2010).

Da bei der Erstellung der Standardlernkurve eine andere Methode verwendet wurde, als bei der Erstellung der „pure-study“-Lernkurve (bspw. besteht letztere aus mehreren Lernaufgaben), können sich die Fehlervarianzen zwischen den Methoden unterscheiden; Das heißt, dass die unterschiedlichen Messprozesse (hier in Form von Lernbedingung) eine Quelle zusätzlicher Varianz in den Daten sein können (Diggle et al., 2013). Hierbei zeigen sich in den Experimenten 1 und 2 jeweils statistisch relevante Unterschiede in den beiden Lernbedingungen. Es zeigen sich größere Unterschiede für die „pure-study“-Bedingung. Nachdem sich dieses Muster auch in Experiment 2 ergibt, bei dem es insgesamt nur eine Lernaufgabe gab, können die Unterschiede nicht allein auf individuelle Lernstrategien der Probanden bei den Aufgaben ohne dazwischenliegende Tests zurückgeführt werden (da in Experiment 2 die Items der Standardlernkurve und die der „pure-study“-Lernkurve im selben Lerndurchgang gelernt wurden). Eine mögliche Erklärung könnte in den Wechselwirkungen von Lern- und Testdurchgängen liegen. Jedoch wurde dieser Zusammenhang in der

vorliegenden Arbeit nicht direkt untersucht und stellt damit ein interessantes Feld zukünftiger Forschung zu Lernkurven dar.

Betrachtet man die zufälligen Effekte, so zeigen sich Unterschiede zwischen den Probanden in dem Ausgangswert, der Lernrate und der Lernbedingung. D.h., dass die Teilnehmer unterschiedlich viel Lernmaterial nach einem Durchgang gelernt haben und unterschiedlich stark von den Effekten des Durchgangs und der Lernbedingung profitierten. Die individuellen Unterschiede in der Lernkurve konnten auch in früheren Studien gefunden werden (Rast & Zimprich, 2009; Zimprich, Rast & Martin, 2008). Dass Probanden unterschiedlich von den Testeffekten profitieren, zeigt, dass es noch weitere Einflussfaktoren für die Effektivität von Testdurchgängen geben müsste (bspw. Organisation des Lernmaterials, siehe Kapitel 4.4.1.). Spannend hierbei ist es zu erforschen, welche Eigenschaften eines Lernalters zu besonders wirksamen Testeffekten führen (wie bspw. hoher initialer Abruf, siehe Karpicke, 2017). Die individuellen Abweichungen vom Ausgangswert hängen in beiden Experimenten im mittleren bis hohen Maße mit den Testeffekten zusammen. Dies unterstützt die Grundannahme des Testeffekts, dass dieser von den Prozessen des erfolgreichen Abrufs abhängt.

Zum Einfluss der Verarbeitungsgeschwindigkeit auf die Lernleistung in den Lernkurven zeigen die beiden Experimente unterschiedliche Ergebnisse. Einerseits erklärt bei Experiment 1 die Verarbeitungsgeschwindigkeit keine bedeutsamen individuellen Unterschiede in der Lernleistung über das Alter hinweg, andererseits wird durch die Berücksichtigung der Verarbeitungsgeschwindigkeit bei Experiment 2 6% mehr Varianz innerhalb und 9% mehr Varianz zwischen den Personen aufgeklärt. In Experiment 2 wirkt sich eine höhere Verarbeitungsgeschwindigkeit so aus, dass die Lernrate ansteigt (siehe Paarassoziationsaufgabe bei Kurtz et al., 2013). Experiment 2 unterscheidet sich im Lerndesign insofern stark von Experiment 1, dass bei den Testdurchgängen jeweils nur ein Teil des gelernten Materials abgefragt wurde (außer beim letzten Testdurchgang). Wie diese

Unterschiede des Einflusses der Verarbeitungsgeschwindigkeit zustande kommen, geht über den Fokus der Arbeit hinaus. Jedoch sollte zukünftig in längsschnittlichen Studien–im Gegensatz zum hier verwendeten Querschnittsansatz–nicht nur der Zusammenhang von Verarbeitungsgeschwindigkeit mit Lernleistung, sondern auch der Einfluss der Verarbeitungsgeschwindigkeit auf das kognitive Altern analysiert werden.

Zusammenfassend zu den Ergebnissen kann man sagen, dass ein wesentlicher Anteil der Leistungssteigerung in der Standardlernkurve in Experiment 1 bei beiden Altersgruppen und in Experiment 2 bei den jüngeren Erwachsenen auf die Effekte zusätzlicher Testdurchgänge zurückgeht. Testen erfüllt daher nicht nur den Zweck der Erfassung des Wissensstandes, sondern beeinflusst diesen auch positiv. Allerdings profitieren jüngere und ältere Probanden nicht im gleichen Maße von den Effekten des Testens, was sich in einem schwächeren Testeffekt (siehe Experiment 1) oder einem nicht vorhandenen Testeffekt (siehe Experiment 2) gezeigt hat. Hierbei kann darüber hinaus die Schwierigkeit sowie die Art des verwendeten Lernmaterials einen Einfluss auf die Wirkung des Testeffekts ausüben.

#### *4.1.2. Testeffekt beim freien Erinnern und hinweisgesteuerten Abruf*

Aufgaben mit freien und hinweisgesteuerten Abruf unterscheiden sich nicht nur in der Struktur des Abrufdurchgangs (frei vs. vorgegebene Struktur), sondern auch in den Prozessen des Abrufversuchs. Bei freien Abrufversuchen spielt das Organisieren des Lernmaterials eine entscheidende Rolle für den Abrufprozess–bspw. durch die Erstellung eines Abrufplans im Rahmen von subjektiver Organisation (Tulving, 1962; mehr dazu in Kapitel 4.4.1.). Hingegen spielt bei hinweisgesteuerten Abfrageformaten, wie bspw. Paarassoziationsaufgaben, die assoziative Verknüpfung zwischen den beiden Einheiten eine zentrale Rolle. Je stärker und/oder vielfältiger die Einheiten miteinander verknüpft sind, desto wahrscheinlicher ist der erfolgreiche Abruf (vgl. Theorie zur ausbreitenden Aktivierung semantischer Prozesse; Collins & Loftus, 1975). Im Rahmen des Ansatzes des episodischen Kontextes wird

angenommen, dass größere Testeffekte bei freiem Abruf im Gegensatz zu Paarassoziationsaufgaben entstehen, da beim freien Abruf keine Hinweisreize den Abruf initiieren (Karpicke, Lehman, & Aue, 2014). Je weniger Hinweisreize durch die Umgebung der Abfragesituation gegeben sind, desto wichtiger ist das Zurückversetzen in vorherige Lern- und Abrufkontexte. Bei Aufgaben mit Hinweisreizen ist der Testeffekt abhängig von den vorgegebenen Hinweisreizen. Je mehr diese beim Abruf helfen und dadurch folglich das Zurückversetzen in den vorherigen Kontext überflüssig machen, desto geringer fallen die Testeffekte aus—wenn bspw. statt einem Buchstaben des Zielwortes drei vorgegeben werden (siehe Carpenter & DeLosh, 2006). Dass der Testeffekt mit der Schwierigkeit des Abrufs zusammenhängt, passt auch zur Theorie der wünschenswerten Schwierigkeiten, wonach ein schwieriger und erfolgreicher Abruf mit einer tieferen Verarbeitung des Lerninhalts einhergeht (vgl. Bjork & Bjork, 2011; Pyc & Rawson, 2009). In Übereinstimmung mit den theoretischen Annahmen des Ansatzes des episodischen Kontextes zeigen sich höhere Testeffekte in der Aufgabe mit freiem Abrufformat (Experiment 1: jung:  $d = -0.75$ ; alt:  $d = -0.21$ ) als bei der Paarassoziationsaufgabe (Experiment 2: jung:  $d = -0.40$ ; alt:  $d = -0.09$ ). Zwar fallen bei Experiment 1 die Altersunterschiede im Testeffekt größer aus, als in Experiment 2, jedoch sind bei älteren Lernern in Experiment 2 kaum Testeffekte zu beobachten. Allerdings haben die Experimente dieser Arbeit nicht das Ziel, die Testeffekte unterschiedlicher Abrufformate zu vergleichen, sondern zu analysieren, ob die Altersunterschiede im Testeffekt bei verschiedenen Lernmaterialien auftreten. Aufgrund der heterogenen Stichprobe, der verschiedenen Designs und der Schwierigkeit der Aufgaben, sollen die Vergleiche der Größe des Testeffekts in den beiden Experimenten als Anstoß für zukünftige Arbeiten bezüglich des Testeffekts und der Altersunterschiede gesehen werden. Hierbei könnte der Frage nachgegangen werden, bei welchem Abfrageformat ältere und jüngere Lerner in ähnlichem Maße von dem Testeffekt profitieren und ob Altersunterschiede

bestehen bleiben, wenn beide Aufgaben für beide Altersgruppen im vergleichbaren Maße schwer sind.

Ein weiterer Unterschied zwischen Experiment 1 und 2 liegt darin, dass sich die Lernkurven in ihrer Form unterscheiden. In Experiment 1 werden die Daten am besten durch eine logarithmische Modellierung des Effekts des Durchgangs vorhergesagt, wohingegen in Experiment 2 die Daten durch eine lineare Modellierung besser repräsentiert werden. Auch bei Roediger und Smith (2012) zeigten sich unterschiedliche Lernkurven in den freien Erinnerungsaufgaben, im Gegensatz zur Wortpaaraufgabe. Ein genauerer Vergleich der Form der Lernkurve soll in der vorliegenden Arbeit ebenfalls nicht näher thematisiert werden. Man beachte jedoch, dass in den Regressionsmodellen der vorliegenden Arbeit die Funktionsform für die Unterschiede der Lernleistung durch den Testeffekt so modelliert wurde, wie die Funktionsform der Standardlernkurve (d.h. in Experiment 1 wird der Unterschied der Standard- zur „pure-study“-Lernkurve anhand des logarithmischen Effekts des Durchgangs modelliert und in Experiment 2 anhand eines linearen Effekts). Fokus zukünftiger Studien zur Differenzierung von Lern- und Testeffekten bei Lernkurven sollte daher ebenfalls die Funktionsform der beiden Effekte sein. Jedoch wird in Lernkurven nicht nur die Gedächtnisleistung abgebildet, sondern ein Konglomerat aus mehreren kognitiven Prozessen. Newell, Mayer-Kress, Hong und Liu (2010) beschreiben, dass die Lernkurve nicht nur von der Gedächtnisleistung beeinflusst wird, sondern auch Einflüsse einer Aufwärmphase, der Inhibition von Störquellen und Erschöpfung beinhalten kann. Wobei bei den verschiedenen Prozessen wiederum Altersunterschiede auftreten können wie bspw. bei Inhibitionsprozessen (Hasher & Zacks, 1988) oder der Erschöpfung (Craik & Byrd, 1982).

Zusammenfassend zu Studienblock 1 kann man sagen, dass die Ergebnisse der Experimente 1 und 2 für Altersunterschiede in der Wirksamkeit des Testeffekts sprechen und die Berücksichtigung dieser große Anteile der Varianz zwischen und innerhalb von Personen aufklären kann. Im Gegensatz zu bisherigen Studien (Coane, 2013; Meyer & Logan, 2013),

welche diese Unterschiede nicht gefunden haben, wurden in den Experimenten 1 und 2 mehrere Lern-Test-Durchgänge ohne Feedback verwendet. Mithilfe dieser Designs ist es möglich, die Einflüsse des Testeffekts genauer zu beobachten und zu beschreiben. Diese Altersunterschiede konnten für Aufgaben mit freiem Abrufformat, ebenso wie für Aufgaben mit Hinweisreizen gezeigt werden. Zwar beziehen sich die Theorien zur Entstehung des Testeffekts nicht explizit auf Altersunterschiede, jedoch lassen sich mit Theorien zu kognitiven Alterungsprozessen Schlüsse ableiten. Demnach passen im Rahmen des Ansatzes zum episodischen Kontext die hier beobachteten Altersunterschiede zu den Überlegungen der Defizite beim Erinnern von Kontextattributen.

#### *4.2. Geschichten als Lernmaterial und Dropout-Design*

Mit den Experimenten 3 und 4 werden mehrere Ziele verfolgt. Hauptziel ist, die Testeffekte bei jüngeren und älteren Erwachsenen für alltagsnäheres Material zu untersuchen. Hierfür wurden kurze, standardisierte Geschichten aus dem WMS-R und dem WMS-IV gelernt und abgefragt. Im Gegensatz zu Experiment 1 und 2, stellen Geschichten ein kohärentes Lernmaterial dar. Geschichten können auf mehreren Ebenen einen Einfluss auf den Lernprozess–und damit auch auf den Abfrageprozess–haben; bspw. ist das gesamte Lernmaterial aufgrund des Rahmens der Geschichte verständlicher und die einzelnen Propositionen sind innerhalb der Geschichte vernetzt (bspw. Bower, 1976). Vorwissen über die allgemeinen Rahmenbedingungen einer Geschichte kann hierbei aktiviert und beim Abfragen genutzt werden (bspw. gibt es typischerweise einen Protagonisten/in, einen Ort, eine Handlung usw.). Dadurch kann nicht nur inhaltliches Vorwissen einer Person beim Lernen (siehe McNamara & Kintsch, 1996), sondern darüber hinaus das Wissen über den Aufbau von Geschichten einen Einfluss auf die Enkodierung und den Abruf haben (vgl. *Schematic Support*, Craik & Bosman, 1992). Im Gegensatz zum häufiger verwendeten hinweisgesteuerten Abruf durch gezielte Fragen zu Untersuchungen zum Testeffekt in

Fließtexten, wird hier ein freier Abruf verwendet (durch Fragen gesteuerter Abruf: bspw. Butler, 2010; Meyer & Logan, 2013; freier Abruf: Roediger & Karpicke, 2006b). Durch den Rahmen der Geschichte (kohärente Erzählweise, alltägliche Themen) unterscheidet sich der freie Abruf von dem freien Abruf von Wortlisten (mit zusammenhangslosen Wörtern). Bei letzterem ist die Erstellung eines Abrufplans, bzw. die subjektive Organisation des Lernmaterials von zentraler Bedeutung, welche jedoch während der Lern-Test-Durchgänge erstellt wird (Tulving, 1962), wohingegen bei Geschichten die subjektive Organisation des Lernmaterials (für den Abrufprozess) durch ein bereits bekanntes Schema über Geschichten übernommen werden könnte. Dies könnte dafür sorgen, dass vor allem ältere–aufgrund von detaillierteren Schemata und breiterem Vorwissen (bspw. Castel, 2005)–von dieser Struktur des Lernmaterials profitieren. Demnach könnte es älteren Erwachsenen möglich sein, die Lerninhalte besser zu enkodieren und abzurufen, wodurch altersspezifische Nachteile beim Lernen von Assoziationen kompensiert werden könnten. Obgleich bei älteren Lernern aufgrund von generellen Defiziten im episodischen Gedächtnis eine geringere Lernleistung bei mehreren Lern- und Testdurchgängen erwartet wird, könnten beide Altersgruppen im vergleichbaren Umfang von Testeffekten profitieren.

Ein weiteres Ziel der Experimente 3 und 4 ist der Vergleich verschiedener Lerndesigns. Neben der Standardlernbedingung werden zusätzlich eine Bedingung mit nur einem finalen Testdurchgang sowie eine Lernbedingung mit Dropout-Design eingeführt. Im Dropout-Design werden während des Experiments die erfolgreich abgerufenen Lerninhalte nicht mehr in dem folgenden Lerndurchgang präsentiert, wodurch sich die Dauer der Lerndurchgänge verkürzt. Karpicke und Roediger (2008) zeigten, dass Lerner im selben Umfang von den Testeffekten in einer Dropout-Bedingung profitieren, wie in der Standardbedingung, wenn der „Dropout“ nur im Lerndurchgang stattfindet, jedoch mit jedem Testdurchgang das gesamte Material abgefragt wird. Inwieweit eine solche Dropout-Bedingung ebenfalls für den freien Abruf von alltagsnahen Geschichten bezüglich des

Testeffekts hilfreich ist, soll in den Experimenten 3 und 4 untersucht werden. Einerseits sind die Lernzeiten in der Dropout-Bedingung kürzer und die Lerner können direkt sehen, welche Teile des Lernmaterials sie noch nicht erfolgreich abgerufen haben; andererseits jedoch, könnte das Wegfallen von Satzteilen das Verständnis beim Lernen erschweren. Die Testeffekte werden unmittelbar während des Lernens der Geschichten und verzögert nach zwei Tagen oder einer Woche getestet.

#### *4.2.1. Ergebnisse Studienblock 2*

Im Folgenden werden zunächst die Testeffekte zum finalen Durchgang des Experiments interpretiert, bevor die Effekte im verzögerten Abruf besprochen werden. Die Ergebnisse aus Experiment 3 und 4 unterstützen die Annahme, dass auch beim freien Abruf von Geschichten zusätzliche Testdurchgänge die Lernleistung verbessern. Dabei erinnerten sich Probanden beider Altersgruppen am Ende des Lernexperiments in den Bedingungen mit mehreren, vollständigen Lern-Test-Durchgängen an mehr Propositionen als in den Bedingungen mit nur einem finalen Abfragedurchgang. Obgleich die ältere Vergleichsstichprobe insgesamt weniger Texteinheiten erinnerte als die jüngere, profitierten sie in einem vergleichbaren Umfang vom Testeffekt. Damit unterscheiden sich die Ergebnisse gegenüber Experiment 1 und 2, in welchen sich Altersunterschiede in der Lernkurve zeigten. Dies ist wahrscheinlich auf das verwendete Lernmaterial zurückzuführen und unterstützt die Annahme, dass vor allem Ältere von dem Lernmaterial profitieren. Man beachte, dass sich die „pure-study“-Bedingungen in Experiment 3 und 4 unterscheiden. In Experiment 4 werden in der „pure-study“-Lernbedingung gleich viele Lerndurchgänge dargeboten, wie in der Standardbedingung (jeweils vier Lerndurchgänge), wohingegen in Experiment 3 in der „pure-study“-Bedingung für jeden Testdurchgang der Standardbedingung ein zusätzlicher Lerndurchgang dargeboten wird (Standardbedingung: drei Lern- und drei Testdurchgänge; „pure-study“: fünf Lerndurchgänge und ein Testdurchgang). Auf diese Weise wird in

Experiment 3 kontrolliert, dass die Effekte des Testens nicht nur durch die mehrmalige Darbietung des Materials in den Testdurchgängen zustande kommt (siehe Roediger & Smith, 2012).

Bezogen auf die Dropout-Bedingung zeigen sich in Experiment 3 Altersunterschiede im Testeffekt. Hierbei profitierten Ältere von dem Testeffekt im finalen Testdurchgang des Experiments nur in der Standardbedingung. Zusätzlich zeigten Ältere eine geringere Lernrate in der Dropout-Bedingung im Gegensatz zur Lernrate der Standardbedingung. In Experiment 4 jedoch, profitierten ältere Personen im vergleichbaren Maße in beiden Bedingungen mit mehreren Testdurchgängen vom Testeffekt. Man beachte jedoch, dass vor allem in der Dropout-Bedingung in Experiment 3 große Unterschiede in der Präsentation des Lernmaterials zur „pure-study“-Bedingung bestehen. Es wird nicht nur mehr Material in den Lerndurchgängen der „pure-study“-Bedingung präsentiert, sondern darüber hinaus zweimal öfter das gesamte Material gezeigt. Für jüngere Probanden stellt die Dropout-Bedingung eine zeitlich effizientere Alternative zur Standardbedingung dar (bezogen auf die Lernleistung während des Experiments). Unter welchen Bedingungen ältere Lerner im gleichen Maße von einer solchen Dropout-Bedingung beim Geschichtenlernen profitieren, gilt es in zukünftigen Experimenten zu untersuchen.

Selbst wenn sich die Probanden im Durchschnitt an mehr Wörter in den Bedingungen mit Testdurchgängen erinnern, werden in Experiment 3 und 4 vom finalen Testdurchgang auf den verzögerten Abruf (nach zwei Tagen oder nach einer Woche) gleich viele Texteinheiten vergessen (vgl. Roediger & Smith, 2012). Einzige Ausnahme stellt hierbei die ältere Altersgruppe im Dropout-Design von Experiment 4 dar. Ältere Lerner vergessen in der Dropout-Bedingung weniger Textpropositionen im Laufe von einer Woche im Vergleich zur Standard- oder „pure-study“-Bedingung. Durchschnittlich zeigen sich bei kurzen Geschichten als Lernmaterial demnach keine Testeffekte in der absoluten Vergessensrate bei zwei Tagen oder eine Woche. Jedoch kann sich unter bestimmten Umständen ein Testeffekt ergeben, wie

in Experiment 4. Hierbei könnte die Kombination aus Testdurchgängen, Schwierigkeit der Aufgabe sowie Zeitintervall bis zum verzögerten Abruf eine Rolle spielen.

Dass in der Standardbedingung vergleichbar viele Lerninhalte vergessen werden wie in der „pure-study“-Bedingung zeigt sich auch beim Lernen von Wortlisten (Roediger & Smith, 2012). Dabei zeigte sich anhand der Ergebnisse keine Interaktion aus Abrufintervall (15 Minuten vs. zwei Tage später: dabei bearbeite eine Person entweder den Test nach 15 Minuten oder zwei Tage später) und Lerndesign (vier Lern-Test-Durchgänge, vier Lerndurchgänge, acht Lerndurchgänge), obgleich in der Standardbedingung zu beiden Zeitpunkten mehr Wörter erinnert wurden. Demnach scheint der Testeffekt beim Vergleich einer Standardlernbedingung zu einer Kontrollbedingung ohne Testdurchgänge keine Effekte auf die Vergessensrate nach mehreren Tagen zu haben. Es zeigen sich die positiven Effekte der Testdurchgänge während der Lern-Test-Durchgänge und bleiben auch nach mehreren Tagen bestehen; jedoch schützen diese nicht vor Vergessen. In Experiment 3 und 4 könnte auch das Lernmaterial einen Einfluss auf den fehlenden Testeffekt beim verzögerten Abruf haben, da aufgrund der Vernetzung der einzelnen Propositionen der Geschichte ein sinnvoll interpretierbarer Kontext bereits vorgegeben ist. Nachdem die Testeffekte abhängig von der Anzahl der vorgegebenen Hinweisreize abnehmen–und dadurch das Zurückversetzen in den vorherigen Kontext überflüssig wird–könnte dies beim verzögerten Abruf dazu führen, dass kein Testeffekt auftritt (es wurden beim verzögerten Abruf Vor- und Zuname der Hauptperson, Ort sowie ein Hinweis auf die Handlung vorgegeben). Auf welche Weise das Material und die Abfragemethode die unmittelbaren und verzögerten Testeffekte beeinflussen, bleibt vorerst offen. Man beachte jedoch, dass die Auswirkungen des Testeffekts während des Experiments gleichzeitig die Interpretation der Testeffekte beim verzögerten Abruf erschweren. Am Ende der Experimente 3 und 4 (vor dem verzögerten Abruf) wurden aufgrund des Testeffekts in den Bedingungen mit Testdurchgängen mehr Lerninhalte erinnert als in der „pure-study“-Bedingung. Somit ist das Leistungsniveau der

Bedingungen am Ende des Experiments verschieden (vor dem verzögerten Abruf). Um die Vergessensrate genauer zu untersuchen, wird für weitere Forschung zum Testeffekt bei Lernkurven empfohlen, das Leistungsniveau am Ende eines Experiments für die Test- sowie für die Kontroll-Lern-Bedingung auf relativ ähnlichem Niveau zu halten (bspw. auf 100%).

Wie in den Experimenten 1 und 2, zeigen sich ebenfalls im 3. und 4. Experiment der vorliegenden Arbeit unterschiedliche Varianzen für die Lernmethoden (mit Testen vs. ohne Testdurchgänge). Und zwar zeigten sich größere Abweichungen von den geschätzten Werten in der „pure-study“-Bedingung. Betrachtet man die Fehlervarianz der verwendeten Methoden für die beiden Altersgruppen getrennt, so zeigen sich in Experiment 3 und 4, dass vor allem in der Gruppe der „pure-study“-Bedingung ältere Lerner die höchsten Varianzen zeigen.

Dadurch wird diese Gruppe durch die Modelle am schlechtesten vorhergesagt, bzw. sind individuelle Unterschiede bei den älteren Lernern in der Bedingung ohne Testdurchgänge am größten. Dies deutet einen unterschiedlichen Umgang mit dem wiederholten Lernen von Geschichten an. Eine mögliche Erklärung ist die Annahme, dass sich in der Standardbedingung aufgrund der Testdurchgänge die verwendeten Strategien der Probanden mehr ähneln als in der „pure-study“-Bedingung (da der Abfrageversuch als Strategie verwendet wird) und dadurch weniger Unterschiede auftreten. Jedoch liegt die Analyse verwendeter Lernstrategien der Probanden außerhalb des Schwerpunkts dieser Arbeit.

Darüber hinaus zeigen sich beim Lernen von Geschichten individuelle Unterschiede. In beiden Experimenten profitieren Lerner in unterschiedlichem Umfang von den drei Lernbedingungen. Zusätzlich unterscheiden sich die Lerner auch darin, wie schnell sie neue Textpropositionen dazu lernen. Diese Ergebnisse unterstützen die Befunde zu individuellen Unterschieden in Lernkurven und deren Wichtigkeit für die Berücksichtigung in der Modellierung der Lernleistung. Dabei hängen vor allem die Effekte der Lernbedingungen (Standard-, Dropout- und „pure-study“-Bedingung) zusammen. Probanden, die überdurchschnittliche Lernleistung in der Dropout-Bedingung zeigen, erzielen ebenfalls eher

überdurchschnittliche Leistungen in der Standard- und in der „pure-study“-Bedingung (in Experiment 4 hängt die Leistung der Standardbedingung zusätzlich mit der Leistung in der „pure-study“-Bedingung zusammen). Diese Ergebnisse sprechen dafür, dass Eigenschaften, die einen guten Lerner kennzeichnen, in den verschiedenen Lernbedingungen von Bedeutung sind. Zu diesen Eigenschaften zählt unter anderem die Verarbeitungsgeschwindigkeit. Die individuellen Unterschiede in der Verarbeitungsgeschwindigkeit klären einen Teil der individuellen Unterschiede in der Lernleistung auf (in Experiment 3: zusätzlich 2.3% innerhalb und zusätzlich 3.7% zwischen Individuen; Experiment 4: zusätzlich 4.3% innerhalb und zusätzlich 11.0% zwischen Individuen). Dabei zeigt sich in Experiment 4 lediglich bei den älteren Lernern ein statistisch signifikanter Effekt der Verarbeitungsgeschwindigkeit auf die Lernleistung. Wie schon in Experiment 1 und 2, zeigt sich auch in Experiment 3 und 4 eine höhere Relevanz der kognitiven Verarbeitungsgeschwindigkeit bei der schwierigeren Lernaufgabe. Demnach scheinen individuelle Unterschiede in der Verarbeitungsgeschwindigkeit vor allem bei schweren Aufgaben eine größere Rolle zu spielen als bei leichteren. Da die Ergebnisse uneindeutig zu interpretieren sind, sollte in zukünftiger Forschung die Rolle der Verarbeitungsgeschwindigkeit für verschiedene Lernkurven im Rahmen des Testeffekts weiter untersucht werden (bspw. mit unterschiedlichen Verarbeitungsgeschwindigkeitsaufgaben und Schwierigkeiten des Lernmaterials).

Die drei verwendeten Geschichten unterscheiden sich in ihrer Schwierigkeit. So ist in Experiment 3 und 4 jeweils die Geschichte 1, diejenige Geschichte mit den wenigsten Wörtern die leichteste und Geschichte 2, mit den meisten Wörtern, die durchschnittlich schwierigste Geschichte. Obgleich alle Geschichten aus gleich vielen Propositionen bestehen, spielt die Anzahl der Wörter eine Rolle für die Lernleistung. Darüber hinaus kann die Reihenfolge der Lerndesigns einen Einfluss auf die Lernleistung in den Lernbedingungen haben. Aufgrund der Kontrolle von Positionseffekten durch griechisch-lateinisch Quadrat

modellierten Versuchsdesigns, zeigen sich diese Effekte der unterschiedlichen Geschichtenschwierigkeiten und der Position des Lerndesigns (an erster, zweiter oder dritter Stelle im Experiment) in der Residualvarianz und ohne systematische Auswirkung auf die geschätzten festen Effekte. Jedoch könnten bei der Erstellung und Empfehlung von Lernplänen eben diese Positionseffekte eine Rolle spielen; bspw. um der Frage nachzugehen, ob ein Dropout-Design bezüglich der Lernleistung effizienter ist, wenn zuvor mit dem Standarddesign ausreichend geübt/gelernt wurde.

Eine Limitation der Experimente 3 und 4 liegt in der Verwendung einer einzigen „pure-study“-Aufgabe. Dadurch, dass lediglich ein Testdurchgang am Ende des Experiments erfolgte, können die Lernkurven nicht wie in Experiment 1 und 2 verglichen werden. Es ist zwar möglich zu beobachten, ob es Unterschiede zwischen den Lernbedingungen gibt, jedoch nicht, wie diese zustande kommen. Demzufolge kann nicht der Unterschied in der Lernrate der Standard- oder Dropout-Bedingung im Gegensatz zur Lernrate der „pure-study“-Bedingung beobachtet werden. Somit werden hier die Auswirkungen von mehreren Testdurchgängen und nicht der Verlauf des Testeffekts während des Experiments betrachtet. Eine weitere Limitation besteht darin, dass es vor den Lernbedingungen in den Experimenten keine Übungsdurchgänge gab. Diese wäre für die Personen vermutlich besonders hilfreich in dem Dropout-Design gewesen, da diese Art des Lernens von Texten im Alltag untypisch ist und aufgrund des ungewohnten Lerndurchgangs zu Verwirrung führen könnte.

#### *4.2.2. Implikationen für Geschichten als Lernmaterial*

Im Rahmen der Analyse des Testeffekts bei lernrelevantem Textmaterial wurde ebenfalls das abrufinduzierte Vergessen untersucht (bspw. Little, Storm & Bjork, 2011). Beim abrufinduzierten Vergessen wird angenommen, dass beim wiederholten Abrufversuch eines spezifischen Teils des Lerninhalts dieses gestärkt wird, aber die Erinnerungsleistung der ungetesteten Items dabei sinkt (siehe Anderson, Bjork, & Bjork, 1994). Beim Lernen von

Texten kann das abrufinduzierte Vergessen auftreten, wenn zur einen Hälfte des Textes Fragen gestellt werden und zur anderen nicht. In einem finalen Test werden die Fragen zu den nicht-getesteten Passagen durchschnittlich schlechter beantwortet als die Fragen eines Kontrolltextes, der nicht getestet wurde. Jedoch fanden Little und Kollegen (2011), dass die Kohärenz eines Textes zur spontanen Zusammenführung des gesamten Lerninhalts führen kann und so gegen abrufinduziertes Vergessen schützen kann—dennoch sind die getesteten Inhalte bei einem finalen Abruf den anderen, nicht-getesteten Inhalten überlegen. Werden jedoch die Sätze des Lernmaterials in zufälliger Reihenfolge zusammengesetzt (dabei bleibt jeder einzelne Satz für sich genommen verständlich), ist wiederum abrufinduziertes Vergessen beobachtbar (Little et al., 2011). Diese Beobachtung verdeutlicht einerseits die Besonderheit von Texten als Lernmaterial im Gegensatz zu zusammenhangslosen Wortlisten und andererseits, wie wichtig der Abruf des gesamten Lerninhalts für den Lernerfolg ist (vgl. Karpicke & Roediger, 2008). Jedoch bleibt beim Lernen von Texten die Frage bezüglich Effizienz des wiederholten Lernens nicht-erinnerter Einheiten offen (Dropout-Design). Bspw. die Frage, ob ein Dropout-Design mit freiem Abrufformat, wie es in der vorliegenden Arbeit verwendet wurde, sich ebenfalls für komplexere Texte eignet, bei welchen neben dem Thema auch die einzelnen Lernpropositionen für den Lerner unbekannt sind und noch in keine bestehenden Schemata eingebaut werden können (bspw. bei Definitionen oder dem Lernen der Bestandteile sowie deren Zusammenhänge bei einem Atommodell).

Um den freien Abruf bei kohärentem Textmaterial in Zusammenhang mit dem Testeffekt zu untersuchen, werden für zukünftige Untersuchungen darüber hinaus Analysen auf Itemebene vorgeschlagen. Durch die interne Struktur des Materials ist es wahrscheinlich, dass die einzelnen Einheiten des Textes nicht in zufälliger Reihenfolge erinnert werden. Dementsprechend sollte nicht nur beschrieben werden, welche Teile einer Geschichte typischerweise als erstes erinnert werden, sondern darüber hinaus, welche Propositionen am wahrscheinlichsten als nächstes genannt werden und wie dies von dem Testeffekt beeinflusst

wird (vgl. Ansatz zur Analyse beim freien Abruf bei Kahana, 1996). Ein weiterer Forschungsansatz ist die Auswirkung von Texteigenschaften auf den Testeffekt. So fanden McDaniel, Waddill, Finstad und Bourg (2000) Zusammenhänge von Interesse am Text, Aufmerksamkeit und Enkodierungsprozessen beim Lernen von Texten, und folgern, dass das Interesse am Text die Effektivität von Lernstrategien beeinflussen kann.

Zusammengenommen zeigt der zweite Studienblock, dass ein unmittelbarer Testeffekt beim freien Abruf von Geschichten auftritt. Durch die besondere Struktur des Lernmaterials profitieren ältere und jüngere Lerner im gleichen Umfang von den Effekten des erfolgreichen Abrufs. Obwohl sich die Lerner im verzögerten Abruf an mehr Wörter aus den Bedingungen mit wiederholten Testdurchgängen erinnerten, haben sie gleich viele Textpropositionen vergessen wie in den Lernbedingungen mit einem Testdurchgang. Die Effekte konnten für unterschiedlich schwere Aufgaben sowie für verschiedene Vergleichsbedingungen gezeigt werden (gleich viele Lernbedingungen, wie in der Standardbedingung sowie je eine Lernbedingung für jeden Lern- und jeden Testdurchgang in der Standardbedingung). Für ältere Lerner stellt die Dropout-Bedingung eine Sonderrolle dar. Je nach Schwierigkeit des Materials und Eigenschaften des Versuchsdesigns können unmittelbare Testeffekte für Ältere ausbleiben, oder sogar langfristige Testeffekte erzielt werden. Wie diese im Spezifischen zusammenhängen und wie man dies für praktische Lernempfehlungen nutzen kann, bleibt vorerst offen.

#### *4.3. Überblick Testeffekt*

In allen vier Experimenten konnten Testeffekte nachgewiesen werden. Die Experimente unterscheiden sich im Lerndesign, der Aufgabenart bzw. Art der Abfrage und der Schwierigkeit der Aufgaben und unterstützen die Annahme eines generalisierbaren Testeffektes auf verschiedene Lernaufgaben. Diese Beobachtungen werden auch im Rahmen des Ansatzes des episodischen Kontextes erwartet. Bei der Verwendung von Wortlisten sowie

Wortpaaren zeigen sich Altersunterschiede in der Wirksamkeit des Testeffekts, wobei Ältere weniger von den Effekten profitieren als Jüngere. Diese Beobachtung ist konform mit etablierten Theorien zur kognitiven Entwicklung im Alter–wie der Assoziations-Defizit-Hypothese (Naveh-Benjamin, 2000) oder der Hypothese zur umweltbezogenen Unterstützung (Craik, 1986)–und altersbedingter Abnahmen im episodischen Gedächtnis (siehe Kapitel 1.2.2.). Eine besondere Rolle kommt der anfänglichen Lernleistung zu. Je mehr Wörter/Wortpaare erfolgreich erinnert wurden, desto mehr profitierten die Lerner von dem Testeffekt (vgl. Marsh et al., 2009; Meyer & Logan, 2013). Dabei sind ältere Probanden doppelt benachteiligt, da ältere Erwachsene nicht nur weniger Inhalte im Durchschnitt erinnern, sondern auch in geringerem Umfang von einem erfolgreichen Abruf profitieren als jüngere Erwachsene. Dies kann sogar dazu führen, dass ältere Erwachsene nicht mehr von den Testeffekten profitieren (siehe Experiment 2) oder gar negative Testeffekte zeigen (Tse et al., 2010).

Jedoch zeigt sich in Experiment 3 und 4 kein langfristiger Testeffekt (beim Vergleich der Standard- mit der „pure-study“-Lernbedingung; siehe auch Roediger & Smith, 2012). Die Ergebnisse aus den Experimenten zum Geschichtenlernen stehen daher teilweise im Widerspruch zum Testeffekt. Jedoch könnte der ausbleibende Testeffekt beim verzögerten Abruf in Zusammenhang mit der Hinweisvorgabe der Abrufsituation sowie mit dem durch den im Material vorgegebenen Kontext stehen (siehe oben). Das in dieser Arbeit verwendete freie Erinnern von Geschichten stellt eine Art des Abfragens dar, welches im Alltag ein selbstgesteuertes, sinnwiedergebendes (bis hin zum zitierenden) Erzählen eines gelesenen Inhaltes widerspiegeln (bspw. genaue Wiedergabe eines Zeitungsartikels oder das Zitieren eines Gedichts). Im Gegensatz dazu würde das häufiger untersuchte, hinweisgesteuerte Abrufen von Geschichten mit Fragen eine Abfragesituation widerspiegeln, wie man sie aus dem Schulkontext kennt. Auch beim Lernen von Geschichten könnten sich zwischen freiem und hinweisgesteuertem Abruf Unterschiede in der Wirksamkeit des Testeffekts und in der

Sensitivität des Testeffekts gegenüber dem Alter ergeben und stellt damit ein spannendes Feld für weitere Forschung dar.

Eine mögliche Kritik an der Methode der vier Experimente ist weiterhin, dass keine falschen Antworten betrachtet wurden (bspw. Nennungen von Items, welche nicht aus dem abgefragten Material stammen, sondern bspw. aus vorherigen Lernaufgaben–hier: Intrusionen). Zwar ist es mit dem verwendeten Design möglich, die Lernleistung, welche durch den Testeffekt positive beeinflusst wird, zu beobachten–im Sinne von Anzahl richtig genannter Antworten. Mit dem Design ist es jedoch nicht möglich zu beobachten, ob durch zusätzliche Testdurchgänge weniger falsche Antworten gegeben werden. Man würde erwarten, dass generell weniger Intrusionen bei den Experimenten zum Lernen von Geschichten auftreten, da diese eventuell die Sinnhaftigkeit der Geschichte stören können und dadurch schneller als falsch vom Lerner entlarvt werden. Durch das Zurückerinnern an den spezifischen Lernkontext würde man weiterhin erwarten, dass weniger Items aus anderen Aufgaben genannt werden.

Die Auswirkungen von Testdurchgängen auf den Lernerfolg können auf verschiedene Arten untersucht werden. So können darüber hinaus erfolglose Abrufversuche den Lernprozess bei folgenden Lerndurchgängen unterstützen, bspw. wenn vor dem ersten Lerndurchgang ein Testdurchgang erfolgt und dabei versucht wird die Antwort zu generieren–welche ohne vorherigen Lerndurchgang sehr wahrscheinlich falsch ist (siehe Kornell, Hays, & Bjork, 2009; Grimaldi & Karpicke, 2012). In dieser Arbeit wird der Testeffekt als Effekt des erfolgreichen Abrufs und im Rahmen des Ansatzes des episodischen Kontextes verstanden und dessen Auswirkung auf die Lernleistung bei mehreren Lern-Test-Durchgängen untersucht. Dies bildet einen essentiellen, jedoch unvollständigen Teil der Forschung zum Testeffekt ab. Beispielsweise fördert der Testeffekt auch Transferleistungen, zeigt jedoch darüber hinaus Grenzen in seiner Effektivität.

Neben der Steigerung der Lernleistung gegenüber reinem, wiederholtem Lernen, gibt es noch weitere Vorteile von Testdurchgängen, welche essentiell für das alltägliche Lernen sind. Eines der großen Ziele beim Lernen ist der Transfer, also die Fähigkeit gelerntes Wissen auf neue Kontexte zu übertragen (für eine Taxonomie der möglichen Dimensionen von Transfer siehe Barnett & Ceci, 2002). In einer Überblicksarbeit präsentiert Carpenter (2012) Befunde für den Zusammenhang des Testeffekts mit der Transferleistung. Dabei wurde ein positiver Einfluss des Testeffekts für die Transferleistung a) über den zeitlichen Kontext, b) über verschiedene Testformate und c) über Wissensdomänen hinweg gefunden. So können Testdurchgänge nicht nur das langfristige Lernen unterstützen, sondern ebenfalls bei der Abfrage durch verschiedene Testformate (bspw. Kang et al., 2007), oder gar bei der Anwendung gelerntes Wissen auf nicht-getestete Inhalte übertragen werden (bspw. Butler, 2010; Chan, McDermott & Roediger, 2006). Inwieweit die durch den Testeffekt geförderten Transferleistungen auf Individuen im höheren Erwachsenenalter übertragbar sind, bleibt allerdings weitestgehend offen.

Unter bestimmten Bedingungen können Testeffekte die Lernleistung negativ beeinflussen (bspw. Butler & Roediger, 2008; Peterson & Mulligan 2013). In einer Studie von Roediger und Marsh (2005) bearbeiteten Probanden unter anderem nach dem Lesen von Texten Multiple-Choice Aufgaben und in einem finalen Test offene Fragen. Hierbei zeigt sich der typische positive Testeffekt–vorheriges Testen mit Multiple-Choice Fragen förderte die Erinnerungsleistung im finalen Fragetest–jedoch ebenfalls ein negativer Testeffekt. Letzterer zeigt, dass in der Testbedingung mehr falsche Antworten beim finalen Fragetest produziert wurden als ohne vorherigen Multiple-Choice-Test. Je mehr Distraktoritems ein Multiple-Choice-Test beinhaltete, desto weniger profitierten die Lerner vom Testeffekt und desto häufiger wurden die Distraktoren als falsche Antworten im finalen Fragetest angegeben. So ergibt sich beim Multiple-Choice-Test nicht ausschließlich die Möglichkeit, die richtige

Antwort nochmals zu lesen, sondern ebenfalls die falschen Antwortalternativen, die wiederum bei einem späteren Abruf fälschlicherweise für richtig gehalten werden.

Darüber hinaus kann durch selektives Testen abrufinduziertes Vergessen auftreten, bei dem sich die Erinnerungsleistung getesteter Einheiten erhöht, dafür werden jedoch ungetestete Einheiten schlechter erinnert (bspw. Aslan & Bäuml, 2011; Little et al., 2011). Dieser Effekt kann ebenso bei älteren Lernern auftreten, wobei die Auswirkungen des abrufinduzierten Vergessens sich im Alter verringern (Aslan & Bäuml, 2012). Die Abnahme könnte auf altersassoziierte Defizite in der Inhibitionskontrolle (siehe Hasher & Zacks, 1988) zurückgehen (Aslan & Bäuml, 2012).

In ihrer Studie untersuchten Storm und Kollegen (2014), wie sich vorherige Test- oder Lerndurchgänge auf zukünftige Lern-Test-Durchgänge auswirken. Hierfür wurden Probanden nach einem Lerndurchgang für einen Teil der Wortpaare sechsmal getestet ohne Feedback, für den anderen Teil folgten sechs Lerndurchgänge. Nach einer Woche wurden die Probanden abgefragt und es zeigte sich der typische Testeffekt—es wurden mehr Wörter aus der Testbedingung erinnert als aus der Lernbedingung. Es folgten fünf weitere Testdurchgänge (alle sechs Testdurchgänge jeweils mit Feedback, was einem Lern-Test-Durchgang ähnelt), in welchen die Abrufleistung der getesteten Items jeweils schlechter ausgefallen ist als die Abrufleistung für die vorher gelernten Items. Wurde jedoch in dem ersten Teil des Experiments während der Testdurchgänge auch Feedback gegeben (Experiment 2 bei Storm et al. 2014), zeigte sich in den Lern-Test-Durchgängen eine Woche später in allen Testungen eine höhere Erinnerungsleistung für die vorher getesteten Items. Im Rahmen des Bifurkationsmodells (Kornell et al., 2011) können die Ergebnisse (bei Testdurchgängen ohne Feedback) so interpretiert werden, dass die Gedächtnisstärke der getesteten Items zwar so groß ist, dass sie im Laufe einer Woche weiter über den Abrufgrenzwert bleibt, jedoch v.a. die Gedächtnisstärke der ungetesteten Items geringer ist, als die durchschnittliche Gedächtnisstärke der Items in der Lernbedingung. Demnach sollte sich das Muster der

höheren Lernleistung bei den wiederholt gelernten Items zu späteren Lern-Test-Durchgängen vor allem dann zeigen, wenn die Aufgabe relativ schwer ist—d.h. wenn im ersten Teil des Experiments nur wenige Abrufprozesse von den wiederholten Testdurchgängen profitieren. Die Ergebnisse des zweiten Experiments von Storm und Kollegen (2014) zeigen, dass der Testeffekt bei Tests mit Feedback bei zukünftigen Lern-Test-Durchgängen beobachtbar ist. Dies lässt darauf schließen, dass ebenso bei Lernkurven bestehend aus mehreren Lern-Test-Durchgängen der Testeffekt bei einem verzögerten Abruf beobachtbar ist.

Der Testeffekt ist eine wertvolle Methode, welche die Lernleistung über längere Zeiträume, bei verschiedenen Abfrageformaten sowie bei Transferaufgaben positiv beeinflusst. Jedoch können unter spezifischen Bedingungen die Effekte des Testens auch negative Auswirkungen auf die Lernleistung haben. Grundsätzlich ist bei der bisherigen Forschung zum Testeffekt sowie bei der vorliegenden Arbeit zu beachten, dass Altersunterschiede im Testeffekt untersucht werden, jedoch nicht Altersveränderungen. Die Daten werden aus Querschnittsmessungen gewonnen, wodurch Kohorteneffekte auftreten können. Durch diese können aufgrund von historischen und soziokulturellen Veränderungen, systematische Unterschiede zwischen den Altersgruppen auftreten (bspw. in der Bildung). Dennoch lassen sich durch diese Arbeiten Schlüsse und Interventionen für heutige Altersgruppen bilden. Jedoch würden Längsschnittstudien Erkenntnisse über Altersveränderungen im Testeffekt bieten, v.a. wenn es bei der Untersuchung um altersassoziierte, kognitive Prädiktoren wie Verarbeitungsgeschwindigkeit oder Arbeitsgedächtnis geht (vgl. Hofer & Sliwinski, 2001).

Dennoch gilt: der Testeffekt konnte für unterschiedliche Materialien, Stichproben und Umgebungen repliziert werden. Er stellt einen essentiellen Anteil des Lernerfolgs der Lernkurve dar und zeigt, dass Testen nicht nur die Leistung misst, sondern diese auch verändert. Doch es bleiben Fragen offen. Daher sollte die zukünftige Forschung zum Testeffekt die Theorien des Testeffekts weiter empirisch überprüfen und diese mit Theorien

zur kognitiven Entwicklung im Alter verbinden, um Aussagen und Empfehlung für verschiedene Altersgruppen abzuleiten. Besonders bezüglich der Schwierigkeit von Aufgaben, sollten Material und Design so aufgebaut sein, dass ältere Lerner möglichst viele Items mit möglichst hoher Schwierigkeit richtig lösen können, um von den Testeffekten zu profitieren. Des Weiteren können durch den Testeffekt Lernstrategien für Lerner verschiedenen Alters und unterschiedlicher kognitiven Fähigkeiten und Vorwissens abgeleitet werden, damit in der Praxis das volle Potential aus Testdurchgängen genutzt werden kann (bspw. Rogalski, Altmann, & Rosenbek, 2014).

#### *4.4. Ausblick*

Die Ergebnisse der Arbeit weisen auf Unterschiede in der Effektivität von Testdurchgängen zwischen Lernenden, sowie zwischen den verwendeten Lerndesigns hin. Im Folgenden werden die Ergebnisse in Zusammenhang mit individuellen Eigenschaften von Lernenden einerseits sowie mit Eigenschaften des Lerndesigns andererseits diskutiert.

##### *4.4.1. Testeffekt und Lernereigenschaften*

Der Testeffekt steht in Zusammenhang mit der subjektiven Organisation des Lernmaterials. Tulving (1962) hat in seiner Studie gezeigt, dass sich über mehrere Lern-Test-Durchgänge die subjektive Organisation des Lerninhalts bei einem freien Abruf erhöht und diese mit der Abrufleistung in Beziehung steht. Die subjektive Organisation zeichnet hierbei aus, wie konsistent die Abrufsequenz der Items über mehrere Abrufdurchgänge hinweg ist (für verschiedene Maße der subjektiven Organisation siehe Sternberg & Tulving, 1977). Dabei können sich die Abrufsequenzen zwischen Personen stark unterscheiden, bspw. beim freien Abruf von Wortlisten nicht verwandter Wörter (Tulving, 1962). Bei den Lern-Test-Durchgängen tragen ebenso die Testdurchgänge zur subjektiven Strukturierung bei. Zaromb und Roediger (2010) fanden heraus, dass durch Testdurchgänge, anstelle von

Lerndurchgängen, sich die subjektive Organisation erhöht. Ähnlich wie bei der Lernleistung, erhöht ein Testdurchgang die subjektive Organisation im größeren Umfang, als ein zusätzlicher Lerndurchgang (Wortlisten bestehen aus verschiedenen Kategorien). Congleton und Rajaram (2011) fanden ebenfalls höhere subjektive Organisation in den Gruppen mit wiederholten Testungen nach sieben Minuten sowie nach zwei Stunden. Im Rahmen der Untersuchung des Ansatzes des episodischen Kontextes haben Whiffen und Karpicke (2017) gezeigt, dass Testen und das damit verbundene Zurückversetzen in die Lernsituation ebenfalls den Abruf organisiert (Items wurden in ähnlicher Abfolge erinnert, wie sie präsentiert wurden). Aus den Studien lässt sich folgern, dass Testen nicht nur mit Lernerfolg, sondern ebenso Abrufversuche mit subjektiver Organisation zusammenhängen; die subjektive Organisation jedoch auch mit dem Lernerfolg in Beziehung steht. Des Weiteren können sich Altersunterschiede nicht nur in der Lernleistung und beim Testeffekt zeigen (Ergebnisse dieser Arbeit), sondern ebenfalls in der subjektiven Organisation von Lernmaterial bei mehreren Lern-Test-Durchgängen (Davis et al., 2003; Sauzéon, Claverie & N’Kaoua, 2006; Witte, Freund & Sebbby, 1990). Selbst im hohen Alter verbessert sich die subjektive Organisation mit weiteren Lern-Test-Durchgängen, wobei individuelle Unterschiede auftreten (Kurtz & Zimprich, 2014b). Dabei können Testeffekt, subjektive Organisation und Alter auf verschiedene Art zusammenhängen. Bspw. könnte die subjektive Organisation bzw. die Veränderung der subjektiven Organisation bei Lernkurven die Effekte des Testens moderieren. Auf der anderen Seite könnte der Testeffekt über subjektive Organisation Einflüsse auf die Lernleistung haben, da konsistente Muster in den Abrufsituationen als Hinweisreize für den Abruf dienen. Der erfolgreiche Abruf eines Wortes kann dabei helfen, sich in eine frühere Abrufsituation zurückzusetzen und dadurch wiederum als Hinweisreiz zu dienen (v.a. bei Wörtern, die im engen zeitlichen Zusammenhang stehen).

Altersspezifische Defizite in der Entwicklung einer subjektiven Organisation könnten weiterhin Einfluss auf den Testeffekt haben. Beispielsweise fanden Sauzéon und Kollegen

(2006) heraus, dass ältere Lerner mehr als jüngere davon profitierten, wenn die Wörter miteinander verwandt sind. Ein ähnliches Bild zeigt sich ebenso in Experiment 3 und 4 dieser Studie beim freien Abruf von zuvor gelernten Geschichten (kohärentes Lernmaterial) in dem Sinne, dass Ältere keine Defizite im Ausmaß des Testeffekts zeigen.

Weiterhin wurde ein möglicher Zusammenhang des Testeffekts mit basalen kognitiven Fähigkeiten von Lernern untersucht. Jedoch zeigt die geringe Anzahl an Studien ein uneindeutiges Muster. Gemischte Ergebnisse zeigen sich für Lerner mit hoher vs. niedriger Arbeitsgedächtniskapazität. Hierbei fanden Studien, dass Lerner mit geringer Arbeitsgedächtniskapazität mehr von den Effekten wiederholter Testdurchgänge profitierten (bspw. Agarwal et al., 2017), dennoch zeigen andere Befunde, dass der Testeffekt unabhängig vom Arbeitsgedächtnis ist (Aslan & Bäuml, 2011; Wiklund-Hörnqvist, Jonsson & Nyberg, 2014; für einen Überblick siehe Karpicke, 2017). Jedoch ist eine hohe Arbeitsgedächtnisleistung mit höherem abrufinduzierten Vergessen bei jüngeren Erwachsenen (Aslan & Bäuml, 2011) sowie bei älteren Erwachsenen assoziiert (Aslan & Bäuml, 2012). Neben dem Arbeitsgedächtnis existieren ebenso zur Rolle der Verarbeitungsgeschwindigkeit beim wiederholten Testen eine geringe Anzahl an Studien (Verarbeitungsgeschwindigkeit wird in der vorliegenden Arbeit als Prädiktor für Lernleistung, nicht jedoch für den Testeffekt verwendet). Karpicke und Kollegen (2016) haben bei Kindern keinen Zusammenhang der Verarbeitungsgeschwindigkeit und dem Ausmaß des Testeffekts beobachtet. Wodurch der Testeffekt eine robuste Methode zur Unterstützung des Lernens darstellt, welche für unterschiedliche Lerner vergleichbare, positive Effekte mit sich bringt. Wie sich jedoch Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit direkt auf den Testeffekt bei Lernern unterschiedlicher Altersgruppen auswirken, bleibt offen. Zwar würde man vermuten, dass die basalen kognitiven Fähigkeiten mit der (anfänglichen) Lernleistung zusammenhängen und dadurch das Ausmaß des Testeffekts beeinflussen, jedoch bleibt unklar, ob diese individuelle Unterschiede im Ausmaß des Testeffekts aufklären.

#### 4.4.2. Testeffekt und Eigenschaften des Lerndesigns

Für die Untersuchung des Testeffekts ist der Testdurchgang eines Experimentes zentral. Dieser kann auf verschiedene Weise gestaltet sein. Dabei kann neben dem Abrufformat–bspw. freier Abruf, mit Hinweisreizen oder offene Fragen–ebenso die Rückmeldung über die Qualität und Quantität der Leistung erfolgen. Ein solches Feedback kann bspw. angeben, wie viele Einheiten korrekt genannt wurden (Meyer & Logan, 2013) oder die richtige Lösung direkt anzeigen (Coane, 2013). Korrigierendes Feedback zeigt dem Lerner dabei nicht nur auf, welche Inhalte schon korrekt erinnert wurden, sondern hilft auch dabei, Aufmerksamkeit auf noch nicht korrekt gelernte Inhalte zu lenken (Kang et al., 2007; Kornell et al., 2009). Dies ist besonders bei schwierigen Aufgaben der Fall, d.h. wenn eine geringe Anzahl an Items erinnert werden (Kang et al., 2007; Tse et al., 2010). Demnach könnten besonders ältere Lerner von einem Feedback profitieren, aufgrund von Defiziten im episodischen Gedächtnis (siehe Coane, 2013; Tse et al., 2010). Jedoch ist die Rolle des Feedbacks kritisch zu sehen, da hier zusätzlich eine erneute Lernmöglichkeit gegeben wird und sich dadurch nicht nur „reine“ Testeffekte zeigen, sondern auch weitere Effekte der Enkodierung. Daher wurde in den hier präsentierten Experimenten kein Feedback in den Testdurchgängen gegeben. Dadurch ist es möglich, den „reinen“ Testeffekt und somit altersspezifische Unterschiede im Testeffekt zu untersuchen. Dennoch ist das Feedback für wirkungsvolle Lerndesigns in der Praxis für jüngere und für ältere Lerner von zentraler Bedeutung und kann die Effekte von Testdurchgängen verstärken.

Ein zentraler Bestandteil für die Erklärung des Testeffekts ist das Zurückversetzen in frühere Lern- und Abrufsituationen (Ansatz des episodischen Kontextes; Karpicke et al., 2014). Je unterschiedlicher die Kontexte in den Durchgängen sind (Lern- und/oder Testdurchgänge), desto mehr Attribute können mit einem Item assoziiert werden und ein breites Suchset für ein Item kann im Gedächtnis entstehen. Diese neuen, spezifischen Kontextattribute können den Prozess beim Abrufversuch unterstützen. Daraus lässt sich

ableiten, dass auch bei zeitlich längeren Abständen (Veränderung des temporalen Kontextes) ein größerer Testeffekt auftreten sollten. Allerdings verändern sich bei den hier verwendeten Lernkurven (v.a. Standardbedingung) der räumliche und zeitliche Kontext kaum, da die Daten alle innerhalb einer Erhebung gesammelt wurden (zwar gibt es in Experiment 3 und 4 einen verzögerten Abruf nach mehreren Tagen, jedoch würden die daraus resultierenden Testeffekte nur Abrufversuche, die nach diesem verzögerten Abrufversuch stattfinden, beeinflussen). Allerdings gibt es vor jedem Abrufversuch (und damit vor jedem Zurückversetzen) einen neuen Lerndurchgang und damit einen neuen Lernkontext, selbst wenn dieser sich temporal und räumlich kaum von den vorherigen unterscheidet. Man würde vermuten, dass Lern-Test-Durchgänge, wie sie bei der Erhebung von Lernkurven durchgeführt werden, durch zeitlich versetzte Wiederholungen effektiver (bspw. nach Tagen oder Wochen) wären, und zwar in dem Sinne, dass eine bessere Lernleistung erzielt wird, die gleichzeitig relativ robust gegen das Vergessen ist. Für dieses langfristige Lernen spielen vor allem die Testdurchgänge eine wesentliche Rolle. Die wiederholten Lern-Test-Durchgänge können dabei nach verschiedenen Plänen gestaltet sein. Bspw. können diese nach immer wieder gleichen Abständen oder mit immer größer werdenden Intervallen durchgeführt werden. Es scheint jedoch vor allem wichtig zu sein, dass die Testungen zeitlich versetzt sind, jedoch weniger der Intervallplan (innerhalb einer Erhebung siehe Karpicke & Bauernschmidt, 2011; bei jüngeren und älteren Lernern: Logan & Balota, 2008; über mehrere Wochen: Kang, Lindsey, Mozer & Pashler, 2014). Für die angewandte Gerontopsychologie könnte sich somit das Ziel ergeben, Lernpläne für verschiedene Altersgruppen sowie individuell für Lerner mit unterschiedlichen kognitiven Fähigkeiten zu erstellen.

#### 4.5. Fazit

Das wiederholte Abfragen von Lerninhalten ist eine effektive Strategie, um die Lernleistung zu steigern. Dabei lassen sich Testeffekte bei einer Vielzahl von Materialien, Abfragearten und verschiedenen Gruppen von Lernern beobachten. Jedoch haben wenige Studien bisher untersucht, ob Erwachsene im höheren Lebensalter trotz altersbedingter kognitiver Veränderungen von den Testeffekten profitieren. Die vorliegende Arbeit knüpft an der Forschung zu Altersunterschieden im Testeffekt an und unterstützt die Hypothesen, dass 1) beim Bearbeiten mehrerer Lern-Test-Durchgänge ältere Lerner im geringeren Umfang von Testeffekten profitieren, 2) diese Unterschiede jedoch von der Art des Lernmaterials abhängen. Durch die Berücksichtigung etablierter Theorien zum Testeffekt und zur kognitiven Entwicklung im Alter, werden Erklärungsansätze für altersensitive Testeffekte präsentiert und die Ergebnisse in Beziehung zur bisherigen Forschung gesetzt. Dabei helfen die Befunde der vier vorgestellten Experimente, die Diskrepanz von theoretischen Überlegungen und bisherigen empirischen Befunden zu alterssensitiven Testeffekten aufzuklären.

Weitere Forschung sollte in diesem Zusammenhang darüber hinaus die Auswirkungen von individuellen kognitiven Eigenschaften und der Gestaltung des Lerndesigns auf die Effekte von Testdurchgängen für Lerner verschiedener Altersgruppen untersuchen. Durch die daraus folgenden Befunde können die Theorien zum Testeffekt um eine altersspezifische Komponente erweitert werden, um dadurch Empfehlungen und Strategien für alle Altersgruppen zu treffen. Denn für jüngere wie für ältere Lerner gilt: Testen ist weitaus mehr als nur die Messung des momentanen Wissenstandes.

**Literaturverzeichnis**

- Adesope, O. O., Trevisan, D. A. & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Agarwal, P. K., Finley, J. R., Rose, N. S. & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory (Hove, England)*, 25(6), 764–771. <https://doi.org/10.1080/09658211.2016.1220579>
- Agarwal, P. K., Roediger III, H. L., McDaniel, M. A. & McDermott, K. B. (2010). Improving student learning through the use of classroom quizzes: Three years of evidence from the columbia middle school project. *Society for Research on Educational Effectiveness*.
- Anderson, M. C., Bjork, R. A. & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087. <https://doi.org/10.1037//0278-7393.20.5.1063>
- Anderson, J. R. & Funke, J. (2007). *Kognitive Psychologie* (6. Aufl.). Berlin: Spektrum Akadem. Verl.
- Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. *Archives of Clinical Neuropsychology: the Official Journal of the National Academy of Neuropsychologists*, 22(8), 1003–1011. <https://doi.org/10.1016/j.acn.2007.08.004>
- Aslan, A. & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 264–269. <https://doi.org/10.1037/a0021324>
- Aslan, A. & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting in old and very old age. *Psychology and Aging*, 27(4), 1027–1032. <https://doi.org/10.1037/a0028379>

- Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, 128(4), 612–637.  
<https://doi.org/10.1037//0033-2909.128.4.612>
- Bartlett, F. C. (1977). *Remembering. A study in experimental and social psychology* (Reprinted.). Cambridge: Cambridge Univ. Press.
- Becker-Carus, C. & Wendt, M. (2017a). Lernen. In C. Becker-Carus & M. Wendt (Hrsg.). *Allgemeine Psychologie. Eine Einführung* (Lehrbuch, 2., vollständig überarbeitete und erweiterte Neuauflage, S. 291–351). Berlin: Springer.
- Becker-Carus, C. & Wendt, M. (2017b). Gedächtnis. In C. Becker-Carus & M. Wendt (Hrsg.). *Allgemeine Psychologie. Eine Einführung* (Lehrbuch, 2., vollständig überarbeitete und erweiterte Neuauflage, S. 353–420). Berlin: Springer.
- Bishara, A. J. & Jacoby, L. L. (2008). Aging, spaced retrieval, and inflexible memory performance. *Psychonomic Bulletin & Review*, 15(1), 52–57.  
<https://doi.org/10.3758/PBR.15.1.52>
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Hrsg.). *Information processing and cognition: The Loyola symposium* (S. 123–144). Oxford, England: Lawrence Erlbaum.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Hrsg.). *Metacognition: Knowing about knowing* (S. 185–205). Cambridge, MA: MIT Press.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough & J. R. Pomerantz (Hrsg.). *Psychology and the real world: Essays illustrating fundamental contributions to society* (S. 56–64). New York, NY, US: Worth Publishers.

- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn & R. Shiffrin (Hrsg.). *From Learning Processes to Cognitive "Processes: Essays in Honor of William K. Estes* (Vol. 2, S. 35–67). Hillsdale, NJ: Erlbaum.
- Bower, G. H. (1976). Experiments on Story Understanding and Recall. *Quarterly Journal of Experimental Psychology*, 28(4), 511–534.  
<https://doi.org/10.1080/14640747608400579>
- Bregman, A. S. & Wiener, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 9(6), 689–698.  
[https://doi.org/10.1016/S0022-5371\(70\)80034-2](https://doi.org/10.1016/S0022-5371(70)80034-2)
- Brockhaus - Die Enzyklopädie. 21. Aufl. Bd 17 LINL – MATG.* (2006) (21. Aufl.). Leipzig, Mannheim: Brockhaus.
- Bundesministerium für Familie, Senioren, Frauen und Jugend.* (2016, 2. November). Siebter Altenbericht. Sorge und Mitverantwortung in der Kommune - Aufbau und Sicherung zukunftsfähiger Gemeinschaften und Stellungnahme der Bundesregierung. Verfügbar unter <https://www.bmfsfj.de/blob/120144/2a5de459ec4984cb2f83739785c908d6/7--altenbericht-bundestagsdrucksache-data.pdf>
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118–1133. <https://doi.org/10.1037/a0019902>
- Butler, A. C. & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604–616.  
<https://doi.org/10.3758/MC.36.3.604>
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569. <https://doi.org/10.1037/a0017021>

- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547–1552. <https://doi.org/10.1037/a0024140>
- Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*(5), 279–283. <https://doi.org/10.1177/0963721412452728>
- Carpenter, S. K., Pashler, H. & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, *13*(5), 826–830. <https://doi.org/10.3758/BF03194004>
- Carpenter, S. K. & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, *92*, 128–141. <https://doi.org/10.1016/j.jml.2016.06.008>
- Carpenter S. K. & DeLosh E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268–276.
- Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633–642. <https://doi.org/10.3758/BF03202713>
- Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, *20*(4), 718–721. <https://doi.org/10.1037/0882-7974.20.4.718>
- Cerella, J., Onyper, S. V. & Hoyer, W. J. (2006). The associative-memory basis of cognitive skill learning: Adult age differences. *Psychology and Aging*, *21*(3), 483–498. <https://doi.org/10.1037/0882-7974.21.3.483>

- Chan, J. C. K., McDermott, K. B. & Roediger, H. L. (2006). Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology. General*, 135(4), 553–571. <https://doi.org/10.1037/0096-3445.135.4.553>
- Cho, K. W., Neely, J. H., Brennan, M. K., Vitrano, D. & Crocco, S. (2017). Does testing increase spontaneous mediation in learning semantically related paired associates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1768–1778. <https://doi.org/10.1037/xlm0000414>
- Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition*, 2(2), 95–100. <https://doi.org/10.1016/j.jarmac.2013.04.001>
- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240–247. [https://doi.org/10.1016/S0022-5371\(69\)80069-1](https://doi.org/10.1016/S0022-5371(69)80069-1)
- Congleton, A. R. & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology. General*, 140(4), 535–551. <https://doi.org/10.1037/a0024308>
- Conrad, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 92(2), 149–154. <https://doi.org/10.1037/h0032072>
- Craik, F. I. & Rabinowitz, J. C. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology*, 40(3), 309–315.

- Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix (Hrsg.). *Human memory and cognitive capabilities. Mechanisms and performance. Symposium in memoriam Hermann Ebbinghaus from July 1-6, 1985 at Berlin Humboldt University* (S. 409–422). Amsterdam: North-Holland.
- Craik, F. I. M. & Schloerscheidt, A. M. (2011). Age-related differences in recognition memory: Effects of materials and context change. *Psychology and Aging, 26*(3), 671–677. <https://doi.org/10.1037/a0022203>
- Craik, F. I. M., & Bosman, B. A. (1992). Age-related changes in memory and learning. In H. Bouma & J. A. M. Graafmans (Hrsg.) *Gerontechnology* (S. 79–92). Amsterdam: IOS Press.
- Craik, F. I. M. & Byrd, M. (1982). Aging and cognitive deficits. The role of attentional resources. In F. I. M. Craik (Hrsg.). *Aging and cognitive processes* (Advances in the study of communication and affect, Bd. 8, S. 191-211). New York: Plenum P.
- Davis, H., Small, S., Stern, Y., Mayeux, R., Feldstein, S. & Keller, F. (2003). Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex, 39*(4-5), 1063–1091.  
[https://doi.org/10.1016/S0010-9452\(08\)70878-5](https://doi.org/10.1016/S0010-9452(08)70878-5)
- Diggle, P., Heagerty, P. J., Liang, K.-Y. & Zeger, S. L. (2013). *Analysis of longitudinal data* (Oxford statistical science series, Vol. 25, 2. Aufl.). Oxford: Oxford University Press.
- Dunlosky, J. & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*(4), 597–607.  
<https://doi.org/10.1037/0882-7974.13.4.597>

- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest: a Journal of the American Psychological Society*, *14*(1), 4–58.  
<https://doi.org/10.1177/1529100612453266>
- Eisenkraemer, R. E., Jaeger, A. & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paidéia (Ribeirão Preto)*, *23*(56), 397–406.  
<https://doi.org/10.1590/1982-43272356201314>
- Ekstrom, R. B., French, J. W., Harman, H. H. & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Educational Testing Service, Princeton.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*(3), 392–399.
- Grimaldi, P. J. & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513.  
<https://doi.org/10.3758/s13421-011-0174-0>
- Härting, C., Markowitsch, H. J., Neufeld, H., Calabrese, P., Deisinger, K., & Kessler, J. (Hrsg.). (2000). *Wechsler Gedächtnistest – Revidierte Fassung*. Bern: Huber.
- Hasher, L. & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193–225.
- Hofer, S. M. & Sliwinski, M. J. (2001). Understanding ageing. An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology*, *47*(6), 341–352. <https://doi.org/10.1159/000052825>
- Houx, P. J. Shepherd, J., Blauw, G. J., Murphy, M. B., Ford, I., Bollen, E. L., Westendorp, R. G. et al. (2002). Testing cognitive function in elderly populations: the PROSPER study. *Journal of Neurology, Neurosurgery & Psychiatry*, *73*(4), 385–389.  
<https://doi.org/10.1136/jnnp.73.4.385>

- Hultsch, D. F., Hertzog, C., Small, B. J., McDonald-Miszczak, L. & Dixon, R. A. (1992). Short-term longitudinal change in cognitive performance in later life. *Psychology and aging*, 7(4), 571–584. <https://doi.org/10.1037/0882-7974.7.4.571>
- Jacoby, L. L., Jones, T. C. & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, 5(4), 705–709. <https://doi.org/10.3758/BF03208849>
- Johnson, R. (2003). Aging and the remembering of text. *Developmental Review*, 23(3), 261–346. [https://doi.org/10.1016/S0273-2297\(03\)00009-1](https://doi.org/10.1016/S0273-2297(03)00009-1)
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24(1), 103–109. <https://doi.org/10.3758/BF03197276>
- Kahana, M. J. (Hrsg.). (2014). *Foundations of human memory*. New York: Oxford University Press.
- Kang, S. H. K., Lindsey, R. V., Mozer, M. C. & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, 21(6), 1544–1550. <https://doi.org/10.3758/s13423-014-0636-z>
- Kang, S. H. K., McDermott, K. B. & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5), 528–558. <https://doi.org/10.1080/09541440601056620>
- Karpicke, J. D. (2017). Retrieval-based learning: A Decade of Progress. In J. H. Byrne (Hrsg.). *Learning and Memory* (2. Aufl., S. 487–514). San Diego: Elsevier Science. <https://doi.org/10.1016/B978-0-12-809324-5.21055-9>
- Karpicke, J. D. & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1250–1257. <https://doi.org/10.1037/a0023436>

- Karpicke, J. D., Blunt, J. R. & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, 7, 350.  
<https://doi.org/10.3389/fpsyg.2016.00350>
- Karpicke, J. D., Lehman, M. & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, 61, 237–284.  
<https://doi.org/10.1016/B978-0-12-800283-4.00007-1>
- Karpicke, J. D. & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151–162.  
<https://doi.org/10.1016/j.jml.2006.09.004>
- Karpicke, J. D. & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science (New York, N.Y.)*, 319(5865), 966–968.  
<https://doi.org/10.1126/science.1152408>
- Koch, I. & Stahl, C. (2017). Lernen und Gedächtnis. In J. Müsseler & M. Rieger (Hrsg.). *Allgemeine Psychologie* (S. 319–355). Berlin, Heidelberg: Springer.
- Kornell, N. & Bjork, R. A. (2008). Optimising self-regulated study: The benefits - and costs - of dropping flashcards. *Memory (Hove, England)*, 16(2), 125–136.  
<https://doi.org/10.1080/09658210701763899>
- Kornell, N., Bjork, R. A. & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85–97.  
<https://doi.org/10.1016/j.jml.2011.04.002>
- Kornell, N., Hays, M. J. & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989–998. <https://doi.org/10.1037/a0015729>

- Kurtz, T., Mogle, J., Sliwinski, M. J. & Hofer, S. M. (2013). Individual differences in task-specific paired associates learning in older adults: The role of processing speed and working memory. *Experimental Aging Research*, *39*(5), 493–514.  
<https://doi.org/10.1080/0361073X.2013.839024>
- Kurtz, T. & Zimprich, D. (2014a). Individual differences in criterion-based dropout learning in old age: the role of processing speed and verbal knowledge. *European Journal of Ageing*, *11*(2), 183–193. <https://doi.org/10.1007/s10433-013-0293-5>
- Kurtz, T. & Zimprich, D. (2014b). Individual differences in subjective organization and verbal learning in old age. *Experimental Aging Research*, *40*(5), 531–554.  
<https://doi.org/10.1080/0361073X.2014.956619>
- Leeming, F. C. (2002). The Exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*(3), 210–212.  
[https://doi.org/10.1207/S15328023TOP2903\\_06](https://doi.org/10.1207/S15328023TOP2903_06)
- Lehman, M. & Karpicke, J. D. (2016). Elaborative retrieval: Do semantic mediators improve memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1573–1591. <https://doi.org/10.1037/xlm0000267>
- Lehman, M., Smith, M. A. & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787–1794.  
<https://doi.org/10.1037/xlm0000012>
- Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest* [Spota-word intelligence test] (MWT-B; 4. neu bearbeitete Aufl.). Balingen, Germany: Spitta.
- Little, J. L., Storm, B. C. & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory (Hove, England)*, *19*(4), 346–359.  
<https://doi.org/10.1080/09658211.2011.569725>

- Logan, J. M. & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: exploring different schedules of spacing and retention interval in younger and older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 15(3), 257–280.  
<https://doi.org/10.1080/13825580701322171>
- Lustig, C., Hasher, L. & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". In D. S. Gorfein & C. M. MacLeod (Hrsg.). *Inhibition in cognition* (Decade of behavior, 1. Aufl., S. 145–162). Washington, DC: American Psychological Association. <https://doi.org/10.1037/11587-008>
- MacCallum, R. C., Kim, C., Malarkey, W. B. & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32(3), 215–253. [https://doi.org/10.1207/s15327906mbr3203\\_1](https://doi.org/10.1207/s15327906mbr3203_1)
- MacDonald, S. W. S., Stigsdotter-Neely, A., Derwinger, A. & Bäckman, L. (2006). Rate of acquisition, adult age, and basic cognitive abilities predict forgetting: New views on a classic problem. *Journal of Experimental Psychology*, 135(3), 368–390.  
<https://doi.org/10.1037/0096-3445.135.3.368>
- Marsh, E. J., Agarwal, P. K. & Roediger, H. L. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology. Applied*, 15(1), 1–11.  
<https://doi.org/10.1037/a0014721>
- Martin, M. & Zimprich, D. (2010). Differentiation-dedifferentiation as a guiding principle for the analysis of lifespan development. *Leben im Alter–Eigen-und Mitverantwortlichkeit in Gesellschaft, Kultur und Politik*, 35–43. <https://doi.org/10.5167/uzh-40639>
- McDaniel, M. A., Anderson, J. L., Derbish, M. H. & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4-5), 494–513.  
<https://doi.org/10.1080/09541440701326154>

- McDaniel, M. A., Waddill, P. J., Finstad, K. & Bourg, T. (2000). The effects of text-based interest on attention and recall. *Journal of Educational Psychology*, 92(3), 492–502.  
<https://doi.org/10.1037//0022-0663.92.3.492>
- McNamara, D. S. & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247–288.  
<https://doi.org/10.1080/01638539609544975>
- Meyer, A. N. D. & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, 28(1), 142–147.  
<https://doi.org/10.1037/a0030890>
- Morris, C. D., Bransford, J. d. & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519–533. [https://doi.org/10.1016/S0022-5371\(77\)80016-9](https://doi.org/10.1016/S0022-5371(77)80016-9)
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1170–1187. <https://doi.org/10.1037//0278-7393.26.5.1170>
- Naveh-Benjamin, M. & Mayr, U. (2018). Age-related differences in associative memory: Empirical evidence and theoretical perspectives. *Psychology and Aging*, 33(1), 1–6.  
<https://doi.org/10.1037/pag0000235>
- Newell, K. M., Mayer-Kress, G., Hong, S. L. & Liu, Y.-T. (2010). Decomposing the performance dynamics of learning through time scales. In K. M. Newell & P. C. M. Molenaar (Hrsg.). *Individual pathways of change. Statistical models for analyzing learning and development* (Decade of behavior, S. 71–86). Washington, DC: American Psychological Association. <https://doi.org/10.1037/12140-005>

- Old, S. R. & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*(1), 104–118.  
<https://doi.org/10.1037/0882-7974.23.1.104>
- Oswald, W. D. & Fleischmann, U. M. (1999). *Nürnberger-Alters-Inventar (NAI)*. NAI Testmanual und-Textband. Verlag für Psychologie, Hogrefe.
- Park, D. C. & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 72*(1), 82–90. <https://doi.org/10.1093/geronb/gbw066>
- Pastötter, B. & Bäuml, K.-H. T. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging, 34*(2), 242–250. <https://doi.org/10.1037/pag0000307>
- Petermann, F. & Lepach, A. C. (2012). Wechsler memory scale–fourth edition (WMS-IV). *Manual zur Durchführung und Auswertung. Deutsche Übersetzung und Adaptation der WMS-IV von David Wechsler*. Frankfurt/Main: Pearson Assessment.
- Peterson, D. J. & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1287–1293. <https://doi.org/10.1037/a0031337>
- Peterson, D. J., Schmidt, N. E. & Naveh-Benjamin, M. (2017). The role of schematic support in age-related associative deficits in short-term and long-term memory. *Journal of Memory and Language, 92*, 79–97. <https://doi.org/10.1016/j.jml.2016.05.007>
- Piccinin, A. M. & Rabbitt, P. M. A. (1999). Contribution of cognitive abilities to performance and improvement on a substitution coding task. *Psychology and aging, 14*(4), 539–551. <https://doi.org/10.1037/0882-7974.14.4.539>
- Pyc, M. A. & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35*(8), 1917–1927.  
<https://doi.org/10.3758/BF03192925>

- Pyc, M. A. & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437–447. <https://doi.org/10.1016/j.jml.2009.01.004>
- Pyc, M. A. & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science (New York, N.Y.)*, 330(6002), 335. <https://doi.org/10.1126/science.1191465>
- Pyc, M. A. & Rawson, K. A. (2011). Costs and benefits of dropout schedules of test-restudy practice: Implications for student learning. *Applied Cognitive Psychology*, 25(1), 87–95. <https://doi.org/10.1002/acp.1646>
- Raaijmakers, J. G. & Shiffrin, R. M. (1981). Search of associative memory. *Psychological review*, 88(2), 93–134. <https://doi.org/10.1037/0033-295X.88.2.93>
- Rast, P. (2011). Verbal knowledge, working memory, and processing speed as predictors of verbal learning in older adults. *Developmental Psychology*, 47(5), 1490–1498. <https://doi.org/10.1037/a0023422>
- Rast, P. & Zimprich, D. (2009). Individual differences and reliability of paired associates learning in younger and older adults. *Psychology and Aging*, 24(4), 1001–1006. <https://doi.org/10.1037/a0016138>
- Rawson, K. A., Vaughn, K. E. & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. *Memory & Cognition*, 43(4), 619–633. <https://doi.org/10.3758/s13421-014-0477-z>
- Recchia, A. (2010). R-squared measures for two-level hierarchical linear models using SAS. *Journal of Statistical Software*, 32, 1-9. <https://doi.org/10.18637/jss.v032.c02>
- Rheinberg, F., Vollmeyer, R. & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. *Diagnostica*, 47(2), 57–66. <https://doi.org/10.1026//0012-1924.47.2.57>

- Rickard, T. C. & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, 25(3), 847–869. <https://doi.org/10.3758/s13423-017-1298-4>
- Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In N. J. Smelser & P. B. Baltes (Hrsg.). *International encyclopedia of the social & behavioral sciences* (S. 8602–8605). Amsterdam: Elsevier.
- Roediger, H. L. (2012). The one-trial learning controversy and its aftermath: Remembering rock (1957). *The American Journal of Psychology*, 125(2), 127. <https://doi.org/10.5406/amerjpsyc.125.2.0127>
- Roediger, H. L. & Karpicke, J. D. (2006a). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science : a Journal of the Association for Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Roediger, H. L. & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roediger, H. L. & Karpicke, J. D. (2018). Reflections on the resurgence of interest in the testing effect. *Perspectives on Psychological Science: a Journal of the Association for Psychological Science*, 13(2), 236–241. <https://doi.org/10.1177/1745691617718873>
- Roediger, H. L. & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155–1159. <https://doi.org/10.1037/0278-7393.31.5.1155>
- Roediger, H. L. & Smith, M. A. (2012). The "pure-study" learning curve: the learning curve without cumulative testing. *Memory & Cognition*, 40(7), 989–1002. <https://doi.org/10.3758/s13421-012-0213-5>

- Rogalski, Y., Altmann, L. J. P. & Rosenbek, J. C. (2014). Retrieval practice and testing improve memory in older adults. *Aphasiology*, 28(4), 381–400.  
<https://doi.org/10.1080/02687038.2013.870965>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463.  
<https://doi.org/10.1037/a0037559>
- Rowland, C. A. & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory (Hove, England)*, 23(3), 403–419.  
<https://doi.org/10.1080/09658211.2014.889710>
- Salthouse, T. A. (1994). Aging associations: Influence of speed on adult age differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1486–1503.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, 103(3), 403–428. <https://doi.org/10.1037/0033-295X.103.3.403>
- SAS Institute Inc. (2016). SAS/STAT 9.4 User's Guide. Cary, NC: SAS Institute Inc.
- Sauz on, H., Claverie, B., & N'Kaoua, B. (2006). Age differences in the organization and acquisition-forgetting processes in a multi-free-recall task. *Current psychology letters*.  
Verf ugbar unter  
<https://pdfs.semanticscholar.org/c8b8/d6ae284dc5e191aeef52c574c473e9facfb.pdf>
- Sliwinski, M. & Buschke, H. (1999). Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychology and aging*, 14(1), 18–33.  
<https://doi.org/10.1037/0882-7974.14.1.18>
- Sternberg, R. J. & Tulving, E. (1977). The measurement of subjective organization in free recall. *Psychological bulletin*, 84(3), 539–556. <https://doi.org/10.1037/0033-2909.84.3.539>

- Storm, B. C., Friedman, M. C., Murayama, K. & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 115–124. <https://doi.org/10.1037/a0034252>
- Tse, C.-S., Balota, D. A. & Roediger, H. L. (2010). The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. *Psychology and Aging*, 25(4), 833–845. <https://doi.org/10.1037/a0019933>
- Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological review*, 69(4), 344–354. <https://doi.org/10.1037/h0043150>
- Umanath, S. & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science: a Journal of the Association for Psychological Science*, 9(4), 408–426. <https://doi.org/10.1177/1745691614535933>
- Veltre, M. T., Cho, K. W. & Neely, J. H. (2015). Transfer-appropriate processing in the testing effect. *Memory (Hove, England)*, 23(8), 1229–1237. <https://doi.org/10.1080/09658211.2014.970196>
- Verhaeghen, P., Marcoen, A. & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journal of gerontology*, 48(4), 157–171. <https://doi.org/10.1093/geronj/48.4.P157>
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J. & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538. <https://doi.org/10.3758/BRM.41.2.534>
- Wheeler, M. A. & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological science*, 3(4), 240–246. <https://doi.org/10.1111/j.1467-9280.1992.tb00036.x>
- Whiffen, J. W. & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1036–1046. <https://doi.org/10.1037/xlm0000379>

- Wiklund-Hörnqvist, C., Jonsson, B. & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, 55(1), 10–16.  
<https://doi.org/10.1111/sjop.12093>
- Witte, K. L., Freund, J. S. & Seiby, R. A. (1990). Age differences in free recall and subjective organization. *Psychology and aging*, 5(2), 307–309.  
<https://doi.org/10.1037/0882-7974.5.2.307>
- Zaromb, F. M. & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995–1008.  
<https://doi.org/10.3758/MC.38.8.995>
- Zimprich, D. (2002). Cross-sectionally and longitudinally balanced effects of processing speed on intellectual abilities. *Experimental Aging Research*, 28(3), 231–251.  
<https://doi.org/10.1080/03610730290080290>
- Zimprich, D. & Kurtz, T. (2013). Individual differences and predictors of forgetting in old age: The role of processing speed and working memory. *Aging, Neuropsychology, and Cognition: A Journal on Normal and Dysfunctional Development*, 20, 195–219.
- Zimprich, D. & Kurtz, T. (2015). Subjective and objective memory changes in old age across five years. *Gerontology*, 61(3), 223–231. <https://doi.org/10.1159/000369927>
- Zimprich, D. & Martin, M. (2002). Can longitudinal changes in processing speed explain longitudinal age changes in fluid intelligence? *Psychology and Aging*, 17(4), 690–695.  
<https://doi.org/10.1037//0882-7974.17.4.690>
- Zimprich, D., Rast, P., Martin, M. (2008). Individual differences in verbal learning in old age. In S. M. Hofer & D. F. Alwin (Hrsg.). *Handbook of cognitive aging. Interdisciplinary perspectives* (S. 224–243). Los Angeles: SAGE Publications.

Die Inhalte der Seite wurden aus Gründen des Datenschutzes entfernt.