

Institut für Bildungsforschung in der School of Education der  
Bergischen Universität Wuppertal

**Kognitive Lernvoraussetzungen von Schüler\*innen mit Lern-  
behinderung: Eine empirische Untersuchung von Prozessen und  
Strukturen des Arbeits- und Langzeitgedächtnisses**

Dissertation  
zur Erlangung des Grades eines  
Doktors der Philosophie (Dr. phil.)

vorgelegt von  
Gunnar Bruns

Wuppertal  
2021

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20210526-101712-2

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3A468-20210526-101712-2>]

DOI: 10.25926/zyyq-nv03

[<https://doi.org/10.25926/zyyq-nv03>]

1. Gutachterin: Prof. Dr. Jasmin Decristan, Bergische Universität Wuppertal
2. Gutachter: Prof. Dr. Jan Kuhl, Technische Universität Dortmund
3. Gutachter: Prof. Dr. Matthias Grünke, Universität zu Köln

Betreuer: Prof. Dr. Michael Grosche, Bergische Universität Wuppertal

Gemäß § 15 der Promotionsordnung des Instituts für Bildungsforschung in der School of Education der Bergischen Universität Wuppertal vom 17.11.2017 wurde die Dissertationsschrift am 02.02.2021 angenommen.

Tag der mündlichen Prüfung (Disputation): 05.03.2021

## Danksagung

Ich danke meinem Doktorvater Prof. Dr. Michael Grosche, der für mich die Rolle des Mentors, Ratgebers und Betreuers großartig ausgefüllt hat. Er bot mir einerseits ein großes Maß an Freiheit, um meine eigenen Ideen zu entwickeln, sie durchzuführen, aber dabei auch mal über Hindernisse zu stolpern und daraus zu lernen. Gleichzeitig fand ich in ihm jederzeit kompetente und verständnisvolle Beratung, in der ich mich ernstgenommen und wertgeschätzt fühlte. Die vielen Diskussionen und das gemeinsame Brennen für die Sache haben mich, mein Verständnis von und meine Begeisterung für die Wissenschaft wesentlich geprägt.

Prof. Dr. Jasmin Decristan und Prof. Dr. Jan Kuhl danke ich sehr herzlich für ihre Bereitschaft, auch in einer durch die ungewöhnliche Pandemie-Situation verstärkt arbeitsintensiven Zeit, meine Arbeit zu begutachten. Zudem gilt mein Dank den Schulen, Lehrkräften und besonders natürlich den Schüler\*innen, die an der Datenerhebung teilgenommen haben, sowie den Studierenden und studentischen Hilfskräften für ihre Hilfe bei der Datenerhebung – sie alle haben mit ihrer Mitwirkung die Realisierung des Projektes erst ermöglicht.

Dem gesamten Team des Arbeitsbereichs („Team Lernen“) mit allen seinen aktuellen und ehemaligen Mitgliedern bin ich für die intensive und spannende Zeit dankbar – wir sind gemeinsam als Team gewachsen und haben inhaltlich und methodisch als auch persönlich viel von- und miteinander gelernt. Birgit, Elisabeth, Gino, Jacquelin, Janine, Jessica, Michèle und Timo, ihr wart und seid mir unschätzbar wertvolle Kolleg\*innen. Ein besonderer Dank geht an Lea Pulst für ihre langjährige und stets sorgfältige und verlässliche Unterstützung, sei es bei Formatierungsfragen oder qualitativen Inhaltsanalysen. Außerdem danke ich den Mitgliedern der Graduate School für die kollegiale Unterstützung, die vielen fachlichen Diskussionen und Beratungen mit hilfreicher Kritik und Ideen, und auch die inspirierenden und ermutigenden Gespräche in allen Höhen und Tiefen des Doktoranden-Daseins.

Schließlich danke ich meiner Familie und Freunden für ihre stetige, warmherzige und Anteilnehmende Unterstützung in all den Jahren. Neben sanftem Druck und interessierten Nachfragen hielten sie mir stets den Rücken frei und boten eine willkommene Ablenkung und einen Ort, ein wenig Abstand zu gewinnen und für Ausgleich zu sorgen. Meinem Vater danke ich für seine sorgfältige Unterstützung in allen verzwickten Fragen der Zeichensetzung und seine hilfreichen Verständnisfragen. Euch allen, die ihr mich in den letzten Jahren in meinem Promotionsprojekt begleitet, ermutigt, geprägt und unterstützt habt, bin ich unendlich dankbar.

## Abstract

Diese Arbeit untersucht das verbale Arbeitsgedächtnis und das semantische Gedächtnis als zwei kognitive Lernvoraussetzungen an einer Stichprobe von  $n = 93$  Kindern mit Lernbehinderung und  $n = 120$  unbeeinträchtigten Kindern daraufhin, ob hinsichtlich der Verfügbarkeit und Funktionalität dieser Aspekte differenzielle Entwicklungsmuster vorliegen. Eine zentrale These dieser Arbeit ist, dass – bedingt durch die Dominanz der Strategiedefizit-Hypothese nach Klauer und Lauth (1997) – kognitiven Lernvoraussetzungen bisher eine zu geringe Bedeutung beigemessen wird.

Daher untersucht diese Arbeit die Fragestellung nach der Verfügbarkeit von Arbeits- und Langzeitgedächtnis in drei Studien. Studie 1 behandelt die Kapazität der phonologischen Schleife sowie den Rehearsal- und den Redintegrationprozess, die beide zur Aufrechterhaltung der Gedächtnisspur dienen. In der zweiten Studie werden die Binnen- und die Außenstruktur des mentalen Lexikons betrachtet. Schließlich sichert Studie 3 einen differenziellen Befund der ersten Studie methodisch ab.

Die Ergebnisse sprechen allgemein für das Muster einer Entwicklungsverzögerung. Schüler\*innen mit Lernbehinderung weisen in Arbeits- und Langzeitgedächtnis keine grundsätzlichen strukturellen Defizite auf, die über ihr allgemeines mentales Alter hinausgehen. Gleichzeitig war im spezifischen Aspekt der Interaktion von Arbeits- und Langzeitgedächtnis ein differenzieller Befund zu beobachten, der weder inhaltlich auf die Struktur des Langzeitgedächtnisses, noch auf ein methodisches Artefakt zurückzuführen ist. Hier schienen Schüler\*innen mit Lernbehinderung schlechter in der Lage zu sein, Wissen im Langzeitgedächtnis effektiv für die Rekonstruktion im Arbeitsgedächtnis zu nutzen.

Die Arbeit trägt zu einem besseren theoretischen und praktischen Verständnis des immer noch unscharf definierten Konstruktes ‚Lernbehinderung‘ bei, indem sie erstmals den Redintegrationprozess sowie Arbeits- und Langzeitgedächtnis gemeinsam in einer Stichprobe untersucht. Die Anwendung der innovativen Auswertungsmethode von Developmental Trajectories (Thomas et al., 2009) ermöglicht eine präzisere Beschreibung des Entwicklungsmusters und die Aufdeckung des differenziellen Befundes. Auf theoretischer Ebene werden die Ergebnisse vor dem Hintergrund der Strategiedefizit-Hypothese diskutiert; eine stärkere wissenschaftliche Aufmerksamkeit auf kognitive Lernvoraussetzungen erscheint durchaus angemessen. Für Unterricht und Intervention ist es sinnvoll, Anforderungen am mentalen Entwicklungsalter auszurichten und spezifisch bei Rekonstruktionsaufgaben zu reduzieren.

## Inhaltsverzeichnis

Einleitung.....	1
1. Eingrenzung der Zielgruppe.....	5
1.1. Begriffsdefinition.....	6
1.2. Begriffliche Unklarheiten.....	9
1.2.1. Unklarheiten hinsichtlich des Lernbehinderungs-Konstruktes .....	10
1.2.2. Verhältnis von Lernbehinderung und sonderpädagogischem Förderbedarf..	12
1.2.3. Abgrenzung zu anderen Begriffen .....	16
1.2.4. Internationale Begriffe .....	18
1.2.5. Fazit – Klarstellungen von Missverständnissen .....	19
1.3. Ursachenerklärung von Lernbehinderung.....	20
1.3.1. Transaktionales Entwicklungsmodell als Grundannahme.....	21
1.3.2. Strategie-Hypothese nach Klauer und Lauth.....	22
1.4. Zusammenfassung und Fazit zur Zielgruppe .....	25
2. Eingrenzung der untersuchten Domäne.....	28
2.1. Arbeitsgedächtnis.....	31
2.1.1. Aufbau der phonologischen Schleife.....	33
2.1.2. Empirische Effekte zur Untersuchung der phonologischen Schleife.....	36
2.1.3. Forschungsstand zum Arbeitsgedächtnis bei Lernbehinderung.....	41
2.1.4. Zusammenfassung Arbeitsgedächtnis .....	45
2.2. Langzeitgedächtnis.....	45
2.2.1. Forschungsstand zum Langzeitgedächtnis bei Lernbehinderung.....	48
2.2.2. Binnenstruktur des Langzeitgedächtnisses (1. Ansatz) .....	50
2.2.3. Außenstruktur des Langzeitgedächtnisses (2. Ansatz) .....	57
2.3. Zusammenfassung und Fazit zu Gedächtniskomponenten .....	62
3. Forschungsmethodischer Hintergrund.....	63
3.1. Mentaler Altersvergleich im Gruppensdesign .....	64
3.1.1. Messung des mentalen Alters.....	65
3.1.2. Inhaltliche Interpretation von MA-Szenarien .....	65
3.1.3. Anwendungsbeispiele für MA-Designs.....	67
3.1.4. Kritik am MA-Design .....	68
3.2. Developmental Trajectory Design.....	69
3.2.1. Anwendungsbeispiele für Developmental Trajectories .....	69
3.2.2. Konstruktion von Developmental Trajectories .....	70
3.2.3. Statistische Effekte in Developmental Trajectories.....	73
3.2.4. Inhaltliche Interpretation der Szenarien von Developmental Trajectories.....	75
3.3. Fazit – Mentaler Altersvergleich .....	78
4. Fragestellungen der Arbeit und ihrer Teilstudien.....	80

5.	Arbeitsgedächtnis-Kapazität und -Prozesse (Studie 1)	83
5.1.	Zusammenfassung von Studie 1: Abstract	83
5.2.	Forschungsfrage und Hypothesen	85
5.3.	Methode	86
5.3.1.	Stichprobe	86
5.3.2.	Sitzungen und Material	87
5.3.3.	Analysen	91
5.4.	Ergebnisse	94
5.4.1.	Kapazität der Phonologischen Schleife	94
5.4.2.	Effektivität des Rehearsalprozesses	96
5.4.3.	Effektivität des Redintegrationprozesses	98
5.5.	Diskussion	100
5.5.1.	Entwicklung der Arbeitsgedächtnis-Prozesse	100
5.5.2.	Limitationen	105
5.5.3.	Implikationen	106
6.	Struktur des Langzeitgedächtnisses (Studie 2)	108
6.1.	Einleitung und Fragestellung	108
6.2.	Methode	111
6.2.1.	Stichprobe und Material	111
6.2.2.	Auswertungsstrategien	114
6.2.3.	Methode zu Fragestellung 1: Binnenstruktur	116
6.2.4.	Methode zu Fragestellung 2: Außenstruktur	124
6.3.	Ergebnisse	128
6.3.1.	Ergebnisse zu Fragestellung 1: Binnenstruktur	128
6.3.2.	Ergebnisse zu Fragestellung 2: Außenstruktur	133
6.4.	Diskussion	136
6.4.1.	Zusammenfassung und Interpretation der Ergebnisse	136
6.4.2.	Limitationen und Ausblick	139
6.4.3.	Fazit Studie 2	143
7.	Absicherung des Redintegration-Effekts (Studie 3)	145
7.1.	Einleitung und Fragestellung	146
7.2.	Methode	149
7.2.1.	Ermittlung von Items mit Differential Item Functioning	150
7.2.2.	Developmental Trajectory Re-Analyse	151
7.3.	Ergebnisse	152
7.3.1.	Ausschluss von Items mit DIF	152
7.3.2.	Re-Analyse des Redintegration-Effekts	157
7.4.	Diskussion	158

7.4.1.	Zusammenfassung und Interpretation der Ergebnisse .....	158
7.4.2.	Limitationen.....	159
7.4.3.	Fazit Studie 3 .....	160
8.	Gesamtdiskussion.....	161
8.1.	Zusammenfassung und Interpretation der Ergebnisse .....	161
8.1.1.	Studie 1 – Arbeitsgedächtnis .....	161
8.1.2.	Studie 2 – Langzeitgedächtnis.....	163
8.1.3.	Studie 3 – Absicherung der Redintegration-Interaktion.....	164
8.2.	Limitationen.....	165
8.3.	Ausblick auf Netzwerkanalyse.....	168
8.4.	Theoretische und praktische Implikationen.....	171
8.4.1.	Theoretische Implikation: Strategie- vs. Processing-Hypothese.....	172
8.4.2.	Praktische Implikation .....	175
8.5.	Schlussbetrachtung .....	176
	Literaturverzeichnis .....	178
	Tabellenverzeichnis .....	206
	Abbildungsverzeichnis.....	207
	Anhang A – Tabellen .....	208
	Anhang B – Fachbeitrag (englischsprachiges Original) .....	218

## Einleitung

Ziel der vorliegenden Arbeit ist, mit den Komponenten des Arbeits- und des Langzeitgedächtnisses zwei kognitive Lernvoraussetzungen von Schüler\*innen mit Lernbehinderung genauer zu untersuchen. Der Begriff Lernbehinderung birgt auch nach fast fünfzig Jahren seit seiner Einführung im schulischen Kontext zahlreiche Probleme und Widersprüche. Besonders problematisch erscheinen insbesondere eine unscharfe Begriffsdefinition, und dass die Annahmen zu Erklärungsmodellen nur auf wenig aktuellen empirischen Grundlagen zu beruhen scheinen. Da konstatiert werden muss, dass zu wenig präzises Beschreibungswissen über die Problemlagen und insbesondere die kognitiven Voraussetzungen vorliegt, treffen diese Erklärungsmodelle häufig implizite und allgemein gehaltene Annahmen (z. B. Kretschmann, 2007; Vernooij, 2007), und es ist daher fraglich, inwiefern diese Erklärungsmodelle die Realität von Kindern mit Lernbehinderung adäquat abbilden können.

Der häufig zitierte und richtungsweisende Beitrag von Klauer und Lauth (1997) misst dem Bereich Strategie-Einsatz und Metakognition eine sehr hohe Bedeutung für die Erklärung der Probleme von Kindern mit Lernbehinderung bei. Dagegen werden kognitive Lernvoraussetzungen, wie Gedächtniskapazität und -struktur, als weniger bedeutsam erachtet. Gleichwohl sind die empirischen Ergebnisse, auf die sich Klauer und Lauth (1997) stützen, nicht sehr aktuell; neuere Forschungsarbeiten im Bereich des Arbeitsgedächtnisses kommen zudem zu abweichenden Ergebnissen, die darauf schließen lassen, dass kognitive Lernvoraussetzungen möglicherweise doch nicht so vollumfänglich und unbeeinträchtigt zur Verfügung stehen, wie von Klauer und Lauth (1997) angenommen. Zu anderen Bereichen wie dem semantischen Gedächtnis und dessen Struktur liegen kaum empirische Arbeiten vor.

Diese Forschungslücke wird in der vorliegenden Arbeit zum Anlass genommen, zwei Aspekte von kognitiven Lernvoraussetzungen bei Schüler\*innen mit Lernbehinderung zu untersuchen, die beide für das schulische Lernen eine wichtige Rolle spielen. Den ersten Aspekt stellt die phonologische Schleife als Teil des Arbeitsgedächtnisses dar, in der verbale Informationen für eine kurze Zeit aufrechterhalten und manipuliert werden können, damit sie dann im Langzeitgedächtnis gespeichert werden können. Der zweite Aspekt ist die Struktur des mentalen Lexikons als Teil des semantischen Langzeitgedächtnisses, da beim Lernen auf gut vernetztes Vorwissen zurückgegriffen wird.

Mit der Untersuchung dieser zwei kognitiven Lernvoraussetzungen wird daher beabsichtigt, aktuelles deskriptives Wissen auf einer empirischen Grundlage bereitzustellen. Aufbauend auf diesem Wissen kann das theoretische Verständnis des Konstruktes Lernbehinderung weiterentwickelt werden, denn es kann besser beurteilt werden, ob die genannten Aspekte der kognitiven Lernvoraussetzungen in ihrer Kapazität und Funktion voll zur Verfügung stehen, oder ob sie



beeinträchtigt sind. Falls eine Beeinträchtigung vorliegt, ist weiter zu differenzieren, ob es sich eher um ein Muster einer Entwicklungsverzögerung oder aber um eine gravierende strukturelle Abweichung (Mähler & Hasselhorn, 1990) handelt. Für eine differenziertere Betrachtung von Entwicklungsmustern eignet sich die innovative Auswertungsmethode der Developmental Trajectories (Thomas et al., 2009; siehe Kapitel 3.2), die in dieser Arbeit im Rahmen der ersten Studie zum Einsatz kommt. Somit ist diese Arbeit für die Weiterentwicklung des theoretischen Verständnisses von Lernbehinderung relevant, da sie die oben genannten allgemeinen Annahmen zu kognitiven Lernvoraussetzungen zum einen präzisieren und zum anderen empirische Hinweise für oder gegen diese Annahmen liefern soll.

Darüber hinaus ist die Untersuchung dieser Annahmen auch auf praktischer Ebene im Unterricht relevant, da sie Lehrkräften fundierteres Wissen für den schulischen Alltag bereitstellen soll. Damit kann es sie dabei unterstützen, die Schwierigkeiten ihrer Schüler\*innen beim Lernen besser nachzuvollziehen. Es erscheint hilfreich, dass Lehrkräfte besser einschätzen können, auf welche kognitiven Lernvoraussetzungen ihre Schüler\*innen zurückgreifen können, und welche Anforderungen sie möglicherweise vor größere Probleme stellen, indem z. B. die Kapazität des Arbeitsgedächtnisses zu stark beansprucht wird. Mit diesen Kenntnissen können Lehrkräfte Instruktionen und Interventionen besser an die Bedürfnisse ihrer Schüler\*innen anpassen und damit Barrieren zu Partizipation und Bildung abbauen.

Der folgenden Abschnitte geben einen Überblick über die Struktur der Arbeit. Die theoretische Grundlage der Arbeit wird in den ersten drei Kapiteln gelegt: das erste Kapitel behandelt die Definition, Eingrenzung und Probleme des Konstruktes Lernbehinderung und geht näher auf die Rolle der kognitiven Lernvoraussetzungen in Erklärungsmodellen ein. Dabei zeigt sich, dass in einschlägigen Publikationen kognitive Prozesse und Strukturen zwar durchaus thematisiert werden, jedoch anscheinend nur wenige und veraltete empirische Studien dazu vorliegen.

Im zweiten Kapitel werden daher zur thematischen Eingrenzung das Arbeits- und das Langzeitgedächtnis als zwei wesentliche am Prozess des Wissenserwerbs beteiligte Komponenten erläutert<sup>1</sup>: Das Arbeitsgedächtnis, welches in Studie 1 untersucht und anhand eines bereits etablierten Modells operationalisiert wird; und das Langzeitgedächtnis, zu dem in Ermangelung eines präzise messbaren Modells die theoretische Plausibilität der Annahmen von Studie 2 hergeleitet werden.

---

<sup>1</sup> Es sei bereits hier angemerkt, dass es trotz dieser Fokussierung keineswegs darum gehen soll, ein defizitorientiertes, rein an kognitiven Einschränkungen orientiertes, womöglich sogar monokausales Erklärungsmodell zu konstruieren – die Lebenswirklichkeit von Schüler\*innen ist deutlich komplexer und darf nicht auf die Betrachtung der Funktionsfähigkeit von Arbeits- und Langzeitgedächtnis reduziert werden.

Schließlich befasst sich das dritte Kapitel mit der methodischen Herangehensweise und dem Hintergrund des mentalen Altersvergleichs, wie die gewonnenen Daten für die Beantwortung der Frage nach möglicherweise beeinträchtigter Entwicklung interpretiert werden können. Die dort dargelegte erweiterte Methode der Developmental Trajectories spielt für die Arbeit (Studien 1 und 3, s.u.) eine zentrale Rolle und verspricht auch allgemein für die Art von Fragestellungen, wie sie in der Sonderpädagogik gestellt werden, ein hilfreiches Auswertungstool zu sein.

Die vorliegende Dissertationsschrift nähert sich der übergeordneten Fragestellung (Kapitel 4) nach der Verfügbarkeit oder spezifischen Beeinträchtigung der beiden kognitiven Lernvoraussetzungen Arbeitsgedächtnis und Langzeitgedächtnis bei Kindern mit Lernbehinderung in drei Teilstudien an.

Studie eins (Kapitel 5) betrachtet mit der Methode der Developmental Trajectories das phonologische Arbeitsgedächtnis und zwei dort verortete Teilprozesse zur Aufrechterhaltung bzw. Rekonstruktion der Gedächtnisspur: Rehearsal- und Redintegrationprozess. Zum Rehearsalprozess liegen bereits einige Befunde vor, die darauf hinweisen, dass in diesem Bereich Defizite oder zumindest eine Entwicklungsverzögerung vorliegen könnten. Redintegration wurde bislang in der Population nicht untersucht, ist aber insbesondere deshalb interessant, weil es die Rekonstruktion von Arbeitsgedächtnis-Inhalten aus dem Langzeitgedächtnis operationalisiert. Im Ergebnis scheint eine spezifische Interaktion zwischen Redintegration und dem Wortschatz (als Maß für das Langzeitgedächtnis) bei Kindern mit Lernbehinderung vorzuliegen, der im Rahmen der anschließenden beiden Studien weiter nachgegangen wird.

In der zweiten Studie (Kapitel 6) wird als weitere kognitive Lernvoraussetzung das Langzeitgedächtnis betrachtet. Die Untersuchung erfolgt in zwei getrennten Ansätzen, mit denen zwischen Binnenstruktur und Außenstruktur unterschieden wird. Neben der generellen Aussagekraft der Ergebnisse für die übergeordnete Fragestellung, dass nämlich die Langzeitgedächtnis-Struktur sich zwischen Kindern mit Lernbehinderung und jüngeren, nach mentalem Alter vergleichbaren unbeeinträchtigten Kindern anscheinend nicht unterscheidet, lassen sich die Ergebnisse auch vor dem Hintergrund der Interaktion aus Studie 1 interpretieren: ist die Interaktion möglicherweise inhaltlich mit einer abweichenden Struktur des Langzeitgedächtnisses erklärbar?

Die dritte Studie (Kapitel 7) ist schließlich als methodische Folgestudie zu Studie 1 angelegt, in der abgesichert wird, dass es sich bei der Redintegration-Interaktion mit dem Langzeitgedächtnis nicht um ein methodisch bedingtes Artefakt handelte. Da die Messäquivalenz auf Seiten des Instrumentes zur Messung des Langzeitgedächtnisses möglicherweise nicht bei allen Items gegeben ist (also evtl. Differential Item Functioning vorliegt), werden problematische Items aus der Skala ausgeschlossen, mit der bereinigten Skala wird dann die Interaktion mit einem Developmental Trajectory einer Re-Analyse unterzogen.

Die Gesamtdiskussion (Kapitel 8) fasst die Ergebnisse zusammen: Insgesamt betrachtet zeigen Schüler\*innen mit Lernbehinderung hinsichtlich der kognitiven Lernvoraussetzungen Arbeits- und Langzeitgedächtnis vorrangig Muster einer verzögerten Entwicklung; strukturelle und über das Maß des generellen mentalen Alters hinausgehende Defizite lassen sich fast flächendeckend nicht feststellen. So sprechen sowohl die Befunde aus Studie 1 zu Arbeitsgedächtnis-Kapazität und -prozessen als auch die Ergebnisse der Studie 2 zu Binnen- und Außenstruktur generell für die Annahme eines Musters von verzögerter Entwicklung.

Insbesondere wird dort auf einen differenziellen Befund zum Redintegrationprozess eingegangen, den die Strategie-Hypothese nicht angemessen zu erklären vermag. Insofern erscheint die Vernachlässigung von kognitiven Lernvoraussetzungen in der Erforschung der Bedingungen von Kindern mit Lernbehinderung nicht gerechtfertigt. Vielmehr zeigt die vorliegende Arbeit, dass sie nicht so bedeutungslos sind, wie aus Klauer und Lauth (1997) herausgelesen werden könnte. Es erscheint also auch für zukünftige Forschung lohnend, diese genauer zu untersuchen, um das Phänomen der Lernbehinderung präziser zu beschreiben. Von einer solchen empirischen Deskription profitieren sowohl das theoretische Verständnis als auch die pädagogische Praxis, da die Ergebnisse die Notwendigkeit von angepassten Instruktionen begründen.

## 1. Eingrenzung der Zielgruppe

In diesem ersten Kapitel wird die Zielgruppe dieser Arbeit definiert und eingegrenzt und das Forschungsinteresse an der Betrachtung von kognitiven Lernvoraussetzungen in dieser Population begründet. Im Fokus der Arbeit stehen diejenigen Schüler\*innen, die eine Lernbehinderung mit einem sonderpädagogischen Förderbedarf im Förderschwerpunkt Lernen aufweisen. Auch wenn das akademische Fach, in seinen Ursprüngen noch „Lernbehindertenpädagogik“ (Klauer, 1977) genannt, auf eine über fünfzigjährige Geschichte zurückblicken kann, bleiben doch wichtige Fragen einer präzisen Definition, eindeutig festgelegter und testbarer diagnostischer Kriterien und eines empirisch solide begründeten Erklärungsmodells bis heute unbeantwortet. Die vorliegende Arbeit soll einen Beitrag dazu leisten, das Konstrukt von Lernbehinderung und die kognitiven Herausforderungen, denen Kinder mit Lernbehinderung in ihrem schulischen Alltag begegnen, näher zu beleuchten und dadurch ein Stück weit besser zu verstehen.

Das erste Kapitel gliedert sich in vier Schritte. Im ersten Schritt wird anhand von Definitionen der Begriffe ‚Lernbehinderung‘ und ‚sonderpädagogischer Förderbedarf‘ die Eingrenzung auf die Zielpopulation vorgenommen, die sich in einer operationalen Setzung von Ein- und Ausschlusskriterien zur Bestimmung der Stichprobe niederschlägt. Diese Eingrenzung erfolgt im Bewusstsein, dass die Begriffe keineswegs unproblematisch sind.

Daher werden im zweiten Schritt einige Probleme der Begrifflichkeiten Lernbehinderung und sonderpädagogischer Förderbedarf und ihres Verhältnisses zueinander beleuchtet. Die historisch gewachsene Begriffsvielfalt trägt ebenfalls zu Unklarheiten und Unschärfen bei, daher wird überblickartig eine knappe Abgrenzung von Lernbehinderung zu anderen Bezeichnungen vorgenommen. Der Schluss dieses zweiten Schrittes geht auf drei Missverständnisse ein und stellt sie für diese Arbeit klar.

Im dritten Schritt wird auf die Ursachenerklärung von Lernbehinderung eingegangen. Ist bereits die empirische Datenlage zur Beschreibung des Phänomens Lernbehinderung als unzureichend und veraltet zu bewerten, so gilt dies insbesondere für die Versuche, ein Erklärungsmodell zur Entstehung von Lernbehinderung zu formulieren. Dabei wird für diese Arbeit besonderes Augenmerk auf die Bedeutung von kognitiven Lernvoraussetzungen gelegt.

Abschließend wird im vierten Schritt als Fazit abgeleitet, dass es relevant ist, kognitive Lernvoraussetzungen bei Kindern mit Lernbehinderung empirisch zu untersuchen, um eine bessere Deskription der möglichen Schwierigkeiten, aber auch Stärken dieser Zielgruppe zu liefern, womit jedoch anderen Erklärungsansätzen ihre Bedeutung keineswegs abgesprochen werden soll.

## 1.1. Begriffsdefinition

In dieser Arbeit werden Schüler\*innen mit Lernbehinderung und sonderpädagogischem Förderbedarf im Förderschwerpunkt Lernen untersucht. Gemäß Grünke und Grosche (2014, S. 76) kann der Begriff Lernbehinderung definiert werden als „schwerwiegende, anhaltende und umfängliche Schwierigkeiten bei der Bewältigung von intellektuellen Leistungsanforderungen“, die sich insbesondere auf das absichtsvolle schulische Lernen „kognitiv-verbaler und abstrakter Inhalte (v. a. Lesen, Rechtschreiben und Rechnen)“ (ebd., S. 76) negativ auswirken. Die Leistungsrückstände sind als schwerwiegend zu bezeichnen, wenn sie „mindestens zwei bis drei Schuljahre“ betragen, als anhaltend, wenn sie „über mehrere Jahre“ persistieren, und als umfänglich, wenn sie „mehrere Unterrichtsfächer“ betreffen (ebd., S. 78).

Zahlreiche weitere Definitionen nehmen ebenfalls Bezug auf die drei Dimensionen schulischer Probleme *Umfang*, *Dauer* und *Schweregrad*, die im Bestreben einer ordnenden Einteilung verschiedener Formen von Lernbeeinträchtigungen bereits früh eingeführt wurden (Kanter, 1980; Klauer, 1977). So wird nach Kanter (1980, S. 47) ein Mensch „als lernbehindert [...] bezeichnet, der in seinem Lernen schwerwiegend, langdauernd und umfänglich so beeinträchtigt ist, daß deutlich normabweichende Verhaltens- und Leistungsformen sichtbar werden.“ Klauer beruft sich wiederum auf Kanter (1974, zit. n. Klauer, 1977, S. 6), dass eine Lernbehinderung durch „eine herabgesetzte schulische Lernleistung gekennzeichnet“ sei, die von schwerwiegendem, umfänglichen und langdauernden Ausmaß sei. Ähnliches gilt für die Definitionen von Lauth und Schlottke (2005) und auch die dimensional angelegte Definition von Klauer und Lauth (1997).

In diesen älteren Ansätzen (Kanter, 1980; Klauer, 1977; vgl. Heimlich, 2009) wurden die drei Dimensionen von Dauer, Schweregrad und Umfang zur Charakterisierung von Lernbeeinträchtigungen im Allgemeinen, und zur Differenzierung zwischen „leichteren“ Lernstörungen<sup>2</sup> und schwerwiegenderen „Lernbehinderungen“ im Besonderen herangezogen (Heimlich, 2009; Kanter, 1980; Klauer & Lauth, 1997; Lauth & Schlottke, 2005). Nach diesem Verständnis war eine Lernbehinderung in allen drei Dimensionen stärker ausgeprägt als eine Lernstörung: die generellen Minderleistungen betrafen mehrere Schulfächer (umfänglicher), waren insgesamt gravierender (schwerwiegender), und persistierten über mehrere Schuljahre hinweg (langdauernder). Hinsichtlich des Schweregrades ordnete Klauer (1977, S. 6f.) die Lernbehinderung als mittlere Form zwischen einer leichten Form der Lernstörung und einer schweren Form der geistigen

---

<sup>2</sup> Es ist zu beachten, dass dieser Begriff der „Lernstörung“ vom heutigen Verständnis einer spezifischen Lernstörung (Gold, 2015; siehe Kapitel 1.2.3) abweicht.

Behinderung ein. Auch wenn eine Lernbehinderung langdauernd ist, bedeute dies aber nicht, dass sie „unabänderlich“ sein müsse (ebd., S. 7).

Klauer und Lauth (1997) öffnen ein zweidimensionales Schema mit den potentiell kontinuierlichen Achsen „Zeit“ und „Umfang oder Breite“ der schulischen Minderleistung. In diesem Schema wird Lernbehinderung als dauerhafte und domänenübergreifende Beeinträchtigung eingeordnet und von zwar ebenfalls überdauernden, jedoch domänenspezifischen Beeinträchtigungen wie Lese-Rechtschreib-Störung oder Dyskalkulie abgegrenzt. Auf das Kriterium einer Intelligenzbeeinträchtigung wird hier verzichtet.

Im Gegensatz dazu gibt der Deutsche Bildungsrat (1976) in den Empfehlungen der Bildungskommission vor, dass eine Lernbehinderung bei Intelligenzwerten im Bereich zwischen 55-85 IQ-Punkten und zugleich vorliegendem oder zu erwartendem „erhebliche[n] Schulversagen“ (ebd., S. 38) anzunehmen sei. In der Entstehung wirken demnach organische und soziale Faktoren zusammen (sodass „die Entstehungsursachen der Lernbehinderung nicht eindeutig aufweisbar sind“, ebd., S. 38) und führen zusammen mit niedriger Intelligenz zu beeinträchtigten Lernleistungen. Die schwachen Lernleistungen äußerten sich darin, „daß die Aufnahme, Speicherung und Verarbeitung von Lerninhalten nicht in altersentsprechender Weise gelingt.“ (ebd., S. 38).

Schröder (2000) verweist auf den Umstand, dass aufgrund der negativen Definition über den Ausschluss von anderen Arten, auf die Lernen ‚behindert‘ werden kann (z. B. aufgrund von Sinnesschädigungen oder anderen Behinderungsformen) Lernbehinderung zu einer „Restkategorie erklärt“ (S. 73) würde, da die Kategorie nur für diejenigen zutrifft, deren Probleme beim Lernen nicht anderweitig verursacht und erklärbar sind.

Der Begriff Lernbehinderung ist mit dem Schulsystem und insbesondere mit dem Begriff des sonderpädagogischen Förderbedarfs historisch eng verwoben (Souvignier, 2008). So wurde der Begriff Lernbehinderung zunächst im Rahmen einer Neuordnung des Schulsystems von der Ständigen Konferenz der Kultusminister der Länder (KMK) geprägt (KMK, 1960) und u. a. von Klauer (1977) aufgenommen, um die Schülerschaft der Hilfsschule zu charakterisieren. Lernbehinderung muss damit deutlich als schulorganisatorische Verwaltungskategorie bezeichnet werden (Grünke & Grosche, 2014; Lauth & Schlottke, 2005), die häufig synonym zum „sonderpädagogischen Förderbedarf im Förderschwerpunkt Lernen“ gebraucht wird (Haeberlin, 1991). Dieses Verhältnis von Lernbehinderung und sonderpädagogischem Förderbedarf birgt eine Reihe von Problemen, die in Abschnitt 1.2 näher untersucht werden.

Ein sonderpädagogischer Förderbedarf ist laut den Empfehlungen der Ständigen Konferenz der Kultusminister der Länder dann gegeben, wenn bei Schüler\*innen „in ihrer Lern- und Leistungsentwicklung so erheblich[e] Beeinträchtigungen [vor]liegen, dass sie auch mit zusätzlichen Lernhilfen der allgemeinen Schulen nicht ihren Möglichkeiten entsprechend gefördert werden können“ (KMK, 1999, S. 4)<sup>3</sup>. Ein sonderpädagogischer Förderbedarf berechtigt Schüler\*innen mit „Beeinträchtigungen des Lernens [...] Hilfe durch Angebote im Förderschwerpunkt Lernen“ (ebd., S. 2) in Anspruch zu nehmen. Im Gegensatz zur früheren Fokussierung auf die Sonder- oder Förderschule als einziger Förderort kann diese sonderpädagogische Förderung sowohl in der Allgemeinen Schule als auch in einer Förderschule angeboten werden (Schmetz, 2004).

Die Häufigkeit eines Phänomens allein ist als Argument für die wissenschaftliche Beschäftigung damit sicherlich nicht hinreichend, denn die Festlegung, ab wann ein Phänomen ‚gravierend genug‘ ist, bleibt normativ und letztlich arbiträr. Der Vollständigkeit halber und um auf deskriptiver Ebene über die Ausmaße der Problematik im Bilde zu sein, sollen die Zahlen zur Auftretenshäufigkeit hier aber Erwähnung finden. Betrachtet man die amtlichen Schulstatistiken (KMK, 2016), so zeigt sich, dass der Förderschwerpunkt Lernen lange Jahre den größten Anteil der Schüler\*innenschaft mit sonderpädagogischem Förderbedarf ausmachte. In absoluten Zahlen wurde im Jahr 2005 insgesamt 234.272 Schüler\*innen und im Jahr 2014 insgesamt 191.546 Schüler\*innen ein sonderpädagogischer Förderbedarf im Schwerpunkt Lernen zuerkannt. Im Jahr 2018 waren es deutschlandweit 192.600 Schüler\*innen mit sonderpädagogischem Förderbedarf Lernen, was einer Förderquote, d. h. einem Verhältnis zur Anzahl aller Schüler\*innen, von 2.61% entspricht (KMK, 2020). Innerhalb der Schülerschaft mit sonderpädagogischem Förderbedarf hat sich die anteilmäßige Dominanz der Schüler\*innen im SPF-Lernen von 48,1% im Jahr 2005 (KMK, 2016) auf 37.7% im Jahr 2014 und weiter auf 34.6% im Jahr 2018 reduziert (KMK, 2020). Diese Angaben beziehen sich jedoch lediglich auf diejenigen Kinder und Jugendlichen, bei denen ein amtlicher sonderpädagogischer Förderschwerpunkt im Rahmen eines AO-SF Verfahrens diagnostiziert wurde.

Insgesamt betrachtet handelt es sich bei dem Begriff Lernbehinderung um eine Beschreibungskategorie für Probleme beim schulischen Lernen (Grünke & Grosche, 2014), die derart gravierend sind, dass sie die Normerwartungen der allgemeinen Schule deutlich unter- und deren Fördermöglichkeiten überschreiten (Schröder, 2000). Das zentrale und augenfälligste Merkmal von Schüler\*innen mit Lernbehinderung manifestiert sich in „deutlich unterdurchschnittliche[n] schulische [n] Leistungen [...], die im Unterricht der Allgemeinen Schule auch bei

---

<sup>3</sup> Weniger klar formuliert aber sinngemäß gleich lautet die neue Version der Empfehlungen aus dem Jahr 2019 (KMK, 2019)

besonderer Förderung nicht ausgeglichen werden können und sich in dieser Weise als schwerwiegende, umfängliche und langdauernde Lernbeeinträchtigung zeigen“ (Schröder, 2000, S. 169).

Über eine solche symptomatische Beschreibung bzw. Verwendung als Label eines Syndroms (Kleber, 1980) sollte die Interpretation von Lernbehinderung nicht hinausgehen; entsprechend wird in dieser Arbeit auch lediglich eine operationale Definition anhand von Ein- und Ausschlusskriterien angelegt, damit die empirische Untersuchbarkeit gewährleistet ist. Diese Setzung erfolgt vor dem Hintergrund und im Bewusstsein, dass die Begrifflichkeiten und ihre wechselseitigen Bezüge nicht zuletzt angesichts vielfältiger Deutungsprozesse und Missverständnisse (Kanter, 2007b) überaus problematisch sind. Hierauf wird im folgenden Abschnitt 1.2 näher eingegangen.

Als operationale Einschlusskriterien für diese Arbeit wurden das Vorliegen eines amtlich festgestellten sonderpädagogischen Förderbedarfs mit einem primären Förderschwerpunkt „Lernen“ sowie der Besuch einer Förderschule für den Förderschwerpunkt „Lernen“ festgelegt. Zudem hatte der Intelligenzquotient (IQ) in einem Bereich zwischen 1-3 Standardabweichungen unter dem Durchschnitt zu liegen, was Werten zwischen 55-85 IQ-Punkten (oder 20-40 *T*-Wert-Punkten) entspricht. Als Ausschlusskriterien wurde das Vorliegen anderer Störungsbilder definiert, wie z. B. spezifische Teilleistungsstörungen (Dyslexie, Dyskalkulie), aber auch Beeinträchtigungen aus dem Autismus-Spektrum oder hyperkinetische Störungen.

Damit entspricht die gewählte Definition den fünf Kriterien zur Differentialdiagnostik einer Lernbehinderung nach Grünke und Grosche (2014): Durch das Einschlusskriterium der Diagnose eines sonderpädagogischen Förderbedarfs im Förderschwerpunkt Lernen wird das Kriterium nicht normentsprechender Schulleistungen (1) und einer Analyse der bisherigen Schulentwicklung (3) als gegeben angenommen; eine unterdurchschnittliche Allgemeinintelligenz (2) wurde als Einschlusskriterium explizit gesetzt, und Sinnesschädigungen (4) sowie komorbide Verhaltensstörungen (5) wurden über die Ausschlusskriterien kontrolliert.

### **1.2. Begriffliche Unklarheiten**

Die oben zitierten Definitionen von Lernbehinderung und sonderpädagogischem Förderbedarf verdeutlichen vor allem das übergreifende Problem, dass weder wirklich trennscharfe und operationalisierbare Begriffsbestimmungen, noch klare und allgemein anerkannte diagnostische Kriterien vorliegen. So wird Lernbehinderung in internationalen Klassifikationssystemen wie dem ICD-10 (Deutsches Institut für Medizinische Dokumentation und Information, DIMDI, 2020) nicht aufgeführt, wodurch es nicht klinisch als psychische Beeinträchtigung diagnostizier-



bar ist, sondern ein schulorganisatorisches Konstrukt bleibt (Kanter, 2007b), das mit dem sonderpädagogischen Förderbedarf eng verwoben ist. Diese Unklarheit hat weitere Probleme zur Folge und führt zu Missverständnissen, auf die nun näher eingegangen werden soll.

Die angesprochene Problematik des Begriffs und des dahinterliegenden Konstrukts (der Sachlage) von Lernbehinderung konnte in den vergangenen gut 50 Jahren seit seiner Einführung trotz zahlreicher Versuche einer Orientierung und Klärung nicht befriedigend gelöst werden. Zudem scheint sich ein Teil des jüngeren Diskurses, was unter verschiedenen Formen des Oberbegriffs Lernbeeinträchtigung zu verstehen sei, in den Bereich der pädagogischen Psychologie verlagert zu haben (z. B. Gold, 2011; 2015; Hasselhorn & Mähler, 2007). Es erscheint daher wenig aussichtsreich oder zweckmäßig, den Diskurs und den historischen Verlauf der Begriffsbildung seit Klauer (1977) und Kanter (1980) vollumfänglich nachzuzeichnen und zu hoffen, dass dadurch eine bessere Klärung des Konstruktes gelingt.

Zudem hat sich, in Ermangelung klarer Kriterien und angesichts einer unscharfen Definition des sonderpädagogischen Förderbedarfs, eine verwirrende Vielfalt an Begriffen zur Beschreibung und Klassifikation von Schwierigkeiten im absichtsvollen, institutionalisierten Lernen herausgebildet. Für eine Einordnung und kritische Reflexion der Begriffsentwicklung sei daher auf die einschlägigen Beiträge von Schröder (2000), Heimlich (2009), Kanter (2007b) und Gold (2015) verwiesen. An dieser Stelle soll nur versucht werden, einige besonders gravierende Probleme des Begriffs und des Konzepts zu ordnen und zusammenzufassen, sowie in Abschnitt 1.2.3 eine knappe Einordnung und Abgrenzung von Lernbehinderung zu anderen Bezeichnungen wie „Lernschwierigkeit“, „Lernbeeinträchtigung“, „Lernstörung“ und „Lernschwäche“ vorzunehmen.

### **1.2.1. Unklarheiten hinsichtlich des Lernbehinderungs-Konstruktes**

Das erste Problem besteht darin, dass Lernbehinderung lediglich als ein hypothetisches, und dazu äußerst unklares, Konstrukt zu verstehen ist (Schröder, 2000). Wie Kanter (2007b, S. 18) aufzeigt, gibt es keinen absoluten „Wahrheitsanspruch“, da auch der zunächst arbiträre Begriff Lernbehinderung Bestimmungs- und Deutungsprozessen von verschiedenen Seiten unterliegt, was angesichts der unterschiedlichen Professionshintergründe der Beteiligten fast unweigerlich zu Missverständnissen oder sogar absichtlichem Missbrauch führte.

Die „Sache“ (der Gegenstand) selbst hinter dem Konstrukt bleibt aber weiterhin unklar. So konnte bisher nicht eindeutig geklärt werden, was genau eine Lernbehinderung ist oder ausmacht, wer von ihr (warum) betroffen ist, und ob bzw. wie sie sich von anderen Formen beeinträchtigten Lernens sinnvoll abgrenzen lässt. Diese Probleme lassen sich auch nicht durch „Sprachspiele“ (Kanter, 2007b, S. 18), die Einführung neuer Begriffe (Schröder, 2000, S. 71)

oder auf „verbalanalytisch[em]“ Wege lösen (Kanter, 1980, S. 17). Vielmehr hat dies dazu beigetragen, dass sich ein veritables Begriffschaos oder „terminologische[s] Wirrwarr“ (Klauer & Lauth, 1997, S. 703) herausgebildet hat.

Bei aller terminologischer Diskussion darf nicht vergessen werden, dass pragmatisch betrachtet hinter dem abstrakten Konstrukt Lernbehinderung eine Vielzahl von Schüler\*innen steht mit „faktisch [beobachtbaren, manifesten] [...] Lern- und Sozialprobleme[n], die unter bestimmten Bedingungen eben mit diesem Begriff belegt werden“ (Schröder 2000, S. 70f.). Daher kann der Vorschlag, die Debatte um den Lernbehinderungsbegriff durch die Abschaffung der „Sonderschule für Lernbehinderte“ zu einem Ende zu bringen (Heimlich, 2009), sicherlich nicht die massiven, realiter auftretenden Lernschwierigkeiten der betroffenen Kinder auflösen. Eine empirische Untersuchung zum Zweck einer präziseren Beschreibung der Ausgangssituation, wie sie diese Arbeit leisten soll, erscheint daher notwendig und hilfreich.

Mit den im Rahmen der Definition eingeführten drei Dimensionen von Dauer, Schweregrad und Umfang steht zwar ein erster Ansatz zur Systematisierung zur Verfügung, der von Schröder (2000, S. 81) als „vernünftige Mitte zwischen Begriffslosigkeit und [...] Wirrwarr“ bewertet wird. Aber auch an diesen Dimensionen und der aus ihnen resultierenden Einteilung von Lernschwierigkeiten ist Kritik laut geworden. So merkt Heimlich (2009) an, dass die anhand der drei genannten Dimensionen getroffene Unterscheidung zwischen Lernbehinderung und Lernstörung<sup>4</sup> sich insbesondere in der Praxis nicht als trennscharf erwiesen hat, sodass die auf den ersten Blick einfache Unterscheidung de facto den Anspruch einer klaren konzeptuellen Aufteilung nicht erfüllte, sondern dass „Überschneidungen zwischen beiden Ausprägungen von Lernbeeinträchtigungen [Lernbehinderung und Lernstörung] eigentlich der Regelfall war“ (Heimlich, 2009, S. 21).

Um dem zu begegnen und eine typologische Gruppenzuschreibung zu vermeiden und stattdessen die graduellen Abstufungen zu betonen, schlagen Klauer und Lauth (1997, S. 703f.) ein dimensionales (kontinuierliches) Modell vor, in dem prinzipiell alle Punkte der zwischen Dauer und Umfang aufgespannten Ebene besetzbar seien. Schröder (2000, S. 82) bewertet dies insofern kritisch, als der Anspruch, jeden Punkt in diesem Kontinuum zu besetzen, nicht einzulösen sei und, insbesondere aufgrund der auf einer unsicheren Prognose beruhenden Einordnung auf der Zeit-Dimension, die Punktschätzung nur eine scheinbare Präzision darstelle.

In der zeitlichen Dimension liegen zwei Probleme begründet. Erstens kann, wie von Schröder (2000) kritisiert, die Einschätzung der Zeitdimension auf einer Prognose beruhen. Dies ist aufgrund der angesprochenen Unsicherheit der Schätzung problematisch. Andererseits

---

<sup>4</sup>Wiederum im alten Verständnis als „leichte Form“ einer Lernbeeinträchtigung.

kann zweitens die zeitliche Verfestigung der Lernprobleme *bisher* betrachtet werden. Darin sieht Heimlich (2009) eine „wait-to-fail“-Problematik, dass also die Probleme eines Kindes erst manifest und so gravierend geworden sein müssen, dass sie als normabweichend gelten, bevor sie „behandelt“ werden können. Der Präventionsgedanke komme hier deutlich zu kurz. Insofern handelt es sich mit Schröder (2000) bei Lernbehinderung um eine ex-post Definition, da man immer erst hinterher mit Sicherheit weiß, ob die Probleme so gravierende Ausmaße angenommen haben, dass von einer Lernbehinderung gesprochen werden kann.

Als äußerst problematisch ist zudem die Rolle der Intelligenzleistung im Zusammenhang mit Lernbehinderung anzusehen. In der Vergangenheit war die Intelligenz ein zentrales Kriterium bei der Identifikation von Lernbehinderung (Heimlich, 2009). Klauer (1977, S. 5) setzt sogar schulische Lernfähigkeit mit Intelligenz gleich. Auch die vom Deutschen Bildungsrat (1976) formulierten Kriterien beinhalten neben den gravierenden Minderleistungen einen Bereich von IQ-Werten von 55-85 Punkten. Gold (2015) setzt den Begriff der „intellectual disabilities“ (S. 125) mit der Lernbehinderung gleich. Demgegenüber wird diese zentrale Rolle des IQ in der Diagnostik in neuerer Zeit durchaus kritisiert (z. B. Heimlich & Wember, 2014), da die einseitige Fokussierung auf den IQ die Gefahr birgt, dass andere Lernvoraussetzungen übersehen werden könnten (Heimlich, 2009).

### **1.2.2. Verhältnis von Lernbehinderung und sonderpädagogischem Förderbedarf**

Ein Kernproblem, das sich auf sämtliche andere Bereiche auswirkt, ist im Verhältnis der Begriffe Lernbehinderung und sonderpädagogischer Förderbedarf, bzw. dessen Vorläufern „Hilfsschul“- und „Sonderschulbedürftigkeit“, zu sehen.<sup>5</sup> Es bleibt offen, inwiefern diese Begriffe deckungsgleich gemeint oder verwendet werden, oder ob sie doch leicht unterschiedliche Aspekte eines Konstruktes betreffen. Um die Problematik des Verhältnisses zu ordnen, werden im folgenden vier Punkte angeführt: (1) die historisch bedingte enge Verwobenheit der Begriffe; (2) die Unklarheit der Begriffe einzeln betrachtet; (3) die Nutzung der Begriffe, um sich wechselseitig zu erklären; (4) widersprüchliche Aussagen, ob die Begriffe deckungsgleich sind oder nicht.

Zum ersten Punkt: Historisch betrachtet sind die Begriffe Lernbehinderung und sonderpädagogischer Förderbedarf eng miteinander verwoben (Grünke & Grosche, 2014; Heimlich, 2009; Schröder, 2000), sodass diese wechselseitig aufeinander verweisen und voneinander abhängig sind (siehe auch Punkt 3). Im Zuge der Neuordnung des Sonderschulwesens (KMK,

---

<sup>5</sup> Im Folgenden ist aus Konsistenzgründen immer von sonderpädagogischem Förderbedarf die Rede, auch wenn möglicherweise zu einem früheren Zeitpunkt ein anderer Begriff aktuell war.

1960, 1972) wurde die Schüler\*innenschaft der Hilfsschule mit dem Begriff der Lernbehinderung belegt (Schröder, 2000). Damit sollte erstens der Zuschnitt der Sonderschulen einheitlich auf ‚Behinderungsformen‘ vorgenommen werden. Zweitens erhoffte man sich eine Abnahme von Diskriminierung der bislang an den Hilfsschulen unterrichteten Schüler\*innen durch das Stigma der „Schwachsinnigen“ (Schröder, 2000). Von Klauer (1977) wurde der Begriff Lernbehinderung für ein Lehrbuch zur „Lernbehindertenpädagogik“ aufgegriffen. Darin zeigt sich das Problem, dass ein schulorganisatorischer, aus praktischen Gegebenheiten entstandener Begriff wissenschaftlich gefüllt werden musste – dies war problematisch und ist es bis heute geblieben.

Zum zweiten Punkt: Jeder Begriff für sich genommen birgt Unschärfen in Form einer unklaren Definition. So ist Lernbehinderung lediglich als ein „Arbeitskonzept“ (Kleber, 1980) zu verstehen, das daher keinen Anspruch auf ‚finale‘ Gültigkeit erheben kann und ständiger Aktualisierung bedarf. Dies führt zu Missverständnissen, von denen hier drei genannt werden und auf die zum Abschluss des Kapitels näher eingegangen wird: Indem Lernbehinderung auf die gleiche Ebene wie z. B. die „Sprachbehinderung“ gesetzt wurde, kam häufig das Missverständnis auf, Lernbehinderung sei ein ursächlicher Faktor (Kanter, 2007b) für die schulischen Minderleistungen. Desweiteren kann Lernbehinderung als eine Beeinträchtigung der „allgemeinen Lernfähigkeit“ (Kanter, 1980, S. 47; Schröder, 2000, S. 73) missverstanden werden. Schließlich wäre es ein Missverständnis, Lernbehinderung als Eigenschaft oder ein Wesensmerkmal einer Person anzusehen (Schröder, 2000, S. 94).

Stattdessen hat der Begriff Lernbehinderung rein beschreibenden Charakter (Kanter, 2007b; Schröder, 2000), um oberflächlich den Symptomkomplex oder das „Syndrom“ (Kleber, 1980) von (nicht erklärbaren) schulischen Minderleistungen begrifflich zu benennen, ohne dass klare diagnostische Kriterien für die Identifikation von Lernbehinderung vorliegen. So konstatiert Schröder (2000), dass Lernbehinderung vor allem eine „Restkategorie“ (S. 73) sei, nämlich das, was übrigbleibe, wenn alle diejenigen Schüler\*innen ausgeschlossen werden, deren schulische Minderleistungen auf andere Faktoren, z. B. durch Sinnesschädigung oder Körperbehinderung, zurückzuführen sind. Für die Kategorie Lernbehinderung kommen diejenigen Schüler\*innen in Frage, deren gravierende schulische Schwierigkeiten nur allgemein mit niedriger Intelligenzleistung erklärbar sind. Schüler\*innen mit spezifischen Lernstörungen<sup>6</sup> sind ebenfalls ausgenommen, da bei dieser Kategorie das Diskrepanzkriterium greift: Die schulischen Leistungen weichen im Verhältnis zum IQ erheblich nach unten ab, sodass eine Diskrepanz zwischen der anhand des IQ erwarteten und der tatsächlich gezeigten Leistung besteht. Die Tatsache, dass

---

<sup>6</sup> z. B. Lese-Rechtschreib-Störung, Dyskalkulie; siehe Abgrenzung von anderen Begriffen Kapitel 1.2.3

Lernbehinderung also vor allem darüber definiert wird, was es nicht ist, dürfte zu den begrifflichen Problemen maßgeblich beitragen. Daher bleiben auch die Erklärungsmodelle zur Ätiologie von Lernbehinderung vage und auf einem abstrakten Level, zudem häufig ohne ausreichende empirische Grundlage. Dieser Punkt der Erklärungsmodelle wird in einem eigenen Abschnitt (Kapitel 1.3) aufgegriffen und genauer analysiert.

Ein wesentliches Problem der Definition des sonderpädagogischen Förderbedarfs liegt in ihrem ans Tautologische grenzenden Kern: Einen solchen Förderbedarf hat, wer ohne besondere Förderung nicht (ausreichend) gefördert werden kann (Schröder, 2000, S. 77). Erkennbar ist die grundlegende Schwäche der Definition, dass sie vage bleibt und keine klaren Grenzen zieht, und somit als beliebig und nicht hilfreich anzusehen ist. Auch die aktualisierte Fassung der KMK-Empfehlungen für den Förderschwerpunkt Lernen (KMK, 2019) kommt nicht ohne diese zirkuläre Verwendung des Förder-Begriffs aus: „Schülerinnen und Schüler mit erheblichen Schwierigkeiten im schulischen Lernen weisen [...] Denk- und Lernmuster auf, die bei der Begegnung und Auseinandersetzung mit schulischen Lerngegenständen zu einer Irritation bzw. Desorientierung führen können, so dass durch Unterstützungs- und Fördermaßnahmen der allgemeinen Schule allein noch keine Basis für den Anschluss an schulisches Lernen gefunden werden kann.“ (KMK, 2019, S. 5).

Zum dritten Punkt: Zu den oben beschriebenen einzelnen Problemen der Begriffe Lernbehinderung und sonderpädagogischer Förderbedarf kommt hinzu, dass diese (aufgrund der langen Tradition des Lernbehinderung-Begriffs als schulorganisatorisch geprägter Kategorie und der daraus resultierenden engen Verwebung) wechselseitig voneinander abhängig sind und nicht ohne einander auskommen. Da für Lernbehinderung kaum klare diagnostische Kriterien vorliegen, muss das Konzept des sonderpädagogischen Förderbedarfs zu Hilfe genommen werden. Andersherum ist auch der sonderpädagogische Förderbedarf „noch immer maßgeblich durch das Konstrukt ‚Lernbehinderung‘ bestimmt“ (Kobmann, 2019, S. 10f.). Dadurch, dass die Begriffe nur ausgetauscht werden und das eine Symptom mit dem anderen und vice versa erklärt wird, nähert man sich jedoch nicht dem inhaltlichen Kern des Problems. Zu dieser begrifflichen Unklarheit und Vermengung hat sicher auch die praktische Verwendung des Begriffs Lernbehinderung als schulverwaltungstechnische Kategorie (Grünke & Grosche, 2014; Kanter, 2007b) beigetragen.

Zum vierten Punkt: Schließlich gibt es in der Literatur Widersprüche, inwiefern Lernbehinderung und sonderpädagogischer Förderbedarf synonym zu verwenden seien. Eine Vielzahl von Autor\*innen vertreten die Ansicht, dass die beiden Begriffe synonym zu verstehen seien. Das zeigt sich u. a. darin, dass die Nominaldefinition von Lernbehinderung (Grünke & Grosche,

2014) nahezu wörtlich deckungsgleich mit der Definition eines sonderpädagogischen Förderbedarfs im Förderschwerpunkt Lernen in der Ausbildungsordnung sonderpädagogische Förderung im Land NRW ist (Ministerium für Schule und Weiterbildung NRW [MfSW], 2016, §4.2). Eine solche Sichtweise würde die oben geschilderten Definitionspraktiken des Begriffsaustauschs erst recht ad absurdum führen. Andererseits weisen Autor\*innen, teilweise in ein und demselben Beitrag darauf hin, dass es sich um distinkte Konstrukte handele, sodass Lernbehinderung und sonderpädagogischer Förderbedarf nicht automatisch gleichzusetzen seien (z. B. Grünke & Grosche, 2014; Kanter, 1980). Auf diese Kontroverse, inwiefern Lernbehinderung und sonderpädagogischer Förderbedarf deckungsgleich sind, wird im Folgenden näher eingegangen.

Es besteht ein besonderes, nicht ganz widerspruchsfreies Verhältnis zwischen Lernbehinderung und sonderpädagogischem Förder- bzw. Unterstützungsbedarf im Förderschwerpunkt Lernen. Insbesondere die alltägliche pragmatische Verwendung des Begriffes Lernbehinderung als Label (Heimlich & Wember, 2014) für Kinder mit sonderpädagogischem Förderbedarf bzw. Kinder, die eine Förderschule mit dem Förderschwerpunkt Lernen besuchen, erscheint weniger stringent als es wünschenswert wäre. So herrscht zwar grundsätzlich, auch aus der historischen Einbettung und Nutzung des Begriffs heraus, die Ansicht vor, dass es sich bei Lernbehinderung um einen „schulorganisatorischen Begriff, geschaffen von der KMK zur Neuordnung des Sonderschulwesens“ (Kanter, 2007a; Lauth & Schlotke, 2005; Haeberlin, 1991; Grünke & Grosche, 2014) handelt. Somit läge es nahe, die Begriffe Lernbehinderung (Grünke & Grosche, 2014) und sonderpädagogischer Förderbedarf im Förderschwerpunkt Lernen (MfSW NRW, 2016, §4.2) synonym zu verwenden, da die Definitionen nahezu wörtlich identisch sind.

Es kommt jedoch in einigen Beiträgen auch zu inkonsistenten Verwendungen, die möglicherweise auf die unterschiedliche Verwendung in der sonderpädagogischen Theorie und Praxis zurückzuführen sind. So führen Grünke und Grosche (2014) zwar zum einen klar an, dass Lernbehinderung und sonderpädagogischer Förderbedarf im Förderschwerpunkt Lernen deckungsgleich seien. Gleichzeitig schränken sie aber ein, dass die analoge Verwendung im Alltag unzutreffend sei, da die Population der Schüler\*innen in Förderschulen mit dem Förderschwerpunkt Lernen diverser seien, und auch Schüler\*innen mit „durchschnittlicher Allgemeinintelligenz und eine[r] hinreichende[n] Lernfähigkeit“ (S. 78) diese Schulen besuchen. Als Grund hierfür nennen sie, dass mitunter Schüler\*innen mit Migrationshintergrund (fälschlicherweise) für den Förderschwerpunkt Lernen diagnostiziert werden. Grundsätzlich scheinen Schüler\*innen mit Migrationshintergrund in Förderschulen mit dem Förderschwerpunkt Lernen überrepräsentiert zu sein (Powell & Wagner, 2014). Wenn neben sozialer Herkunft und kognitiven Grundfertigkeiten auch Schulleistungen statistisch kontrolliert werden, ist diese Überrepräsentation jedoch nicht

mehr gegeben (Kölm, Gresch & Kuhl, 2019). Auch Heimlich und Wember (2014) attestieren praktische Probleme bei der Anwendung des sonderpädagogischen Förderbedarfs, wie die Zuweisung von Schüler\*innen mit durchschnittlichen Intelligenzwerten. Gleichzeitig sehen sie in der Dominanz des psychometrischen IQ als Diagnosekriterium sowie der personenzentrierten Defizitorientierung ein grundsätzliches Problem des Begriffs „Lernbehinderung“ und plädieren dafür, auf diesen (z. B. zugunsten von „Lernschwierigkeiten“) gänzlich zu verzichten.

Gegen die Gleichsetzung von Lernbehinderung und sonderpädagogischem Förderbedarf (bzw. vormals „Sonderschulbedürftigkeit“) sprechen weitere Quellen. So nimmt Kanter (1980, S. 57) eine Abgrenzung vor, welche Kinder als „sonderschulbedürftig lernbehindert“ anzusehen sind. Die Sonderschulbedürftigkeit steht dabei am Ende des Bedingungsgefüges: Zunächst sei das Lernen umfassend beeinträchtigt, was zu deutlichen Minderleistungen führe, die sich schließlich im dritten Schritt als so gravierend erweisen, dass die Kinder nicht an einer allgemeinen Schule gefördert werden können. Demzufolge müsste es auch Kinder geben, die lernbehindert, aber nicht sonderschulbedürftig sind; die Begriffe sind also nicht deckungsgleich.

Dies ist insofern bemerkenswert, da Kanter wenige Seiten zuvor Lernbehinderung lediglich als Beschreibungskategorie für einen „real-anthropologischen Sachverhalt“ (Kanter, 1980, S. 34) bezeichnet, nämlich als „schwerwiegende, umfängliche und langdauernde Beeinträchtigung der Lernprozesse und des Lernaufbaues eines Menschen“ (Kanter, 1980, S. 34). Angesichts dessen und auch mit Blick auf die vehemente Kritik am Verständnis von Lernbehinderung als Ursache für Minderleistungen (Kanter, 2007b) überrascht diese oben dargestellte kausale Einbettung.

Obwohl Kanter (2007b) und auch Schröder (2000, S. 70) betonen, dass Lernbehinderung lediglich beschreibenden Charakter hat und nicht als „physische[r] Tatbestand oder gar ein[e] Art Krankheit“ zu verstehen sei, hält sich die Zuschreibung einer kausalen Rolle von Lernbehinderung als Ursache für sonderpädagogischen Förderbedarf auch in der bis heute gültigen Ausbildungsordnung für sonderpädagogischen Förderbedarf (MfSW NRW, 2016, §3). Dort wird Lernbehinderung (als Erscheinungsform einer Lern- und Entwicklungsstörung) als eine mögliche Ursache für die Zuweisung eines sonderpädagogischen Förderbedarfs im Förderschwerpunkt Lernen explizit benannt. Diese Vermischung von deskriptiven und explanativen Ebenen und allgemein die widersprüchlichen Sichtweisen zur Äquivalenz von Lernbehinderung und sonderpädagogischem Förderbedarf sind überaus problematisch und sorgen für eine mangelnde Präzision der Begriffe und eine Verstetigung des „Begriffs-Wirrwarr[s]“ (Klauer & Lauth, 1997).

### **1.2.3. Abgrenzung zu anderen Begriffen**

Das Konstrukt der „Lernbehinderung“ und seine Abgrenzung zu anderen Begriffen von Lernbeeinträchtigungen ist wissenschaftlich hoch umstritten (pointiert dargestellt z. B. in Gold,

2015). In diesem Abschnitt soll in Anlehnung an Gold (2011; 2015) ein Überblick über die gängigen Begrifflichkeiten gegeben werden, um sie vom Konstrukt Lernbehinderung abzugrenzen. Die Verortung des Konstruktes im internationalen Kontext erfolgt im nächsten Abschnitt 1.2.4. Die naheliegende Bezeichnung „learning disability“ ist nicht (bzw. nur in wenigen Ausnahmen) mit Lernbehinderung gleichzusetzen, sondern mit spezifischen Lernstörungen. Dagegen scheint das Konstrukt *Mild and Borderline Intellectual Disabilities* dem deutschen Verständnis einer Lernbehinderung am nächsten zu kommen.

In Anlehnung an die von Gold (2011; 2015) vorgeschlagene Systematisierung aus der Perspektive der pädagogischen Psychologie werden die folgenden Begriffe erläutert: (1) Lernschwierigkeit und Lernbeeinträchtigung, (2) Lernstörung, (3) Lernschwäche sowie (4) Lernbehinderung.

Die Bezeichnungen *Lernschwierigkeit* und *Lernbeeinträchtigung* werden als Oberbegriffe verstanden, unter die alle Formen von erschwertem Lernen fallen. Analog zu den „massiven Lernproblemen“ (Heimlich & Wember, 2014, S. 51) ist als Kernsymptom das schulische Lernen in den Bereichen Lesen, Rechtschreiben und/oder Rechnen zumindest für eine Weile maßgeblich beeinträchtigt.

Eine (spezifische) *Lernstörung* ist gemäß ICD-10 (Codes F81.0 – F81.3) als umschriebene Entwicklungsstörung schulischer Fertigkeiten nach international verbindlichen Kriterien diagnostizierbar. Zentral ist, dass die festzustellende schulische Minderleistung insofern ‚erwartungswidrig‘ zu sein hat, als dass eine deutliche Diskrepanz zu den kognitiven Voraussetzungen vorliegt. Das bedeutet, dass der IQ im Normbereich bzw. zwei Standardabweichungen über der schulischen Leistung liegt. Eine Intelligenzminderung wird als Ursache explizit ausgeschlossen.

Eine *Lernschwäche* bezeichnet eine ähnliche Minderleistung wie die Lernstörung, jedoch ist das Kriterium der individuellen Diskrepanz zu den kognitiven Voraussetzungen nicht erfüllt. Zur Differenzierung von Lernstörung und Lernschwäche gibt es eine lebhaftige Kontroverse, auf die im übernächsten Absatz eingegangen wird.

Bei einer *Lernbehinderung* schließlich liegt zusätzlich zu den schulischen Minderleistungen eine intellektuelle Beeinträchtigung vor. Von einer intellektuellen Beeinträchtigung wird im ICD-10 ab einem  $IQ < 70$  ausgegangen; vom Deutschen Bildungsrat wurde in den 1960er Jahren ein erweiterter Begriff der „grenzwertigen Intelligenz“ (Gold, 2015, S. 126) auch für einen IQ-Bereich  $IQ < 85$  geprägt.

Als Randnotiz sei auf die kontroverse Diskussion des Diskrepanzkriteriums für das Konstrukt der Lernstörung verwiesen (Beaujean, Benson, McGill & Dombrowski, 2018; Ehlert, Schroeders & Fritz-Stratmann, 2012; Stanovich, 2005). Die empirische Befundlage zur Berech-



tigung der Differenzierung zwischen Lernstörung (Diskrepanzkriterium erfüllt) und Lernschwäche (Diskrepanzkriterium nicht erfüllt) spricht eher nicht dafür, dass eine solche Unterteilung haltbar und sinnvoll ist: So untersuchten u.a. Brandenburg et al. (2013) Kinder mit Lernschwäche und Lernstörung bei sonst gleichen schulischen Leistungen, und konnten hinsichtlich der Arbeitsgedächtniskapazität keine Unterschiede feststellen. Auch Kuhn, Raddatz, Holling und Dobel (2013) konnten keine Unterschiede in Mathematik-Leistungsvariablen zwischen Kindern mit Dyskalkulie (Lernstörung, Diskrepanzkriterium erfüllt) und mit Rechenschwäche (Lernschwäche, Diskrepanzkriterium nicht erfüllt) finden. Auch beim Vergleich von Kindern mit einer Lernstörung (F81.3 Kombinierte Störung schulischer Fertigkeiten) und Lernbehinderung wurde die zeitliche Stabilität der Lernschwierigkeiten in beiden Gruppen gleichermaßen in einer längsschnittlichen Studie von Skowronek, Schuchardt & Mähler (2018) nachgewiesen.

Die exemplarisch angeführten Befunde zur mangelnden empirischen Differenzierbarkeit von Lernstörung und Lernschwäche verdeutlichen, dass auch ein nach ICD-10 vermeintlich klar definiertes Konstrukt nicht immer empirisch trennscharf ist – umso mehr dürfte dies für das vage definierte Konstrukt Lernbehinderung gelten, was die Notwendigkeit von empirischen Studien zur Beschreibung der Bedingungen von Schüler\*innen mit Lernbehinderung unterstreicht.

#### 1.2.4. Internationale Begriffe

Zur begrifflichen Verwirrung trägt zudem maßgeblich bei, dass im internationalen, insbesondere im anglo-amerikanischen Raum die vermeintlich analoge Bezeichnung *learning disability* für eine andere Subgruppe Verwendung findet. Im US-Amerikanischen Sprachraum trifft der Begriff am ehesten auf Kinder mit spezifischen Lernstörungen zu, während im UK eher das Phänomen der deutschen Lernbehinderung gemeint zu sein scheint. In einem Themenheft der Zeitschrift *Learning Disabilities: A Contemporary Journal* lassen Grünke und Cavendish (2016) Forscherinnen und Forscher aus acht verschiedenen Nationen zu Wort kommen, indem diese auf die Frage antworten, wie „Learning Disability“ in ihrem Land verstanden wird und welche Schüler\*innen unter diese Definition fallen. Diesen Faden greifen Grosche, Grünke, Scanlon und Sideridis (2018, Juli) im Rahmen einer Roundtable-Diskussion einer internationalen Tagung auf mit dem Ziel, daraus Kerndefinitionen oder ‚Lesarten‘ herauszuarbeiten, um in Zukunft die internationale Verständigung bei der Verwendung des Begriffes „learning disability“ zu erleichtern (Williams, Miciak, McFarland & Wexler, 2016). Mittlerweile liegt von Grosche et al. (2019, Juni) ein Codebuch vor, mit dem Aussagen in den Beiträgen des Themenheftes klassifiziert und Leitfragen für die Experteninterviews generiert werden können, um schließlich mittels Typenbildung im Rahmen einer qualitativen Inhaltsanalyse diese Lesarten herauszuarbeiten.

Im internationalen Kontext hat sich für die Bezeichnung des Phänomens Lernbehinderung die Begriffskombination der *Mild and Borderline Intellectual Disabilities* (MBID, z. B. in Poloczek, Büttner & Hasselhorn, 2012; Poloczek et al., 2016; van der Molen, Henry & van Luit, 2014) etabliert. So liegt eine leichte (*mild*) Intelligenzbeeinträchtigung bei Werten zwischen 55-70 IQ-Punkten und eine grenzwertige (*borderline*) Intelligenzbeeinträchtigung zwischen 70-85 IQ-Punkten vor, wenn zusätzlich die adaptiven Funktionen, als welche schulische Minderleistungen durchaus betrachtet werden können, eingeschränkt sind (zur weiteren Diskussion von Borderline Intellectual Functioning siehe auch Ferrari, 2009; Greenspan, 2017; Hassiotis, 2015). Mit dieser Bezeichnung von MBID, die also leichte und grenzwertige Intelligenzminderung umfasst, ist die Zielgruppe der vorliegenden Arbeit somit international am ehesten anschlussfähig (Eser, 2006).

### 1.2.5. Fazit – Klarstellungen von Missverständnissen

Im Diskurs zum Begriff der Lernbehinderung scheinen drei Aspekte gerade für diese Arbeit von Relevanz zu sein: (1) Lernbehinderung ist nicht als Ursachenkomplex für die manifesten Lernprobleme anzusehen (Kanter, 2007b); (2) Lernbehinderung bedeutet nicht, dass ein „globaler Mangel an Lernfähigkeit“ vorläge (Kanter, 1980, S. 47); (3) das Phänomen Lernbehinderung und die betreffenden Schüler\*innen sind heterogen, die Lernprobleme komplex (Heimlich, 2009, S. 17).

Zum ersten Punkt: Lernbehinderung darf nicht als Ursachenfaktor verstanden werden, der für sich genommen schon erklären könnte, warum das umfassende Lern- und Leistungsversagen auftritt (Kanter, 2007b; Schröder, 2000). Vielmehr ist „der Begriff *Lernbehinderung* [...] demnach in der Begriffssystematik des Sonderschulwesens ursprünglich eine Beschreibungskategorie für umfängliche, schwerwiegende und langfristige Beeinträchtigungen des Lernens vielfältiger Ursache- und Genesemöglichkeiten.“ (Kanter, 2007b, S. 24, Hervorhebung i. O.). Daher plädiert Kanter (2007b, S. 26) für die Verwendung des Begriffs allein als „funktionaler Terminus technikus“, mit dem vorrangigen Ziel, „Kindern und Jugendlichen in Situationen von Lern- und Entwicklungserschwernissen effektive Hilfe zu bieten“ (ebd.). Dieser Sichtweise schließt sich die vorliegende Arbeit explizit an, indem Lernbehinderung als Symptombeschreibung oder Syndrom (Kleber, 1980) verstanden wird, aber keinesfalls als etwas, das die Lernprobleme ursächlich begründen könnte.

Zum zweiten Punkt: Nach Klauer (1977) ist Intelligenz zwar eng mit (schulischer) *Leistungsfähigkeit* verknüpft, nicht jedoch als Persönlichkeitsmerkmal einer allgemeinen *Lernfähigkeit* zu verstehen. Inwieweit Schüler\*innen unterschiedliche Lernleistungen zeigten, sei vor allem aufgabenspezifisch (ebd., S. 3). Indem Klauer (1977) aber die „*schulische Lernfähigkeit*“ (ebd., S. 5, Hervorhebung i. O.) mit Schulleistung gleichstellt und auf den starken Zusammenhang von

Intelligenz und Schulleistung (d. h. niedrige Intelligenz ist eine „ungünstige Startbedingung“, ebd., S. 5) verweist, leitet er eine Gleichsetzung von Intelligenz und schulischer Lernfähigkeit ab. Dass aber bei Kindern mit Lernbehinderung die allgemeine, globale Lernfähigkeit nicht beeinträchtigt sei, betonen auch Kanter (1980, S. 47) und Schröder (2000, S. 73). Ein solcher allgemeiner Anspruch wird in dieser Arbeit nicht gestellt. Da hier lediglich zwei sehr spezifische kognitive Lernvoraussetzungen untersucht werden (siehe Kapitel 2), können sich die Schlussfolgerungen auch nur auf diese ausgewählten Aspekte beziehen; ein Rückschluss auf etwaige allgemeine Lernfähigkeiten (unabhängig davon, ob es sie gibt oder nicht) kann schon allein aufgrund der Begrenztheit der Daten nicht gezogen werden.

Zum dritten Punkt: Hinsichtlich der Schüler\*innenschaft herrscht große Heterogenität vor (Heimlich, 2009; Mähler, 2007), auch die auftretenden Lernprobleme sind äußerst komplex und nicht monokausal erklärbar (Heimlich, 2009, S. 17). Eine medizinische und individuums- und defizitzentrierte Sichtweise auf Lernbehinderung als Persönlichkeitsmerkmal ist angesichts der multifaktoriellen Erklärungsansätze (Schröder, 2000) nicht haltbar. Insofern verbergen sich sehr wahrscheinlich hinter dem beschreibenden Label „Lernbehinderung“ wiederum ganz unterschiedliche Subgruppen von Schüler\*innen, mit jeweils eigenen und individuellen Voraussetzungen und Bedingungsgefügen; im Sinne der „Äquifinalität“ können unterschiedliche Ursachen zu den gleichen Ergebnissen führen (siehe Abschnitt 1.3.1; Pinquart & Silbereisen, 2007).

### **1.3. Ursachenerklärung von Lernbehinderung**

Der Mangel an aktueller empirischer Forschung hat ebenfalls zur Folge, dass es kaum präzise Vorstellungen darüber gibt, welche Faktoren wie bei der Entstehung und Aufrechterhaltung einer Lernbehinderung zusammenwirken (Schröder, 2000). Zudem hat Lernbehinderung in den vergangenen Jahren deutlich weniger Aufmerksamkeit erfahren als noch in den 1980er und 1990er Jahren (Mähler & Hasselhorn, 1990; Merrill et al., 1987; Wember, 1987). Neuere Forschungsmethoden des Feldes sowie Erkenntnisse in den relevanten Grundlagenfächern wie der kognitiven Psychologie konnten, mit Ausnahme der Arbeitsgedächtnisforschung, somit kaum Einzug finden. Gründe für den Mangel an empirischer Forschung sind sicherlich in der oben angeführten unklaren und (schulverwaltungs-)pragmatischen Definition sowie der geringen internationalen Anschlussfähigkeit zu suchen.

Als Konsens für die Erklärung psychischer Beeinträchtigungen allgemein hat sich ein Transaktionales Entwicklungsmodell etabliert (Pinquart & Silbereisen, 2007), demzufolge eine Vielzahl von Faktoren und deren komplexe Wechselwirkungen herangezogen werden müssen. Daher werden zunächst dessen Grundannahmen knapp dargelegt, um die Fokussierung dieser Arbeit auf kognitive Lernvoraussetzungen in dem Modell zu verorten. Anschließend wird die

Strategie-Hypothese nach Klauer und Lauth (1997) vorgestellt, die zu einem prominenten Erklärungsansatz geworden ist. Auch wenn die Bedeutung von Strategien nicht von der Hand zu weisen und nicht grundsätzlich in Frage zu stellen ist, wird in dieser Arbeit dafür argumentiert, dass die Bedeutung kognitiver Lernvoraussetzungen dadurch zu stark vernachlässigt wird.

### **1.3.1. Transaktionales Entwicklungsmodell als Grundannahme**

Ausgehend von der Gen-Umwelt Debatte und dem Diathese-Stress-Modell (Wittchen, 2011) wird Entwicklung im Transaktionalen Entwicklungsmodell als komplexes Muster von Interaktionen auf mehreren Ebenen und Transaktionen über die Zeit hinweg beschrieben (Beelmann & Raabe, 2007; Pinquart & Silbereisen, 2007). So wirken zahlreiche Risiko- und Schutzfaktoren auf biologischer, psychischer und sozialer Ebene zusammen. Die Kumulation von Risikofaktoren und ihr Verhältnis zu den kumulierten Schutzfaktoren führt zu einem mehr oder weniger wahrscheinlichen Auftreten von manifesten Lern- und Verhaltensproblematiken (Pinquart & Silbereisen, 2007; Yates, Egeland & Sroufe, 2003). Die Aussage des Transaktionalen Entwicklungsmodells ist probabilistisch, das bedeutet, sie kann nur eine Wahrscheinlichkeit beziffern. Eine Ableitung von deterministischen Aussagen („wenn Risikofaktor ‚X‘ vorliegt, führt dies auf jeden Fall zu einer Lernbehinderung“) ist nicht möglich. „Fehlentwicklung (gleich welcher Art) kann nach dieser Konzeption als das Resultat eines dynamischen Entwicklungsprozesses verstanden werden, der durch ein relativ ungünstiges Verhältnis von Vulnerabilitäten (als personale Risiken) und Stressoren (als soziale Risiken) zu Resilienz [...] und Ressourcen [...] gekennzeichnet ist.“ (Beelmann & Raabe, 2007, S. 55).

Eine Lernbehinderung kann gemäß dem Transaktionalen Entwicklungsmodell aus einer Vielzahl von Ursachen und deren Kumulation und komplexen Wechselwirkungen entstehen (Schröder, 2000). Biologische Faktoren umfassen genetisch und organisch bedingte Fehlfunktionen (Suhrweier, 1993, zit. n. Werning & Lütje-Klose, 2012) oder Hirnschädigungen (Zielinski, 1980). Als psychische Faktoren sind z. B. Gedächtnisleistungen (Schröder, 2000), metakognitive Fähigkeiten und Strategien (Klauer & Lauth, 1997; Grünke & Grosche, 2014) oder Motivation und Kausalattribution (Schröder, 2000; Zielinski, 1995) zu nennen. Soziale Faktoren stellen z. B. elterlicher Erziehungsstil (Zielinski, 1995), sozio-ökonomische und -kulturelle Benachteiligung (Schröder, 2000; Werning & Lütje-Klose, 2012; Zielinski, 1995), sowie aus der Zuweisung eines Förderbedarfs resultierende Stigmatisierung (Werning & Lütje-Klose, 2012) dar.

Keiner der genannten Faktoren kann für sich alleine genommen die Entstehung von Lernbehinderung zufriedenstellend erklären. Im Verständnis von Schröder (2000) stellt Lernbehinderung das (momentane) Resultat eines negativen und misserfolgsbehafteten Bildungswegs dar. Auf diesem Bildungsweg kommen zahlreiche Faktoren zusammen und wirken sich wechselseitig aus. Als Wechselwirkungen seien nur einige exemplarisch genannt: Von einem schwierigen

sozialen Umfeld können auch leichte kognitive Probleme („minimale Restriktionen“, Zielinski, 1995, S. 64) mangels Anregungspotential nicht ausgeglichen werden. Misserfolge werden häufiger erlebt, deren stabil-interne Attribuierung sich zu einem ungünstigen Selbstkonzept verfestigt und in dysfunktionale Strategien wie einer erlernten Hilflosigkeit mündet (Werning & Lütje-Klose, 2012).

In einer der wenigen aktuelleren Studien zum Thema führten Lauth, Hammes-Schmitz und Lebens (2014) eine empirische Bedingungsanalyse von (spezifischen) Lernstörungen durch, in die fünf verschiedene Ebenen von Faktoren einfließen, um deren Einfluss auf das Auftreten von Lernstörungen bzw. schulischem Misserfolg zu untersuchen. Die ersten drei Ebenen umfassten Prädiktoren aus drei Bereichen: Metakognition, Basisfertigkeiten und bereichsspezifisches Wissen/Vorkenntnisse. Als Moderator-Ebene<sup>7</sup> führten die Autor\*innen motivational-emotionale Variablen von nicht-kognitiven Persönlichkeitsmerkmalen an. Schließlich flossen auf der fünften Ebene neben Alter und Geschlecht weitere Moderatoren aus dem sozialen Umfeld in Form von „familiäre[n] und schulische[n] Kontextbedingungen“ (ebd. S. 353) ein. Die zugrundeliegenden Daten stammten zum einen aus Patientenakten, die durch verhaltensanalytische Interviews ergänzt und validiert wurden, sowie zusätzlichen Daten aus Fragebögen zu Arbeitsverhalten und motivationalen Variablen. Insgesamt wurden je nach Kriterium (verbale bzw. mathematische Schulleistung sowie Schwere der Störung) zwischen 60% und 72% der Varianz durch das Modell erklärt. Insbesondere das bereichsspezifische Wissen schien sich insgesamt als starker Prädiktor zu erweisen, zudem waren familiäre und schulische Kontextfaktoren mit mittleren Effektstärken mit Lernstörungen assoziiert. Die Güte und Verlässlichkeit dieser Aussagen bleibt jedoch unklar, da „Daten zum Modell-Fit noch ausstehen“ (S. 361); dennoch stellt die Studie einen ersten Ansatz dar, verschiedene Bereiche von möglichen Einflussvariablen in einen empirisch prüfbareren Kontext zu bringen. Zudem muss offen bleiben, inwiefern diese Ergebnisse zu spezifischen Lernstörungen auch auf Kinder mit Lernbehinderung übertragbar sind.

### 1.3.2. Strategie-Hypothese nach Klauer und Lauth

Mit der Hypothese eines ineffektiven und schlecht überwachten Strategieeinsatzes (im Folgenden ‚Strategie-Hypothese‘) stellen Klauer und Lauth (1997) einen Erklärungsansatz für die Entstehung von Lernbehinderung auf, der sehr plausibel erscheint und als einflussreicher und prominenter Ansatz gilt: Leistungsschwache/lernbehinderte Kinder zeichnen sich weniger durch dauerhafte Fähigkeitsdefizite (etwa im Bereich des Gedächtnisses oder des Denkens) als vielmehr durch die Art aus, wie sie Lernvorgänge bewältigen (z. B.: ‚Raten‘ statt systematischer

---

<sup>7</sup> Aufgrund der Modellskizze (S. 353) eher als Mediatoren zu bezeichnen

Inspektion der Materialvorlage).“ (Klauer & Lauth, 1997, S. 707). Und weiterhin: „Strategiedefizite im Sinne eines mangelnden Einsatzes günstiger Strategien [...] sind bei Lernbehinderung und Leistungsschwierigkeiten unstrittig.“ (ebd., S. 710).

Bei Strategien handelt es sich um planvolle Herangehensweisen an Aufgaben, welche wiederum verschiedene Taktiken beinhalten (Belmont & Mitchell, 1987). Klauer und Lauth (1997) kommen zu dem Schluss, dass die Probleme von Kindern mit Lernbehinderung weder auf der Ebene der Informationsverarbeitung (des „fundamentalen Prozessors“, Belmont & Mitchell, 1987, S. 97) noch der Taktiken zu suchen sei, sondern bei der zielgerichteten und adäquaten Auswahl und Überwachung der Strategien. So seien Kinder mit Lernbehinderung sehr wohl in der Lage, Strategien anzuwenden und dann auch zu guten Ergebnissen zu kommen, jedoch wendeten sie diese Strategien nicht eigenständig ohne aktive Instruktion dazu an.

Damit nehmen Klauer und Lauth (1997) eine deutliche Positionierung zugunsten der Erklärungskraft von Strategien und Metakognition vor, während sie kognitiven Lernvoraussetzungen für die Entstehung von Lernbehinderung eine untergeordnete Rolle zuweisen. In der vorliegenden Arbeit soll diese untergeordnete Rolle hinterfragt und empirisch überprüft werden, womit jedoch nicht impliziert werden soll, dass Strategien generell bedeutungslos wären. Es erscheint jedoch aus insgesamt vier Gründen gerechtfertigt, kognitiven Lernvoraussetzungen trotz der klaren Positionierung von Klauer und Lauth (1997) Aufmerksamkeit im Rahmen der vorliegenden Arbeit zu widmen.

(1) Obwohl die Strategie-Hypothese kognitive Lernvoraussetzungen als Ursache so klar ablehnt, werden in Erklärungsmodellen häufig zumindest implizit eine funktionierende Informationsverarbeitung und Vorwissen als mit-ursächlich für Lernbehinderung angenommen (z. B. Grünke & Grosche, 2014; Kretschmann, 2007; Vernooij, 2007; Zielinski, 1980), meist jedoch nur wenig präzisiert und ohne aktuelle empirische Belege. Eser (2006) spricht offen von spekulativen Vermutungen über kognitive Lernvoraussetzungen: „Wenn man bereits für umschriebene Lernstörungen schulischer Fertigkeiten von einer ‚andersartigen Entwicklung des zentralen Nervensystems‘ ausgehen muss, um wie viel mehr wird das für die Lernbehinderung als umfassende und tief greifende Entwicklungsstörung zutreffen.“ (ebd., S. 59). Zudem führen Klauer und Lauth (1997, S. 710) selbst an, dass „für lernbehinderte/leistungsschwache Kinder [...] als gesichert [gilt], daß sie nur in eingeschränkter Weise über das relevante bereichsspezifische Wissen verfügen. Sie gehen von einer unzureichenden Wissensbasis aus“. Insgesamt betrachtet wird kognitiven Lernvoraussetzung also anscheinend mehr Bedeutung beigemessen, als von der Strategie-Hypothese postuliert wird.

(2) Insgesamt liegen nur wenig aktuelle Ergebnisse zu kognitiven Aspekten von Lernbehinderung vor, da die meisten Forschungsarbeiten zu kognitiven Lernvoraussetzungen bereits

mehrere Jahrzehnte zurückliegen. So zeigte Wember (1987), dass Schüler\*innen mit Lernbehinderung (Primarstufe,  $n = 84$ ) in konkret-operatorischen Aufgaben deutliche Schwächen aufwiesen, die einem Entwicklungsrückstand von ein bis zwei Jahren entsprachen. Zwar kam Hasselhorn (1987) zu dem Schluss, dass bei lernschwachen Kindern (sehr breit definiert,  $IQ > 70$ ) Probleme weniger durch strukturelle Probleme in basalen kognitiven Bereichen (wie Wahrnehmung und Aufmerksamkeit), sondern durch strategische und metakognitive sowie nicht-strategische Verarbeitungsprozesse (Verarbeitungsgeschwindigkeit) begründet seien. Diese Ergebnisse sind seitdem jedoch nicht weiter empirisch überprüft worden.

(3) Als Ausnahme vom und im Gegensatz zum sonstigen Mangel an aktueller Forschung zu kognitiven Lernvoraussetzungen wurden in der Domäne des Arbeitsgedächtnisses eine Reihe von Studien bei Schüler\*innen mit Lernbehinderung berichtet (z. B. Mähler, 2007; Mähler & Hasselhorn, 2003; Poloczek et al., 2016; van der Molen, van Luit, Jongmans & van der Molen, 2007). Eine ausführliche Darstellung der Forschungsergebnisse erfolgt in Kapitel 2.1.3. An dieser Stelle kann schon einmal festgehalten werden, dass in den genannten und weiteren Studien durchaus spezifische differenzielle Entwicklungsmuster bei Kindern mit Lernbehinderung nachgewiesen wurden. Damit widerlegen diese neueren Forschungen zumindest für den Aspekt des Arbeitsgedächtnisses die Aussage von Klauer und Lauth (1997), dass kognitive Lernvoraussetzungen nicht beeinträchtigt seien, bzw. wäre zu diskutieren, inwiefern es möglich ist, diese Ergebnisse im Rahmen der Strategie-Hypothese zu erklären. Dies ist Anlass für die vorliegende Studie, das Arbeitsgedächtnis genauer zu untersuchen.

(4) Schließlich formuliert Swanson (2008) einige Kritikpunkte an der Strategie-Hypothese, die zwar auf Lernstörungen bezogen sind, aber auch grundsätzlich auf Lernbehinderung übertragen werden können: ein Defizit in der Anwendung von Strategien bzw. metakognitiver Überwachung ist allein genommen nicht ausreichend, um Unterschiede zwischen Gruppen zu erklären; Strategien, die für unbeeinträchtigte Lernende hilfreich sein können, mögen für Schüler\*innen mit Lernbeeinträchtigung unwirksam sein; selbst wenn eine Strategie von beeinträchtigten Lernenden (oberflächlich) eingesetzt wird, bedeutet dies noch nicht, dass sie im intendierten Sinne genutzt wird (Bsp. Ordnen von Lernstoff, ohne jedoch die Bedeutung zu erfassen). Daher konstatiert Swanson (2008, S. 44), dass in der neueren Forschung der Fokus verschoben werde auf „nicht strategisch[e] Prozesse, welche nicht unbedingt bewusst eingesetzt werden“. Als weiteren, möglicherweise fruchtbaren, Ansatz schlägt er die Überprüfung der Wissens- und der Kapazitätshypothese vor, die bei Lernstörungen noch nicht ausreichend getestet seien (S. 72f.). Die Wissenshypothese betont die Rolle von Qualität und Quantität des bereichs-

spezifischen Vorwissens, während die Kapazitätshypothese von einer Beeinträchtigung der kognitiven Verarbeitung ausgeht, also von Unterschieden in der (reifungsbedingten) Verarbeitungskapazität bzw. begrenzten Ressourcen, insbesondere im Arbeitsgedächtnis.

Aus den oben angeführten vier Gründen, nämlich (1) impliziter Nennung von kognitiven Lernvoraussetzungen, (2) insgesamt veralteten Forschungsergebnissen, (3) der Strategie-Hypothese zuwiderlaufenden neueren Forschungen zum Arbeitsgedächtnis sowie (4) der von Swanson (2008) formulierten Kritikpunkte, erscheint es gerechtfertigt und relevant, kognitive Lernvoraussetzungen bei Schüler\*innen mit Lernbehinderung differenzierter empirisch zu untersuchen. Indem in den vorangegangenen Abschnitten die Relevanz der kognitiven Lernvoraussetzungen begründet wurde, soll jedoch nicht die Bedeutung oder Beeinflussbarkeit der Strategieranwendung und -vermittlung als Intervention in Abrede gestellt werden. Jedoch erscheint die Aussage von Klauer und Lauth (1997), dass auf Ebene der kognitiven Prozesse keine Erklärungskraft zu vermuten sei, nur unzureichend belegt und angesichts neuerer Forschungsergebnisse zum Arbeitsgedächtnis auch bereits als teilweise widerlegt.

### **1.4. Zusammenfassung und Fazit zur Zielgruppe**

Die Ausführungen des ersten Kapitels belegen deutlich genug, dass Lernbehinderung ein vages und äußerst problembehaftetes Konstrukt darstellt. Gleichwohl müssen diese Probleme aber in Kauf genommen und zumindest im Hinterkopf behalten werden, wenn dennoch eine empirische Untersuchung zum Zwecke einer präziseren Deskription ermöglicht werden soll. Denn auch wenn die begriffliche Diskussion (zu Recht) als unfruchtbar ad acta gelegt wird, lösen sich damit die pragmatischen realen Probleme von Schüler\*innen beim schulischen Lernen nicht in Luft auf (Schröder, 2000). Die gewonnenen Daten können dann dazu beitragen, das theoretische und praktische Verständnis von Lernbehinderung zu erweitern.

Daher wurde für diese Arbeit eine operationale und replizierbare Definition gewählt: In die Untersuchung werden diejenigen Kinder aufgenommen, bei denen ein sonderpädagogischer Förderbedarf im Förderschwerpunkt Lernen (als primärer Förderschwerpunkt) festgestellt wurde und die an einer entsprechenden Förderschule unterrichtet werden. Außerdem wurde, der operationalen Definition in Lauth und Schlottke (2005) sowie den Empfehlungen des Deutschen Bildungsrates (1976) folgend, ein IQ im Bereich zwischen 55-85 IQ-Punkten als Einschlusskriterium festgelegt und andere Formen ausgeschlossen, um die Kriterien der Differentialdiagnose einer Lernbehinderung zu erfüllen (vgl. Grosche & Grünke, 2014, S. 80f.).

Die operationale Definition von Lernbehinderung für die vorliegende Studie ist, angesichts der oben genannten Probleme mit der Definition und der Heterogenität der Zielgruppe, zwar angreifbar; auch ist der Kritikpunkt von Haerberlin (1991, S. 169f.) durchaus berechtigt, dass



durch die Orientierung am Förderschwerpunkt Lernen eine Festschreibung am status quo stattfindet. Da jedoch im Rahmen einer empirischen Studie eine Entscheidung über die Zuordnung zur Untersuchungsgruppe gefällt werden muss, wird hier eine klar mess- und nachvollziehbare, jedoch recht breite Definition bzw. Ein- und Ausschlusskriterien zugrundegelegt.

Gestützt auf der Argumentation von Kanter (2007b) soll die vorliegende Arbeit einen Beitrag auf deskriptiver Ebene zur Aktualisierung von wissenschaftlicher Erkenntnis über die Problemlage und die Art und Weise der Beeinträchtigung des Lernens leisten: Welche Einschränkungen im kognitiven Bereich tragen möglicherweise dazu bei, dass Schüler\*innen mit Lernbehinderung solch gravierende Schwierigkeiten im schulischen Lernen aufweisen? Gemäß Kanter (2007b) besteht das primäre Ziel der Sonderpädagogik darin, Kindern mit solchen Erschwernissen möglichst wirksame Hilfen zu bieten, und diese „Hilfe zu organisieren und auf hohem Qualitätsniveau sicherzustellen“ (S. 23). Um dieses Ziel erreichen zu können, ist eine aktuelle, fundierte und präzise Kenntnis der kognitiven Lernvoraussetzungen bei Kindern mit Lernbehinderung notwendig.

Um den Anspruch der Arbeit genauer zu umreißen und mögliche Missverständnisse zu vermeiden, sei an dieser Stelle darauf verwiesen, dass sie weder versucht (1) ein monokausales Bedingungs- oder Erklärungsmodell für Lernbehinderung zu konstruieren, noch postuliert, dass (2) unmittelbare präskriptive Anweisungen (Klauer, 1980) aus den Ergebnissen abgeleitet werden können. Auf diese beiden Punkte wird im Folgenden eingegangen.

Die Fokussierung der Forschungsfrage auf kognitive Lernvoraussetzungen könnte den Eindruck erwecken, dass diese als alleinige Ursache für die Entstehung einer Lernbehinderung angesehen würden; ein solches monokausales Erklärungsmodell ist jedoch keinesfalls intendiert. Vielmehr ist die Zielsetzung der Arbeit, im Sinne von (angewandter) Grundlagenforschung möglichst präzise die kognitiven Voraussetzungen von Schüler\*innen mit Lernbehinderung mit dem sonderpädagogischen Förderbedarf im Förderschwerpunkt Lernen zu beschreiben, um bisher vorherrschende (zwar plausible, aber auf Vermutungen beruhende) Annahmen dazu auf eine solidere empirische Basis zu stellen: Können Schüler\*innen mit sonderpädagogischem Förderbedarf im Förderschwerpunkt Lernen bei der Bewältigung von Anforderungen des schulischen Lernens auf ähnliche, altersäquivalente, oder nur auf verzögerte oder gar strukturell abweichende kognitive Ressourcen zurückgreifen?

Gegen die Annahme eines monokausalen Modells sprechen zudem zwei Gründe. Erstens kann schon aufgrund des Designs keine Aussage über Ursachen (*kausale* Aussage) getätigt werden. Die Daten wurden querschnittlich erhoben, zu einem Zeitpunkt, an dem die Schüler\*innen bereits eine längere Zeit eine Förderschule mit dem Förderschwerpunkt Lernen besucht haben.

So bleibt die Richtung der Wirkweise offen: Zwar könnten einerseits kognitive Beeinträchtigungen durchaus die Ursache für gravierende schulische Probleme und die Feststellung eines sonderpädagogischen Förderbedarfs darstellen; jedoch könnte aber genauso gut der Besuch einer Förderschule mit dem Förderschwerpunkt Lernen zu niedrigeren kognitiven Leistungen führen (Schröder, 2000). Da keine längsschnittlichen Daten vorliegen, die auch den Zeitpunkt *vor* Feststellung des sonderpädagogischen Förderbedarfs umfassen, wird das Ziehen eines kausalen Schlusses von vornherein ausgeschlossen.

Zweitens schließt darüber hinaus die Annahme des Transaktionalen Entwicklungsmodells erst recht aus, dass ein Faktor der alleinige Grund (*mono-kausale Aussage*) für die Entstehung einer Lernbehinderung sein könnte. Wie in Abschnitt 1.3.1 aufgezeigt, müssen neben einer Vielzahl an Faktoren auch deren komplexe Wechselwirkungen für die Erklärung herangezogen werden. Das Transaktionale Entwicklungsmodell widerspricht damit ausdrücklich einer individuals- und defizitorientierten Sichtweise (Schröder, 2000). Wenn nun in dieser Arbeit ausschließlich kognitive Lernvoraussetzungen untersucht werden, soll damit keinesfalls suggeriert werden, dass diese die Entstehung von Lernbehinderung erklären könnten. Vielmehr werden sie lediglich als eine von vielen Teilkomponenten auf der psychischen Ebene des Transaktionalen Entwicklungsmodells angesehen (Pinquart & Silbereisen, 2007). Anderen Erklärungsansätzen, die z. B. die Relevanz von Strategien oder sozialen Faktoren in den Vordergrund stellen, soll damit ausdrücklich nicht ihre Gültigkeit abgesprochen werden.

Die zweite Einschränkung bezieht sich auf die deskriptive Aussageebene: Die Mittel und Ergebnisse der vorliegenden Arbeit sind nicht dazu geeignet, Implikationen auf präskriptiver Ebene über die Erfolgsaussichten von Interventionen abzuleiten. Die geleistete Deskription muss bzw. darf nur als erster notwendiger Schritt betrachtet werden, bevor man dieses Wissen in einem nächsten Schritt in präskriptives Handlungswissen überführen und Maßnahmen entwickeln kann, die dann ihrerseits empirisch auf ihre Wirksamkeit hin überprüft werden müssen.

Zum Beispiel können sich diese Maßnahmen darauf beziehen, wie auf mögliche Defizite kompensatorisch reagiert werden kann und – nicht weniger wichtig – wie vorhandene Stärken (besser) genutzt werden können. Inwiefern daraus auch rechtliche Konsequenzen erwachsen dürfen, bspw. in Form von zieldifferentem Lernen, dem Einsatz von Nachteilsausgleich oder dem formalen Anspruch auf individuelle Förderung, ist auf der normativen Ebene anzusiedeln: zu dieser Debatte können die Ergebnisse der Arbeit als Argumente oder Belege herangezogen werden. Die dazu notwendigen bildungstheoretischen und -philosophischen Prämissen, wie z. B. die Frage nach Bildungsgerechtigkeit (Gold, 2015) oder nach dem Ziel von Differenzierung (Wember, 2007) würden den Rahmen und Anspruch der vorliegenden Arbeit allerdings deutlich übersteigen.

## 2. Eingrenzung der untersuchten Domäne

Die Arbeit untersucht die Fragestellung, ob bei Kindern mit Lernbehinderung hinsichtlich ihrer kognitiven Lernvoraussetzungen differenzielle Entwicklungsmuster zu beobachten sind. Nachdem im ersten Kapitel die Zielgruppe auf Schüler\*innen mit Lernbehinderung eingegrenzt und operational definiert wurde, folgt in diesem Kapitel die Fokussierung auf kognitive Lernvoraussetzungen. Unter kognitiven Lernvoraussetzungen werden hier individuelle „interne, personale Bedingungen“ (Kretschmann, 2007, S. 13) in der kognitiven Domäne verstanden (siehe auch Hasselhorn, 2005; Hasselhorn & Gold, 2013). Die vorliegende Arbeit fokussiert die beiden Aspekte Arbeits- und das Langzeitgedächtnis, und damit zwei ausgewählte kognitive Lernvoraussetzungen, deren Beeinträchtigung einem erfolgreichen Lernen von Kindern mit Lernbehinderung möglicherweise im Wege steht.

Wie in Kapitel 1.4 argumentiert wurde, ist es relevant, die Funktionsweise von Arbeits- und Langzeitgedächtnis bei Kindern mit Lernbehinderung genauer zu untersuchen. Wenn sich die kognitiven Lernvoraussetzungen tatsächlich als unbeeinträchtigt herausstellen, kann dies als empirischer Hinweis für die Gültigkeit der Strategie-Hypothese interpretiert werden. Lassen sich Defizite nachweisen, ermöglicht das ein besseres theoretisches und praktisches Verständnis des Phänomens Lernbehinderung, da die Probleme in einer kognitiven Lernvoraussetzung präziser beschrieben und im Unterricht von Lehrkräften besser adressiert werden können.

Daher wird im weiteren Verlauf dieses Abschnitts die Rolle von Arbeits- und Langzeitgedächtnis für das schulische Lernen im Modell der guten Informationsverarbeitung verortet. In den beiden folgenden Unterkapiteln werden beide Komponenten näher betrachtet und die für die jeweiligen Teilstudien relevanten theoretischen Grundlagen gelegt. Der Bereich des Arbeitsgedächtnisses (Kapitel 2.1) wird auf die Phonologische Schleife und die Prozesse Rehearsal und Redintegration eingegrenzt, die zur Aufrechterhaltung von Gedächtnisspuren dienen. Für das Langzeitgedächtnis (Kapitel 2.2) wird als Untersuchungsgegenstand der Abruf aus dem mentalen Lexikon fokussiert.

Auch wenn, wie im Rahmen des Transaktionalen Entwicklungsmodells (siehe Kapitel 1.3.1) postuliert, vielfältige Ursachen und deren Wechselwirkungen für die Entstehung und Aufrechterhaltung einer Lernbehinderung verantwortlich sind, so macht doch den Kern einer Lernbehinderung aus, dass das Lernen als zentrale Tätigkeit im Ergebnis beeinträchtigt ist (Schröder, 2000; Zielinski, 1995). Schüler\*innen mit Lernbehinderung gelingt es weniger effektiv als unbeeinträchtigten Schüler\*innen, ohne zusätzliche sonderpädagogische Unterstützung im Kontext Schule einen systematischen und aktiven Aneignungsprozess zum Ziel des Wissenserwerbs zu durchlaufen.

Unter Lernen wird allgemein gesprochen die „dauerhafte Veränderung des Verhaltens bzw. des potenziellen Verhaltens (Verhaltensdisposition)“ (Bodenmann, 2006, S. 1) verstanden. Anderson (2000) ergänzt den Aspekt der Erfahrung als Ursache des Lernens und verweist damit auf die enge Verbindung und grundlegende Bedeutung des Gedächtnisses für Lernprozesse: „Memory is the record of the experience that underlies learning“ (Anderson, 2000, S. 4).

Das Spektrum an wissenschaftlichen Perspektiven auf das Lernen ist außerordentlich breit (Schermer, 2013; Wember & Heimlich, 2014). Sie reichen von basalen Annahmen der Reiz-Reaktions-Verknüpfung im Rahmen der klassischen und operanten Konditionierung (siehe Anderson, 2000; Hasselhorn & Gold, 2013) über Modelllernen (Bandura, 1965) bis hin zu geisteswissenschaftlich geprägten Konzeptionen, die versuchen, die Komplexität des Lernprozesses vollumfänglich zu erfassen und den Bildungsaspekt für ein mündiges Subjekt in den Vordergrund stellen (z. B. Künkler, 2008). Auf diese Bandbreite an Perspektiven kann hier allerdings nicht detailliert eingegangen werden; bevor die Eingrenzung auf die Sichtweise auf Lernen als Wissenserwerb vorgenommen wird, sei im folgenden Absatz auf den Bereich des impliziten Lernens verwiesen, welches ebenfalls als kognitive Lernvoraussetzung verstanden, aber aufgrund der Untersuchungsanlage dieser Arbeit nicht näher betrachtet werden kann.

Als implizites Lernen wird die unbewusste Ableitung von Regelmäßigkeiten aus einer Reihe von Ereignissen ohne explizite Kenntnis darüber (Reber, 1967) bezeichnet. Forschungsarbeiten zum Zusammenhang von impliziten Lernleistungen mit psychometrischer Intelligenz und Lernbeeinträchtigungen haben z. T. inkonsistente Befunde erbracht: Während einige Studien keinen Zusammenhang zwischen IQ und der impliziten Lernleistung nachweisen konnten (z. B. Atwell, Connors & Frances, 2003; Gebauer & Mackintosh, 2007; Vinter & Detable, 2003; für LRS siehe Roodenrys & Dunn, 2008), fanden andere Autor\*innen Effekte für einen solchen Zusammenhang bzw. systematische Unterschiede im impliziten Lernen zwischen Kindern mit und ohne Lernbeeinträchtigung (z. B. Fletcher, Mayberry & Bennett, 2000; Wilbert & Grosche, 2012; für LRS siehe Ise & Schulte-Körne, 2012; Sperling, Lu & Manis, 2004). Diese widersprüchliche Befundlage deutet auf weitere Forschungsdesiderata hin, die aber hier nicht bearbeitet werden können. Zugleich stellen sie die Aussage von Kanter (1980) und Klauer und Lauth (1997) in Frage, dass Schüler\*innen mit Lernbehinderung in basalen kognitiven Lernprozessen *un*beeinträchtigt seien. Dies unterstreicht die Relevanz der Betrachtung von kognitiven Lernvoraussetzungen, die in dieser Arbeit verfolgt wird.

Für diese Arbeit wird zur Bestimmung und Eingrenzung der relevanten kognitiven Lernvoraussetzungen eine Perspektive mittleren Komplexitätsgrades ausgewählt, indem Lernen in einem instruktionspsychologischen Verständnis als expliziter und bewusster Wissenserwerb verstanden wird (siehe z. B. Hasselhorn & Gold, 2013; Renkl, 2015), wie er im Kontext von

schulischen Institutionen angebahnt wird. Lernen dient nach Wember und Heimlich (2014) dem Erwerb von Kompetenzen, unter die allgemein „kognitive Kenntnisse, Einstellungen, Emotionen, motorisches und soziales Verhalten“ (Wember & Heimlich, 2014, S. 54) gefasst werden.

Lauth, Brunstein und Grünke (2014, S. 21ff.) zeichnen nach, was beim Lernen geschieht: Neues Wissen muss von Lernenden in ein bestehendes Wissenssystem integriert und gespeichert werden, um es zu einem späteren Zeitpunkt „zum Zweck der Verhaltenssteuerung“ (ebd., S. 21) abrufen zu können. Nach neurowissenschaftlichen Annahmen basiert Lernen dabei auf der Verknüpfung von Nervenzellen, welche wiederholt ähnliche Erregungsmuster durchlaufen. Ursprünglich geht diese Annahme auf die sogenannte Hebb'sche Regel („what fires together wires together“, Kaiser, 2017, S. 713) zurück. Unterdessen wird aber eine Erklärung für das Lernen nicht nur in der Verknüpfung einzelner Neuronen auf der Mikroebene, sondern auch in der funktionalen Vernetzung von Hirnarealen und -modulen (Kaiser, 2017) auf der Makroebene durch ein gemeinsames Oszillieren in einem ähnlichen Rhythmus gesehen (Bassett et al., 2011; Klimesch, 2012; Kraut, Pitcock & Hart, 2004).

Lauth, Brunstein et al. (2014, S. 21) identifizieren zwei Ursachenbereiche, weswegen Beeinträchtigungen beim Lernen auftreten können, nämlich (1) „unzureichende Informationsverarbeitung“ und (2) „mangelnde und unzureichende Lernaktivität“. Damit verweisen sie auf das Modell der guten Informationsverarbeitung (Pressley, Borkowski & Schneider, 1989), das in der Instruktionspsychologie häufig für die Beschreibung von erfolgreichem Wissenserwerb herangezogen wird (z. B. Büttner, 2007; Hasselhorn, 2005; Klauer & Leutner, 2012). Die dort dargestellten Charakteristika und Vorgehensweisen erfolgreicher Lerner\*innen weisen Schüler\*innen mit Lernbehinderung nur in unzureichendem Maße auf. Die genannten zwei Ursachenbereiche (Lauth, Brunstein et al., 2014) finden sich etwas differenzierter auch bei Hasselhorn und Gold (2013, S. 69), die ebenfalls in Anlehnung an das Modell der guten Informationsverarbeitung die Charakteristika erfolgreicher Lerner\*innen in vier Bereiche einordnen: Aufmerksamkeit und Arbeitsgedächtnis-Funktionen, Vorwissen, Nutzung und Regulation von Lernstrategien, sowie emotionale und motivationale Faktoren.

In dieser Arbeit liegt der Betrachtungsfokus auf den ersten beiden Bereichen, da funktionierende Gedächtnisprozesse als eine Voraussetzung für erfolgreiche Strategienutzung angesehen werden können. Im Arbeitsgedächtnis werden neue Informationen für eine kurze Dauer bereitgehalten und falls nötig manipuliert, bevor sie im Langzeitgedächtnis mit dem bestehenden Vorwissen verknüpft und damit dauerhaft verankert werden. Auch in den Lern-Modellen bei Klauer und Leutner (2012) sowie Mietzel (2017) stellen Arbeits- und Langzeitgedächtnis zentrale Komponenten dar.

Borkowski, Schneider und Pressley (1989, S. 171ff.) führen einige Voraussetzungen an, in denen Defizite bei Kindern mit Lernbehinderung vermutet werden können: unter „neurological integrity“ fassen sie intakte kognitive Prozesse zusammen, zu denen sie auch das Kurz- bzw. Arbeitsgedächtnis zählen. Die Basis des konzeptuellen Vorwissens (*conceptual knowledge*) stellt eine weitere Voraussetzung dar, damit Kinder Strategien effizient nutzen können (Borkowski et al., 1989, S. 175), da weniger mentale Ressourcen benötigt werden, wenn die Aufgabeninhalte bereits bekannt sind. Auch Zielinski (1980; 1995) und Büttner (2003; 2007) sprechen dem Vorwissen eine hohe Bedeutung für die Erklärung von Lernbeeinträchtigungen zu. Da es sich beim Vorwissen um Inhalte handelt, die im Langzeitgedächtnis gespeichert sind, wird damit die Relevanz des Langzeitgedächtnisses als kognitive Lernvoraussetzung unterstrichen.

Zur Zusammenfassung: Der Funktionsweise von Gedächtniskomponenten wird im Modell der guten Informationsverarbeitung eine zentrale Rolle zugewiesen. Daher wird für die empirische Untersuchung von kognitiven Lernvoraussetzungen in dieser Arbeit der Fokus auf zwei individuelle kognitive Voraussetzungen gelegt, die beide für den Wissenserwerb eine hohe Relevanz haben (Hasselhorn, 2005): (1) das Arbeitsgedächtnis, da hier neue Informationen kurzfristig aufrechterhalten und manipuliert werden, und (2) das Langzeitgedächtnis, da hier neue Informationen mit bestehendem Wissen verknüpft und dadurch langfristig gespeichert werden. Da die beiden Komponenten beim Wissenserwerb in engem Zusammenspiel stehen, ist ihre Funktionalität für ein effektives Verankern von Informationen im Gedächtnis (sprich, Lernen) essentiell. Bislang ist keine Studie bekannt, die beide Komponenten (Arbeits- und Langzeitgedächtnis) gemeinsam in einer Stichprobe bei Kindern mit Lernbehinderung untersucht. In den folgenden Kapiteln werden die für die vorliegende Arbeit relevanten Aspekte von Arbeits- und Langzeitgedächtnis theoretisch eingeführt.

### **2.1. Arbeitsgedächtnis**

Allgemein definiert dient das Arbeitsgedächtnis zur kurzfristigen Aufrechterhaltung von Gedächtnisinhalten und deren mentaler Manipulation (Adams, Nguyen & Cowan, 2018; Baddeley, 2012). Das hier verwendete und in der Gedächtnisforschung sehr prominente Konzept des Arbeitsgedächtnisses geht auf Forschungsarbeiten von Baddeley und Hitch (1974) und Baddeley (1986) zurück, die damit auf Kritik am Mehrspeichermodell nach Atkinson und Shiffrin (1968) eingingen.

In der ursprünglichen Konzeption des Arbeitsgedächtnis-Modells nimmt Baddeley (1986) drei Komponenten an: eine zentrale Exekutive dient zur Steuerung von Aufmerksamkeitsprozessen und nimmt Manipulation von Gedächtnisinhalten vor. Dafür greift sie auf zwei modalitätsspezifische Subsysteme zurück, nämlich die phonologische Schleife (*phonological loop*, PL) für

akustische Informationen, und den visuo-räumlichen Notizblock (*visuo-spatial sketchpad*, VSS) für visuelle und räumliche Informationen. Beide Subsysteme bestehen wiederum aus einer Speicherkomponente und einem Mechanismus zur Aufrechterhaltung der im Speicher enthaltenen Information. Dieses ursprüngliche Modell wurde von Baddeley (2000) um den sog. episodischen Puffer zur Integration von Informationen aus verschiedenen Subsystemen erweitert. Auch hinsichtlich der Wirkungsweise einzelner Subkomponenten wurde das Modell weiter ausdifferenziert (Hasselhorn, Grube & Mähler, 2000). In der vorliegenden Arbeit steht aufgrund ihrer Bedeutung für den sprachlich vermittelten Wissenserwerb die phonologische Schleife im Fokus, daher wird im folgenden Unterkapitel detaillierter auf deren Funktionsweise und Prozesse eingegangen.

Die oben angeführte theoretisch plausible Relevanz des Arbeitsgedächtnisses für das Lernen aufgrund seiner zentralen Rolle im Modell der guten Informationsverarbeitung hat sich auch empirisch bestätigen lassen (für einen Überblick siehe Holmes, Gathercole & Dunning, 2010). So erwies sich das Arbeitsgedächtnis in zahlreichen deutschen und internationalen Studien als ein von der psychometrischen Intelligenz unabhängiger Prädiktor für schulische Leistungen (Alloway, 2009; Alloway & Alloway, 2010; Poloczek et al., 2012), sowie für den Spracherwerb (Baddeley, Gathercole & Papagno, 1998; Gathercole, Hitch, Service & Martin, 1997) und mathematische (Vorläufer-)Fertigkeiten (Röhm, Starke & Ritterfeld, 2017; für einen Überblick siehe Schulze & Kuhl, 2019).

Das Arbeitsgedächtnis wurde insbesondere für das Verständnis und die Erklärung von verschiedenen Beeinträchtigungsformen untersucht. So konnten Studien häufig spezifische Profile von Defiziten in unterschiedlichen Arbeitsgedächtnis-Bereichen sowie differenzielle Zusammenhänge mit schulischen Leistungen nachweisen. Auf den Forschungsstand zum Arbeitsgedächtnis bei Kindern mit Lernbehinderung wird weiter unten (Abschnitt 2.1.3) genauer eingegangen, im Folgenden seien einige Studien zu verschiedenen Störungsformen exemplarisch genannt:

- im Bereich des Schriftspracherwerbs (Brandenburg et al., 2013; 2015; Gathercole, Alloway, Willis & Adams, 2006; Schmidt et al., 2020; Schuchardt, Mähler & Hasselhorn, 2008; für Meta-Analysen siehe O’Shaughnessy & Swanson, 1998; Peng & Fuchs, 2016).
- bei Schwierigkeiten in Mathematik (Kluszczewski et al., 2015; 2018; Schuchardt, Mähler et al., 2008)
- bei Sprachentwicklungsstörungen (Alloway & Archibald, 2008; Jackson, Leitao & Claessen, 2016; Röhm, 2020)
- und bei Intelligenzminderungen (Henry & MacLean, 2003; Henry & Winfield, 2010; Mähler & Schuchardt, 2016; Poloczek et al., 2012).

Zwar lässt sich ein enger Zusammenhang vom Arbeitsgedächtnis zu Maßen der Intelligenz (Alloway, 2010; Cornoldi & Giofrè, 2014;) sowie zu Aufgaben zur Kategorisierung und Inferenz verzeichnen (Craig & Lewandowsky, 2013; Lewandowsky, Yang, Newell & Kalish, 2012); jedoch betonen Conway, Kane und Engle (2003), dass Arbeitsgedächtnis und fluide Intelligenz ( $g$ ) nicht identisch sind. Entsprechend dazu sehen Shipstead, Harrison und Engle (2016) in Intelligenz und Arbeitsgedächtnis zwei komplementär wirkende Prozesse, die für das Lösen von komplexen Problemen beide notwendig und wechselseitig aufeinander angewiesen sind.

In Bezug auf beeinträchtigte Schulleistungen scheinen Arbeitsgedächtnis und Intelligenz eher unabhängig voneinander zu wirken: Mähler und Schuchardt (2016) konnten mit einem gekreuzten Design (vier Gruppen: Schüler\*innen mit durchschnittlichen vs. unterdurchschnittlichen IQ-Werten und mit durchschnittlichen vs. unterdurchschnittlichen Schulleistungen) zeigen, dass geringere Leistungen im Arbeitsgedächtnis mit niedrigen Schulleistungen assoziiert waren, und zwar unabhängig vom Intelligenzlevel. Auch die Befunde von Brandenburg et al. (2013) sprechen dafür, dass bei Kindern mit Problemen im Schriftspracherwerb die Arbeitsgedächtnis-Leistungen vor allem mit den niedrigen Schulleistungen zusammenhängen und dies unabhängig davon, ob das IQ-Diskrepanzkriterium (siehe Abschnitt 1.2.3) erfüllt war oder nicht.

Für diese Arbeit wird aus den Komponenten des Arbeitsgedächtnisses die phonologische Schleife fokussiert, da sie für das häufig auf verbal vermittelten Informationen basierte schulische Lernen besonders relevant erscheint. Für die Untersuchung in der ersten Studie wird daher in den folgenden Abschnitten genauer auf deren Funktionsweise und Komponenten eingegangen und der Untersuchungsbereich auf die folgenden drei Aspekte eingegrenzt: (1) die *Gesamtkapazität* der PL (Hasselhorn et al, 2000), (2) den *Rehearsalprozess* zur Aufrechterhaltung einer Gedächtnisspur (Hasselhorn et al., 2000) und (3) den *Redintegrationprozess* zur Rekonstruktion von bereits teilweise zerfallenen Spuren aus dem Langzeitgedächtnis (Schweickert, 1993). Diese drei Aspekte werden im Rahmen je einer eigenen Teilfragestellung in Studie 1 untersucht.

Im Folgenden werden zunächst die phonologische Schleife und die Prozesse von Rehearsal und Redintegration auf konzeptueller Ebene eingeführt und im Anschluss die empirischen Effekte, an denen sie messbar sind, näher erläutert. Es folgt eine Darstellung von Befunden zum Arbeitsgedächtnis bei Kindern mit Lernbehinderung, und abschließend wird auf empirische Ergebnisse zur Trainierbarkeit des Arbeitsgedächtnisses eingegangen.

### **2.1.1. Aufbau der phonologischen Schleife**

Die phonologische Schleife ist ein modalitätsspezifisches Subsystem des Arbeitsgedächtnisses, das für die kurzfristige Aufrechterhaltung von verbalen Informationen dient. Die Bewältigung der Anforderung, sich kurzfristig einen akustisch-verbalen Gedächtnisinhalt zu merken,



kann anhand folgender Analogie veranschaulicht werden: Eine Information wird zunächst im phonetischen Speicher wie auf einem „Tonband in Endlosschleife mit sehr kurzer Aufnahmekapazität“ (Hasselhorn & Gold, 2013, S. 79) für eine eng begrenzte Dauer von ungefähr zwei Sekunden vorgehalten (Baddeley, Thomson & Buchanan, 1975). Danach zerfällt die Spur rasch bzw. wird durch neu aufzunehmende Informationen überschrieben. Nach der Konzeption des Arbeitsgedächtnisses von Hasselhorn et al. (2000) lässt sich der phonetische Speicher innerhalb der PL weiter differenzieren in zwei Teilkomponenten: die Reproduktionsleistung in einer Behaltensaufgabe von akustischem Material ist zum einen durch die Größe des Speichers, d. h. gewissermaßen die „Länge des Tonbands“, bestimmt; zum anderen hat der Präzisionsgrad der Verarbeitung bzw. Einspeicherung der akustischen Information Einfluss auf die Leistung beim Abruf, also die „Qualität der Tonbandaufnahme“, oder „Wie ‚klar‘ ist das phonetische Sprachmuster repräsentiert?“ (Hasselhorn & Gold, 2013, S. 81).

### 2.1.1.1. Rehearsalprozess

Dem Verfall von Informationen stehen mehrere Prozesse entgegen, wie Rehearsal, attentionales Refreshing, Elaboration, Konsolidierung oder Updating im Sinne einer Bereinigung des Speichers (siehe z. B. Bartsch, Singman & Oberauer, 2018; Bayliss, Bogdanovs & Jarrold, 2015). Im ursprünglichen Modell nach Baddeley (1986) und Hasselhorn et al. (2000) spielt der phonologisch basierte Rehearsalprozess eine zentrale Rolle, daher wird auf diesen im Folgenden vertieft eingegangen.

Beim Rehearsalprozess werden die Inhalte des phonetischen Speichers als „inneres Sprechen“ subvokal artikuliert und dadurch aufrechterhalten. Der Rehearsalprozess hat nicht nur die Aufgabe der Aufrechterhaltung von Information innerhalb der PL, sondern wird auch für die modalitätsübergreifende Aufbereitung und Rekodierung von visuellem Material (z. B. geschriebenen Wörtern) in akustische Information benötigt. Die Wirksamkeit von Rehearsal machten Baddeley et al. (1975) an einer Reihe von empirisch messbaren Effekten wie dem Wortlängeneffekt und dem Effekt für artikulatorische Suppression fest. Gemäß Hasselhorn et al. (2000) kann der Rehearsalprozess wiederum in zwei Aspekte ausdifferenziert werden: zum einen die Geschwindigkeit, d. h. die innere Sprechrate, zum anderen die automatische Aktivierung des Rehearsalprozesses. Die Rehearsalgeschwindigkeit wird empirisch an der Korrelation von Gedächtnisleistung mit der Sprechrate, die automatische Aktivierung des Rehearsalprozesses am Auftreten eines Wortlängeneffektes festgemacht. Auf beide Effekte wird in Abschnitt 2.1.2.2 näher eingegangen.

## 2.1.1.2. Redintegrationprozess

Neben dem phonologisch basierten Rehearsalprozess zur Aufrechterhaltung von Gedächtnisspuren in der PL wird auch auf zahlreiche Einflüsse aus dem Langzeitgedächtnis hingewiesen (Stuart & Hulme, 2009; Thorn, Frankish & Gathercole, 2009). So kann eine teilweise zerfallene Gedächtnisspur aus dem Langzeitgedächtnis rekonstruiert werden, was als Redintegrationprozess bezeichnet wird (Grube, Lingen & Hasselhorn, 2008; Roodenrys & Miller, 2008; Schweickert, 1993). Dieser Prozess ist nicht direkt in der Konzeption des Arbeitsgedächtnisses nach Baddeley verankert, kann aber die dort bislang unterspezifizierte Interaktion mit dem Langzeitgedächtnis abbilden und erscheint gut mit dem Modell vereinbar. Man kann sich Redintegration so vorstellen, dass die noch im phonetischen Speicher erhaltenen Fragmente der Gedächtnisspur als Hinweisreize für eine Suche im Langzeitgedächtnis genutzt (Roodenrys & Miller 2008) und somit durch das Vorwissen ergänzt werden. Diese Rekonstruktion der Gedächtnisspur macht den korrekten und vollständigen Abruf wieder möglich.

Schweickert (1993, S. 169f.) konzipiert den Abrufprozess in zwei Schritten, nämlich erstens den direkten Abruf aus dem phonetischen Speicher und zweitens die Rekonstruktion aus dem Langzeitgedächtnis. In beiden Schritten können mit einer bestimmten Wahrscheinlichkeit Fehler auftreten, was Schweickert (1993) in einem Entscheidungsbaum (*multinomial processing tree*, MPT) darstellt. Mit einer Wahrscheinlichkeit  $I$  tritt der Fall ein, dass der Gedächtnisinhalt noch vollständig intakt aus dem phonetischen Speicher abrufbar ist, also korrekt wiedergegeben werden kann (Ereignis  $C$ ). Dies ist die schnellste und auch am wenigsten fehleranfällige Variante, ihr würde also in einem effizienten System der Vorzug gegeben. Mit entsprechender Gegenwahrscheinlichkeit  $(1 - I)$  ist die direkte Wiedergabe jedoch aufgrund der bereits zu stark zerfallenen Spur nicht möglich. Dann muss anhand der verbleibenden Information das Item zeitaufwändiger und fehleranfälliger aus dem Langzeitgedächtnis unter Rückgriff auf das phonemische und lexikalische Hintergrundwissen rekonstruiert werden (ebd., S. 170f.). Diese Rekonstruktion ist mit einer Wahrscheinlichkeit  $R$  erfolgreich und ermöglicht einen korrekten Abruf, mit entsprechender Gegenwahrscheinlichkeit  $(1 - R)$  ist dies nicht der Fall, sodass ein Fehler (Ereignis  $E$ ) produziert wird.

Roodenrys und Miller (2008) formulieren die Annahmen von Redintegration als ein restringiertes Rasch-Modell, bei dem die Unversehrtheit der Spur im phonetischen Speicher als Personenparameter und die Verfügbarkeit des Wortes im Langzeitgedächtnis als Itemparameter betrachtet werden. Wenn also ein teilweise zerfallenes Item rekonstruiert werden muss, ist der Erfolg zum einen davon abhängig, wie gut dieses Item überhaupt bekannt ist, zum anderen davon, wie distinktiv es aktiviert werden kann. Bei einer Liste mit sehr ähnlich klingenden Items werden durch eine nur teilweise vorhandene Spur viele ähnliche Items im Langzeitgedächtnis

aktiviert, sodass die Wahrscheinlichkeit für den Abruf des korrekten Wortes sinkt (Roodenrys & Miller, 2008). Im Vergleich zu dem ursprünglichen MPT-Modell konnten sie für das restriktierte Raschmodell eine bessere Modellpassung zu experimentell erhobenen Daten einer seriellen Recallaufgabe nachweisen. Anders als das MPT-Modell gehen Roodenrys & Miller (2008) davon aus, dass Redintegration immer als automatisch ablaufender Prozess stattfindet, sich aber auf intakte Spuren weniger stark auswirkt.

### 2.1.2. Empirische Effekte zur Untersuchung der phonologischen Schleife

#### 2.1.2.1. Immediate Serial Recall Design

Die Leistung der phonologischen Schleife kann über Aufgaben zum Nachsprechen von Kunstwörtern und über Wortspannen-Aufgaben erfasst werden (Hasselhorn et al., 2012). In der vorliegenden Arbeit wurden Wortspannen unter vier verschiedenen Wortlängen- und Lexikalitätsbedingungen eingesetzt (Grube et al., 2008). Zum besseren Verständnis des Versuchsaufbaus wird zunächst das Wortspannen-Paradigma der unmittelbaren seriellen Gedächtnisspanne (*immediate serial recall*) eingeführt. Die oben genannten Effekte ergeben sich aus der systematischen Variation der Stimuli bzw. deren Zusammensetzung in Listen, die von den Proband\*innen unmittelbar seriell reproduziert werden sollen.

Mit Aufgaben zum immediate serial recall können Rückschlüsse auf die Gesamtkapazität der PL und die Funktionsweise ihrer Aufrechterhaltungs-Prozesse gezogen werden. Dabei werden die Memoranda (d. h. die zu erinnernden Wort-, Kunstwort- oder Zahlenreihen) in einem festen Rhythmus (zwischen 0.5-2 Sekunden) akustisch oder schriftlich dargeboten. Die Proband\*innen werden instruiert, dass sie die gesamte Serie bis zum Ende anhören und im Anschluss vollständig und in der korrekten Reihenfolge wiedergeben sollen (Hasselhorn et al., 2012). Je länger eine Sequenz ist, also je mehr (Kunst-)Wörter oder Zahlen sie enthält, desto schwieriger wird es, sie korrekt wiederzugeben. Außerdem können die Zusammensetzung der Sequenzen (z. B. alle Wörter einer Sequenz sind einander sehr ähnlich) als auch (lexikalische) Charakteristika der Stimuli selbst die Behaltensleistung beeinflussen.

Die Erfassung des korrekten Recalls kann auf Listen- und auf Itemebene erfolgen. Bei der Listenebene wird nur bewertet, ob die gesamte Sequenz korrekt wiedergegeben wurde; bei der Itemebene wird für jedes Element einer Sequenz einzeln bewertet. Dadurch kann differenziert werden, ob das Element überhaupt erinnert wurde (*item-information*) und ob es zudem an der korrekten Position genannt wurde (*order-information*). Auch wenn diese feinere Unterscheidung Vorteile bietet, indem Primacy- und Recency-Effekte (Burgess & Hitch, 2006; Gupta, 2005; 2009) oder differenzielle Auswirkungen von Redintegration für Item- vs. Order-Information (Roodenrys & Miller, 2008) betrachtet werden können, sind diese Effekte für die vorliegende

Arbeit zu detailliert und eher für grundlagenorientierte Forschungsfragen geeignet. Daher wurden in Studie 1 die Arbeitsgedächtnis-Leistungen auf Listenebene erfasst.

Weiterhin ist bei der Konstruktion von Gedächtnisspannen-Aufgaben zwischen sog. *closed-pool* und *open-pool* Varianten zu unterscheiden. Bei einem *closed-pool* werden die Sequenzen für jeden Trial aus ca. 8-15 Wörtern neu zusammengesetzt, sodass sich die Wörter über die verschiedenen Trials hinweg wiederholen und nur die Zusammensetzung und Reihenfolge variiert. Bei einem *open-pool* werden die Sequenzen aus Stimuli aus einer Sammlung von bis zu 200 Wörtern jeweils neu und einzigartig zusammengestellt. Bei der ersten Variante ist die Wiedergabeleistung insgesamt besser (Roodenrys & Quinlan, 2000), da aufgrund der begrenzten Anzahl an Wörtern die Repräsentation im Langzeitgedächtnis leichter aufrechterhalten werden kann. Dadurch kann der Redintegrationprozess besser genutzt werden, und die Intrusion durch listenfremde Stimuli ist weniger wahrscheinlich. In Anlehnung an die Vorgehensweise in der Arbeitsgedächtnis-Testbatterie für 5-12 Jährige (Hasselhorn et al., 2012) als standardisiertes Testinstrument wurden in dieser Arbeit die Sequenzen unterschiedlicher Länge aus einem *closed-pool* von jeweils neun Elementen zusammengestellt.

Mit den dargestellten Wortspannen-Aufgaben im *immediate serial recall* Format werden primär Speicherkapazität und Prozesse zur Aufrechterhaltung der Gedächtnisspur in der modalitätsspezifischen phonologischen Schleife untersucht. Für komplexere Anforderungen, auch an die zentrale Exekutive zur Manipulation der Gedächtnisinhalte, können z. B. N-Back- und Complex-Span Aufgaben genutzt werden, auf die aber hier nicht näher eingegangen wird (siehe Hasselhorn et al., 2012; Redick, et al., 2012).

Nachdem nun das generelle Design vorgestellt wurde, werden im Folgenden die für die Konzeption der PL relevanten empirischen Effekte erläutert, gegliedert in phonologische und lexikalische Effekte. Unter die phonologisch basierten Effekte sind der Wortlängeneffekt, der Zusammenhang zwischen Sprechrate und Gedächtnisspanne und der Effekt von artikulatorischer Suppression zu fassen. Der Begriff ‚lexikalisch‘ ist hier sehr breit gewählt, um Effekte zu bezeichnen, die nicht rein auf Ebene der phonologischen Schleife zu erklären sind, sondern einen Rückgriff auf Wissen im Langzeitgedächtnis erfordern (siehe Thorn & Frankish, 2005; Thorn et al., 2009). Darunter fallen der Lexikalitätseffekt, die Neighborhood-Size und semantische Effekte.

### 2.1.2.2. Phonologische Effekte

Die empirische Operationalisierung des Rehearsalprozesses, genauer gesagt dessen automatische Aktivierung (Hasselhorn et al., 2000), wird im Wortlängeneffekt gesehen: In einer *immediate serial recall*-Aufgabe ist die Wiedergabeleistung für Sequenzen, die aus kurzen Wörtern bestehen, deutlich höher als für Sequenzen mit langen Wörtern (Baddeley et al., 1975). Das zeigt

sich darin, dass längere Sequenzen von kurzen Wörtern in der korrekten Reihenfolge wiedergegeben werden können. Am augenfälligsten ist dies bei einem Kontrast von einsilbigen mit mehrsilbigen Wörtern, wie es bspw. im Rahmen der AGTB 5-12 (Hasselhorn et al. 2012) umgesetzt wird. Aber auch bei gleicher Silbenanzahl scheint eine unterschiedlich lange Artikulationsdauer und die phonologische Komplexität für einen Wortlängeneffekt zu sorgen (Baddeley et al., 1975; Service, 1998).

Der Wortlängeneffekt kann gemäß der Modellvorstellung von Hasselhorn et al. (2000) als die automatische Aktivierung des phonologisch basierten Rehearsalprozesses interpretiert werden (Mähler & Hasselhorn, 2003): Wie oben beschrieben kann im phonetischen Speicher nur eine zeitlich eng begrenzte Menge an Information bereitgehalten werden, danach verblässen die Informationen von alleine (*decay*) oder werden durch neue Informationen überschrieben (*interference*; Berman, Jonides & Lewis, 2009; Farrell et al., 2016). Durch den Rehearsalprozess des inneren Vorsprechens werden diese Informationen nun ‚dichter gepackt‘ und durch Wiederholung im phonetischen Speicher aufrechterhalten. Da in der gleichen begrenzten Zeit (bis ein neuer Stimulus im Rhythmus der Aufgabenpräsentation erfasst werden muss) mehr kurze als lange Wörter subvokal artikuliert werden können, kommt es zum Wortlängeneffekt.

Zwei weitere empirisch beobachtbare Effekte sprechen für die Annahme, dass der Rehearsalprozess phonologisch basiert ist und als eine Form von innerem Sprechen angesehen werden kann. Erstens konnten Baddeley et al. (1975) einen positiven korrelativen Zusammenhang der Behaltensleistung mit der Artikulationsrate feststellen, welche bemisst, wie viele Wörter eine Person pro Minute durchschnittlich aussprechen kann. Hasselhorn et al. (2000) sehen in der Sprechrate einen Indikator für die Geschwindigkeit des Rehearsalprozesses, die einen Einfluss auf die Gesamtkapazität der phonologischen Schleife hat.

Zweitens ist eine Reduktion bzw. ein vollständiges Wegfallen des Wortlängeneffektes bei gleichzeitiger konkurrierender Artikulation zu beobachten (Baddeley et al., 1975; Jacquemot, Dupoux & Bachoud-Lévi, 2011): Wird eine Testperson aufgefordert, zeitgleich mit der Bearbeitung der Behaltensaufgabe eine (sinnfreie) Silbe leise auszusprechen, so verringert sich der Wortlängeneffekt deutlich oder verschwindet vollkommen (Baddeley et al., 1975; Baddeley, Lewis & Vallar, 1984; siehe jedoch Souza & Oberauer, 2018 für dem widersprechende Befunde). Damit kann die Wirkweise des Rehearsalprozesses folgendermaßen erklärt werden: Da der Rehearsalprozess auf der Artikulation der Memoranda beruht, ist er durch das Aussprechen der zusätzlichen, aufgaben-irrelevanten Silbe blockiert. Die Testperson kann also in beiden Wortlängenbedingungen gleichermaßen nur die wenigen im phonetischen Speicher verbliebenen Wörter wiedergeben. Durch den vorgegebenen Rhythmus der Stimulusdarbietung ist die Menge

der Gedächtnisinhalte im phonetischen Speicher in beiden Bedingungen (kurz vs. lang) gleich, folglich ist kein Wortlängeneffekt mehr feststellbar.

Insofern ist es plausibel, sich den Rehearsalprozess als inneres, subvokales Sprechen vorzustellen, bei dem die Inhalte des phonetischen Speichers stumm wiederholt werden, um sie auf diese Weise aufrechtzuerhalten. Je schneller eine Testperson artikulieren kann, desto mehr Wörter können in der gleichen begrenzten Zeit wiederholt werden; wenn diese subvokale Artikulation der Inhalte des phonetischen Speichers durch eine konkurrierende offene Artikulation blockiert ist, kann der Rehearsalprozess nicht stattfinden.

Angesichts neuerer Erkenntnisse ist es jedoch offen, ob der Wortlängeneffekt vollständig durch den Rehearsalprozess erklärt werden kann (Campoy, 2008; Guitard, Gabel, Saint-Aubin, Surprenant & Neath, 2018). Es konnte in einer Reihe von Studien gezeigt werden, dass der für das Rehearsal so zentrale Wortlängeneffekt auch anderweitig durch lexikalische Faktoren erklärt werden kann. So stellten Jalbert, Neath, Bireta & Surprenant (2011) fest, dass in den meisten Studien die Wortlänge mit der Neighborhood-Size (der phonologischen oder orthographischen Überlappung eines Wortes mit anderen Wörtern der Sprache, s.u.) konfundiert waren. Bei Kontrolle der Neighborhood-Size wurde der Wortlängeneffekt nicht mehr signifikant (Derraugh, Neath, Surprenant, Beaudry & Saint-Aubin, 2017; Jalbert, Neath, Bireta et al., 2011); und Jalbert, Neath und Surprenant (2011) konnten sogar zeigen, dass der Neighborhood-Size Effekt den Wortlängeneffekt überlagert. Andererseits plädieren Jacquemot et al. (2011) angesichts dissoziativer Ergebnisse in zwei neurologischen Fallstudien für die Wirksamkeit des Rehearsalprozesses.

### 2.1.2.3. Lexikalische Effekte

Die bisher genannten Effekte basieren vornehmlich auf der lautlichen, phonologischen Ebene, und sind ohne den Einfluss des Langzeitgedächtnisses erklärbar. Häufig wurde jedoch auch ein Effekt für lexikalische und sublexikalische Variablen auf die Reproduktionsleistung von Wortspannen und Kunstwörtern nachgewiesen, für den das Langzeitgedächtnis zur Erklärung herangezogen werden muss (Gathercole, Pickering, Hall & Peaker, 2001; Thorn & Page, 2009). Lexikalische Variablen beziehen sich auf Wörter als Ganzes, z. B. die Lexikalität („echtes Wort vs. Kunstwort“)<sup>8</sup> während sublexikalische Variablen kleinere Einheiten betreffen (Janse & Newman, 2013), z. B. die Häufigkeit von Buchstabenkombinationen in einem Wort (Bi- oder Trigram-Frequency; Peereboom & Content, 1995). Zu den hier aufgeführten Effekten zählen der

---

<sup>8</sup> Im eigentlichen Sinne ist auch die Wortlänge eine lexikalische Variable. Da im Rahmen der AG-Konzeption der Wortlängeneffekt jedoch rein phonologisch basiert und nicht auf das Langzeitgedächtnis angewiesen ist, wird die Wortlänge im vorigen Kapitel angeführt.

Lexikalitätseffekt, der Einfluss von Neighborhood-Size und Häufigkeit sowie semantische Variablen.

Analog zur Operationalisierung des Rehearsalprozesses anhand des Wortlängeneffektes wird der Redintegrationprozess anhand des Lexikalitätseffektes gemessen: echte Wörter können besser wiedergegeben werden als Kunstwörter, bzw. Wortspannen mit echten Wörtern besser als Wortspannen mit Kunstwörtern (Gathercole et al., 2001; Grube et al., 2008; Ritchie, Tolan & Tehan, 2015; Turner, Henry, Smith & Brown, 2004). Dieser Effekt lässt sich in der oben eingeführten Konzeption des Redintegrationprozesses (Schweickert, 1993) plausibel damit erklären, dass für echte Wörter ein Eintrag im Langzeitgedächtnis vorhanden ist, der für die Rekonstruktion einer teilweise zerfallenen Gedächtnisspur genutzt werden kann; bei Kunstwörtern ist dies nicht der Fall, sodass keine direkte Rekonstruktion des Stimulus als Ganzem möglich ist.

Aber auch subtilere Effekte von Häufigkeit und Neighborhood-Size<sup>9</sup> lassen darauf schließen, dass Proband\*innen je nach Stimulus unterschiedlich gut auf ihr Wissen im Langzeitgedächtnis zurückgreifen können. So kann z. B. der Effekt der Wort-Häufigkeit (Allen & Hulme, 2006) nur durch eine Beteiligung des Langzeitgedächtnisses erklärt werden, da die Häufigkeit der Nutzung nicht den Wörtern selbst inhärent ist, sondern sich im Langzeitgedächtnis niederschlägt und damit für eine stärkere Repräsentation und Abrufbereitschaft sorgt (Majerus, 2009). Auch werden Kunstwörter besser behalten, wenn sie in phonotaktischer Hinsicht echten Wörtern ähnlicher sind, sie also grundsätzlich mit den Mustern echter Wörter in der jeweiligen Sprache vereinbar und aus diesen bekannt sind (Gathercole, 1995; Jones & Witherstone, 2011; Thorn & Frankish, 2005; Thorn, Gathercole & Frankish, 2005). Die phonotaktische Ähnlichkeit betrifft z. B. die Silbenstruktur, die Häufung von Konsonanten oder das Auftreten von morphologischen Endungen, wie sie in der verwendeten Sprache üblicherweise auftreten.

Messbar ist der Effekt der phonotaktischen Plausibilität für Kunstwörter bspw. an der Bi- oder Trigramm-Frequenz oder der Neighborhood-Size. Aber auch auf die Wiedergabe von echten Wörtern hat die Neighborhood-Size einen positiven Einfluss (Allen & Hulme, 2006; Clarkson, Roodenrys, Miller & Hulme, 2017; Derraugh et al., 2017). Der Effekt dieser Variablen wird damit erklärt, dass unter Rückgriff auf vorhandenes Hintergrundwissen über die lexikalischen und sublexikalischen Eigenschaften von Wörtern die Stimuli möglichst plausibel rekonstruiert werden (Thorn & Frankish, 2005; Jalbert, Neath, Bireta et al., 2011; Jalbert, Neath & Surprenant 2011). So kann eine Testperson bspw. anhand der Vertrautheit mit einem Stimulus, sei es über die Häufigkeit eines Wortes oder die phonotaktische Plausibilität eines Kunstwortes,

---

<sup>9</sup> Die Neighborhood-Size ist ein Maß für die (orthographische) Überlappung eines Wortes mit anderen Wörtern der Zielsprache. Eine genauere Erläuterung findet sich in Kapitel 2.2.3.

aus mehreren Kandidaten den wahrscheinlichsten auswählen, was Lewandowsky und Farrell (2000) ebenfalls auf die Wirkung von Redintegration zurückführen.

Schließlich können auch semantische Effekte auf die Behaltensleistung beobachtet werden. So können konkrete Wörter besser erinnert werden als abstrakte (Acheson, Postle & MacDonald, 2010; Allen & Hulme, 2006; Meltzer et al., 2016; Walker & Hulme, 1999). Sogar die affektive Valenz scheint sich auf die Behaltensleistung auszuwirken, da Monnier und Syssau (2008) einen positiven Effekt für angenehme im Gegensatz zu neutralen Wörtern nachweisen konnten. Savill, Metcalfe, Ellis und Jefferies (2015) verglichen semantische und phonologisch basierte Kategorisierung der Stimuli in Wortspannen-Aufgaben. Zwar kostete die zusätzliche Anforderung insgesamt kognitive Ressourcen, was in beiden Bedingungen zu einer Verschlechterung im Vergleich zu einer Aufgabe ohne Kategorisierung führte; die semantische Kategorisierung zeigte jedoch deutlich geringere abträgliche Effekte (siehe auch Acheson, MacDonald & Postle, 2011).

Um zusammenzufassen: Die Fähigkeit, Sequenzen für eine kurze Dauer aufrechtzuerhalten, scheint von Merkmalen des Arbeitsgedächtnisses selbst abzuhängen, wie an den Effekten von Wortlänge und konkurrierender Artikulation sowie dem Zusammenhang mit der Sprechrate gezeigt wurde. Gleichzeitig greift aber eine rein phonologisch basierte Konzeption des verbalen Arbeitsgedächtnisses zu kurz, da das Langzeitgedächtnis einen maßgeblichen Einfluss auf die Wiedergabeleistung hat, wie sich an zahlreichen lexikalischen Effekten beobachten lässt (für eine umfassende Aufstellung an Variablen siehe Jalbert, Neath, Bireta et al., 2011). Besonders die Interaktion zwischen Arbeits- und Langzeitgedächtnis, wie sie der Redintegrationprozess abbildet, erscheint für das Verständnis von Lernen als Wissenserwerb zentral; jedoch bleibt sie im Modell nach Baddeley deutlich unterspezifiziert und wurde bislang bei Kindern mit Lernbehinderung nicht genauer untersucht. Die vorliegende Arbeit setzt an dieser Forschungslücke an, indem sie neben der Gesamtkapazität der PL und der Aktivierung des Rehearsalprozesses auch den Redintegrationprozess berücksichtigt.

### **2.1.3. Forschungsstand zum Arbeitsgedächtnis bei Lernbehinderung<sup>10</sup>**

Die grundsätzliche Bedeutung des Arbeitsgedächtnisses für verschiedene Störungsformen und schulische Leistungen wurde bereits oben (Abschnitt 2.1) angeführt. Dass auch bei Kindern mit Lernbehinderung Beeinträchtigungen allgemein im Arbeitsgedächtnis und auch speziell in der phonologischen Schleife vorliegen, ist ebenfalls gut belegt (Büttner & Hasselhorn, 2007;

---

<sup>10</sup> Der in diesem Abschnitt 2.1.3 geleistete Forschungsüberblick orientiert sich sinngemäß an der Darstellung aus dem bereits publizierten Beitrag und baut darauf auf (Studie 1, Seite 3: Bruns G., Ehl B. & Grosche M. (2019). Verbal Working Memory Processes in Students With Mild and Borderline Intellectual Disabilities: Differential Developmental Trajectories for Rehearsal and Redintegration. *Frontiers in Psychol.* 9:2581. doi:10.3389/fpsyg.2018.02581). Mit dieser Kenntlichmachung wird den Bestimmungen der Promotionsordnung vom 17.11.2017 der Bergischen Universität Wuppertal entsprochen.



Pickering & Gathercole, 2004, für einen Überblick siehe Lifshitz, Kilberg & Vakil, 2016). Dagegen sind die Befunde zu den Entwicklungsmustern, insbesondere die Unterscheidung von Entwicklungsverzögerung vs. qualitatives Defizit (die im Rahmen des nächsten Kapitels 3 genauer erläutert wird), uneinheitlich – häufig werden innerhalb einer Studie je nach untersuchter Arbeitsgedächtnis-Domäne unterschiedliche Ergebnisse berichtet.

Einerseits sprechen die Ergebnisse einer Reihe von Studien dafür, dass das Arbeitsgedächtnis bei Kindern mit Lernbehinderung (bzw. intellektuellen Beeinträchtigungen) in seiner Entwicklung lediglich verzögert ist, d. h. dass es nicht über den Rahmen des allgemeinen mentalen Alters hinaus beeinträchtigt ist (Danielsson, Henry, Messer & Rönnerberg, 2012; Mähler, 2007, van der Molen, van Luit, Jongmans, van der Molen, 2009). Spezifischer werden dabei entsprechend geringfügige Defizite sowohl hinsichtlich der Kapazität der phonologischen Schleife (Hasselhorn & Mähler, 2007; Henry & MacLean, 2002; Rosenquist, Conners & Roskos-Ewoldsen, 2003), als auch hinsichtlich des Rehearsalprozesses (van der Molen et al., 2007; Schuchardt, Mähler & Hasselhorn, 2011) berichtet. Eine Entwicklungsverzögerung des Rehearsalprozesses ist auch mit anderen, non-verbalen Erhebungsverfahren nachweisbar (Poloczek et al., 2016) und scheint bis ins junge Erwachsenenalter zu persistieren (Mähler & Hasselhorn, 2003). Weiterhin wurden Hinweise für eine Entwicklungsverzögerung in der visuo-räumlichen Komponente (Alloway, 2010; Rosenquist et al., 2003) und zusätzlich dazu in der zentralen Exekutive erbracht (Schuchardt, Gebhardt & Mähler, 2010; Mähler, 2007). Bezogen auf den theoretischen Rahmen der exekutiven Funktionen schien für den raschen Wechsel zwischen verschiedenen Aufgaben (*switching*), eine Entwicklungsverzögerung vorzuliegen (Danielsson et al., 2012). Bei einem Vergleich von Kindern mit Lernbehinderung und spezifischen Lernstörungen kamen Mähler und Schuchardt (2009) zu dem Ergebnis, dass beide Gruppen gleichermaßen Defizite im Arbeitsgedächtnis aufwiesen, sich aber nicht voneinander unterschieden.

Andererseits lassen sich auch Hinweise dafür finden, dass die Entwicklung des Arbeitsgedächtnisses bei Kindern mit Lernbehinderung spezifisch und qualitativ beeinträchtigt ist, und zwar dergestalt, dass die Beeinträchtigung deutlich stärker ausfällt, als anhand des mentalen Alters zu erwarten wäre. Dies ist z. B. hinsichtlich der phonologischen Schleife bei Kindern mit Lernbehinderung im Alter von ca. 10 Jahren der Fall (Mähler, 2007). Henry und MacLean (2002) sowie van der Molen et al. (2007; 2014) wiesen hinsichtlich der Kapazität der phonologischen Schleife eine spezifische Beeinträchtigung nach. Es zeigten sich zum einen gravierende Schwierigkeiten in der Subkomponente des phonetischen Speichers bei Kindern (Schuchardt et al., 2010; 2011) und jungen Erwachsenen (Mähler & Hasselhorn, 2003); zum anderen liegen auch für den Rehearsalprozess Ergebnisse vor, die für eine spezifische Abweichung der Entwicklung

sprechen (Hasselhorn & Mähler, 2007; Rosenquist et al., 2003). Poloczek, Büttner und Hasselhorn (2014) wiesen zwar bei Kindern mit intellektuellen Beeinträchtigungen einen geringfügigen Wortlängeneffekt nach, führten diesen aber auf einen sog. *output delay* zurück, da längere Wörter mehr Zeit zur Artikulation beanspruchen, was sich bei der Wiedergabe insbesondere auf die hinteren Positionen der Sequenz negativ auswirkt (Doshier & Ma, 1998).

Weitere Studien lassen darauf schließen, dass das Ausmaß der Beeinträchtigung im Arbeitsgedächtnis mit dem Schweregrad der intellektuellen Beeinträchtigung (Henry, 2001) und dem generellen mentalen Alter (Mähler, 2007) einhergeht. Schuchardt et al. (2011) untersuchten drei Gruppen gleichen mentalen Alters mit unterschiedlichem Schweregrad einer Intelligenzminderung; hier wurde lediglich für die schwerer beeinträchtigte Gruppe der Kinder mit mild intellectual disability (IQ 50-69) ein qualitatives Defizit festgestellt, während die Gruppe der Kinder mit borderline intellectual disability (IQ 70-84) Leistungen im Rahmen des mentalen Alters aufwies. Bezüglich der exekutiven Funktionen zeigten Kinder Defizite qualitativ-struktureller Art vor allem in Aufgaben der Inhibition und in non-verbalen Aufgaben mit hohen Anforderungen an die exekutive Steuerung (Danielsson et al., 2012; van der Molen et al., 2009). Eine andere Herangehensweise zur Überprüfung der Struktur wählte van der Molen (2010), die mit einer konfirmatorischen Faktorenanalyse die Passung des Drei-Komponenten-Modells nach Baddeley bei Kindern mit intellektuellen Beeinträchtigungen testete und keine gute empirische Modellpassung feststellen konnte. Dagegen schien bei Kindern mit spezifischen Lernstörungen eine solche Struktur-Äquivalenz vorzuliegen, wie Schuchardt, Roick, Mähler und Hasselhorn (2008) anhand einer Überprüfung der Messinvarianz über zwei Gruppen nachwiesen.

Zur Effektivität des Redintegrationprozesses bei Kindern mit Lernbehinderung oder Intelligenzminderungen liegen keine Studien vor. Daher werden hier nur einige Studien angeführt, die andere Subgruppen untersuchen oder auf ähnliche Konzepte (aber eben nicht explizit Redintegration) zurückgreifen: Bei Schüler\*innen mit isolierten spezifischen Störungen im Lesen Schriftsprachbereich konnten Hasselhorn, Schuchardt und Mähler (2010) einen Trend für einen stärkeren Redintegrationprozess nachweisen, wohingegen Kinder mit isolierten Rechtschreib-Schwierigkeiten<sup>11</sup> eher einen stärkeren Rehearsalprozess zeigten. Grube et al. (2008) fanden bei unbeeinträchtigten Kindern im Alter von fünf und neun Jahren Unterschiede im phonologischen Ähnlichkeitseffekt und machten hierfür den Redintegrationprozess verantwortlich. Marton und Eichorn (2014) untersuchten die Interaktion von Arbeitsgedächtnis und Langzeitgedächtnis bei Kindern mit spezifischen Sprachentwicklungsstörungen (ohne dabei aber explizit

---

<sup>11</sup> Im veröffentlichten Beitrag (Bruns et al., 2019, S. 3) ist an der Stelle fälschlicherweise von „Dyskalkulie“ die Rede.

auf das Konzept von Redintegration zu verweisen). Schließlich zeigte Henry (2010), dass die Interaktion zwischen Langzeit- und Arbeitsgedächtnis bei Schüler\*innen mit leichten bis moderaten Intelligenzbeeinträchtigungen (IQ 39-70) im Rahmen der allgemeinen mentalen Entwicklung lag, dass also eine Entwicklungsverzögerung vorzuliegen scheint.

Die oben bereits zitierte Studie von Hasselhorn et al. (2010) soll an dieser Stelle noch etwas detaillierter beschrieben werden, da ihr Design dem von Studie 1 sehr ähnlich ist. Auf eine weitere relevante Studie von Mähler (2007) wird im Rahmen der Erläuterung des Gruppendesigns (in Kapitel 3.1.3) detaillierter eingegangen.

Hasselhorn et al. (2010) untersuchten insgesamt  $n = 68$  Kinder mit isolierten und kombinierten Lese- und Rechtschreibstörungen sowie eine nach Alter und IQ vergleichbare unbeeinträchtigte Kontrollgruppe ( $n = 30$ ). Als Testmaterial wurden Wortspannen mit ansteigender Sequenzlänge in insgesamt vier Bedingungen dargeboten: einsilbige echte Wörter, dreisilbige echte Wörter, einsilbige Kunstwörter und dreisilbige Kunstwörter. Neben signifikanten Haupteffekten für alle drei Faktoren zeigten sich zweifach-Interaktionen für Wortlänge  $\times$  Lexikalität, Gruppe  $\times$  Lexikalität und marginal für Gruppe  $\times$  Wortlänge ( $p < .08$ ), was auf eine insgesamt differenzielle Wirkung sowohl von Rehearsal- als auch Redintegrationprozess in den Gruppen schließen lässt. Die anschließende Posthoc-Analyse kontrastierte Kinder mit isolierter Lesestörung und mit isolierter Rechtschreibstörung, bei der Redintegration und Rehearsal sich differenziell in den Gruppen auswirkten.

Schließlich sei noch knapp auf Studien zur Effektivität von Trainingsmaßnahmen eingegangen. Für das Arbeitsgedächtnis kommen diese zu überwiegend ernüchternden Ergebnissen. Auch wenn einzelne Untersuchungen positive Effekte für Trainings berichten (Clerc & Courbois, 2017; Norris, Hall & Gathercole, 2019; Peng & Fuchs, 2017), ergeben Meta-Analysen ein insgesamt eher negatives Bild (Melby-Lervåg & Hulme, 2013; 2016; Sala & Gobet, 2017; 2020). Auch bei Drittklässlern mit und ohne LRS zeigten sich wenn überhaupt nur kurzfristige Effekte (Mähler, Jörns, Radtke & Schuchardt, 2015; Mähler, Jörns & Schuchardt, 2019). Speziell für Personen mit intellektuellen Beeinträchtigungen (damit am ehesten der Zielgruppe von Schüler\*innen mit dem Förderschwerpunkt Lernen äquivalent) trugen Danielsson, Zottarel, Palmqvist und Lanfranchi (2015) insgesamt zehn Studien aus den Jahren 2001-2014 zusammen. Es zeigte sich eine durchschnittliche kleine bis mittlere Effektstärke von Prä- zu Posttest von  $d = 0.42$ , bzw. um Effekte in den Kontrollgruppen bereinigte Effektstärke von  $d = .24$ . Damit scheinen Trainings zumindest für Kinder mit Lernbehinderung zwar niedrige, aber immerhin gewisse Erfolgsaussichten zu versprechen.

#### **2.1.4. Zusammenfassung Arbeitsgedächtnis**

In diesem Abschnitt wurden, orientiert am Modell von Baddeley (1986), das Arbeitsgedächtnis und dessen Subkomponente der phonologischen Schleife eingeführt, die als phonetischer Speicher und Rehearsalprozess konzipiert ist. In ihrer Funktion für die kurzfristige Aufrechterhaltung verbaler Information hat die phonologische Schleife gerade für das schulische Lernen eine hohe Bedeutung, der Zusammenhang zwischen einem funktionierenden Arbeitsgedächtnis und schulischen Leistungen wurde empirisch überzeugend belegt. Insofern ist auch davon auszugehen, dass Kinder mit Lernbehinderung und anderen gravierenden schulischen Problemen generell Defizite im Arbeitsgedächtnis aufweisen, wie in auch zahlreichen Studien gezeigt werden konnte. Dies gilt jedoch vorrangig für die Konzeption des Arbeitsgedächtnisses im ursprünglichen Sinne des Modells nach Baddeley (1986) bzw. Hasselhorn et al. (2000). Der Redintegrationprozess, der nicht Teil dieses Modells ist, wurde bislang nicht bei Kindern mit Lernbehinderung untersucht – dabei ist dieses Zusammenspiel von Arbeits- und Langzeitgedächtnis für das Lernen von zentraler Bedeutung. Zudem herrscht Uneinheitlichkeit in den bisherigen Befunden zu Arbeitsgedächtnis-Kapazität und Rehearsalprozess; zum einen darüber, ob alle Subkomponenten des Arbeitsgedächtnisses gleichermaßen beeinträchtigt sind, zum anderen über das Ausmaß der Beeinträchtigung (Entwicklungsverzögerung oder qualitative Abweichung) in phonetischem Speicher oder Rehearsalprozess.

Aus diesem Grund erscheint es sinnvoll, in Studie 1 neben der Gesamtkapazität der phonologischen Schleife (1. Teilfragestellung) die beiden Prozesse von Rehearsal (2. Teilfragestellung) und Redintegration (3. Teilfragestellung) gemeinsam in einer Stichprobe zu untersuchen, um eine Vergleichbarkeit zwischen den Prozessen herzustellen und Stichprobeneffekte zu vermeiden. Zudem wurde in den bisherigen Studien häufig ein Design zum Vergleich auf Gruppenebene eingesetzt, das weniger detaillierte Analysen ermöglicht als dessen Weiterentwicklung, der Developmental Trajectories, die in dieser Studie eingesetzt werden. Auf die Hintergründe zu diesen Designs wird in Kapitel 3 näher eingegangen.

### **2.2. Langzeitgedächtnis**

Nachdem im vorigen Unterkapitel die für die erste Studie dieser Arbeit relevanten Aspekte der phonologischen Schleife eingeführt wurden, wird im Folgenden das Langzeitgedächtnis betrachtet, das in dieser Arbeit die zweite relevante Komponente für den Wissenserwerb im Modell der guten Informationsverarbeitung darstellt und in der zweiten Studie untersucht wird. Dabei wird eine Eingrenzung auf das mentale Lexikon als Speicher für Wörter und Konzepte im semantischen Gedächtnis vorgenommen. Während im Arbeitsgedächtnis nur eine begrenzte

Menge an Informationen für einen kurzen Zeitraum bereitgehalten werden kann, ist das Langzeitgedächtnis ein langfristiger Informationsspeicher mit potentiell unbegrenzter Kapazität (Hasselhorn & Gold, 2013, S. 53).

Gemäß den etablierten Gedächtnistheorien kann das Langzeitgedächtnis grob in ein deklaratives und ein prozedurales Gedächtnis unterteilt werden (Anderson, 2013; Purves et al., 2013). Im deklarativen Gedächtnis werden bewusste Gedächtnisinhalte gespeichert, wohingegen im prozeduralen Gedächtnis eher unbewusste Formen wie Priming und Konditionierung (Purves et al., 2013), aber auch Fertigkeiten (*skills*; Renkl, 2015) zusammengefasst werden. Weiterhin wird häufig auch metakognitives Wissen als eigene Wissensart im Langzeitgedächtnis verstanden, das Wissen und Fertigkeiten zur Überwachung und Steuerung von kognitiven Strategien umfasst (Renkl, 2015).

Das bewusste deklarative Gedächtnis lässt sich wiederum unterteilen in episodisches und semantisches Gedächtnis (Duff, Covington, Hilverman & Cohen, 2020; Tulving, 1972): Im episodischen Gedächtnis werden einzelne Erlebnisse und Erfahrungen einer Person aus autobiographischer Perspektive abgebildet. Im semantischen Gedächtnis wird dagegen Wissen über allgemeine Fakten und sprachliche Konzepte repräsentiert (Büttner, 2003). Dazu zählt insbesondere das mentale Lexikon als Speicher für den Wortschatz einer Person (Dilkina, McClelland & Plaut, 2010).

Nach Tulving (1972, S. 386) stellt sich das semantische Gedächtnis als großes mentales Lexikon („mental thesaurus“) dar, in dem alles Wissen, das zur Verarbeitung von Sprache notwendig ist, gespeichert wird. Somit kann das mentale Lexikon als Speicher von sprachlichen Begriffen und den dazugehörigen Konzepten verstanden werden (Glück, 2011). Häufig wird daher das mentale Lexikon auch mit dem semantischen Gedächtnis gleichgesetzt (Bjorklund, 1987; Kenett, Beaty, Silvia, Anaki & Faust, 2016). Daher wird für diese Arbeit der Wortschatz-Abruf zur Operationalisierung für die empirische Erfassung des Langzeitgedächtnisses in Studie 2 herangezogen. Durch die Fokussierung auf die Prozess-Stufe des Abrufes bleiben die vorgelagerten Stufen von Einspeicherung und Konsolidierung (Purves et al., 2013) jedoch unberücksichtigt.

Nicht nur die reine Größe (Quantität) des mentalen Lexikons wird für gute Erinnerungsleistungen verantwortlich gemacht (Simonsmeier, Flaig, Deiglmayr, Schalk & Schneider, 2018), sondern insbesondere für den Informationsabruf auch dessen Struktur und Organisation (Qualität), um Informationen im semantischen Gedächtnis bzw. dem mentalen Lexikon schneller und effektiver wieder auffinden zu können (Vernooij, 2007). Hier hat sich bereits seit längerem die Vorstellung etabliert, dass Wissen vernetzt ist (Anderson, 2013; Büttner, 2003; Collins & Loftus, 1975); der Vernetzungsgrad des Wissens wirkt sich also als Qualitätsmerkmal auf erfolgreichen Wissenserwerb aus (Renkl, 2015). Gemäß dieser Vorstellung lässt sich konzeptuelles

Wissen als propositionales Netzwerk fassen, bei dem ein Wort oder Konzept einen Knoten darstellt, und dessen Bezüge zu anderen Konzepten die Verbindungen zwischen Knoten abbilden; z. B. wird in der Aussage „Ein Hund ist ein Tier“ das Konzept „Hund“ mit dem Oberbegriff „Tier“ durch die Relation „ist ein“ verknüpft. Informationsverarbeitung findet statt, indem sich die Aktivierung eines Knotens auf dessen Nachbarn ausbreitet (*spreading activation*, Collins & Loftus, 1975; Beckage & Colunga, 2016).

Bislang handelte es sich jedoch bei diesen Vorstellungen eher um plausible Vermutungen, wie Wissen als netzartige Struktur repräsentiert ist (Hartley & Barnden, 1997). Erst neuere netzwerktheoretische Ansätze (z. B. Baronchelli, Ferrer-i-Cancho, Pastor-Satorras, Chater & Christiansen, 2013; Bassett & Sporns, 2017) sind dazu übergegangen, in der funktionalen Vernetzung von Modulen und Arealen des Gehirns nach empirisch prüfbareren Erklärungen für Wissensrepräsentation und Lernen zu suchen (Bassett et al., 2011; Kraut et al., 2004). So scheint die Vernetzung und Organisation auf dieser Makro-Ebene (Kaiser, 2017) auf einem gemeinsamen Oszillieren der Neuronenaktivität in einem ähnlichen Rhythmus (Klimesch, 2012) zu basieren. Durch die gemeinsame Aktivität entsteht ein Zusammenspiel von Hirnarealen über verschiedene Modalitäten hinweg, welches wiederum die Repräsentation von einzelnen Erfahrungen zu abstrakteren Konzepten ermöglicht („distributed, modality-independent cortical representation“, Handjaras et al., 2016, S. 232).

Bassett und Mattar (2017) sehen Lernen allgemein als eine Formung, Entfernung oder Veränderung von neuronalen Verbindungen an, die den behavioral beobachtbaren Assoziationen (z. B. zwischen Wörtern und Bedeutungen) zugrunde liegen. Lernen lässt sich in diesem Rahmen als graduelle Veränderung der Vernetzungsdichte oder -qualität verstehen, oder dass sich die Zuordnung eines Knotens zu einem Modul-Cluster ändert (*reconfiguration*). Diese Flexibilität scheint ein allgemeines Charakteristikum des Gehirn-Netzwerkes zu sein, und es lassen sich Zusammenhänge zwischen der Flexibilität der modularen Netzwerkstruktur und der Lernrate nachweisen (Bassett & Mattar, 2017; Bassett et al., 2011).

Neben der Netzwerkanalyse von neurophysiologischen Daten können auch semantische Netzwerke aus assoziativen Antworten auf eine große Anzahl von Cue-Wörtern (z. B. Kenett, Gold & Faust, 2016) erzeugt und mithilfe der Netzwerk- und Graphentheorie analysiert werden; hier stehen die Knoten für einzelne Wörter, deren Verbindungen auf assoziativer oder syntaktischer Nähe beruhen (Borge-Holthoefner & Arenas, 2010; de Deyne, Verheyen & Storms, 2016). Beiden Ansätzen ist gemeinsam, dass sie auf sehr erhebungintensiven Methoden beruhen und damit über den Rahmen der vorliegenden Arbeit hinausgehen würden.

Gleichwohl unterstreichen die neueren Ergebnisse, dass die Netzwerk-Annahmen zu Wissensrepräsentation und Lernen nicht nur theoretisch plausibel sind, sondern auch als funktionale und dynamische Netze modelliert werden können und einer empirischen Überprüfung standhalten. Für die vorliegende Arbeit wird aus diesen Ergebnissen vor allem abgeleitet, dass die (Netzwerk-)Struktur und Organisation für die Wissensrepräsentation relevant ist. Die Struktur kann sich zum einen im Verhältnis der Elemente des mentalen Lexikons zueinander ausdrücken, und zum anderen in deren Verhältnis zu äußeren Variablen, d. h. zu Personenmerkmalen oder Eigenschaften der Wörter, die z. B. Einfluss auf den Erwerb (Hills, Maouene, Maouene, Sheya & Smith, 2009) oder deren korrekte Benennung haben (Winters & Brzoska, 1975).

Für die zweite Studie dieser Arbeit wird aus dem semantischen Langzeitgedächtnis das mentale Lexikon fokussiert, da dessen Struktur für die begriffliche Speicherung von Wissen wichtig erscheint (Büttner, 2003). Der Abruf aus dem mentalen Lexikon wird hier anhand des expressiven Wortschatzes in einem Wortschatztest (Wortschatz- und Wortfindungstest für 6- bis 10-Jährige, WWT 6-10, Glück, 2011) operationalisiert. Da die Forschungslage nicht ausreicht, um die in Kapitel 1.3 angeführten Vermutungen zur Rolle des Langzeitgedächtnisses bei Kindern mit Lernbehinderung empirisch zu stützen (wie im folgenden Unterkapitel 2.2.1 aufgezeigt wird), werden in der zweiten Studie zwei Ansätze für Strukturmerkmale des mentalen Lexikons betrachtet und in den darauf folgenden Abschnitten theoretisch begründet: (1) die Binnenstruktur in Abschnitt 2.2.2, bezogen auf die Verhältnisse der Elemente des mentalen Lexikons zueinander; (2) die Außenstruktur in Abschnitt 2.2.3, bezogen auf das Verhältnis der Einträge des mentalen Lexikons zu äußeren personen- und itembezogenen Variablen.

### **2.2.1. Forschungsstand zum Langzeitgedächtnis bei Lernbehinderung**

Auch wenn häufig davon ausgegangen wird, dass Kinder mit Lernbehinderung von einer „reduzierten Wissensbasis ausgehen“ (Lauth, Brunstein & Grünke, 2014, S. 21), liegen zum semantischen Gedächtnis und insbesondere zu dessen Struktur keine aktuellen Studien vor. Zwar existieren einige ältere Studien zum Langzeitgedächtnis, sie sind aber aus zwei Gründen für die vorliegende Fragestellung nicht sehr ergiebig. Zum einen wird häufig das episodische Gedächtnis untersucht (z. B. Ceci, 1984), indem Listen vorgelegt werden, die nach einem Intervall von einigen Minuten reproduziert werden sollen. Zum anderen basieren diese Studien zum Teil auf veralteten Annahmen, wie derjenigen, dass Zahl und Stärke der Verknüpfungen von Wissens-elementen mit dem Alter stetig ansteigt (Hasselhorn, 2005). Dies steht im Widerspruch zur *interactive specialization*-Hypothese (Johnson, 2011), nach der sich (neuronale) Verknüpfungen bei Kindern mit zunehmendem Alter von einem eher chaotischen, breiten Vernetzungsstadium zu einer spezifischen, geordneten Struktur entwickeln, wobei durchaus Verknüpfungen entfernt werden.

Daher werden an dieser Stelle lediglich zwei ausgewählte Studien vorgestellt, die sich mit der Auswirkung von semantischen Kategorien auf das episodische Erinnern bei Jugendlichen mit intellektuellen Beeinträchtigungen befassen sowie Hinweise auf den positiven Einfluss von Vorwissen unabhängig von den kognitiven Fähigkeiten liefern. In den anschließenden verbleibenden Unterkapiteln wird aufgrund der oben attestierten Forschungslücke daher primär die Plausibilität der Herangehensweise und der theoretischen Annahmen hergeleitet, welche der Untersuchung der Struktur des mentalen Lexikons bei Kindern mit Lernbehinderung in Studie 2 zugrunde liegen.

Mit der Verfügbarkeit und dem Zugriff auf semantische Kategorien befasste sich eine Studie von Glidden und Mar (1978). Die Stichprobe bestand aus 40 Kindern mit intellektuellen Beeinträchtigungen (*mental retardation*, durchschnittlicher IQ = 60) im Alter von durchschnittlich 15 Jahren und einer gleichgroßen Kontrollgruppe, die nach chronologischem Alter vergleichbar war; eine nach mentalem Alter vergleichbare Kontrollgruppe lag nicht vor. Die Probanden sollten in einer Abrufaufgabe aus zwei Kategorien (Tiere und Sportarten) in vier Minuten möglichst viele Elemente wiedergeben, worin Kinder mit intellektueller Beeinträchtigung insgesamt eine geringere Leistung zeigten, da sie weniger Wörter wiedergeben konnten. Anschließend wurden ihnen 24 Stimuli dieser Kategorien zur semantischen Entscheidung vorgelegt, in welche Kategorie sie gehörten. Bei Items mit mittlerer und niedriger Auftretens-Häufigkeit schnitten Kinder mit intellektueller Beeinträchtigung signifikant schlechter ab. Im zweiten Experiment (n = 63 Kinder mit intellektueller Beeinträchtigung, keine Kontrollgruppe) wurden einer Gruppe beim Abruf Subkategorien (z. B. bei Tieren: Farm, Zoo, Insekten...) als zusätzliche Hinweisreize vorgegeben. Gegenüber einer einfachen Kontrollbedingung ohne Hinweisreiz wurden signifikant mehr Wörter reproduziert, jedoch transferierte dieser Effekt nicht auf eine weitere Abruf-Phase, bei der die Hinweise nicht mehr gegeben wurden. Neben dem Alter der Studie und dem Fehlen einer nach mentalem Alter vergleichbaren Kontrollgruppe (siehe Kapitel 3.1) bzw. einer Kontrollgruppe in Experiment 2 ist zu bemängeln, dass die genaue theoretische Begründung und Vorgehensweise der Messung („retrieval“ und „retrieval phases“) zumindest unklar bleibt und die Schlussfolgerungen nicht unmittelbar aus den berichteten Daten nachvollziehbar sind. Die Ergebnisse zeigen lediglich, dass Kinder mit intellektuellen Beeinträchtigungen insgesamt weniger Wörter im episodischen Gedächtnis behalten konnten, und dass ihnen niedrig-frequente Wörter nicht bekannt waren, sodass sie diese nicht korrekt den Kategorien zuordnen konnten.

Die Rolle des bereichsspezifischen Vorwissens für Textverständnis untersuchten Schneider, Körkel und Weinert (1989) in einem Experten-Novizen Vergleich. Dazu legten sie Kindern aus der dritten, fünften und siebten Jahrgangsstufe mit unterschiedlichem Vorwissen und unter-



schiedlicher kognitiver Leistungsfähigkeit eine Geschichte zum Thema Fußball vor. Die Leistungen der Kinder, sich an Details aus der Geschichte zu erinnern (sowohl in der Beantwortung von geschlossenen Fragen als auch bei freier und unterstützter Wiedergabe), sowie Widersprüche zu erkennen und Inferenzen zu ziehen, hingen vor allem vom Alter und dem Vorwissen ab, nicht jedoch von der kognitiven Leistungsfähigkeit. Dabei unterschieden sich die beiden Expertengruppen (diejenige mit hohen und die mit niedrigen kognitiven Fähigkeiten) in keiner der abhängigen Variablen. Das Vorwissen scheint also Erinnerungsleistungen – wiederum Aufgaben aus dem Bereich des episodischen Gedächtnisses, für die auf Wissen im semantischen Gedächtnis zurückgegriffen wurde – positiv zu beeinflussen, und dies unabhängig von den kognitiven Fähigkeiten.

### 2.2.2. Binnenstruktur des Langzeitgedächtnisses (1. Ansatz)

Im ersten Ansatz wird die Organisation des Langzeitgedächtnisses anhand der Binnenstruktur des WWT 6-10 betrachtet. Die Binnenstruktur zwischen den manifesten (d. h. empirisch messbaren) Variablen äußert sich in Korrelationsmustern, die sich als konfirmatorische Faktorenanalyse (CFA) formalisieren lassen (Backhaus, Erichson & Weiber, 2013). Üblicherweise werden einzelne Items als manifeste Variablen betrachtet, in dieser Arbeit werden jedoch aufgrund der Stichprobengröße mehrere Items zu sog. Item-Parcels zusammengefasst (Little, Rhemtulla & Schoemann, 2013; eine ausführliche Begründung und Erläuterung der Parceling Strategie wird in Kapitel 6.2.3.3 gegeben).

Manifeste Variablen, die etwas ähnliches messen, weisen höhere Korrelationen untereinander auf als solche, die unterschiedliche Dinge messen. Mit der Untersuchung der Korrelationsmuster wird die Frage beantwortet, wie sich die Item-Parcels im WWT 6-10 zueinander verhalten, wie also die Binnenstruktur des mentalen Lexikons bezogen auf das Strukturmerkmal *Wortarten* aussieht. Da eine Aussage über den Vergleich der Gruppen getroffen werden soll, muss geprüft werden, ob dieses Verhältnis bei Kindern mit und ohne Lernbehinderung gleich ausgeprägt ist. Das für diesen Vergleich gewählte Vorgehen einer Multigroup-CFA wird im Folgenden dargelegt.

Allgemein gesprochen lassen sich Korrelationsmuster zwischen manifesten Variablen in einem Messmodell formalisiert ausdrücken, in dem die manifesten Variablen anhand theoretischer Vorannahmen übergeordneten Fähigkeitskonstrukten, sog. latenten Variablen, zugeordnet werden. Von Items bzw. Parcels, die einer latenten Variable zugeordnet werden, wird angenommen, dass sie ein ähnliches Fähigkeitskonstrukt messen. Die Parcels eines Konstruktes müssten also, wenn das angenommene Modell zutrifft, höhere Zusammenhänge untereinander aufweisen, und geringere Zusammenhänge zu Parcels aus anderen Konstrukten. In dieser Arbeit

werden die Wortarten als latente Variablen angenommen, denen die einzelnen Parcels der WWT-Items entsprechend zugeordnet werden.

Für die Fragestellung der Studie 2 ist der Gruppenvergleich von zentraler Bedeutung: Gilt das Korrelationsmuster (d. h. das angenommene Messmodell) in beiden Gruppen gleichermaßen, oder gibt es zwischen den Gruppen systematische Unterschiede in den Messmodellen? Die Frage, ob diese Struktur zwischen den Gruppen äquivalent ist (siehe Schuchardt, Roick et al., 2008), wird mithilfe einer konfirmatorischen Faktorenanalyse im Multigruppenvergleich (MG-CFA, z. B. Putnick & Bornstein, 2016; van de Schoot, Schmidt, de Beuckelaer, Zondervan-Zwijenburg, 2015) untersucht. Diese prüft, ob das angenommene Messmodell in beiden Gruppen gleichermaßen gut zu den Daten passt, was auch als Messinvarianz bezeichnet wird. Das methodische Vorgehen einer MG-CFA zur Überprüfung der Messinvarianz wird im Methodenteil von Studie 2 unter Kapitel 6.2.3 genauer erläutert.

Der hier eingesetzte expressive Subtest des WWT 6-10 (Glück, 2011) besteht aus 95 Items, bei denen ein Bild präsentiert wird und vom Kind benannt, also aus dem Langzeitgedächtnis abgerufen, werden soll. Die Items gehören zu vier verschiedenen Wortarten: Nomen, Verben, Adjektive und Kategorien<sup>12</sup>. Zwar ist das Vorgehen selbst (Präsentation eines visuellen Cues und ein auditiver Frage-Hinweis) bei allen Items gleich, jedoch unterscheiden sich die Wortarten hinsichtlich der Formulierung der gestellten Hinweis-Frage: 1) Nomen: Es werden Gegenstände oder Teile von Gegenständen gezeigt, die Frage lautet „Was ist das?“; 2) Verben: Es werden Tätigkeiten gezeigt, die Frage lautet „Was tut er/sie?“ bzw. „Was ist hier passiert?“; 3) Kategorien: Es werden verschiedene Gegenstände gezeigt, die zu einer Kategorie gehören, die Frage lautet „Was sind das alles?“ bzw. „Wie nennt man das alles zusammen?“; 4) Adjektive: Es wird ein Symbolbild gezeigt (schwarz-weißes Quadrat), die Frage lautet „Was ist das Gegenteil von ...?“<sup>13</sup>. Damit liegt insbesondere bei den Adjektiven eine wesentliche Abweichung der Stimuluspräsentation im Vergleich zu den anderen Wortarten vor, auf die weiter unten (Kapitel 2.2.2.2) näher eingegangen wird.

In dieser Studie werden die genannten vier Wortarten des WWT 6-10 als latente Variablen für die MG-CFA angenommen. Für diese Annahme gibt es mehrere Gründe. Der erste Grund liegt bereits in der Konstruktion des Instrumentes, da im Testmanual die Wortarten als sog. „Item-Subsets“ angeführt werden (Glück, 2011). Neben dem offensichtlichen oberflächlichen Unterschied in den Instruktionsbedingungen zwischen den Wortarten sprechen darüber hinaus

---

<sup>12</sup>Im linguistischen Sinne ist „Kategorie“ selbstverständlich keine Wortart; der Begriff „Wortart“ wird hier zur besseren Verständlichkeit als Oberbegriff für die latenten Variablen der CFA verwendet

<sup>13</sup> Beispiel-Items sind in Tabelle 4 (Seite 118) und Abbildung 5 (Seite 119) angeführt

zwei theoretisch fundierte Gründe für die Annahme: (1) Reihenfolge der Wortarten im Wortschatzerwerb und (2) unterschiedliche Aufgabenanforderungen bei der Bildbenennung. Zunächst wird knapp auf den Aspekt der Itemauswahl bei der Testkonstruktion (Glück, 2011) eingegangen, im Anschluss folgen ausführlicher die beiden theoretischen Gründe.

In der Normierungsstichprobe (Glück, 2011) zeigten sich unterschiedliche Entwicklungsverläufe je nach Wortart, insbesondere ein steilerer Anstieg für die Adjektive und Kategorien (ebd., S. 35), sodass auch pro Wortart-Itemsubset eigene Normwerte angegeben werden (ebd., S. 133ff.). Eine solche zusätzliche differenzierte Auswertung wird zwar als Tabelle im Manual angeboten, jedoch erfolgt die PC-gestützte Auswertung lediglich auf dem Niveau des Gesamtscores. Auch werden im Manual keine Ergebnisse zu einer Faktorenanalyse berichtet, die empirisch begründen würden, weshalb nur ein Faktor als Summenscore angenommen werden darf. Dennoch erscheint es auch angesichts der im Folgenden anzuführenden beiden theoretischen Gründe plausibel, die Wortarten als latente Variablen in der MG-CFA zu betrachten.

### 2.2.2.1. Erstes Argument: Wortschatzerwerb

Das erste theoretische Argument dafür, die Wortarten differenziert zu betrachten, bezieht sich darauf, dass verschiedene Wortarten zu verschiedenen Phasen im Verlauf des Spracherwerbs erlernt werden (Glück, 2011; Kauschke, 2003). Erkenntnisse zum kindlichen Wortschatzerwerb zeigen über verschiedene Sprachen ein Muster bei der Abfolge, welche Wörter zuerst erlernt werden. Bates et al. (1994; siehe auch David & Wei, 2005) fassen die Abfolge in drei Wellen zusammen:

1. In der ersten Welle der ersten 100 Wörter werden vorrangig (konkrete) Nomen erlernt.
2. In der zweiten Welle von 100 bis 400 Wörtern wird ein stärkerer Zuwachs an Verben verzeichnet.
3. In der dritten Welle folgen grammatische Funktionswörter (z. B. Präpositionen, Konjunktionen, Pronomen).

Diese drei Stufen finden sich inhaltlich auch in den Stufen 2-4 von Caselli, Casadio und Bates (1999) wieder; sie schalten lediglich als erste Stufe die ‚Sprachspiele‘ vor, die sich keiner Wortart zuordnen lassen. In diesem vierstufigen Modell nehmen insbesondere konkrete handlungsbezogene Nomen und Verben eine prominente Position ein: Nach den Sprachspielen stehen als zweite Stufe das Referenzieren und als dritte Stufe die ‚Prädikation‘ (ebd., S. 71). Referenzierung geschieht vor allem durch den Gebrauch von Nomen (im weitesten Sinne), die auf Objekte Bezug nehmen (Rothweiler & Kauschke, 2007). Die Fähigkeit, relationale Bedeutungen zu übermitteln, hängt stark von der Verfügbarkeit von Verben und Adjektiven ab. Diese werden in der darauffolgenden Stufe der Prädikation verstärkt erworben und ermöglichen die ersten Kombinationen von Wörtern in einem ‚Satz‘ (Caselli et al., S. 71f.). Als vierte Stufe folgt der Erwerb

von grammatischen Funktionswörtern, analog zur Stufe 3 in Bates et al. (1994). Die Bedeutung abstrakter Nomen sowie die Einordnung in ein Begriffssystem von Ober- und Unterbegriffen werden im Verlauf der Sprachentwicklung erst deutlich später als der Erwerb konkreter Nomen und Verben (Rothweiler & Kauschke, 2007) erlernt.

Angesichts der unterschiedlichen Stufen im Worterwerb (Kauschke, 2003) erscheint es sinnvoll, eine Unterteilung des WWT 6-10 nach den vier Wortarten in einzelne Subskalen anstelle eines globalen Summenscores über alle 95 Items anzunehmen. Im Vergleich zu den Wortarten Kategorien und Adjektive spiegeln die Wortarten Nomen und Verben im WWT 6-10 eine frühe Phase des Wortschatzerwerbs wider, wobei Nomen noch vor Verben erworben werden (Spanoudis & Natsopoulos, 2011). Daher ist es plausibel, eine größere Bekanntheit der Stimuli bei Nomen und Verben zu vermuten, sodass ein direkter Abruf des Begriffs aus dem mentalen Lexikon besonders gut möglich ist.

Dagegen dürfte die Benennung von Kategorien voraussetzungsreicher sein, da zusätzlich die kognitive Fähigkeit der Abstraktion und die Kenntnis von Taxonomien nötig sind (Glück, 2011). Der Punkt, dass zusätzliche Fähigkeiten zum Abruf benötigt werden, leitet zum zweiten Argument über: das Benennen von Kategorien und Adjektiven stellt andere aufgabenspezifische Anforderungen über den direkten Abruf hinaus. Zuerst wird auf die spezifischen Anforderungen bei der Benennung der Kategorien und im Anschluss auf die Adjektive eingegangen.

### 2.2.2.2. Zweites Argument: Aufgabenspezifische Anforderungen

Für die Wortarten Kategorien und der Adjektive wird vermutet, dass die aufgabenspezifischen Anforderungen von denjenigen der Benennung von Nomen und Verben abweichen und sie somit in der MG-CFA einen eigenen Faktor darstellen. Bei den Kategorien liegt dies vor allem daran, dass sie die Kenntnis bzw. sichere Anwendung taxonomischer Relationen voraussetzen, bei den Adjektiven liegt dies vor allem an der andersartigen Präsentation der Stimuli.

Im Falle der Wortart ‚Kategorien‘ im WWT 6-10 muss die taxonomische Beziehung der Hyperonomie, d. h. „die semantische Relation der Über- und Unterordnung, die lexikalische Ausdrücke hierarchisch ordnet“ (Meibauer et al., 2015, S. 353) erkannt und benannt werden. Hier basiert die Verbindung auf der Ähnlichkeit bzw. der Zugehörigkeit zu einer Kategorie. Eine andere Art von Beziehung zwischen Wörtern ist in der thematischen Relation zu sehen. Eine thematische Verbindung entsteht assoziativ, also nur aufgrund des häufigen gemeinsamen Auftretens (Mirman, Landrigan & Britt, 2017), ohne dass zwei Wörter zur gleichen Kategorie gehören müssen. Ein Beispiel für eine Hyperonymie als taxonomische Relation ist „Tier“ als Oberbegriff für „Hase“. Ein thematisch relatives Beispiel für den Begriff „Hase“ wäre „Osterei“, da hier die inhaltliche, assoziative Nähe im Vordergrund steht.

Die Benennung taxonomischer Relationen ist voraussetzungsreicher und stellt höhere Anforderungen an die Informationsverarbeitung als die direkte Benennung eines abgebildeten Gegenstandes. So zeigten frühe Studien, dass jüngere Kinder eine Präferenz für thematische Relationen zu haben schienen, und Kinder erst im Schulalter taxonomisch relationale Vergleiche vorzuziehen schienen (Smiley & Brown, 1979). In einer Studie von Fenson, Vella und Kennedy (1989) wurde gezeigt, dass zweijährige Kinder thematische Relationen zuverlässig erkennen konnten, in der taxonomischen Relation allerdings die jüngeren Kinder (im Alter von 26 Monaten) diese bei wenig offensichtlich geteilten Merkmalen schlechter identifizierten als ältere Kinder (im Alter von 34 Monaten); z. B. dass „Schuh“ und „Hemd“ beide zur Kategorie „Kleidung“ gehören, obwohl sie sich wenig ähnlich sehen.

Des Weiteren wurde in einem Priming-Experiment von Sachs et al. (2011) die Auswirkung der Relation des Primes zum Zielwort anhand der Reaktionszeit in einer lexikalischen Entscheidungsaufgabe bei Erwachsenen wie folgt untersucht. Zunächst wurde ein Prime gezeigt, der entweder eine taxonomische, eine assoziative oder in der Kontrollbedingung keine Verbindung zum Zielwort hatte. Anschließend wurde entweder ein echtes Wort oder ein Pseudowort präsentiert, und die Testperson sollte möglichst schnell entscheiden, ob es sich um ein echtes Wort handelte oder nicht. Eine kürzere Reaktionszeit bei einer existierenden Verbindung zwischen Prime und Zielwort wurde damit auf die Aktivierung durch den Prime zurückgeführt. Die Autoren konnten zwar keine Unterschiede in der Reaktionszeit zwischen taxonomisch und assoziativ relationierten Primes finden, jedoch ergaben sich auf neuronaler Ebene im fMRI-Bild Unterschiede in der Aktivierung von Hirnarealen zwischen den Prime-Bedingungen. Weiterhin zeigt ein Experiment von Hashimoto, McGregor und Graham (2007), dass Sechsjährige im Vergleich zu Achtjährigen zwar durch taxonomisch und thematisch relationierte Primes die gleichen Steigerungen der Reaktionszeit aufwiesen, sie jedoch größere Schwierigkeiten bei der konkreten Beschreibung taxonomischer Relationen hatten.

Es kann also davon ausgegangen werden, dass die Benennung von Kategorien voraussetzungsreicher als die Benennung eines direkt abgebildeten Gegenstandes ist und andere, möglicherweise kognitiv anspruchsvollere und weniger automatisierte Verarbeitungsprozesse verlangt (Glück, 2011). Da erstens konkrete Inhalte leichter gelernt werden können als abstrakte (Büttner & Hasselhorn, 2007), und zweitens Kinder mit Lernbehinderung besonders häufig Probleme bei abstrakten Aufgabenstellungen haben (Eser, 2006), ist zu vermuten, dass die Bearbeitung der WWT-Aufgabenform der Kategorien bei Kindern mit Lernbehinderung andersartig abläuft als bei unbeeinträchtigten Kindern, was sich auf die Binnenstruktur der mentalen Repräsentation auswirken könnte.

Für die Wortart Adjektive wird angesichts einer abweichenden Präsentation der Stimuli vermutet, dass die Aufgabenanforderungen anders sind als bei den übrigen Wortarten, sodass sie in der MG-CFA einen eigenen Faktor darstellen. Unterschiede in der Präsentation bestehen zwischen den Adjektiven und den anderen Wortarten hinsichtlich dreier Gesichtspunkte: (1) soll das *Gegenteil* eines vorgegebenen Wortes benannt werden, (2) gibt es *keine visuelle* Unterstützung und (3) erfolgt die Präsentation des kritischen Hinweises *nur einmal*. Daher wird vermutet, dass die andersartige Präsentation auch eine im Vergleich zu den anderen drei Wortarten nach Art und Schwierigkeit abweichende Aufgabenanforderung mit sich bringt.

Aus Sicht der Testkonstruktion ist die Erfassung des expressiven Wortschatzes bei Adjektiven zugegebenermaßen eine Herausforderung. Begründet ist die abweichende Präsentationsform darin, dass „diese Begriffe [Adjektive und Adverbien] in möglichen Abbildungen keine eindeutigen Antworten evozieren.“ (Glück, 2011, S. 22). Dennoch sollte berücksichtigt werden, dass die Anforderungen an die Probanden hier im Vergleich zu den Nomen und Verben abweicht. Zur Erläuterung der Unterschiede der Präsentationsform werden im nächsten Absatz die drei Gesichtspunkte für die anderen Wortarten (Nomen, Verben, Kategorien) dargestellt, und im Anschluss daran, inwiefern die Präsentation der Adjektive davon abweicht.

Bei den anderen drei Wortarten (Nomen, Verben, Kategorien) wird erstens dem Kind ein Bild präsentiert, welches es direkt benennen soll. Der Abruf des Zielwortes kann (mit Ausnahme der Kategorien) unmittelbar erfolgen, da keine weiteren mentalen Schritte notwendig sind. Zweitens liegt die wesentliche Information zum Zielwort im gezeigten Bild, es gibt also eine visuelle Unterstützung für den Abruf. Drittens wird das Bild solange gezeigt, bis das Kind eine Antwort gibt (bei maximal 15 Sekunden Wartezeit; wenn bis dahin keine Antwort erfolgt ist, wird das Item im Test als falsch gewertet). Diese drei Aspekte sind bei den Adjektiven anders.

Erstens wird bei den Adjektiven das Zielwort nur indirekt über sein Gegenteil präsentiert; die Instruktion lautet, dass nicht der Stimulus selbst, sondern dessen Gegenteil benannt werden soll („Was ist das Gegenteil von ‚interessant?‘“). Zur Beantwortung ist also ein zusätzlicher gedanklicher Schritt notwendig, denn zunächst muss das auditiv vorgegebene Hinweiswort vom Kind verstanden werden und ihm bekannt sein, anschließend muss das Gegenteil gefunden und benannt werden. Das setzt voraus, dass dem Kind erstens das Konzept von ‚Gegenteil‘ allgemein und zweitens das jeweilige dargebotene Hinweiswort bekannt sein müssen. Wenn bereits das Hinweiswort dem Kind nicht bekannt ist, kann das Zielwort nur schwerlich produziert werden – so liegt z. B. anekdotische Evidenz der Versuchsleitungen vor, dass das Wort „rau“ vielen Kindern nicht bekannt war, insofern konnten sie auch „glatt“ als Zielwort nicht benennen. Dass aber diese Kinder das Wort „glatt“ nicht produzieren können, lässt sich daraus nicht zweifelsfrei schließen.

Zweitens ist die Modalität des Hinweises auf das Zielwort bei den Adjektiven eine andere als bei den übrigen Wortarten: bei den Adjektiven wird im visuellen Stimulus keine Information zum Zielwort übermittelt, sondern es wird nur eine schematische Abbildung (ein entlang der Diagonalen geteiltes schwarz-weißes Quadrat, siehe Abbildung 5, S. 114) bei allen Adjektiven gleichermaßen präsentiert. Dem gezeigten Bild kann also keine Information über das Zielwort entnommen werden. Vielmehr liegt die Information, auf die es für eine korrekte Benennung des Adjektivs wesentlich ankommt, im auditiven Hinweis. Daher muss bei den Adjektiven zuerst eine konkrete Vorstellung des Zielwortes als mentale Repräsentation (Schnotz & Bannert, 1999) z. B. des Konzept-Paares („rau“ – „glatt“) aufgebaut werden, ohne dass dies durch bildliche Information unterstützt würde.

Drittens und schließlich wird bei den Adjektiven der auditive Hinweis auf das Zielwort nur einmal präsentiert, die wesentliche Information zum Zielwort steht also nicht für längere Zeit zur Verfügung. Das Kind muss somit im richtigen Moment genügend aufmerksam sein, um den Hinweis bei der einmaligen auditiven Präsentation auf Anhieb verstehen zu können. Daraufhin muss es ihn im Arbeitsgedächtnis bereithalten, um die oben genannte Suche nach dem Gegenteil im mentalen Lexikon zu vollziehen. Eine Wiederholung des Hinweises, z. B. wenn er akustisch nicht verstanden wurde, oder wenn das Kind eine längere Zeit zum Überlegen benötigt, ist laut Manual nicht vorgesehen.

Daraus schlussfolgernd lässt sich zunächst vermuten, dass Adjektive allgemein schwieriger zu lösen sind, da sie aufgrund der dargestellten andersartigen Präsentationsform komplexere Anforderungen stellen als die übrigen Wortarten und damit eine tiefere Verarbeitung voraussetzen. Da die Anforderungen über die reine Benennung hinausgehen, ist es plausibel, für die Wortart der Adjektive einen eigenen Faktor in der MG-CFA zu vermuten. Bezogen auf Kinder mit Lernbehinderung könnten sich die Verarbeitungsanforderungen differenziell auswirken, da sie die komplexeren Aufgaben möglicherweise andersartig bearbeiten und damit eine abweichende Struktur im Langzeitgedächtnis aufweisen.

### 2.2.2.3. Zusammenfassung und Fazit zur Binnenstruktur

Im vorangegangenen Abschnitt wurde gezeigt, dass die Annahme der vier Wortarten (Nomen, Verben, Kategorien, Adjektive) als latente Variablen, neben dem oberflächlichen Grund der Aufteilung im Testmanual, aus zwei Gründen plausibel ist. Erstens zeigen Forschungsergebnisse zum Wortschatzerwerb, dass die Wortarten zu unterschiedlichen Zeitpunkten des Lebensalters erworben werden: Nomen werden noch vor Verben sehr früh erworben, Adjektive und Oberbegriffe („Kategorien“) dagegen erst später. Es ist daher mit einer unterschiedlichen Vertrautheit und Automatisierung des Abrufes zu rechnen. Zweitens stellen die Items der Ka-

tegorien und Adjektive höhere aufgabenspezifische Anforderungen bei der Benennung. Bei Kategorien werden Abstraktionsvermögen und eine Vertrautheit mit der Verwendung der taxonomischen Relation von Oberbegriffen vorausgesetzt; bei Adjektiven weicht die Präsentationsform in drei Aspekten von derjenigen der übrigen Wortarten ab.

Damit kann also die Aufteilung der Item-Parcels nach Wortarten für die Fragestellung zur Binnenstruktur des Langzeitgedächtnisses theoretisch als sinnvoll erachtet werden. Die Zuordnung zu latenten Variablen kann im Folgeschritt als Messmodell mathematisch spezifiziert werden, anhand dessen die Korrelationsmuster der Item-Parcels untereinander abgebildet, und somit die Binnenstruktur des WWT operationalisiert werden kann.

Der Vergleich der Binnenstruktur zwischen den Gruppen erfolgt anhand des Verfahrens zur Feststellung der Messinvarianz: Wenn die Messmodelle zwischen den Gruppen messinvariant (d. h. gleich) sind, kann daraus geschlossen werden, dass die beiden Gruppen eine ähnliche Binnenstruktur in der Organisation des mentalen Lexikons aufweisen. Das Vorgehen bei einer MG-CFA als Verfahren zur Überprüfung der Messinvarianz wird im Methodenteil (Kapitel 6.2.3) genauer erläutert.

Im vorangegangenen Abschnitt wurde als ein Ansatz zur Untersuchung der Struktur des mentalen Lexikons der Vergleich der Binnenstruktur eingeführt. Darüber hinaus kann auch die Außenstruktur des mentalen Lexikons, d. h. das äußere Verhältnis der WWT-Items zu anderen Variablen, betrachtet werden. Daher werden auch derartige Variablen, die die Lösungswahrscheinlichkeit eines Items möglicherweise systematisch beeinflussen können, im Rahmen der zweiten Teilfragestellung untersucht und im folgenden Kapitel eingeführt.

### **2.2.3. Außenstruktur des Langzeitgedächtnisses (2. Ansatz)**

Im zweiten Ansatz von Studie 2 wird die Organisation des Langzeitgedächtnisses anhand der Außenstruktur der WWT-Items betrachtet, d. h., es wird untersucht, wie sich die Items zu anderen, äußeren Variablen verhalten: Wie gut lässt sich die Wahrscheinlichkeit, ein Item im WWT zu lösen, durch personen- und itembezogene Prädiktoren vorhersagen? Da eine Aussage über den Vergleich der Gruppen getroffen werden soll, muss geprüft werden, ob die äußeren Variablen die Lösungswahrscheinlichkeit bei Kindern mit und ohne Lernbehinderung gleichermaßen vorhersagen; anders als im ersten Ansatz ist die Bildung von Item-Parcels hier nicht notwendig. Die theoretischen Grundlagen für diesen Vergleich werden im Folgenden beschrieben.

Die Vorhersage der Lösungswahrscheinlichkeit der WWT-Items anhand äußerer Prädiktoren wird in einem verallgemeinerten gemischten linearen Modell (GLMM, z. B. Baayen, Davidson & Bates, 2008) formalisiert. Die abhängige Variable ist die Bewertung der Antwort des Kindes als korrekt oder nicht. Die hier untersuchten unabhängigen Variablen (Prädiktoren) lassen sich in personenbezogene und itembezogene Variablen unterteilen (Gilbert, Compton &



Kearns, 2011). Eine genauere Beschreibung der Prädiktoren erfolgt in den nächsten Abschnitten nach der Beschreibung des Gruppenvergleichs.

Der Gruppenvergleich ist auch für die zweite Teilfragestellung von zentraler Bedeutung: Gelten die Zusammenhänge zwischen der Lösungswahrscheinlichkeit und den äußeren Prädiktoren in beiden Gruppen gleichermaßen, oder gibt es zwischen den Gruppen systematische Unterschiede in der Vorhersagekraft der personen- und itembezogenen Variablen? Diese Frage kann anhand der Interaktionsmuster der Variablen mit dem Gruppenfaktor untersucht werden.

Damit stellt der hier gewählte Ansatz der GLMM im Prinzip eine Variante eines Developmental Trajectory (Thomas et al., 2009; siehe Kapitel 3.2) dar: Die Leistung in einer Zielaufgabe (hier die Wortbenennung) wird durch Variablen zur Schätzung des (mentalen) Alters vorhergesagt. Eine Interaktion mit dem Faktor Gruppe würde dann darauf hinweisen, dass sich die Vorhersage anhand des (mentalen) Alters bei Kindern mit Lernbehinderung von der der unbeeinträchtigten Kontrollgruppe unterscheidet. Dies wäre inhaltlich so zu interpretieren, dass Kinder mit Lernbehinderung ein abweichendes, differenzielles Entwicklungsmuster zeigen. Das methodische Vorgehen zur Überprüfung der Interaktionsmuster mithilfe eines GLMM wird im Methodenteil unter Kapitel 6.2.4 genauer erläutert, im Folgenden wird konzeptuell auf die Variablen eingegangen.

#### 2.2.3.1. Personen- und itembezogene Variablen

In den folgenden Abschnitten werden die für die zweite Teilfragestellung zum semantischen Gedächtnis relevanten äußeren Prädiktoren eingeführt, für die geprüft werden soll, ob sie einen (differenziellen) Einfluss auf die Lösungswahrscheinlichkeit bei der Wortbenennung im WWT haben. Die *personenbezogenen* Variablen umfassen das chronologische Alter, die kognitive Kapazität (Rohwert des CFT 1-R, Weiß & Osterland, 2012), sowie die Arbeitsgedächtnis-Kapazität (AG-Kapazität). Als *itembezogene* Variablen werden die Worthäufigkeit (Frequency) und die Ähnlichkeit mit anderen Wörtern (Neighborhood Size) eingeführt. Auf beide Variablengruppen wird in den folgenden Abschnitten näher eingegangen.

Die personenbezogenen Variablen können analog zu Entwicklungsindikatoren in Developmental Trajectories (siehe Kapitel 3.2 und Studie 1, Kapitel 5) interpretiert werden. Indem diese Variablen als Prädiktoren in das Modell aufgenommen werden, wird untersucht, ob das (mentale) Alter bei Kindern mit Lernbehinderung einen anderen Einfluss auf die Lösungswahrscheinlichkeit als bei unbeeinträchtigten jüngeren Kindern hat.

Die folgenden Annahmen begründen die Aufnahme von personenbezogenen Variablen in das Modell: Zunächst ist bei zunehmendem chronologischem Alter zu erwarten, dass die Wissensbasis wächst, sodass ältere Kinder mehr Wörter korrekt benennen können. Sodann gilt ähn-

liches für die Variable der kognitiven Kapazität (CFT-Rohwert), mit der das mentale Alter geschätzt wird (eine ausführliche Erläuterung des mentalen Altersvergleichs erfolgt in Kapitel 3). Schließlich sprechen für den Einbezug der AG-Kapazität zwei Gründe: Erstens spielt das Arbeitsgedächtnis auch für den Wortschatzerwerb eine wesentliche Rolle (Baddeley et al., 1998; Gathercole et al., 1997) und wird von Gilbert et al. (2011) beispielsweise als Prädiktor für die Dekodierfähigkeit beim Lesen angeführt. Zweitens erscheint es aufgrund der in Studie 1 gefundenen Interaktion zwischen Wortschatzumfang und Redintegrationprozess sinnvoll, die AG-Kapazität zu kontrollieren.

Die Lösungswahrscheinlichkeit eines Items hängt jedoch nicht nur von Eigenschaften der Testpersonen ab, sondern auch von (linguistischen) Faktoren, die die Itemschwierigkeit beeinflussen. Von diesen werden als itembezogene Variablen zwei relevante Variablen ausgewählt und in das Modell aufgenommen: Häufigkeit (Frequency) und Neighborhood-Size (NBS).

Die *Frequency* ist die Häufigkeit, mit der ein Wort im Sprachgebrauch Verwendung findet. Die Verwendungshäufigkeit lässt sich z. B. über Corpora ermitteln, die große Sammlungen von (häufig schriftsprachlicher) Sprachverwendung darstellen. Die Items des WWT 6-10 entstammen dem Korpus „Deutscher Wortschatz“ der Leipzig Corpora Collection (Goldhahn, Eckart & Quasthoff, 2012). Für die Ermittlung der Frequency wurde in dieser Studie das ChildLex Corpus (Schroeder, Würzner, Heister, Geyken & Kliegl, 2015) verwendet, da es im Gegensatz zur Leipzig Corpora Collection nicht auf Erwachsenen-Normen basiert, sondern altersangemessen ist (bei der Konstruktion des WWT 6-10 war es jedoch noch nicht verfügbar). Das ChildLex Corpus basiert auf ca. 500 Kinderbüchern und bildet damit die Wörter ab, mit denen Kinder im Alter von 6-12 Jahren im Rahmen des typischen Lesens mit großer Wahrscheinlichkeit in Berührung kommen. Die Frequency kann als Schätzer für die Wahrscheinlichkeit herangezogen werden, dass Kinder den Wörtern im schulischen oder familiären Alltag begegnen, und ihnen die Wörter deshalb bekannt sind.

Die *Neighborhood-Size* (NBS) gibt die Ähnlichkeit als Überlappung eines Wortes mit anderen Wörtern der Sprache an. In Anlehnung an Jalbert, Neath, Bireta et al. (2011) wird hier gegenüber der phonologischen NBS die orthographische NBS bevorzugt. Diese ist eindeutiger bestimmbar, da keine (regional) abweichenden Aussprachevarianten berücksichtigt werden müssen. Die NBS kann auf zwei Weisen gemessen werden: (1) als Anzahl der gültigen Wörter, die durch die Änderung des Ausgangswortes an einer Position (Eintauschung, Einschub, Löschung) gefunden werden können (Coltheart-N; Jalbert, Neath, Bireta et al., 2011); (2) als die Anzahl der nötigen Änderungen, um aus dem Ausgangswort ein neues gültiges Wort zu erhalten (Levenshtein-Abstand; Yarkoni, Balota & Yap, 2008). In der vorliegenden Studie wird zur präziseren

Schätzung die Methode des durchschnittlichen Levenshtein-Abstandes der 20 nächsten Nachbarn gewählt (OLD20), wie von Yarkoni et al. (2008) vorgeschlagen. Die NBS kann als Schätzer dafür herangezogen werden, wie leicht einerseits ein Wort (im Sinne der spreading-activation Hypothese, Anderson, 2013) bei der Sprachproduktion aktiviert werden kann, und wie groß andererseits die Verwechslungsgefahr mit anderen Wörtern bei der Spracherkennung ist.

Sowohl zu Frequency als auch zu NBS als Prädiktoren für die Sprachrezeption und -produktion existiert eine Vielzahl an Studien, die daher hier nur knapp genannt werden, um die Relevanz der beiden Parameter zu verdeutlichen. Allgemein nimmt die Frequency eine zentrale Stellung in der linguistischen Forschung ein (McClelland & Rumelhart, 1981). So wirkt sie sich bspw. auf die Augenbewegungen beim Lesen (Kliegl, Grabner, Rolfs & Engbert, 2004; Kretschmar, Schlesewsky & Staub, 2015) und die damit korrespondierende Aktivierung von sprachbezogenen Hirnarealen (Schuster, Hawelka, Hutzler, Kronbichler & Richlan, 2016) aus. Für eine kritische Betrachtung der Rolle der Frequency siehe Baayen (2010). In Aufgaben zur Arbeitsgedächtniskapazität bei Wortspannen hat sich Frequency alleine (Hulme et al., 1997) bzw. gemeinsam mit der Neighborhood-Size als prädiktiv für die Behaltensleistung erwiesen (Allen & Hulme, 2006; Roodenrys, Hulme, Lethbridge, Hinton & Nimmo, 2002). Von Jalbert, Neath, Bireta et al. (2011) werden die beiden Parameter infolgedessen in eine umfangreiche Liste von Faktoren aufgenommen, welche bei der Konstruktion von Arbeitsgedächtnis-Stimuli berücksichtigt werden sollten.

Auch bezogen auf das Aufgabenformat des WWT 6-10, die Benennung von Bildern (*picture naming*), liegen Befunde zur Auswirkung von Frequency vor, die allerdings gemischte Ergebnisse aufweisen. So wurde bereits früh von Oldfield und Wingfield (1965) der Zusammenhang zwischen Antwortzeit und Frequency nachgewiesen. Barry, Morrison und Ellis (1997) konnten einen Effekt für Frequency auf die Benennleistung nachweisen, wobei einschränkend gesagt werden muss, dass dieser bei Kontrolle des Erwerbsalters (*age of acquisition*) der Wörter nicht mehr signifikant war (Barry, Hirsh, Johnston & Williams, 2001; Carroll & White, 1973). Der Frequency-Effekt hatte ebenfalls Bestand bei Almeida, Knobel, Finkbeiner und Caramazza (2007), die zusätzlich die Bekanntheit mit dem Konzept (d. h. vorgelagerte Prozesse der Bildererkennung) kontrollierten, indem sie als Distraktoraufgabe eine Kategorisierungsaufgabe stellten, sodass Probanden in der Phase der Bildererkennung noch nicht wussten, ob sie das Bild benennen oder kategorisieren sollten. Graves, Grabowski, Mehta und Gordon (2007) wiesen darüber hinaus neuronale Korrelate in der Aktivierung mit dem Frequency-Effekt nach.

Die Neighborhood-Size führt zu teils unerwarteten Ergebnissen bei der Vorhersage von linguistischen Leistungen, die mit einem unterschiedlichen Einfluss bei rezeptiven im Gegensatz zu produktiven Sprachaufgaben erklärt werden (Chen & Mirman, 2012; Roodenrys, 2009). So

ist bei rezeptiven Aufgaben, wie z. B. lexikalischen Entscheidungsaufgaben, ein negativer Einfluss der NBS auf die Reaktionszeit nachweisbar (Janse & Newman, 2013): Eine höhere NBS führt zu langsamerer Reaktionszeit, was möglicherweise mit zusätzlich aktivierten, konkurrierenden Antwortoptionen in Verbindung stehen könnte (Luce & Pisoni, 1998). Dazu passen Hinweise auf differenzielle Befunde für die Inhibitionsleistung von Personen: Geringe Inhibitionsleistungen erschwerten insbesondere die Erkennung von ‚schwierigen‘ Wörtern (d. h. solche mit einer höheren Neighborhood-Size), während leichtere Wörter davon weniger beeinträchtigt waren (Sommers & Danielson, 1999). Im Gegensatz dazu scheint eine höhere NBS einen positiven Effekt bei der Sprachproduktion zu haben (Vitevitch & Sommers, 2003). Bei Benennungsaufgaben hatte NBS positive Auswirkungen auf die Leistung bei niedrig-frequenten Wörtern (Andrews, 1989; Peereman & Content, 1995). Dagegen fand Bormann (2011) keinen NBS-Effekt für die Benennung von Wörtern bei unbeeinträchtigten Probanden, nur bei Personen mit einer Wortabrufstörung äußerte sich der Einfluss der NBS in einer erhöhten Fehlerrate beim Benennen.

### 2.2.3.2. Zusammenfassung und Fazit zur Außenstruktur

Zusammengefasst wurde im vorangegangenen Kapitel die Plausibilität der Annahme hergeleitet, dass auch äußere Variablen einen Einfluss auf die Wahrscheinlichkeit haben, ein WWT-Item korrekt zu lösen. Wirken sich diese äußeren Variablen für die Gruppen der Kinder mit und ohne Lernbehinderung differenziell auf die Lösungswahrscheinlichkeit aus, so zeigt sich dies in Interaktionsmustern zwischen den äußeren Variablen und dem Gruppenfaktor. Wenn solche differenziellen Interaktionsmuster vorliegen, dann kann dies als abweichendes Entwicklungsmuster gedeutet werden.

Auf Ebene der personenbezogenen Variablen verläuft die Argumentation analog zum Ansatz der Developmental Trajectories: die personenbezogenen Variablen messen das chronologische und das mentale Alter. Auf Ebene der itembezogenen Variablen ist es angesichts der angeführten empirischen Untersuchungen plausibel zu erwarten, dass auch diese sich auf die Lösungswahrscheinlichkeit bei den Bildbenennungsaufgaben des WWT auswirken.

Damit wurde gezeigt, dass diese äußeren Variablen geeignet sind, die Außenstruktur von Kindern mit und ohne Lernbehinderung vergleichend zu untersuchen. Die Überprüfung erfolgt durch die Spezifikation eines GLMM, in dem die Vorhersage der WWT-Items anhand dieser äußeren Variablen abgebildet wird. In den Interaktionsmustern mit dem Gruppenfaktor wird operationalisiert, ob sich die Außenstruktur der WWT-Items bei Kindern mit und ohne Lernbehinderung unterscheidet; das methodische Vorgehen hierfür wird im Methodenteil (Kapitel 6.2.4) genauer erläutert.

### 2.3. Zusammenfassung und Fazit zu Gedächtniskomponenten

In diesem Kapitel wurden die für diese Arbeit relevanten kognitiven Lernvoraussetzungen anhand des Modells der guten Informationsverarbeitung auf die phonologische Schleife als Teil des Arbeitsgedächtnisses und das mentale Lexikon als Teil des semantischen Langzeitgedächtnisses eingegrenzt. Beide nehmen zur Erklärung des schulischen Wissenserwerbs eine zentrale Rolle ein. So erhält die phonologische Schleife sprachliche Informationen für einen kurzen Zeitraum aufrecht und ermöglicht damit erst die langfristige Speicherung im semantischen Gedächtnis (für sprachliche Begriffe das mentale Lexikon).

Die phonologische Schleife wird im Rahmen der ersten Studie in drei Teilfragestellungen untersucht. Die erste Teilfragestellung bezieht sich auf die Kapazität der phonologischen Schleife insgesamt, die aus dem Zusammenspiel des phonetischen Speichers und des Rehearsalprozesses resultiert. Zudem werden mit Rehearsal und Redintegration zwei Prozesse zur Aufrechterhaltung der Gedächtnisspur genauer betrachtet. Die zweite Teilfragestellung behandelt den Rehearsalprozess, der als inneres Nachsprechen dazu beiträgt, Inhalte über die eng begrenzte Zeit des phonetischen Speichers hinaus aufrechtzuerhalten. In der dritten Teilfragestellung wird der Redintegrationprozess betrachtet, der über die Rekonstruktion von teilweise zerfallenen Inhalten aus dem Langzeitgedächtnis die Aufrechterhaltung in der phonologischen Schleife unterstützt. Zu den Aspekten der ersten beiden Teilfragestellungen liegen bereits Befunde zu Kindern mit Lernbehinderung vor, die jedoch uneinheitlich sind. Zur Verfügbarkeit des Redintegrationprozesses bei Kindern mit Lernbehinderung liegen bislang keine empirischen Untersuchungen vor. Alle drei Teilfragestellungen zielen darauf ab, die Art des Entwicklungsmusters hinsichtlich der genannten Aspekte bei Kindern mit Lernbehinderung zu bestimmen. Hierfür kommt in Studie 1 die Methode der Developmental Trajectories zum Einsatz; die Entwicklungsmuster und das Design werden im folgenden Kapitel 3 eingeführt und erläutert.

Das mentale Lexikon als wesentlicher Teil für die Speicherung sprachlichen Wissens im semantischen Gedächtnis wird im Rahmen der zweiten Studie anhand des Wortschatzabrufs untersucht. Da die Struktur und Organisation des mentalen Lexikons für die Einspeicherung mindestens ebenso wichtig ist wie der reine Wortschatzumfang, werden zwei Strukturmerkmale betrachtet. Die erste Teilfragestellung prüft die Binnenstruktur, also das Verhältnis der Einträge im mentalen Lexikon zueinander, was am Strukturmerkmal der Wortart-Zugehörigkeit festgemacht wird. In der zweiten Teilfragestellung wird die Außenstruktur überprüft, die sich im Verhältnis der Einträge des mentalen Lexikons zu personen- und itembezogenen Variablen darstellt. Zur Bestimmung des Entwicklungsmusters wird in Studie 2 der methodische Zugang des Gruppen-Matching Verfahrens gewählt, der, da er historisch älter ist, im folgenden Kapitel als erstes vorgestellt wird.

### 3. Forschungsmethodischer Hintergrund

In diesem Kapitel wird der mentale Altersvergleich (*mental age matching*) eingeführt und begründet, da dieser Ansatz für das Studiendesign der vorliegenden Arbeit genutzt wird, um zwischen verschiedenen Entwicklungsmustern zu differenzieren. Die Anwendung des mental age Matchings hat Auswirkungen auf die Stichprobenauswahl, da für ein vergleichbares mentales Alter jüngere Kinder ebenfalls in die Kontrollgruppe aufgenommen werden. Zunächst wird die Begründung des Ansatzes erläutert, und anschließend zwei Designs eingeführt: (1) das klassische Matching auf Gruppenebene („Gruppenmatching“) und (2) dessen Weiterentwicklung zu Developmental Trajectories, die gegenüber dem klassischen Ansatz einige Vorteile bieten.

Der inhaltliche Anspruch an das Erhebungsdesign besteht darin, dass es in der Lage sein sollte, zwischen Entwicklungsmustern zu differenzieren, die auf der developmental-difference Kontroverse (Zigler & Balla, 1982) zur Unterscheidung von *developmental delay* und *structural difference* beruhen. Die Frage lautet also: Liegt hinsichtlich der untersuchten Domäne des Arbeits- bzw. Langzeitgedächtnisses bei Kindern mit Lernbehinderung eine Beeinträchtigung vor, und falls ja, welcher Art? Handelt es sich bei einer solchen Beeinträchtigung eher um einen Rückstand, der im Rahmen der allgemein verzögerten Entwicklung liegt (developmental delay), oder eher um eine spezifische Beeinträchtigung von Arbeits- und Langzeitgedächtnis, die über die allgemeine Entwicklungsverzögerung hinausgeht (structural difference)? (siehe auch Brown, Parrikh & Patel, 2020).

Zur Unterscheidung dieser Entwicklungsmuster kann der Ansatz des mental age Matchings genutzt werden, für den in der Regel zwei Kontrollgruppen heranzuziehen sind: eine Gruppe ist gemessen am chronologischen Alter (CA) vergleichbar, und eine weitere nach mentalem Alter (MA). Die MA-Gruppe ist in der Regel deutlich jünger, aber dadurch in relevanten Kontrollvariablen der mentalen Entwicklung vergleichbar.

Der MA-Matching Ansatz beruht auf der Annahme, dass sich der Einfluss des allgemeinen mentalen Entwicklungsalters bereits durch den Versuchsaufbau kontrollieren lässt, indem eine im Vergleich zur Untersuchungsgruppe jüngere Kontrollgruppe rekrutiert wird. Der Vergleich mit einer gleichaltrigen Kontrollgruppe wäre an der Stelle nicht besonders informativ, da per definitionem bereits von Unterschieden auszugehen ist, welche durch eine experimentelle Untersuchung lediglich bestätigt würden ohne neue Erkenntnisse zu gewinnen. Eine jüngere Kontrollgruppe bietet dagegen die Möglichkeit, das mentale Alter bzw. den allgemeinen (kognitiven) Entwicklungsstand als eine weitere Dimension für den Vergleich heranzuziehen. Wenn sich hier immer noch Unterschiede in Arbeits- und Langzeitgedächtnis zeigen, kann dies als empirischer Hinweis interpretiert werden, dass die Beeinträchtigung der Zieldomäne in der Zielgruppe eine besondere Rolle spielt.

Vereinfacht ausgedrückt geht der Vergleich davon aus, dass bspw. Kinder mit Lernbehinderung im Alter von zehn Jahren auf dem Stand eines mentalen Entwicklungsalters von sechs bis sieben Jahren sind; als Referenzgruppe werden dementsprechend sechs- bis sieben-jährige unbeeinträchtigte Kinder herangezogen. Aus dem Gruppenvergleich lässt sich die Aussage zu der Frage ableiten, ob Kinder mit Lernbehinderung im Arbeits- und Langzeitgedächtnis auf dem Stand wie entsprechend jüngere Kinder ohne Beeinträchtigung sind, oder ob die Probleme in Arbeits- und Langzeitgedächtnis noch gravierender sind und über die allgemeine Entwicklungsverzögerung hinausgehen.

In den folgenden zwei Unterkapiteln wird je ein Verfahren des mentalen Altersvergleichs vorgestellt. Beide beruhen auf der identischen, oben eingeführten Grundidee eines Vergleichs mit einer jüngeren Kontrollgruppe. Das Gruppenmatching (also der Vergleich auf Gruppenebene) ist die ältere Variante, als deren Weiterentwicklung die Methode der Developmental Trajectories verstanden werden kann.

#### **3.1. Mentaler Altersvergleich im Gruppensdesign**

In der Forschung zur Leseentwicklung stellen Backman, Mamen und Ferguson (1984) das *reading level design* vor, in dem die Vergleichsgruppe auf dem Level der allgemeinen Lesefähigkeit rekrutiert wird. Sie begründen das Design damit, dass Unterschiede zwischen den Gruppen hinsichtlich der Erfahrung mit Schriftsprache, Stufen im Schriftspracherwerb sowie Schwierigkeit des Aufgabenmaterials auf diese Weise minimiert werden (ebd., S. 561). Insofern liegt insbesondere im Auffinden eines nicht-signifikanten Unterschiedes mit der jüngeren Kontrollgruppe ein besonderer Informationsmehrwert (Bryant & Goswami, 1986; Mamen, Ferguson & Backman, 1986; für eine kritische Einordnung siehe Goswami & Bryant, 1989; Jackson & Butterfield, 1989).

Wenn es keinen Unterschied zwischen der Untersuchungsgruppe und der jüngeren MA-Gruppe gibt, könne dies so interpretiert werden, dass sich Kinder mit Lesestörung qualitativ nicht wesentlich von jüngeren unbeeinträchtigten Lesern unterscheiden, sondern lediglich in ihrem Leseerwerb verzögert seien (Backman et al., 1984). Dagegen wäre es als qualitatives, strukturelles ‚Defizit‘ zu deuten, wenn ein Unterschied zur MA-Gruppe beobachtet werden kann.

Zunächst werden im Folgenden zwei Varianten zur Messung des mentalen Alters vorgestellt, danach die Szenarien des Gruppenmatchings zur Unterscheidung von Entwicklungsverzögerung und qualitativem Defizit genauer erläutert. Nach Anwendungsbeispielen des Gruppensdesigns wird die Kritik am Gruppenmatching dargelegt, die zur Weiterentwicklung der Developmental Trajectories geführt hat.

### 3.1.1. Messung des mentalen Alters

Prinzipiell kann jede Variable, die einen für die Fragestellung relevanten Aspekt der Entwicklung misst, als Schätzer für das mentale Alter genutzt werden, mit dem Ziel, eine nach diesem mentalen Alter parallelisierte Kontrollgruppe zu erheben. So kam im Kontext der Erforschung von intellektuellen Beeinträchtigungen und Lernbehinderung der Ansatz des MA-Matchings zum Zweck der Parallelisierung des Intelligenzalters zum Einsatz (Zielinski, 1980). Im Kontext der Leseforschung wurde dagegen bevorzugt das Leselevel als mental age Variable genutzt (Backman et al., 1984). Bei der Erforschung von Lernbehinderung wird als Variable, für welche kontrolliert werden soll, häufig die kognitive Kapazität eingesetzt (Mähler, 2007; van der Molen et al., 2007).

Das mentale Alter kann auf zwei Vorgehensweisen ermittelt werden: das Norm-Entwicklungsalter oder das Rohwert-Entwicklungsalter. Für das Norm-Entwicklungsalter wird die Normtabelle eines standardisierten Leistungstests gewissermaßen ‚rückwärts‘ gelesen. Es wird diejenige Alterssubgruppe gesucht, für die der vom Kind erzielte Rohwert einem durchschnittlichen Wert ( $T = 50$  bzw.  $PR = 50$ ) entspricht, bzw. diesem am nächsten liegt. Die zweite Variante, hier als ‚Rohwert-Entwicklungsalter‘ bezeichnet, nimmt den Vergleich zwischen den Gruppen direkt anhand der Rohwert-Scores vor, ohne diese zunächst in das Norm-Entwicklungsalter zu übertragen. Beide Varianten basieren also auf den in absoluten Punkten gemessenen Leistungen der Probanden in einem geeigneten standardisierten Testverfahren.

Das Norm-Entwicklungsalter setzt voraus, dass die Normtabellen des standardisierten Verfahrens ausreichend differenziert sind, da ansonsten der mit der Varianzreduktion verbundene Informationsverlust zu stark ins Gewicht fällt. Außerdem ist dieses Vorgehen nur für einen mittleren Leistungsbereich sinnvoll – jenseits der jüngsten bzw. ältesten Normgruppe kann keine Differenzierung vorgenommen werden, was zu Boden- und Deckeneffekten führt.

Zwar bringt das Rohwert-Entwicklungsalter den Nachteil mit sich, dass es nicht unmittelbar als ‚Alter‘ in Jahren und Monaten abgebildet und interpretiert werden kann. Für die vorliegende Arbeit überwog aber das Argument, eine differenziertere Schätzung ohne Informationsverlust durch eine reduzierte Varianz und auch an den Rändern der Verteilung interpretierbare Werte zu erhalten. Daher wurde für das Matching der MA-Gruppe auf Gruppenebene das Rohwert-Entwicklungsalter genutzt.

### 3.1.2. Inhaltliche Interpretation von MA-Szenarien

Bei der Anwendung des MA-Matchings auf Gruppenebene zur Einschätzung der Rolle des Arbeitsgedächtnisses können mehrere Ergebnis-Szenarien auftreten, deren Interpretation im Folgenden erläutert werden. Dazu werden die folgenden drei Gruppen betrachtet: LB = Kinder mit Lernbehinderung, CA = nach chronologischem Alter vergleichbare Kontrollgruppe, und



MA = nach mentalem Alter vergleichbare Kontrollgruppe. Anhand der Szenarien, die aus dem Vergleich zwischen den drei Gruppen resultieren können, können die oben eingeführten Positionen von *developmental delay* und *structural difference* unterschieden werden (Zigler & Balla, 1982). Insbesondere wird dafür auf den Vergleich mit der MA-Gruppe rekuriert (z. B. in Mähler, 2007), der im zweiten Abschnitt angesprochen wird.

Zunächst wird der Vergleich mit der CA-Gruppe betrachtet. Tritt zwischen LB und CA kein Unterschied ( $LB = CA$ ) auf, deutet dies darauf hin, dass das Arbeitsgedächtnis bei Lernbehinderung nicht beeinträchtigt ist und somit als möglicher kausaler Faktor ausgeschlossen werden kann. Lässt sich ein Unterschied feststellen ( $LB < CA$ ), so ist dies isoliert nicht interpretierbar (Goswami & Bryant, 1989), da bereits per Definition bei Kindern mit Lernbehinderung eine schlechtere Leistung im Verhältnis zu gleichaltrigen unbeeinträchtigten Kindern zu erwarten ist.

Anders verhält es sich mit dem Vergleich mit der MA-Gruppe: Schneidet die LB-Gruppe in Arbeitsgedächtnis-Aufgaben selbst im Vergleich zu einer jüngeren Kontrollgruppe schlechter ab ( $LB < MA < CA$ ), dann scheint ein spezifisches Defizit vorzuliegen, das sich in Leistungen zeigt, die deutlich unterhalb des Niveaus liegen, das aufgrund des mentalen Alters zu erwarten wäre (Mähler, 2007). Somit könnte die gemessene Variable möglicherweise auch eine Rolle bei der Erklärung von Lernbehinderung spielen (Backman et al., 1984; Goswami & Bryant, 1989). Dieser deutliche Unterschied wird dann entsprechend als strukturelle Abweichung interpretiert, die auf eine *difference position* (Zigler & Balla, 1982) bzw. auf eine ‚Differenzhypothese‘ (Zielinski, 1980, S. 68) hinweist.

Findet sich kein Unterschied zwischen den Gruppen ( $LB = MA < CA$ ), so wird dies als eine Verzögerung in der gemessenen Domäne interpretiert (Backman et al., 1984), die im Rahmen der allgemeinen verzögerten Entwicklung liegt. Entsprechend würde die Einschränkung in der Zieldomäne als Entwicklungsverzögerung (Mähler, 2007) bzw. *developmental position* (Zigler & Balla, 1982) oder weniger eindeutig formuliert als ‚Defizithypothese‘ (Zielinski, 1980, S. 68) interpretiert. Goswami und Bryant (1989) weisen jedoch darauf hin, dass ausbleibende Unterschiede zwischen MA und LB auch durch andere, nicht erfasste Variablen erklärt werden können, wie z. B. Einsatz kompensatorischer Strategien oder unterschiedlicher Lebenserfahrung.

Daher betonen sie die Relevanz von längsschnittlich und experimentell erfassten Daten, bevor Schlüsse über die kausale Rolle der gemessenen Variablen gezogen werden können. In dieser Arbeit wird eine solche kausale Interpretation jedoch nicht angestrebt (siehe Kapitel 1.4); für die rein deskriptive Beschreibung der kognitiven Lernvoraussetzungen sind die vorliegenden querschnittlichen Daten ausreichend aussagekräftig.

### 3.1.3. Anwendungsbeispiele für MA-Designs

Am Beispiel der oben bereits angeführten Studie von Mähler (2007) seien das Matching-Vorgehen und die Interpretation der Ergebnisse illustriert. In der genannten Studie wurde ein fünf-Gruppen Design gewählt: Es gab zwei Gruppen mit Lernbehinderung, eine in der 4. Klasse und eine in der 9. Klasse. Als CA-Kontrollgruppen wurden jeweils unbeeinträchtigte Kinder aus der 4. und 9. Klasse getestet. Die 4. Klasse diente gleichzeitig als MA-Kontrollgruppe für die älteren Kinder mit Lernbehinderung der 9. Klasse. Als MA-Kontrollgruppe für die jüngeren Kinder mit Lernbehinderung der 4. Klasse wurden zusätzlich unbeeinträchtigte Kinder der 1. Klasse getestet. Mit diesen fünf Gruppen konnten also für jüngere und ältere Kinder mit Lernbehinderung sowohl Vergleiche mit einer nach chronologischem als auch nach mentalem Alter gematchten Kontrollgruppe gezogen werden.

Die Ergebnisse zeigten für die Arbeitsgedächtnis-Bereiche der Zentralen Exekutive und des visuell-räumlichen Notizblocks ein durchgängiges Muster von Entwicklungsverzögerung sowohl für jüngere als auch für ältere Kinder mit Lernbehinderung, was sich am gefundenen Muster ( $MA = LB < CA$ ) ablesen lässt. Nur in den Untertests zur phonologischen Schleife wurde ein strukturelles Defizit bei jüngeren Kindern mit Lernbehinderung festgestellt ( $LB < MA < CA$ ), das jedoch nicht für die älteren Kinder mit Lernbehinderung galt. Dieser Befund wurde dahingehend interpretiert, dass die „substanzielle qualitative Abweichung von der für das jeweilige mentale Alter typischen Funktionsweise“ (Mähler, 2007, S. 105) mit der Zeit abnimmt. Dies wiederum wurde auf ein Einsetzen der automatischen Aktivierung des Rehearsal-Prozesses mit fortgeschrittenem mentalem Alter zurückgeführt.

Als weitere Studien, in denen das MA-Matching auf Gruppenebene im Kontext der Arbeitsgedächtnis-Forschung bei Lernbehinderung zum Einsatz gekommen ist, seien die folgenden angeführt. Hasselhorn und Mähler (2003) untersuchten den Rehearsal-Prozess bei jungen Erwachsenen mit Lernbehinderung. Der strategische Einsatz des Rehearsal-Prozesses bei Kindern mit Lernbehinderung (bzw. *mild intellectual disabilities* im internationalen Kontext) wurde von Polczek et al. (2016) betrachtet, wobei in dieser Untersuchung jedoch auf die CA-Gruppe verzichtet wurde. Van der Molen et al. (2007) untersuchten das Arbeitsgedächtnis bei Jugendlichen mit Lernbehinderung in einem klassischen drei-Gruppen Design, wohingegen Schuchardt et al. (2011) für den Vergleich der phonologischen Schleife den Ansatz wählten, drei Gruppen unterschiedlichen chronologischen Alters mit verschiedenen Schweregraden einer Intelligenzbeeinträchtigung zu erheben, sodass das mentale Alter für alle Gruppen gleich war. Weiterhin wurde die Arbeitsgedächtnis-Komponente des episodischen Puffers von Henry (2010) bei Kindern mit Intelligenzbeeinträchtigung untersucht. Auch für ältere Stichproben mit Jugendlichen und

ungen Erwachsenen kam das MA-Matching zum Einsatz (Mähler & Hasselhorn, 2003; Pochon, Touchet & Ibernou, 2017).

#### 3.1.4. Kritik am MA-Design

Auch wenn sich der Ansatz des MA-Matchings auf Gruppenebene als sehr nützlich erwiesen und, wie oben skizziert, einer Vielzahl an empirischen Studien zur Untersuchung von Entwicklungsmustern gedient hat, unterliegt er einigen methodischen Einschränkungen, die im Folgenden basierend auf der Kritik in Thomas et al. (2009) zusammengefasst werden. Als hilfreiche Weiterentwicklung schlagen Thomas et al. (2009) das Developmental Trajectory (DT) Design vor, welches die Grundlage für die Auswertungen im Rahmen von Studie 1 ist und daher im folgenden Unterkapitel eingeführt und näher erläutert wird.

Ein wesentlicher Kritikpunkt am ursprünglichen Gruppendesign betrifft die Tatsache, dass die Auswertung auf einem Mittelwertvergleich zwischen den Gruppen beruht (Carney et al., 2013). Der informative Gehalt, der in der Varianz innerhalb der Gruppen steckt, werde dadurch übersehen und der Aspekt der Entwicklung möglicherweise sogar überdeckt. Thomas et al. (2009) skizzieren beispielhaft zwei Szenarien, bei denen sich unterschiedliche DT-Muster ergeben, nämlich einerseits das eines verspäteten bzw. niedrigeren Entwicklungseinsatzes (*delayed onset*) und andererseits das einer verlangsamten Entwicklungsrate (*slowed rate*; eine detaillierte Darstellung der DT-Muster erfolgt in Abschnitt 3.2.4). Im traditionellen Gruppendesign wären diese beiden differenziellen Muster jedoch durch nahezu identische Gruppenmittelwerte repräsentiert und dementsprechend kaum unterscheidbar (siehe Abbildung 1f, Seite 76).

Auch stellt es einen Nachteil des MA-Matchings dar, dass die Variablen zur Schätzung des mentalen Alters bereits bei der Untersuchungsplanung vorab festgelegt werden, und daher für jede dieser Variablen eine eigene, parallelisierte Kontrollgruppe erhoben werden muss. Dadurch sind sowohl die Auswahl der MA-Variablen als auch die Flexibilität in der Auswertung im Nachhinein eingeschränkt (Thomas et al., 2009). Dies bringt auch mit sich, dass der Altersbereich innerhalb der Gruppen eher eng zu wählen ist (Thomas et al., 2009), um die Varianz innerhalb der Gruppen zu minimieren. Dadurch ist es zwar möglich, Unterschiede zwischen Gruppen einfacher aufzudecken, jedoch umfasst der Altersbereich damit nur einen kurzen Ausschnitt aus einem vermutlich längerfristigen Entwicklungsverlauf. Allein aufgrund eines nicht-signifikanten Unterschiedes zur MA-Kontrollgruppe ist die Interpretation einer Entwicklungsverzögerung (*developmental lag*) nicht eindeutig, da offen bleibt, ob sich diese Verzögerung über die gesamte Entwicklungsspanne ähnlich darstellt (Goswami & Bryant, 1989).

Thomas et al. (2009) plädieren daher für den Einsatz von DTs, mit denen der Aspekt der Entwicklung stärker in den Fokus der Analyse gerückt werden kann und die sich außerdem als

flexibler in der Auswertung, sowie vielfältiger hinsichtlich der Abbildung von Entwicklungsmustern erweisen. Das Konzept sowie die Konstruktion von DTs ist Thema des folgenden Unterkapitels.

### 3.2. Developmental Trajectory Design<sup>14</sup>

Developmental Trajectories können als Erweiterung des klassischen MA-Matchings auf Gruppenebene verstanden werden. Sie stellen ein vielfältiges methodisches Auswertungstool dar, mit dem auch komplexere Entwicklungszusammenhänge<sup>15</sup> dargestellt werden können, wo der klassische Gruppen-Ansatz an seine Grenzen gelangt.

Die Parallele zum Gruppendesign liegt in der Grundidee des Vergleichs mit einem Indikator für das mentale Alter. Der Unterschied besteht in erster Linie darin, dass das mentale Alter als kontinuierliche, intervallskalierte Variable in ein lineares Modell aufgenommen wird. Thomas et al. (2009, S. 336) fassen das Ziel von DTs folgendermaßen zusammen: „The aim is to construct a function linking performance with age on a specific experimental task and then to assess whether this function differs between the typically developing group and the disorder group.“. Dagegen ist beim klassischen Gruppendesign das mentale Alter konstant zu halten, damit die Gruppen vergleichbar sind.

Indem also im DT-Design die Variabilität innerhalb der Gruppen über ein breiteres (mentales) Altersspektrum zusätzlich betrachtet werden kann, und nicht nur kurze Ausschnitte eines schmalen Altersbereiches, rückt der Entwicklungsaspekt stärker in den Fokus der Analysen (Karmiloff-Smith, 1998; Thomas et al., 2009). Dadurch wird erstens eine eher statische Betrachtungsweise von Entwicklungsbeeinträchtigungen zugunsten einer Sichtweise von (evtl. differenziellen) Veränderungen über die Zeit überwunden (Thomas et al., 2009). Zweitens ermöglicht die Berücksichtigung der Varianz innerhalb der Gruppen eine vielfältigere und differenziertere Beschreibung von Entwicklungsmustern: „The use of trajectories provides a descriptively more powerful empirical vocabulary.“ (Thomas et al., 2009, S. 345; für eine kritische Diskussion des *delay*-Begriffs siehe Levy, 2018).

#### 3.2.1. Anwendungsbeispiele für Developmental Trajectories

Developmental Trajectories kamen in den vergangenen zwei Dekaden zum einen bei uneinträchtigten Gruppen für allgemein-psychologische Fragestellungen zum Einsatz, wurden

---

<sup>14</sup> Die Vorstellung des DT-Designs in diesem Unterkapitel 3.2 orientiert sich sinngemäß an der Darstellung aus dem bereits publizierten Beitrag und baut darauf auf (Studie 1, Seite 3f.: Bruns, Ehl & Grosche, 2019). Mit dieser Kenntlichmachung wird den Bestimmungen der Promotionsordnung vom 17.11.2017 der Bergischen Universität Wuppertal entsprochen.

<sup>15</sup> Angesichts der in dieser Arbeit verwendeten querschnittlich erhobenen Daten darf „Entwicklung“ hier lediglich als Zusammenhangsmuster der Ziel-Aufgaben mit dem chronologischen Alter und Variablen zur Schätzung des mentalen Alters interpretiert werden.

aber vor allem im Kontext von beeinträchtigter Entwicklung für differenziell-psychologische Fragen verwendet. Zwar ohne expliziten Bezug auf das oben dargestellte Rahmenmodell, aber methodisch sehr ähnlich, untersuchten Hasenäcker, Schröter und Schroeder (2017) morphologische Effekte bei lexikalischen Entscheidungsaufgaben in einer großen, querschnittlich angelegten Stichprobe von über 500 Kindern, und Arciuli & Simpson (2011) betrachteten das statistische Lernen in Abhängigkeit von der Stimulus-Präsentationsdauer. DTs wurden auch für die Untersuchung von Gesichter- und Emotionserkennung in den klinisch klar diagnostizierbaren Entwicklungsbeeinträchtigungen des Williams Syndrom (Annaz, Karmiloff-Smith, Johnson & Thomas, 2009; Ibernou, Touchet & Pochon, 2018; Karmiloff-Smith et al., 2004) sowie Down Syndrom (Annaz et al., 2009; Arango, Aparicio & Tenorio, 2018; Pochon et al., 2017) eingesetzt.

Weitere Untersuchungen wurden zum verbalen und visuo-räumlichen Kurzzeitgedächtnis (Carney et al., 2013) und zu Sprachproduktion und -verständnis (Chapman, Hesketh & Kistler, 2002 in einem längsschnittlichen Design) unter Einsatz von DTs durchgeführt. Bei Kindern mit und ohne Lernbeeinträchtigung untersuchten Sullivan, Kohli, Farnsworth, Sadeh und Jones (2016) die Lesefähigkeiten in einem längsschnittlichen Design. In einer ebenfalls längsschnittlich angelegten Erhebung mit einer großen Stichprobe von 261 Kindern zwischen 3-8 Jahren testeten Lei et al. (2011) die Lesekompetenzen von chinesischen Kindern und leiteten daraus vier verschiedene Typen von Entwicklungsmustern ab.

Im Folgenden soll die Konstruktion der Developmental Trajectories anhand eines Regressionsmodelles und die Interpretation der daraus resultierenden Szenarien erläutert werden. Mit der Kenntnis der Funktionsweise von DTs ist es im Anschluss besser möglich, die daraus erwachsenden spezifischen Vorteile des Modellierungsansatzes für die Interpretation und Unterscheidung von Entwicklungsmustern nachzuvollziehen.

### **3.2.2. Konstruktion von Developmental Trajectories**

Die Konstruktion von Developmental Trajectories erfolgt statistisch im Rahmen eines Allgemeinen Linearen Modells. Das Vorgehen orientiert sich an den sehr detaillierten Hinweisen im Beitrag und dem elektronischen Anhang von Thomas et al. (2009)<sup>16</sup> für die technische Umsetzung mit der Software SPSS (siehe auch Peugh & Enders, 2005). Mithilfe der ANOVA Mixed-Prozedur wird ein allgemeines lineares Modell erstellt. Die Leistung in einer Arbeitsgedächtnis-Aufgabe als abhängige Variable wird von verschiedenen Prädiktoren erklärt, auf welche in den folgenden drei Abschnitten schrittweise eingegangen wird. Die erhobenen Daten

---

<sup>16</sup> Der in der Publikation angegebene Link ist veraltet, die Website ist über das Archiv zugänglich: [http://www.bbk.ac.uk/psychology/dnl/old\\_site/stats/Thomas\\_trajectories.html](http://www.bbk.ac.uk/psychology/dnl/old_site/stats/Thomas_trajectories.html) (Zugriff am 02.06.2020).

sollten für dieses Verfahren einen breiten Altersbereich umfassen; außerdem sollte die unbeeinträchtigte Kontrollgruppe den Bereich des jüngsten gemessenen mentalen Alters bis zum ältesten gemessenen chronologischen Alter in der Untersuchungsgruppe abdecken (Thomas et al., 2009).

In der vorliegenden Arbeit wird das mentale Alter durch zwei verschiedene Variablen geschätzt: erstens wird die kognitive Kapazität durch ein non-verbales Screening-Verfahren (Grundintelligenztest Skala 1 – Revision, CFT 1-R, Weiß & Osterland, 2012) erfasst, das sich für Kinder mit und ohne Lernbeeinträchtigung als messinvariant erwiesen hat (Heine et al., 2018). Weiterhin wird der Umfang des aktiven Wortschatzes (Wortschatz- und Wortfindungstest für 6- bis 10-Jährige, WWT 6-10, Glück, 2011) als Variable für die Schätzung des mentalen Alters eingesetzt.

Schritt (1) stellt das Herzstück der DT-Analyse dar, das in der Varianzaufklärung der AG-Kapazität durch einen kontinuierlichen Altersprädiktor liegt. Als Altersprädiktor lässt sich das chronologische Alter aber auch Maße zur Schätzung des mentalen Alters heranziehen. Da es sich um eine kontinuierliche Variable handelt, wird diese als „Covariate“ im SPSS-Mixed-ANOVA Fenster eingesetzt. Damit wird die Fragestellung untersucht, inwiefern die AG-Kapazität mit zunehmendem Alter ansteigt. Wenn anstelle des CA weitere Variablen zur Schätzung des mentalen Alters (MA) als Prädiktor in das Modell eingesetzt werden, z. B. die kognitive Kapazität (gemessen am Rohwert des CFT 1-R) oder der Wortschatzumfang (gemessen am Rohwert des WWT 6-10), so kann untersucht werden, inwiefern die AG-Kapazität mit steigender kognitiver Kapazität bzw. steigendem Wortschatzumfang zusammenhängt.

In Schritt (2) kann als zusätzlicher Zwischensubjekt-Faktor (in das entsprechend bezeichnete Feld im ANOVA-Mixed Fenster) eine kategoriale Variable der Gruppenzugehörigkeit in das Modell aufgenommen werden. Dadurch lässt sich überprüfen, ob die Untersuchungsgruppe im Vergleich zur unbeeinträchtigten Kontrollgruppe ein differenzielles Entwicklungsmuster (näher erläutert in Abschnitt 3.2.4) aufweist, z. B. ob der Zuwachs der AG-Kapazität mit zunehmendem CA in beiden Gruppen ähnlich stark ausfällt.

Schließlich kann in Schritt (3) ein Innersubjekt-Faktor für die Aufgabenart im Modell berücksichtigt werden, indem die entsprechenden Aufgaben-Variablen als Stufen eines messwiederholten Faktors eingesetzt werden. Dadurch ist es möglich, mit einem DT auch komplexere Zusammenhänge auf Ebene der Ziel-Aufgabe darzustellen und möglicherweise differenzielle Auswirkungen von verschiedenen Aufgabenbedingungen miteinander zu modellieren. Hierfür muss jede\*r Proband\*in eine Aufgabe mehrfach unter verschiedenen Bedingungen bearbeitet haben, z. B. Wortspannenaufgaben mit kurzen und mit langen Wörtern.

Anhand der Variable „Wortlänge“ als Aufgaben-Faktor lässt sich untersuchen, ob es (a) einen Wortlängeneffekt gibt (also grundsätzlich kurze Wörter besser behalten werden als lange Wörter), und (b) ob sich die Stärke des Wortlängeneffektes mit zunehmendem CA oder MA ändert. In Interaktion mit dem Gruppenfaktor kann schließlich die Frage beantwortet werden, ob sich Kinder mit und ohne Lernbehinderung hinsichtlich der Größe des Wortlängeneffektes und dessen Änderung in Abhängigkeit von chronologischem und mentalem Alter unterscheiden. Damit würde die Frage beantwortet, ob in Bezug auf den Wortlängeneffekt differenzielle Muster zwischen den Gruppen vorliegen.

Zu den Altersvariablen ist anzumerken, dass deren Skalen-Nullpunkt im Rahmen des DT-Modells in der Regel nicht sinnvoll zu interpretieren ist. Dies liegt daran, dass i. d. R. die Stichprobe in einem Altersbereich liegt, der nicht bereits mit der Geburt beginnt, sondern bspw. eine Altersspanne von sechs bis zehn Lebensjahren umfasst. Da der Intercept-Koeffizient eines linearen Modells sich aber auf den Schnittpunkt mit der y-Achse, d. h. den Nullpunkt der kontinuierlichen Variable, bezieht, würde die in der Stichprobe von 6- bis 10-Jährigen gefundene Beziehung extrapoliert für Neugeborene. Formuliert am Beispiel einer Fragestellung zum Arbeitsgedächtnis hieße dies: „Wie hoch wäre erwartungsgemäß die Arbeitsgedächtniskapazität bei einem Neugeborenen angesichts der Entwicklung von 6-10-Jährigen einzuschätzen?“, was offensichtlich keine sinnvolle Fragestellung darstellt.

Um erstens diese Extrapolation zu vermeiden und zweitens den Intercept-Effekt auch inhaltlich gewinnbringend interpretieren zu können, empfehlen Thomas et al. (2009) eine lineare Transformation der Altersvariablen, so dass die (mental) jüngste Person den Wert Null erhält. Dafür wird für jede Person in der Altersvariable der Wert der jüngsten Person subtrahiert. Auf diese Weise ist der Intercept inhaltlich interpretierbar zur Beantwortung der Frage, ob sich die Leistung in der Zielvariable zum Beginn der gemessenen Entwicklung signifikant von Null unterscheidet. An der Interpretation der Steigungskoeffizienten (*slope*) ändert sich durch die Transformation nichts.

Auf die Interpretation der Haupteffekte und Interaktionen in den DT-Modellen wird im Folgenden vertieft eingegangen, da sie davon abhängig ist, ob eine kategoriale oder kontinuierliche Variable betrachtet wird. Im Gegensatz zur bei Regressionsmodellen häufig üblichen Dummy-Kodierung werden kategoriale Variablen in der von Thomas et al. (2009) erläuterten ANOVA-Prozedur in SPSS effekt-kodiert.

Die Wahl der Kodierungsart hat Auswirkungen auf die Interpretation der Intercept-Koeffizienten: Bei einer Dummy-Kodierung gäbe der Intercept-Koeffizient den Wert einer Referenzgruppe wieder, und der Haupteffekt für z. B. Gruppe bezöge sich auf die Abweichung der zweiten Gruppe von der Referenzgruppe. Bei der hier gewählten Effekt-Kodierung bezieht sich der

Intercept-Koeffizient jedoch nicht auf den Unterschied zu einer Referenzgruppe, sondern auf den Unterschied zwischen den Gruppen im Verhältnis zum Gesamtmittelwert. Dies wirkt sich z. B. in Schritt (2) auf die Interpretation des Haupteffektes für Alter aus: bei einer Dummy-Codierung der Gruppenvariable bezöge sich der Steigungskoeffizient des Faktors Alter nur auf die Referenzgruppe, im vorliegenden Fall der Effekt-Kodierung gibt der Steigungskoeffizient aber die durchschnittliche Zunahme der AG-Kapazität mit steigendem Alter gemittelt über beide Gruppen hinweg an (nähere Erläuterung im folgenden Abschnitt).

### 3.2.3. Statistische Effekte in Developmental Trajectories

Im Folgenden werden daher die Interpretationen der einzelnen Effekte am Beispiel der abhängigen Variable AG-Kapazität und der Prädiktoren Alter, Gruppenzugehörigkeit und im letzten Schritt Wortlänge (als Aufgaben-Faktor) illustriert. Die Reihenfolge der Interpretationen orientiert sich an den oben dargestellten Schritten der Konstruktion (Abschnitt 3.2.2).

Im einfachen Fall (1), in dem die AG-Kapazität durch das chronologische Alter vorhergesagt wird, gibt es einen Intercept-Effekt und einen Slope-Effekt. Der Intercept-Effekt bezieht sich auf den Schnittpunkt der Regressionsgeraden mit der y-Achse. Wenn der Intercept-Effekt signifikant ist, weicht dieser Schnittpunkt überzufällig von Null ab – die Leistung in der Zielvariable ist also bereits zu Beginn der gemessenen Entwicklung erkennbar ausgeprägt. Für die inhaltliche Interpretation ist dieser Effekt in der Regel nicht weiter von Bedeutung (Thomas et al., 2009). Der Slope-Effekt besagt, ob sich die Steigung der Regressionsgeraden zur Vorhersage der AG-Kapazität anhand des Alters signifikant von Null unterscheidet. Dies beantwortet die Frage: Gibt es eine Steigerung der AG-Kapazität mit zunehmendem Alter?

Die Interpretation der Intercept- und Slope-Effekte bei Einsatz von MA-Variablen als Prädiktoren in das Modell anstelle des chronologischen Alters verläuft analog: Ist die Leistung zu Beginn der gemessenen Entwicklung bereits erkennbar ausgeprägt (also von Null verschieden), und gibt es einen signifikanten Anstieg mit zunehmendem mentalem Alter?

Wenn in Schritt (2) der Gruppenfaktor als kategoriale, effekt-kodierte Variable eingeführt wird, gibt das Modell drei Koeffizienten aus: die beiden Haupteffekte für Alter und Gruppe sowie den Interaktionseffekt. Der Haupteffekt für Alter gibt die durchschnittliche Veränderung der AG-Kapazität mit steigendem Alter an; aufgrund der Effekt-Kodierung des Gruppenfaktors ist die Veränderung mit dem Alter über beide Gruppen hinweg gemittelt. Der Haupteffekt für den Faktor Gruppe bezieht sich auf den Unterschied zwischen den Gruppen zum Beginn der gemessenen Entwicklung (also wo die linear-transformierte Altersvariable den Wert „0“ annimmt). Die Interaktion von Gruppe  $\times$  Alter gibt an, ob sich die Regressionsgeraden der beiden



Gruppen voneinander signifikant unterscheiden, d. h., ob die Veränderung in der AG-Kapazität mit zunehmendem Alter in den Gruppen differenziell verläuft.

Mit Aufnahme des Aufgaben-Faktors (beispielhaft hier Wortlänge) in Schritt (3) kommt eine weitere Interaktionsebene hinzu. Inhaltlich steht der Wortlängeneffekt für das Auftreten des Rehearsalprozesses (siehe Abschnitt 2.1.2.2). Der Faktor Wortlänge ist, wie auch der Gruppenfaktor, effekt-kodiert. Das Modell enthält nun insgesamt sieben Effekte: je einen Haupteffekt für Alter, Wortlänge und Gruppe, die Zweifach-Interaktionen von Gruppe  $\times$  Alter, Aufgabe  $\times$  Alter, und Gruppe  $\times$  Aufgabe, sowie schließlich die Dreifach-Interaktion von Gruppe  $\times$  Aufgabe  $\times$  Alter.

Die genannten Haupteffekte und Interaktionen können folgendermaßen interpretiert werden:

1. Der Haupteffekt für (mentales) Alter bezieht sich auf die durchschnittliche Steigung der Regressionsgeraden. Aufgrund der Effekt-Kodierung von Gruppe und Wortlänge gilt die Aussage gemittelt über beide Gruppen und beide Wortlängen-Bedingungen hinweg. Somit sagt der Haupteffekt Alter aus, ob die AG-Kapazität in beiden Gruppen und unabhängig von der Wortlänge insgesamt ansteigt.
2. Der Haupteffekt für Wortlänge bezieht sich auf die Differenz in der Behaltensleistung bei kurzen gegenüber langen Wörtern (d. h. den Wortlängeneffekt), gemittelt über beide Gruppen hinweg, bezogen auf die Schnittstelle der Regressionsgeraden mit der y-Achse. Der Effekt sagt aus, ob es unabhängig von der Gruppe einen Wortlängeneffekt zum Beginn der gemessenen Entwicklung gibt.
3. Der Haupteffekt für Gruppe bezieht sich auf den Unterschied zwischen den Gruppen, gemittelt über beide Wortlängen-Bedingungen hinweg, bezogen auf die Schnittstelle der Regressionsgeraden mit der y-Achse. Der Effekt sagt aus, ob es unabhängig von der Wortlängenbedingung einen Gruppeneffekt zum Beginn der gemessenen Entwicklung gibt.
4. Die Interaktion Gruppe  $\times$  Alter bezieht sich auf den Gruppenunterschied in den durchschnittlichen Steigungen der Regressionsgeraden, gemittelt über beide Wortlängen-Bedingungen hinweg. Somit sagt diese Interaktion aus, ob die AG-Kapazität unabhängig von der Wortlänge mit zunehmendem Alter in den Gruppen unterschiedlich stark ansteigt.
5. Die Interaktion Wortlänge  $\times$  Alter bezieht sich auf den Wortlängen-Unterschied in den durchschnittlichen Steigungen der Regressionsgeraden, gemittelt über beide Gruppen hinweg. Somit sagt diese Interaktion aus, ob die AG-Kapazität unabhängig von der

Gruppenzugehörigkeit mit zunehmendem Alter in den Gruppen je nach Wortlänge unterschiedlich stark ansteigt, bzw. anders ausgedrückt, ob der Wortlängeneffekt mit zunehmendem Alter insgesamt stärker oder schwächer ausgeprägt ist.

6. Die Interaktion Gruppe  $\times$  Wortlänge bezieht sich auf den Gruppenunterschied im Wortlängeneffekt, bezogen auf die Schnittstelle der Regressionsgeraden mit der y-Achse. Der Effekt sagt aus, ob sich die Gruppen in der Stärke ihres Wortlängeneffektes zum Beginn der gemessenen Entwicklung unterscheiden.
7. Die Dreifach-Interaktion Gruppe  $\times$  Wortlänge  $\times$  Alter bezieht sich auf den Gruppenunterschied zwischen den Gruppen in den Steigungsdifferenzen der Regressionsgeraden für lange gegenüber kurzen Wörtern. Diese Interaktion sagt aus, ob die Gruppen hinsichtlich des Wortlängeneffektes unterschiedlich starke Zusammenhänge aufweisen, oder ob die Entwicklung des Wortlängeneffektes in beiden Gruppen parallel verläuft.

Mit den hier angeführten Interpretationen der Haupt- und Interaktionseffekte aus dem dritten Schritt können drei von insgesamt sieben inhaltlichen Szenarien aus Thomas et al. (2009) charakterisiert werden, die sich im Rahmen von DTs ergeben können. Auf diese drei Varianten wird im Folgenden detaillierter eingegangen, um die technischen Interpretationen durch bedeutungshaltige Labels besser greifbar zu machen. Die weiteren vier Szenarien beinhalten u.a. non- und kurvi-lineare Regressionsmodelle, die in dieser Arbeit nicht zur Anwendung kamen und daher nur der Vollständigkeit halber knapp skizziert werden.

### 3.2.4. Inhaltliche Interpretation der Szenarien von Developmental Trajectories

Die drei für diese Arbeit relevanten Szenarien stellen a) ein verspäteter Entwicklungseinsatz (*delayed onset*), b) eine verlangsamte Entwicklungsrate (*slowed rate*) und c) eine Kombination aus beiden Szenarien dar. Diese drei entsprechen den ersten drei Szenarien (a-c) in Abbildung 1.

Das erste Szenario, (a) ein verspäteter Entwicklungseinsatz<sup>17</sup>, ist in Abbildung 1a dargestellt: Hier ist eine Differenz zwischen den beiden Regressionsgeraden der Gruppen zum Zeitpunkt des gemessenen Entwicklungsbeginns (abzulesen an der y-Achse) erkennbar: Die Untersuchungsgruppe weist einen niedrigeren Wert als die Kontrollgruppe an der y-Achse auf. Ob dieser Unterschied signifikant ist, sodass verlässlich von einem delayed onset gesprochen werden kann, gibt der Gruppen-Haupteffekt (Effekt 3) an. Wird zusätzlich eine Aufgaben-Variable (z. B. Wortlänge) mit in das Modell aufgenommen, so informiert die Interaktion von Gruppe  $\times$  Aufgabe (Effekt 6) darüber, ob der Intercept-Unterschied signifikant ist.

---

<sup>17</sup> Genauer und korrekter wäre eigentlich von einem ‚niedrigeren‘ Entwicklungseinsatz zu sprechen, wie sich auch in Abbildung 1 zeigt.

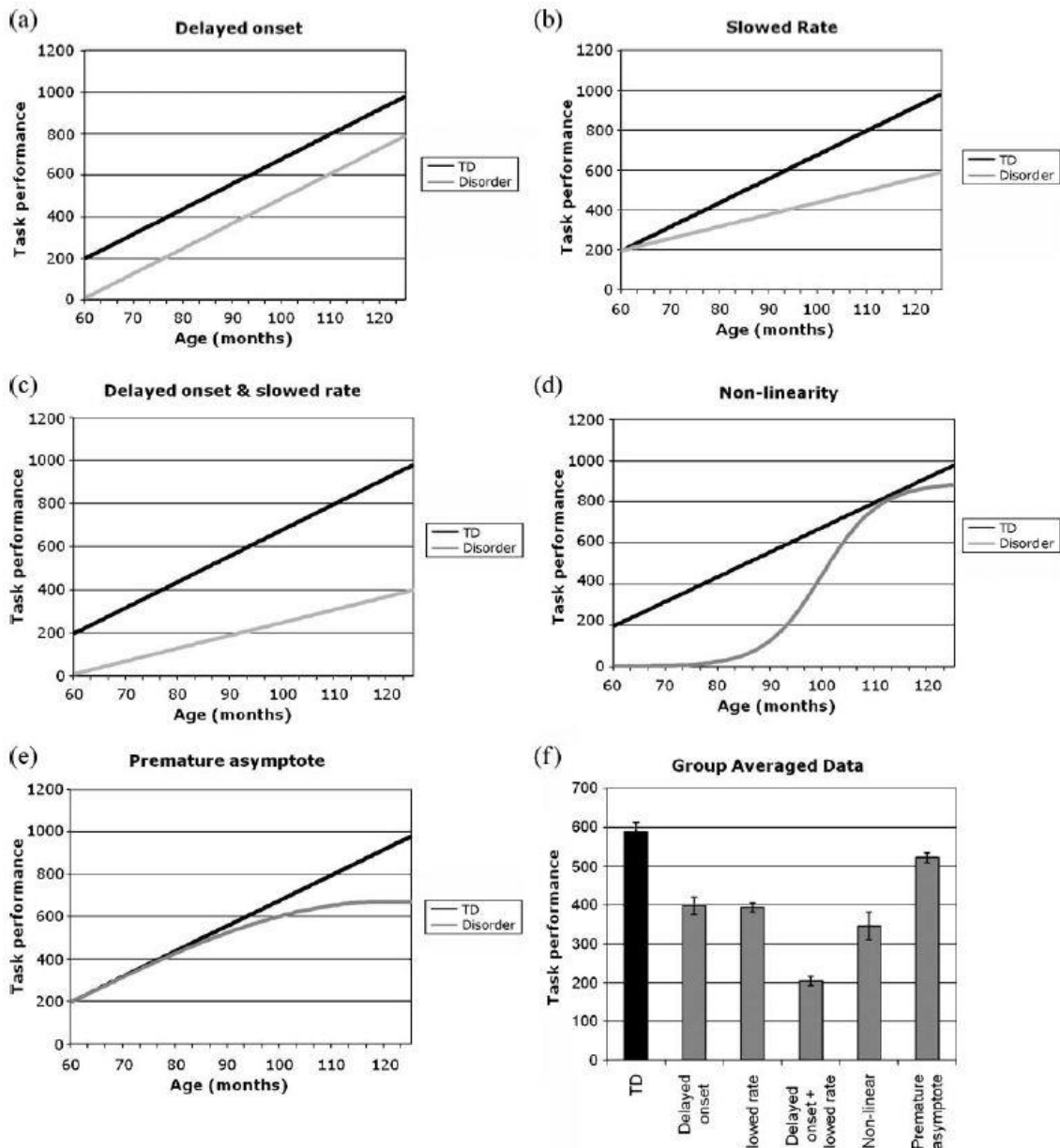


Abbildung 1. Szenarien im Developmental Trajectory Design (Thomas et al., 2009, S. 346)

Das zweite Szenario, (b) eine verlangsamte Entwicklungsrate, ist in Abbildung 1b dargestellt: Hier ist zu sehen, dass die Steigung der Regressionsgeraden für die Untersuchungsgruppe deutlich flacher verläuft als die der Kontrollgruppe. Mit zunehmendem Alter wachsen die Fähigkeiten der Untersuchungsgruppe also weniger stark an als in der Kontrollgruppe. Statistisch zeigt sich dies in der Interaktion von Gruppe  $\times$  Alter (Effekt 4). Bei Berücksichtigung einer zusätzlichen Aufgaben-Variable ist die Information über einen differenziellen Entwicklungsverlauf in der Dreifach-Interaktion von Gruppe  $\times$  Aufgabe  $\times$  Alter (Effekt 7) enthalten.

In Szenario (c) treten beide Merkmale in Kombination auf (siehe Abbildung 1c): ein verspäteter Entwicklungseinsatz gemeinsam mit einer verlangsamten Entwicklungsrate. In diesem Fall sind sowohl die genannten Intercept- als auch Slope-Effekte signifikant: Effekte (3) und (4),

wenn keine Aufgaben-Variable mit in das Modell aufgenommen wird, und Effekte (6) und (7), wenn dies der Fall ist.

Alle drei hier angeführten Szenarien stellen Formen von *developmental delay* dar (Thomas et al., 2009). Entwicklungsmuster, die eher einer *deviance*-Position entsprechen, lassen sich in komplexeren Szenarien finden (siehe Abbildung 1d/e), die im folgenden Abschnitt skizziert werden, jedoch im Rahmen dieser Arbeit nicht zur Anwendung kamen. Zur Identifikation eines *deviance*-Entwicklungsmusters können auch Befunde zu den MA-Prädiktoren herangezogen werden, wenn diese differenzielle Ergebnisse (d. h. Szenarien a-c) hervorrufen. Hierauf wird im darauffolgenden Abschnitt näher eingegangen.

Als komplexere Szenarien seien knapp zusammengefasst zwei non- bzw. kurvilineare Verläufe genannt, wenn z. B. in der Untersuchungsgruppe im Gegensatz zur Kontrollgruppe eine S-förmige logistische Funktion die Daten besser beschreibt oder ein vorzeitiges Abflachen der Entwicklungskurve anhand eines quadratischen Terms zu beobachten ist (siehe Abbildung 1d und Abbildung 1e). Außerdem können zwei Szenarien mit horizontal verlaufenden Regressionsgeraden, und dementsprechend zunächst nicht-signifikanter Varianzaufklärung, anhand eines Rotationsverfahrens unterschieden werden: wenn *no systematic relationship* vorliegt, zeigt sich auch nach Rotation um  $45^\circ$  kein reliabler Zusammenhang. Ist dagegen nach Rotation ein Zusammenhang zu beobachten, wird dies als *zero trajectory* bezeichnet, was bedeutet, dass sich die AG-Kapazität tatsächlich nicht mit zunehmendem Alter verändert (Thomas et al., 2009). Diese komplexeren Szenarien von non- bzw. kurvi-linearen und horizontalen Verläufen in der Untersuchungsgruppe würden nach der Interpretation von Thomas et al. (2009) für eine *deviance*- oder *atypicality*-Position sprechen.

Mit Betrachtung von differenziellen Befunden für verschiedene MA-Variablen lassen sich ebenfalls empirische Hinweise auf ein *deviance*-Entwicklungsmuster finden. Für die Schätzung des mentalen Alters können verschiedene Variablen flexibel herangezogen werden (Thomas et al., 2009). Eher allgemein wäre dies z. B. die kognitive Kapazität, aber auch für spezifischere Variablen, wie das Leselevel bei Kindern mit Dyslexie (vgl. Backman et al., 1984), kann kontrolliert werden. Im klassischen Gruppenmatching-Ansatz müsste zu diesem Zweck für jede MA-Variable eine eigene Kontrollgruppe erhoben werden. Dagegen ist beim DT-Ansatz noch im Zuge der Auswertung deutlich mehr Flexibilität möglich (Thomas et al., 2009), denn unter der Voraussetzung, dass die Spannweite der MA-Variable in der Kontrollgruppe diejenige der Untersuchungsgruppe umfasst und dass keine Boden- oder Deckeneffekte vorliegen, kann jede theoretisch aussagekräftige Variable als Prädiktor in das DT-Modell eingesetzt werden.

In den auf diese Weise untersuchbaren vielfältigen „developmental relations“ (Thomas et al., 2009, S. 341) liegt eine weitere Stärke des DT-Ansatzes. Es lässt sich damit überprüfen, ob die

Gruppen hinsichtlich der Beziehung zwischen einer MA-Variable und der Ziel-Aufgabe Gruppen vergleichbar sind, oder ob die Gruppen ein differenzielles Zusammenhangsmuster aufweisen. Wenn solche differenziellen Muster im Zusammenhang mit verschiedenen MA-Variablen beobachtet werden können, werden diese ebenfalls als atypische Entwicklungsverläufe im Sinne einer deviance-Position interpretiert (Thomas et al., 2009). Begründen lässt sich das damit, dass in diesem Fall anscheinend verschiedene Bereiche oder Domänen der (allgemeinen oder kognitiven) Entwicklung unterschiedlich stark ausgeprägt sind und mit der Ziel-Aufgabe in Zusammenhang stehen, was sich nicht mit einer allgemein verzögerten Entwicklung erklären lässt.

Um den vorangegangenen Abschnitt zusammenzufassen, wurde gezeigt, dass mit dem DT-Ansatz eine differenziertere Möglichkeit zur Beschreibung von Entwicklungsmustern zur Verfügung steht, die sich von einer „monolithic descriptive partition between ‚delay‘ and ‚deviance““ (Thomas et al., 2009, S. 345) des traditionellen Gruppenmatching-Verfahrens abhebt. Mit DTs können drei Varianten von delay unterschieden werden, von denen die beiden Szenarien delayed onset und slowed rate mittels des Gruppenmatching-Ansatzes aufgrund ähnlicher resultierender Mittelwerte nicht differenzierbar wären. Deviance-Entwicklungsmuster lassen sich anhand differenzieller Zusammenhänge bei der Betrachtung weiterer MA-Variablen als Prädiktoren aufdecken. Durch die Verwendung eines breiteren Altersspektrums wird zusätzlich zum Mittelwert der informative Gehalt der Varianz innerhalb der Gruppen genutzt, anstatt sie künstlich zu minimieren.

### 3.3. Fazit – Mentaler Altersvergleich

Mit dem oben beschriebenen Design, das neben dem chronologischen Alter auch das mentale Alter berücksichtigt, kann somit die Fragestellung nach differenziellen Entwicklungsmustern in Arbeits- und Langzeitgedächtnis für Kinder mit Lernbehinderung beantwortet werden. In diesem Kapitel wurde gezeigt, dass die Berücksichtigung des mentalen Alters notwendig ist, um verschiedene Entwicklungsmuster differenzieren zu können. Es wurden zwei Methoden eingeführt, die im Rahmen dieser Arbeit zur Anwendung kommen: Developmental Trajectories (in Studie 1) und das klassische Gruppenmatching (in Studie 2).

Durch Berücksichtigung der MA-Variablen wird geprüft, ob die betrachteten kognitiven Komponenten (Arbeitsgedächtnis-Kapazität und -Prozesse sowie Langzeitgedächtnis-Struktur) überhaupt beeinträchtigt sind, und wenn ja, in welchem Maße. Liegt die Beeinträchtigung im Rahmen des allgemein niedrigeren mentalen Alters, bzw. lassen sich in den DTs vorrangig delay-Szenarien finden? Oder geht die Beeinträchtigung über die allgemein verzögerte Entwicklung hinaus, was sich in einem Unterschied zur MA-Gruppe, bzw. in deviance-Szenarien der DTs

zeigt, welche sich insbesondere in Form differenzieller Muster bei verschiedenen MA-Variablen darstellen könnten?

Die Variablen zur Schätzung des mentalen Alters werden im weiteren Verlauf der Arbeit als ‚Entwicklungsindikatoren‘ (EI) bezeichnet, da sie je verschiedene Aspekte der (allgemeinen und kognitiven) Entwicklung abbilden sollen. Im Fall der vorliegenden Arbeit werden die folgenden drei Entwicklungsindikatoren als Vergleichsmaßstab zwischen den Gruppen herangezogen: (a) das chronologische Alter in Monaten; (b) die kognitive Kapazität, gemessen an den erzielten Rohwert-Punkten im CFT 1-R, womit die absolute Fähigkeit zur Lösung von Aufgaben zu kognitiven Operationen erfasst wird; und (c) der Wortschatzumfang, gemessen an den erzielten Rohwert-Punkten im WWT 6-10, womit die absolute Fähigkeit zum Abruf aus dem mentalen Lexikon erfasst wird.

Es sei an dieser Stelle noch einmal betont, dass die vorliegenden Daten querschnittlich, also zu einem Zeitpunkt über eine gewisse Altersspanne bei unterschiedlichen Personen, erhoben wurden. Daher können aus diesen Daten keine (kausalen) Aussagen über individuelle Entwicklungsverläufe abgeleitet werden. Thomas et al. (2009) verweisen aus diesem Grund darauf, dass DTs mit querschnittlich erhobenen Daten als ein erster Schritt zu verstehen sind, die zwar dem klassischen Gruppenmatching überlegen sind, denen aber eine Validierung durch längsschnittliche Daten folgen sollte (Willett, Singer & Martin, 1998; siehe z. B. Chapman et al., 2002; Isordia & Ferrer, 2016; Sullivan et al., 2016).

Insofern ist auch in Bezug auf die vorliegende Arbeit explizit einschränkend darauf hinzuweisen, dass wegen der rein querschnittlichen Erhebung der zugrundeliegenden Daten die Aussagen allein korrelativ als Zusammenhänge zwischen der Ziel-Aufgabe, d. h. den Arbeitsgedächtnismaßen und dem chronologischen bzw. mentalen Alter zu verstehen sind. Entsprechende längsschnittliche Datenerhebungen wären Aufgabe zukünftiger Forschungsprojekte.

#### 4. Fragestellungen der Arbeit und ihrer Teilstudien

Nachdem in den Kapiteln des Theorieteils die Zielgruppe bestimmt (Schüler\*innen mit Lernbehinderung, Kapitel 1), die inhaltliche Fokussierung auf die Teilkomponenten von Arbeitsgedächtnis-Kapazität und -Prozessen sowie der Langzeitgedächtnis-Struktur (Kapitel 2) eingegrenzt, und schließlich die Rationale des mentalen Altersvergleichs im Gruppenmatching und Developmental Trajectory-Ansatz zur Identifikation von differenziellen Entwicklungsmustern (Kapitel 3) vorgestellt wurden, soll nun darauf aufbauend die Fragestellung der Arbeit abgeleitet werden.

Die vorliegende Arbeit hat zum Ziel, einen Beitrag zum Verständnis des Phänomens „Lernbehinderung“ zu leisten. Sie soll empirische Erkenntnisse dazu liefern, inwiefern und in welchem Maße Kinder mit Lernbehinderung auf Gedächtnisstrukturen und -prozesse, die für das schulische Lernen zentral sind, zurückgreifen können. Sollten Kinder mit Lernbehinderung in einzelnen oder allen Aspekten Defizite aufweisen, so ist es von Interesse, weiter differenzieren zu können, ob diese Schwierigkeiten im Rahmen ihres allgemeinen mentalen Entwicklungsrückstandes liegen oder aber eine spezifische Problematik darstellen, die über eine Entwicklungsverzögerung hinausgeht. Den Vergleichsmaßstab hierfür bietet die jüngere, nach mentalem Alter gematchte Kontrollgruppe bzw. die auftretenden differenziellen Muster im Rahmen der DT-Analysen. Die übergreifende Fragestellung nach der Verfügbarkeit bzw. den Entwicklungsmustern kognitiver Lernvoraussetzungen wird im Rahmen der vorliegenden Arbeit in drei Teilstudien beantwortet, entsprechende Hypothesen finden sich in den Abschnitten zur Fragestellung der jeweiligen Teilstudie.

Die erste Teilstudie (Kapitel 5) untersucht das Arbeitsgedächtnis als eine der zentralen Komponenten im Wissenserwerb (Klauer & Leutner, 2012). Mithilfe des Developmental Trajectory-Ansatzes werden Kinder mit und ohne Lernbehinderung hinsichtlich ihrer phonologischen Arbeitsgedächtniskapazität verglichen. Außerdem wird die Effektivität von zwei Prozessen zur Aufrechterhaltung von Gedächtnisspuren im Arbeitsgedächtnis untersucht: Rehearsal- und Redintegrationprozess.

Operationalisiert wird die AG-Kapazität anhand der durchschnittlichen Merkleistung in Wortspannen-Aufgaben, die im Format des *immediate serial recall* dargeboten werden. Für den Rehearsal-Prozess wird die Größe des Wortlängeneffektes in den Wortspannen-Aufgaben herangezogen, was dem durchschnittlichen Vorteil in der Merkleistung für kurze gegenüber langen Wörtern entspricht. Analog wird für den Redintegrationprozess die Größe des Lexikalitätseffektes in den Wortspannen-Aufgaben herangezogen, was dem durchschnittlichen Vorteil in der Merkleistung für echte Wörter gegenüber Kunstwörtern entspricht. Die Gruppen werden an-

hand von insgesamt drei Entwicklungsindikatoren zur Einschätzung des (mental)en Alters verglichen: a) dem chronologischen Alter, b) der kognitiven Kapazität, gemessen am Rohwert des CFT 1-R (Weiß & Osterland, 2012), und c) dem Wortschatzumfang, gemessen am Rohwert des WWT 6-10 (Glück, 2011).

In der ersten Studie wird anhand des Ansatzes der Developmental Trajectories die Fragestellung untersucht, ob Kinder mit Lernbehinderung im Vergleich zu unbeeinträchtigten Kindern in ihrer AG-Kapazität, sowie der Rehearsal- und Redintegration-Effektivität ähnliche oder aber differenzielle Muster im Zusammenhang mit den drei oben genannten Entwicklungsindikatoren aufweisen.

Die zweite Teilstudie (Kapitel 6) vergleicht Kinder mit und ohne Lernbehinderung hinsichtlich der Struktur beim Abruf aus dem mentalen Lexikon als Teilaspekt des Langzeitgedächtnisses, welches als weitere zentrale Komponente beim Wissenserwerb betrachtet werden kann. Dazu werden die bereits im Rahmen der ersten Studie präsentierten Daten zum Wortschatzumfang (WWT 6-10) genauer betrachtet.

Im ersten Schritt wird mithilfe einer Multigroup-CFA die Messäquivalenz des Instrumentes zwischen den Gruppen untersucht. Hierbei steht weniger die psychometrische Beurteilung des Instrumentes im Vordergrund, sondern vielmehr die inhaltliche Frage nach der Vergleichbarkeit der repräsentierten Faktoren- und Ladungsstruktur zwischen den Gruppen. Damit wird die Fragestellung beantwortet, ob Muster von latenten Zusammenhängen zwischen Wörtern (in diesem Fall basierend auf der Zugehörigkeit zu verschiedenen Wortarten) für beide Gruppen gleich sind oder voneinander abweichen. Die zweite Fragestellung gibt Aufschluss darüber, durch welche personen- und itembezogenen Faktoren die Lösungswahrscheinlichkeit für die korrekte Benennung der gezeigten Items erklärt werden kann, und ob bzw. inwiefern die Gruppen sich hierin systematisch unterscheiden.

Bei der dritten Teilstudie (Kapitel 7) handelt es sich um eine Reanalyse, mit der ein Effekt aus der ersten Studie gegen das methodische Artefakt des Differential Item Functioning (DIF) abgesichert werden soll. Vorgreifend sei bemerkt, dass in der ersten Studie eine überraschende Interaktion zwischen dem Wortschatzumfang und der Redintegration-Effektivität bei Kindern mit Lernbehinderung nachgewiesen werden konnte, die sich in der Kontrollgruppe nicht feststellen ließ. Auf der Suche nach möglichen Erklärungen (Bruns & Grosche, 2018, Juni) wurde auch die Möglichkeit in Betracht gezogen, dass ein systematisches differenzielles Funktionieren der WWT-Items zwischen den Gruppen diese Interaktion hervorrufen könnte. Um diese Erklärung als mögliche Ursache auszuschließen, wird in der dritten Studie dieser Arbeit zunächst die Fragestellung untersucht, ob Items mit DIF vorlagen. Wenn dies der Fall ist, schließt sich die Fragestellung an, ob die in Studie 1 gefundene Interaktion auch erhalten bleibt, wenn anstelle



des vollständigen WWT-Summenscores ein um die DIF-Items bereinigter WWT-Score als Entwicklungsindikator genutzt wird.

Insgesamt haben die Ergebnisse der drei Teilstudien sowohl theoretische als auch praktische Relevanz. Das Wissen über die Verfügbarkeit bzw. Vergleichbarkeit von kognitiven Voraussetzungen kann auf theoretischer Ebene dazu genutzt werden, die Konzeption und Definition des Konstruktes Lernbehinderung zu schärfen. Auf praxisbezogener Ebene können daraus Implikationen abgeleitet werden, inwiefern es angezeigt ist, z. B. Aufgabenstellungen und Instruktionen anzupassen oder zusätzliche Fördermaßnahmen zu entwickeln.

## 5. Arbeitsgedächtnis-Kapazität und -Prozesse (Studie 1)

Die erste empirische Studie befasst sich mit der allgemeinen Kapazität und zwei Prozessen zur Aufrechterhaltung von Gedächtnisspuren im verbalen Arbeitsgedächtnis (nämlich Rehearsal und Redintegration) bei Schüler\*innen mit Lernbehinderung. Da die Studie international publiziert wurde, wurde der Begriff *mild and borderline intellectual disabilities* gewählt, der am ehesten dem deutschen Konstrukt der Lernbehinderung entspricht (siehe Kapitel 1; Poloczek et al., 2016; van der Molen et al., 2007). Mithilfe der in Kapitel 3.2 beschriebenen Methode der Developmental Trajectories (DT) wurde geprüft, inwiefern die Muster der (wohlgemerkt querschnittlich erhobenen) Entwicklung zwischen Kindern mit Lernbehinderung und einer unbeeinträchtigten Kontrollgruppe voneinander abwichen. Da der Beitrag in englischer Sprache erschienen ist, wird in diesem Kapitel eine Übersetzung der relevanten Abschnitte angeführt; der vollständige englische Artikel ist dem Anhang beigelegt. Im Folgenden gibt die Übersetzung des Abstracts eine Zusammenfassung. Der Theorieteil des publizierten Beitrags ist bereits in ausführlicher Form in den Kapiteln 1 und 2 dieser Arbeit eingeflossen und dort entsprechend kenntlich gemacht, und wird daher hier nicht wiederholt. Die Fragestellung der Studie, sowie Methode, Ergebnisse und Diskussion sind ebenfalls Übersetzungen des Beitrags. Ergänzungen sind in eckigen Klammern eingefügt; Informationen zur Beschreibung der deutschen Schullandschaft für den internationalen Beitrag werden hier ausgelassen.

### 5.1. Zusammenfassung von Studie 1: Abstract<sup>18</sup>

Im Arbeitsgedächtnis dienen [mindestens] zwei Prozesse dazu, eine verblässende Gedächtnisspur aufrechtzuerhalten: der subvokale Rehearsalprozess und der lexikalische Redintegrationsprozess. Während aktuelle Studien zu Rehearsal bei Schüler\*innen mit Lernbehinderung gemischte Befunde hervorgebracht haben, wurde Redintegration bislang für diese Schüler\*innengruppe nicht untersucht. Hinzu kommt, dass die meisten bisherigen Studien ein Matching der Untersuchungs- und Kontrollgruppe auf Gruppenebene durchführen; aufgrund dieser methodischen Beschränkung können nur zwei verschiedene Entwicklungsmuster [nämlich Entwicklungsverzögerung und Entwicklungsstörung] unterschieden werden. Daher wurden in dieser Studie beide Arbeitsgedächtnis-Prozesse, nämlich Rehearsal und Redintegration, bei Schüler\*innen mit Lernbehinderung untersucht und mit der Methode der Developmental

---

<sup>18</sup> Der folgende Abschnitt ist eine Übersetzung des Abstracts (S. 1) aus Bruns G., Ehl B. & Grosche M. (2019). Verbal Working Memory Processes in Students With Mild and Borderline Intellectual Disabilities: Differential Developmental Trajectories for Rehearsal and Redintegration. *Frontiers in Psychol.* 9:2581. doi:10.3389/fpsyg.2018.02581

Trajectories ausgewertet, denn diese bietet mehr Möglichkeiten zur Identifizierung von differenziellen Entwicklungsmustern als die herkömmlichen Ansätze mit Matching auf Gruppenebene.

Wir untersuchen, ob sich bei Schüler\*innen mit Lernbehinderung im Vergleich zu unbeeinträchtigten (*typically developing*) Schüler\*innen eine abweichende Entwicklung hinsichtlich dreier Aspekte des Arbeitsgedächtnisses zeigen lässt: (a) die generelle Kapazität der phonologischen Schleife, sowie die Effektivität von (b) Rehearsal und (c) Redintegration. Es kamen drei verschiedene Entwicklungsindikatoren (*developmental indicators*) zum Einsatz, anhand derer die Trajectories der Gruppen verglichen wurden: das chronologische Alter, die kognitive Kapazität und der Wortschatzumfang.

Insgesamt  $N = 210$  Schüler\*innen (87 Schüler\*innen mit Lernbehinderung, 123 unbeeinträchtigte Schüler\*innen) bearbeiteten Aufgaben zu Arbeitsgedächtnis-Wortspannen mit je kurzen und langen (einsilbigen vs. dreisilbigen) echten Wörtern und Kunstwörtern [sodass sich insgesamt vier Wortbedingungen ergaben: echt-kurz, echt-lang, kunst-kurz, kunst-lang]. Der Wortlängeneffekt (kurze vs. lange Wörter) war das Maß für den Rehearsalprozess, und der Lexikalitätseffekt (echte Wörter vs. Kunstwörter) maß den Redintegrationprozess.

Die Ergebnisse zeigen, dass die Developmental Trajectories bei Rehearsal einen Unterschied im Intercept aufwiesen, jedoch keine verlangsamte Entwicklungsrate, während bei Redintegration keine Beeinträchtigung vorzuliegen schien. Jedoch zeigte sich bei Schüler\*innen mit Lernbehinderung hinsichtlich der entwicklungsbezogenen Relation (*developmental relation*) zwischen Wortschatzumfang und Redintegration ein differenzielles Muster: Die Effektivität von Redintegration bei Schüler\*innen mit geringem Wortschatzumfang war höher, reduzierte sich aber entgegen der Erwartung mit zunehmendem Wortschatz.

Wir ziehen den Schluss, dass Schüler\*innen mit Lernbehinderung einen verspäteten Einsatz (*delayed onset*) in der Entwicklung der PL-Kapazität und im Rehearsal aufweisen, und dass sie im Lauf der Entwicklung diesen Rückstand auch nicht gänzlich aufholen. Der Redintegrationprozess erscheint im Verhältnis zum chronologischen Alter und der kognitiven Kapazität nicht beeinträchtigt zu sein, jedoch besteht angesichts der differenziellen Relation zwischen Redintegration und Wortschatzumfang weiterer Forschungsbedarf. Während also möglicherweise die geringere Effektivität des subvokalen Rehearsalprozesses mit den entwicklungsbezogenen schulischen Problemen von Schüler\*innen mit Lernbehinderung in Verbindung steht, erweist sich der lexikalische Redintegrationprozess [auf den ersten Blick] als intakt (im Verhältnis zum chronologischen Alter und kognitiver Kapazität) und könnte damit eine mögliche Stärke und Resource für diese Schüler\*innen darstellen.

## 5.2. Forschungsfrage und Hypothesen<sup>19</sup>

Die [erste] Studie hat zum Ziel, die folgende Forschungsfrage zu beantworten: „Entwickelt sich das verbale Arbeitsgedächtnis bei Schüler\*innen mit Lernbehinderung anders als bei unbeeinträchtigten Schüler\*innen?“. Wir untersuchen, ob drei verschiedene Entwicklungsindikatoren (chronologisches Alter, kognitive Kapazität und Wortschatzumfang) die Kapazität der PL sowie die Effektivität von Rehearsal- und Redintegrationprozessen für Schüler\*innen mit Lernbehinderung im Vergleich zu unbeeinträchtigten Schüler\*innen vorhersagen können. Da Schüler\*innen mit Lernbehinderung per definitionem eine reduzierte kognitive Kapazität aufweisen, ist es notwendig, [diese mitzuerheben, um statistisch] dafür zu kontrollieren. Der Wortschatzumfang wird als Indikator für kristallines Wissen, das im Langzeitgedächtnis gespeichert ist, herangezogen. Da Redintegration auf der Interaktion von Arbeits- und Langzeitgedächtnis beruht, kann durch die Erfassung des Wortschatzumfangs die entwicklungsbezogene Relation zwischen Wortschatz und Redintegration untersucht werden.

- PL Kapazität: Wir erwarten entweder einen verspäteten Einsatz oder eine verlangsamte Entwicklungsrate der PL Kapazität bei Schüler\*innen mit Lernbehinderung, denn aus der Literatur sind Hinweise bekannt, dass Defizite im Arbeitsgedächtnis mit dem mentalen Alter einhergehen. Wie jedoch oben [Kapitel 3] angeführt, sind die Szenarien [verspäteter Entwicklungseinsatz vs. verlangsamte Entwicklungsrate] mit den herkömmlichen ANOVA-Ansätzen nicht unterscheidbar.
- Rehearsalprozess: Auch für die Effektivität des Rehearsalprozesses wird angenommen, dass diese bei Schüler\*innen mit Lernbehinderung beeinträchtigt ist. Da jedoch die bisherige Befundlage gemischt ist, kann keine klare Hypothese gebildet werden, ob ein verspäteter Einsatz, eine verlangsamte Entwicklungsrate oder beides anzunehmen ist. Hinsichtlich der entwicklungsbezogenen Relation mit dem Wortschatzumfang als lexikalische Variable erwarten wir keinen Zusammenhang mit dem phonologisch basierten Rehearsalprozess.
- Redintegrationprozess: Zur Effektivität des Redintegrationprozesses sind uns keine Studien über Schüler\*innen mit Lernbehinderung bekannt, aus denen sich theoretische Annahmen ableiten ließen. Da Redintegration bedeutet, dass Informationen aus dem Langzeitgedächtnis genutzt werden, um Items im Kurzzeit-Speicher zu inferenzieren, könnten [Defizite im Redintegrationprozess] durchaus eine Quelle von kognitiver Beeinträchtigung bei Schüler\*innen mit Lernbehinderung darstellen. Wenn dies der Fall

---

<sup>19</sup> Der folgende Abschnitt ist eine Übersetzung des Abschnitts „Research Questions and Hypotheses“ (S. 4) aus Bruns, Ehl & Grosche (2019).

ist, dann wäre eine reduzierte Effektivität von Redintegration ein differenzieller Befund, der eine Besonderheit bei Lernbehinderung darstellen könnte. Hier erwarten wir, einen Zusammenhang zwischen dem Wortschatzumfang und dem Redintegrationprozess zu finden.

### 5.3. Methode<sup>20</sup>

#### 5.3.1. Stichprobe

Die Stichprobe umfasste  $N = 210$  deutschsprachige Schüler\*innen: 87 gehörten zur Gruppe der Schüler\*innen mit Lernbehinderung (unterteilt in *mild intellectual disability* bei einem IQ von 55-70 mit 24 Schüler\*innen und *borderline intellectual functioning* bei einem IQ von 70-85 mit 63 Schüler\*innen), von denen 48 männlich waren und deren durchschnittliches Alter  $M = 12$  Jahre und 11 Monate betrug, mit einer Standardabweichung von  $SD = 2.39$  und einem Range von 7;4-17;1 Jahren. Zur unbeeinträchtigten Kontrollgruppe gehörten  $n = 123$  Schüler\*innen (von denen 51 männlich waren; das durchschnittliche Alter lag bei  $M = 8$  Jahren und 5 Monaten,  $SD = 1.87$ , Range 6;0-13;5). Die IQ Normdaten wurden aus unterschiedlichen Quellen ermittelt: Die Werte der Schüler\*innen mit Lernbehinderung wurden den diagnostischen Informationen in den Schulakten entnommen. Die IQ-Werte der Kontrollgruppe wurden mit einem altersangemessenen Culture Fair Test Verfahren erhoben (CFT 1-R, Weiß & Osterland, 2012 für Schüler\*innen der Primarstufe; und CFT 20-R, Weiß, 2008 für Schüler\*innen der Sekundarstufe). Die detaillierte Stichprobenbeschreibung findet sich in Tabelle 1. Für die Vergleiche von kognitiver Kapazität und Wortschatzumfang auf Rohwert-Ebene wurden alle Schüler\*innen mit Lernbehinderung getestet sowie diejenigen Schüler\*innen der Kontrollgruppe, die in den normierten Altersbereich der unten angeführten Tests fielen. Daraus ergab sich für die Vergleiche auf Basis des mentalen Alters eine Teilstichprobe von  $n = 102$  Kindern in der unbeeinträchtigten Kontrollgruppe.

Um in die Untersuchungsgruppe (Schüler\*innen mit Lernbehinderung bzw. mild and borderline intellectual disabilities) aufgenommen zu werden, mussten die folgenden Einschlusskriterien erfüllt sein: eine formale Diagnose eines sonderpädagogischen Förderbedarfs (im Förderschwerpunkt Lernen); ein IQ-Wert unter 85, wie er aus den Daten des formalen Feststellungsverfahrens des sonderpädagogischen Förderbedarfs ermittelt wurde; Ausschluss anderer [umschriebener] Entwicklungsstörungen wie ADHS, Autismus, oder spezifische Lernstörungen (LRS, Dyskalkulie) gemäß den Aussagen der Lehrkräfte. Alle Schüler\*innen mit Lernbehinderung wurden an insgesamt vier Förderschulen mit dem Förderschwerpunkt Lernen in einem

---

<sup>20</sup> Der folgende Abschnitt ist eine Übersetzung des Abschnitts „Methods“ (S. 4-7) aus Bruns, Ehl & Grosche (2019). Ebenfalls übernommen und übersetzt wurden Tabelle 1 und Tabelle 2.

städtischen Umfeld im Westen Deutschlands getestet. [...] Die Schüler\*innen der unbeeinträchtigten Kontrollgruppe besuchten eine Regelschule (Primar- oder Sekundarstufe), es lag keine Diagnose eines sonderpädagogischen Förderbedarfs oder einer Entwicklungsstörung vor, und die IQ-Werte mussten mindestens durchschnittlich (d. h. IQ > 85; Wortschatz T-Wert > 40) sein. [...] Wenn Teilnehmer\*innen den Instruktionen [erkennbar] nicht folgen konnten, wurden sie von der Erhebung ausgeschlossen. Es wurde nicht erfasst, ob die Schüler\*innen monolingual waren; jedoch wies gemäß der Einschätzung der Versuchsleiter\*innen keines der teilnehmenden Kinder Schwierigkeiten im Verständnis der Instruktionen auf.

Die Studie wurde unter Beachtung der Empfehlungen der Deutschen Gesellschaft für Psychologie e.V. durchgeführt. Der Erklärung von Helsinki folgend, wurde eine schriftliche Einverständniserklärung von den Erziehungsberechtigten der teilnehmenden Kinder eingeholt. Alle teilnehmenden Schüler\*innen gaben unmittelbar vor der Testung ihre mündliche Einwilligung. Außerdem wurden Eltern oder Erziehungsberechtigte über die Ziele und die Art der durchgeführten Tests in der Studie informiert, sowie darüber, dass die Einwilligung jederzeit widerrufen werden könne. Von den jeweiligen Schulleitungen wurde außerdem die Genehmigung eingeholt, die Testungen an den Schulen durchzuführen.

### **5.3.2. Sitzungen und Material**

Alle Schüler\*innen absolvierten den Test zur kognitiven Kapazität in einer Gruppensitzung. Die Experimentalaufgaben [zum Arbeitsgedächtnis und Wortschatz] bearbeiteten die Schüler\*innen in zwei einzelnen Sitzungen von jeweils ca. 30 Minuten. In der Kontrollgruppe umfasste die erste Sitzung die Arbeitsgedächtnis-Aufgaben. Die zweite Sitzung beinhaltete ein Bildbenennungs-Verfahren zur Messung des Wortschatzumfangs. Für die Untersuchungsgruppe (Schüler\*innen mit Lernbehinderung) wurde die Reihenfolge leicht modifiziert, um die Konzentration und Mitarbeitsbereitschaft zu erhöhen: Die erste Sitzung umfasste nur eine Arbeitsgedächtnis-Aufgabe, die zweite Sitzung eine weitere Arbeitsgedächtnis-Aufgabe und den Wortschatztest. Die älteren Schüler\*innen der Kontrollgruppe (der Sekundarstufe) bearbeiteten einen Wortschatztest im Gruppenverfahren.

#### **5.3.2.1. Wortspannen-Aufgaben**

Es kamen vier verschiedene Wortspannen-Aufgaben zum Einsatz, unter je 2 (Länge: kurz vs. lang) × 2 (Lexikalität: echte Wörter vs. Kunstwörter) Bedingungen. Beispielsweise lautete eine Wortspanne für kurze echte Wörter „Haus – Stern – Schuh“, oder für lange Kunstwörter „karflumen – franulich – wuralten“. Für die Bedingung mit echten Wörtern wurden die Stimuli aus der standardisierten deutschen Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren verwendet (AGTB 5-12, Hasselhorn et al., 2012). Die Kunstwörter wurden aus einer Studie von

Hasselhorn et al. (2010, siehe Tabelle A-1 im Anhang) entnommen. Die Konstruktion der Wortspannen-Sequenzen erfolgte nach den Prinzipien der AGTB 5-12. Die Sequenzen variierten in ihrer Länge von 2 bis 8 Wörtern und wurden getaktet (alle 1.5 Sekunden ein Wort) präsentiert. Jede Sequenz war zufällig aus einem Pool von je neun Wörtern zusammengesetzt, die von einer weiblichen Stimme eingesprochen wurden. Bei der Zusammensetzung wurde darauf geachtet, dass die Reihenfolge der Wörter innerhalb einer Sequenz nicht zu ähnlich zu der anderer Sequenzen war. Für jede Sequenzlänge (d. h. die Anzahl der Wörter pro Sequenz) wurden 11 verschiedene Sequenzen mit pseudorandomisierter Wortreihenfolge in einer Liste von Audiodateien auf einem PC zur Verfügung gestellt, von welchem sie in der Testung abgespielt wurden. Die Anzahl der Wörter pro Sequenz wurde adaptiv ausgewählt, basierend auf der bisherigen Leistung des Kindes (das genaue Vorgehen wird im nächsten Absatz beschrieben). Den Schüler\*innen wurde gesagt, dass sie die vollständige Sequenz bis zu einem Signal am Schluss anhören und im Anschluss die gesamte Sequenz so korrekt wie möglich wiedergeben sollten. Dieses Vorgehen wurde in drei Übungsdurchläufen vor den eigentlichen Aufgaben geprobt. Die Reihenfolge der Lexikalität (echte Wörter oder Kunstwörter) wurde zufällig zugewiesen, dagegen war die Reihenfolge der Bedingung Wortlänge für alle Personen identisch: Die kurzen Wörter wurden zuerst präsentiert, im Anschluss die langen Wörter. Zu den Kunstwörtern wurde den Kindern gesagt, dass diese Wörter ungewöhnlich klingen würden, „wie in einer geheimen Sprache von Aliens“.

Die Wortspannen-Aufgaben waren adaptiv. [Das bedeutet, die Länge der Sequenzen richtete sich nach der Leistung des Kindes in den Sequenzen davor.] In zwei Kalibrierungstests am Anfang einer jeden Bedingung wurde die Länge der Sequenz nach jedem Item angepasst. Wenn das Kind das Item korrekt wiedergegeben hatte, wurde die Länge des darauffolgenden Items um ein Wort erhöht; wenn es nicht korrekt war, wurde die Länge um ein Wort reduziert. Die minimale Länge einer Sequenz betrug zwei Wörter. Die auf die Kalibrierungstests folgenden acht Testitems wurden zur Berechnung des Wortspannen-Wertes herangezogen, und die Anpassung der Sequenzlänge erfolgte nach jedem zweiten Item: Wenn [z. B.] die Sequenzen 3 und 4 beide korrekt wiedergegeben wurden, waren die darauffolgenden zwei Sequenzen um ein Wort länger; wenn beide falsch wiedergegeben wurden, waren die nächsten beiden Sequenzen um ein Wort kürzer; wenn eine Sequenz korrekt und eine falsch wiedergegeben wurde, blieben die nächsten beiden Sequenzen bei der gleichen Länge wie zuvor. Die Punktzahl für die korrekte Wiedergabe einer Sequenz betrug die Anzahl der Wörter in der Sequenz. Bei einer falschen Wiedergabe wurde von der Sequenzlänge ein Punkt abgezogen (z. B. führte die fehlerhafte Wiedergabe einer 4-Wort Sequenz zu drei Punkten). Pro Bedingung wurde der Gesamtwert anhand

des Mittelwertes der acht Testitems berechnet. Wenn alle Sequenzen falsch wiedergegeben worden waren, betrug der Wert „1“; das theoretische Maximum lag bei „7,5“ Punkten, wenn alle Testitems korrekt wiedergegeben wurden. Die interne Konsistenz über die Werte der vier Bedingungen (echte kurze Wörter, echte lange Wörter, kurze Kunstwörter, lange Kunstwörter) wurde mit  $r = .83$  für die Kontrollgruppe und  $r = .84$  für die Untersuchungsgruppe berechnet.

Tabelle 1

*Stichprobenmerkmale und deskriptive Angaben aller Variablen (M und SD) in Studie 1*

	LB (n = 87)		KG (n = 123)		Signifikanz	
	M	SD	M	SD		
Geschlecht (m/w)	48/39		51/72		$\chi^2(1) = 3.31$ ; $p = .069$ ; $\phi = .14$	
Alter (Jahre; Monate)	12;11	2;6	8;5	1;10	$t(155.74)^a = 14.9$ ; $p < .001$ ; $d = 2.1$	
Kognitive Kapazität (CFT 1-R Rohwerte)	63.7	13.2	65.4 <sup>b</sup>	13.5	$t(180) = 0.85$ $p = .395$ ; $d = 0.13$	
Wortschatzumfang (WWT 6–10 Rohwerte)	53.1	15.9	54.7 <sup>b</sup>	14.6	$t(180) = 0.70$ ; $p = .483$ ; $d = 0.10$	
Intelligenz (IQ Normwerte)	74.3	7.3	105.3	10.2	$t(207.98)^b = 25.7$ ; $p < .001$ ; $d = 3.60$	
<b>Wortspannen Performance</b>						
<b>Bedingung</b>						
Wortspannen Echte Wörter	kurz	3.27	0.57	3.47	0.64	$t(208) = 2.43$ ; $p = .016$ ; $d = 0.35$
	lang	2.71	0.44	2.79	0.55	$t(208) = 1.18$ ; $p = .240$ ; $d = 0.17$
	echte Wörter gesamt	2.99	0.46	3.13	0.55	$t(208) = 2.03$ ; $p = .043$ ; $d = 0.29$
Wortspannen Kunstwörter	kurz	2.62	0.66	2.76	0.74	$t(204)^c = 1.38$ ; $p = .169$ ; $d = 0.20$
	lang	1.71	0.43	1.79	0.54	$t(203)^c = 1.09$ ; $p = .277$ ; $d = 0.16$
	Kunstwörter gesamt	2.16	0.51	2.28	0.59	$t(203)^c = 1.44$ ; $p = .151$ ; $d = 0.21$

*Anmerkungen.* LB: Gruppe Schüler\*innen mit Lernbehinderung, KG: Kontrollgruppe. Die Rohwerte des Culture-Fair-Test 1 Revision (CFT 1-R; Weiß und Osterland, 2012) wurden zur Erfassung der kognitiven Kapazität als Entwicklungsindikator (EI) verwendet; die Rohwerte des Wortschatz- und Wortfindungstest für 6-10-Jährige (WWT 6-10, Glück 2011) wurden zur Erfassung des Wortschatzumfangs als EI verwendet; Intelligenznormwerte wurden aus verschiedenen Quellen ermittelt: für Schüler\*innen mit LB aus den offiziellen Schulakten, für jüngere KG-Schüler\*innen wurden die Normwerte des CFT 1-R herangezogen, und für ältere KG-Schüler\*innen (zehn Jahre und älter) wurden die Normwerte des CFT 20-R (Weiß 2008) herangezogen.

<sup>a</sup> *dfs* werden aufgrund von ungleichen Varianzen korrigiert. <sup>b</sup> Teilstichprobe von  $n = 102$  Schüler\*innen in der KG-Gruppe, um Deckeneffekte zu vermeiden. <sup>c</sup> Ausschluss von fünf Schüler\*innen (2 LB, 3 KG), die die Sitzung zu Wortspannenaufgaben mit Pseudowörtern nicht abgeschlossen haben.



### 5.3.2.2. Kognitive Kapazität

Der CFT 1-R (Weiß & Osterland, 2012) wurde bei allen Schüler\*innen mit Lernbehinderung und bei einer jüngeren Teilstichprobe von  $n = 102$  der Kontrollgruppe durchgeführt. Der CFT 1-R ist ein sprachfreies Intelligenzmessverfahren und besteht aus den folgenden sechs Subtests: Substitutionen, Labyrinth, Ähnlichkeiten, Reihenfortsetzen, Klassifikationen und Matrizen. Der Test ist für den Altersbereich von 5;4-9;11 Jahren normiert. Die Retest-Reliabilität liegt laut Manual bei  $r_{tt} = .90$  und die interne Konsistenz bei  $r = .97$ . Da für das Matching [der Gruppen] und den Vergleich der entwicklungsbezogenen Relation der Rohwert herangezogen wurde, war die Berechnung von Normwerten jedoch nicht notwendig, sodass es auch nicht problematisch war, dass das Alter der Kinder in der Untersuchungsgruppe den normierten Bereich des CFT 1-R überstieg. Stattdessen war es wichtig, dass die Rohwerte in einen sensitiven Bereich fielen, d. h. dass keine Boden- oder Deckeneffekte vorlagen. In den Primar-Regelschulen wurde der CFT 1-R als Gruppentest mit bis zu 15 Schüler\*innen durchgeführt. In den Förderschulen (Untersuchungsgruppe) fand diese Testung einzeln oder in Kleingruppen (maximal 4 Schüler\*innen) statt. Die Teilgruppe der älteren unbeeinträchtigten Schüler\*innen (Sekundarstufe) wurden in einer Gruppensitzung mit dem CFT 20-R (Weiß, 2008) getestet, welcher für diese Altersspanne normiert ist (8;5-60;0 Jahre), um sicherzustellen, dass diese Kinder die Einschlusskriterien erfüllten (keine intellektuelle Beeinträchtigung). Die Retest-Reliabilität liegt bei  $r_{tt} = .80$  und die interne Konsistenz bei  $r = .95$ .

### 5.3.2.3. Wortschatzumfang

Die Schüler\*innen der Untersuchungsgruppe und die jüngeren Kinder der Kontrollgruppe wurden in einer Einzelsitzung mit dem PC-basierten expressiven Wortschatz- und Wortfindungstest 6-10 (WWT 6-10, Glück, 2011) getestet. Der Test besteht aus 95 Bildern, die (Teile von) Objekte(n), Aktivitäten, Gegensätze und Oberkategorien darstellen. Das Kind wird verbal zur Benennung aufgefordert: „Was ist das? Was macht er/sie? Was ist das Gegenteil von ...? Was sind das alles zusammen?“. Jedes Item wird für maximal 15 Sekunden auf dem Bildschirm gezeigt. [Wenn nach Ablauf der 15 Sekunden keine Antwort erfolgt ist, wird das Item als falsch gewertet.] Die Retest-Reliabilität wird im Manual mit  $r_{tt} = .96$  angegeben. Die älteren Kinder der Kontrollgruppe (Sekundarstufe) führten den [rezeptiven] Wortschatztest im Rahmen des CFT 20-R (Weiß, 2008) während der Gruppensitzung durch. Hier soll für insgesamt 30 Items für jedes angegebene Wort das korrekte Synonym aus einer Auswahl von 5 Wörtern ausgewählt werden (z. B. Zielwort „Fantasie“. Antwortoptionen: „Form“, „Grundsatz“, „Trugbild“, „Vorstellungsgabe“, „Verstand“); die Reliabilität liegt laut Manual bei  $r_{tt} = .87$ . Es sei darauf hingewiesen, dass der Wortschatztest bei den älteren Schüler\*innen nur zum Zwecke der Stichprobenselektion eingesetzt wurde, um sicherzustellen, dass auch die älteren Schüler\*innen der

Kontrollgruppe einen unbeeinträchtigten Wortschatz aufwiesen. Die Gruppenvergleiche zum Wortschatzumfang wurden ausschließlich anhand der Rohwerte des expressiven Subtests im WWT 6-10 angestellt, die für eine Teilstichprobe von  $n = 102$  unbeeinträchtigten Schüler\*innen gemessen wurden.

Tabelle 2  
Korrelationen zwischen Entwicklungsindikatoren in Studie 1

		KG ( $n = 102$ )		
		Chronologisches Alter	Kognitive Kapazität	Wortschatzumfang
LB ( $n = 80$ )	Chronologisches Alter	—	0.74	0.73
	Kognitive Kapazität	0.42 <sup>a</sup>	—	0.67
	Wortschatzumfang	0.45 <sup>a</sup>	0.49 <sup>b</sup>	—

Anmerkungen. LB: Gruppe Schüler\*innen mit Lernbehinderung (Werte unterhalb der Diagonalen), KG: Kontrollgruppe (Werte oberhalb der Diagonalen). Pearson-Korrelationen zwischen Entwicklungsindikatoren (alle  $p < .001$ ). Aufgrund unsystematisch fehlender Daten in der LB-Gruppe ist die Stichprobengröße kleiner als die vollständige Stichprobe von  $n = 87$  Kindern.

<sup>a</sup>  $n = 80$ . <sup>b</sup>  $n = 73$

### 5.3.3. Analysen<sup>21</sup>

Die Analysen der Developmental Trajectories folgten einer Reihe von Schritten von zunehmender Komplexität. Die Ergebnisse wurden mithilfe von Scatterplots mit Regressionslinien und dazugehörigen Konfidenzintervallen graphisch veranschaulicht. Zur Absicherung der statistischen Signifikanz der Effekte wurden Regressionsanalysen durchgeführt (Thomas et al., 2009). In den folgenden Analysen ist immer der Wortspannen-Wert die abhängige Variable [bzw. in den Abbildungen werden die Differenzwerte abgebildet: Die Differenz von ‚kurz – lang‘ für den Rehearsalprozess und die Differenz ‚echt – kunst‘ für den Redintegrationprozess]. In allen Analysen führten wir drei verschiedene Arten von Prädiktoren in das Modell ein: (a) drei verschiedene kontinuierliche Prädiktoren für das (mentale) Alter, die wir Entwicklungsindikator (EI) nennen; (b) einen Zwischensubjekt-Faktor (Gruppenfaktor); sowie (c) zwei Innersubjekt-Faktoren, Aufgabenbedingungen genannt. Insgesamt wurden neun Developmental Trajectories über beide Gruppen konstruiert: pro Entwicklungsindikator je ein Trajectory für jede der drei abhängigen Variablen.

Die drei Entwicklungsindikatoren [a] umfassten das chronologische Alter (CA) und zwei mentale Altersvariablen: kognitive Kapazität (KOG) und Wortschatzumfang (VOC). Die EI

<sup>21</sup> Siehe hierzu ausführlicher Kapitel 3.

[als Prädiktor im Regressionsmodell] geben wieder, ob das jeweilige Maß für chronologisches oder mentales Alter den Wortspannen-Wert zuverlässig vorhersagen kann. Der Gruppenfaktor [b] zur Unterscheidung von Schüler\*innen mit und ohne Lernbehinderung ermöglichte zwei Vergleiche: Zum einen konnte geprüft werden, ob sich die Gruppen in ihren generellen Wortspannen-Werten unterscheiden, zum anderen konnten die Interaktionen mit EI und Aufgabenbedingungen untersucht werden. Schließlich [c] umfassten die Prädiktoren der Aufgabenbedingung die  $2 \times 2$  Wortlängen- (kurz vs. lang) und Lexikalitätsbedingungen (echt vs. kunst). Anhand eines Aufgabeneffektes für Wortlänge (kurze Wörter werden besser behalten als lange Wörter) wurde Rehearsal gemessen; der Aufgabeneffekt für Lexikalität (echte Wörter werden besser behalten als Kunstwörter) diente zur Schätzung für Redintegration.

[Im folgenden Abschnitt wird das allgemeine Vorgehen bei der Konstruktion von Developmental Trajectories kurz dargestellt, eine ausführliche Erläuterung erfolgte in Kapitel 3.2]. Im ersten Schritt der Konstruktion eines DT wird ein Regressionsmodell angepasst, in dem die Leistung in der Wortspannen-Aufgabe die abhängige Variable darstellt und der EI ein Prädiktor ist. Dies [dient zur Manipulationskontrolle und] zeigt an, ob sich grundsätzlich ein reliabler Zusammenhang zwischen dem EI und der Wortspannen-Leistung feststellen lässt. Im zweiten Schritt wird der Gruppenfaktor in das Modell aufgenommen, um die Gruppen hinsichtlich ihrer Intercept- und Slope-Koeffizienten vergleichen zu können, welche getrennt pro Gruppe geschätzt werden. Der Haupteffekt für den Gruppenfaktor zeigt hier an, ob ein [signifikanter] Unterschied im Intercept zwischen den Gruppen existiert. Die Interaktion von Gruppe  $\times$  EI betrachtet den Unterschied der Slope-Koeffizienten [die Steigung der Regressionsgeraden] um festzustellen, ob sich die Gruppen hinsichtlich des Zusammenhangs zwischen EI und Wortspannen-Leistung unterscheiden (d. h. ob Schüler\*innen mit Lernbehinderung mit der Zeit aufholen oder ob die „Schere“ weiter auseinander geht). Im dritten Schritt kann eine oder mehrere Aufgabenbedingung(en) in das Modell aufgenommen werden, um zu bestimmen, ob sich Intercept- und Slope-Koeffizienten je nach Aufgabenbedingung unterscheiden. Der Haupteffekt für Aufgabenbedingung zeigt an, ob es im Intercept einen generellen Vorteil für z. B. kurze Wörter im Vergleich zu langen Wörtern (Wortlängeneffekt) oder für echte Wörter gegenüber Kunstwörtern (Lexikalitätseffekt) gibt. Die Interaktion von Aufgabe  $\times$  EI zeigt den Unterschied in den Slope-Koeffizienten an, was bedeutet, dass sich der Zusammenhang von EI und Wortspannen-Leistung je nach Aufgabenbedingung unterscheidet (z. B. wenn der Vorteil zugunsten der kurzen Wörter mit zunehmendem Alter stärker wird). Im vierten und letzten Schritt wird das gesamte Modell mit allen drei Prädiktor-Arten (Aufgabenbedingung, Gruppe, und EI) analysiert, um zu überprüfen, ob je nach Aufgabenbedingung differenzielle Entwicklungsmuster in

den Gruppen vorliegen. Die Interaktion von Aufgabe  $\times$  Gruppe zeigt einen Unterschied der Intercepts an, z. B. ob der Wortlängeneffekt (Vorteil von kurzen über lange Wörter) sich zwischen den Gruppen unterscheidet. [Diese Interaktion] zeigt an, ob Schüler\*innen mit Lernbehinderung einen verspäteten Einsatz der Entwicklung [in Rehearsal- oder Redintegrationprozess] aufweisen. Schließlich verweist die dreifach-Interaktion von Aufgabe  $\times$  Gruppe  $\times$  EI auf den Unterschied in den Slopes. Diese gibt an, ob sich der Zusammenhang zwischen EI und Aufgabenbedingung zwischen den Gruppen unterscheidet (z. B. ob der Vorteil für kurze gegenüber langen Wörtern in beiden Gruppen mit höherem Alter in gleichem Maße zunimmt oder nicht). Dieser Effekt zeigt an, ob Schüler\*innen mit Lernbehinderung im Vergleich zur unbeeinträchtigten Kontrollgruppe eine verlangsamte Entwicklungsrate aufweisen.

Die Intercept- und Slope-Effekte der Gruppen wurden gemäß der Hypothesen für die jeweiligen abhängigen Variablen verglichen: die PL-Kapazität (Hypothese 1) sowie die Effektivität des Rehearsalprozesses (Hypothese 2) bzw. des Redintegrationprozesses (Hypothese 3). Für jede abhängige Variable wurde die entwicklungsbezogene Relation zu den drei EI berechnet. Von besonderem Interesse war die entwicklungsbezogene Relation zwischen dem Wortschatzumfang [als Indikator für das Langzeitgedächtnis] und Redintegration (siehe Hypothese 3). Für die Analyse der PL-Kapazität wurde die abhängige Variable der Wortspannen-Werte von den drei verschiedenen EI und dem Gruppenfaktor ohne Berücksichtigung der Aufgabenbedingung vorhergesagt; daher wurde der Mittelwert der Wortspannen-Werte über alle  $2 \times 2$  Aufgabenbedingungen berechnet. Rehearsal wird als Wortlängeneffekt operationalisiert. Daher wurde im Abschnitt zu Rehearsal die Aufgabenbedingung Wortlänge als Prädiktor mit in das Modell aufgenommen, und die Wortspannen-Werte wurden über beide Lexikalitätsbedingungen hinweg gemittelt. Analog wurde die Aufgabenbedingung Lexikalität als Prädiktor in das Modell aufgenommen, um Redintegration zu untersuchen. Hier wurden die Wortspannen-Werte über beide Wortlängenbedingungen hinweg gemittelt.

Für die Berechnung [und leichtere Interpretation] der Regressionskoeffizienten wurden die Entwicklungsindikatoren linear transformiert, sodass deren Minimum in der Untersuchungsgruppe dem Wert 0 entsprach. Die Transformation beeinflusst ausschließlich die Größe des Intercept-Effektes, ermöglicht aber eine sinnvolle Interpretation des Unterschieds im Intercept zwischen den Gruppen (Thomas et al., 2009), da eine Extrapolation über den Bereich der Daten hinaus vermieden wird. Der Slope-Effekt ist von dieser Transformation nicht betroffen. Die Transformation wird allein zum Zweck der statistischen Berechnung der Intercept-Effekte durchgeführt, in den Abbildungen werden dagegen weiterhin die ursprünglichen [mentalen] Altersangaben [auf der x-Achse] gezeigt. Die vollständigen statistischen Angaben ( $F$ ,  $p$ ,  $\eta^2$ ) zu den

Regressionsmodellen können im Anhang in Tabelle A-2 eingesehen werden. Zwecks besserer Lesbarkeit werden im Text nur die  $p$ -Werte berichtet.

Die Aufbereitung der Daten sowie die Erstellung der Abbildungen erfolgte in R (R Studio Team, 2016; R Core Team, 2018) unter Anwendung der Pakete `ggplot2` (Wickham, 2011) und `ggpubr` (Kassambara, 2017). Die Regressionsanalysen für die DTs wurden in SPSS Version 25 (IBM Corp., 2017), gemäß der im elektronischen Anhang zu Thomas et al. (2009) dargestellten Vorgehensweise, durchgeführt

## 5.4. Ergebnisse<sup>22</sup>

Die Merkmale zur Beschreibung der Stichprobe sowie deskriptive Ergebnisse der durchschnittlichen Wortspannen-Werte [getrennt nach den vier] Bedingungen sind in Tabelle 1 dargestellt. Die Korrelationen zwischen den drei EI (chronologisches Alter, kognitive Kapazität, Wortschatzumfang) finden sich in Tabelle 2. [Bei der Datenerhebung wurde] die Reihenfolge der Lexikalität in der Darbietung der Wortspannen-Aufgaben randomisiert zugewiesen und für eine Teilstichprobe ausgewertet; es ergab sich kein signifikanter Effekt dafür, ob echte Wörter als erstes oder als zweites dargeboten wurden,  $t(118) = 0.20$ ;  $p = .844$ . Um das Vorliegen möglicher Geschlechtseffekte zu überprüfen, wurden zwei lineare gemischte Modelle für jeden EI geschätzt. Ein baseline-Modell beinhaltete alle [oben beschriebenen] Prädiktoren (d. h. EI, Gruppe, Länge und Lexikalität), das volle Modell enthielt zusätzlich den Faktor Geschlecht. Mittels eines Chi-Quadrat-Tests wurde die Anpassungsgüte der Modelle verglichen. In keinem der drei EI-Modellvergleiche zeigten sich Hinweise auf einen signifikanten Geschlechtseffekt: für chronologisches Alter  $\chi^2(16) = 14.525$ ;  $p = .56$ ; kognitive Kapazität  $\chi^2(16) = 13.418$ ;  $p = .642$ ; und Wortschatzumfang  $\chi^2(16) = 12.642$ ;  $p = .699$ . Folglich erschien es angemessen, die Daten über beide Geschlechtergruppen zusammenzufassen [und damit in den weiteren Analysen Geschlecht nicht weiter zu berücksichtigen].

### 5.4.1. Kapazität der Phonologischen Schleife

In der ersten Analyse (Abbildung 2) untersuchten wir, ob Schüler\*innen mit Lernbehinderung differenzielle Entwicklungsmuster hinsichtlich der Kapazität der phonologischen Schleife aufwiesen. Damit wird die Frage beantwortet, inwiefern die Entwicklung der PL in Schüler\*innen mit Lernbehinderung auf dem gleichen Level einsetzte (*onset level*) wie bei unbeeinträchtigten Kindern. Außerdem ermöglichte die Analyse festzustellen, ob sich ein bei den Schüler\*innen mit Lernbehinderung möglicherweise vorliegendes Defizit der PL-Kapazität mit zunehmendem

---

<sup>22</sup> Der folgende Abschnitt ist eine Übersetzung des Abschnitts „Results“ (S. 7-10) aus Bruns, Ehl & Grosche (2019). Ebenfalls übernommen jedoch nicht übersetzt wurden Abbildungen 2-4, die im veröffentlichten Fachbeitrag als Abbildungen 1-3 bezeichnet sind.

Alter, kognitiver Kapazität (KOG) oder Wortschatzumfang (VOC) reduzierte, oder ob sie weiterhin hinter den unbeeinträchtigten Schüler\*innen zurückblieben. Die PL-Kapazität wurde operationalisiert als der durchschnittliche Wortspannen-Wert über alle  $2 \times 2$  Aufgabenbedingungen hinweg. Daher wurden lediglich der Gruppenfaktor und die EI als Prädiktoren in das Modell aufgenommen, ohne zwischen verschiedenen Stufen von Wortlänge und Lexikalität zu unterscheiden. Für jeden EI (chronologisches Alter, kognitive Kapazität, Wortschatzumfang) wurden die Gruppen hinsichtlich ihrer Intercept- (gemessen am Haupteffekt für den Faktor Gruppe) und der Slope-Koeffizienten (gemessen an der Interaktion von Gruppe  $\times$  EI) verglichen.

Mit dem *chronologischen Alter* als EI zur Vorhersage der PL-Kapazität ließ sich ein signifikanter Intercept-Effekt zwischen den Gruppen ( $p < .001$ ) nachweisen, da Schüler\*innen mit Lernbehinderung einen niedrigeren Intercept hatten. Die Interaktion von Gruppe  $\times$  Alter, die auf eine verlangsamte Entwicklungsrate hinweisen würde, wurde nicht signifikant ( $p = .068$ ). Dies bedeutet, dass Schüler\*innen mit Lernbehinderung einen verspäteten Einsatz in der Entwick-

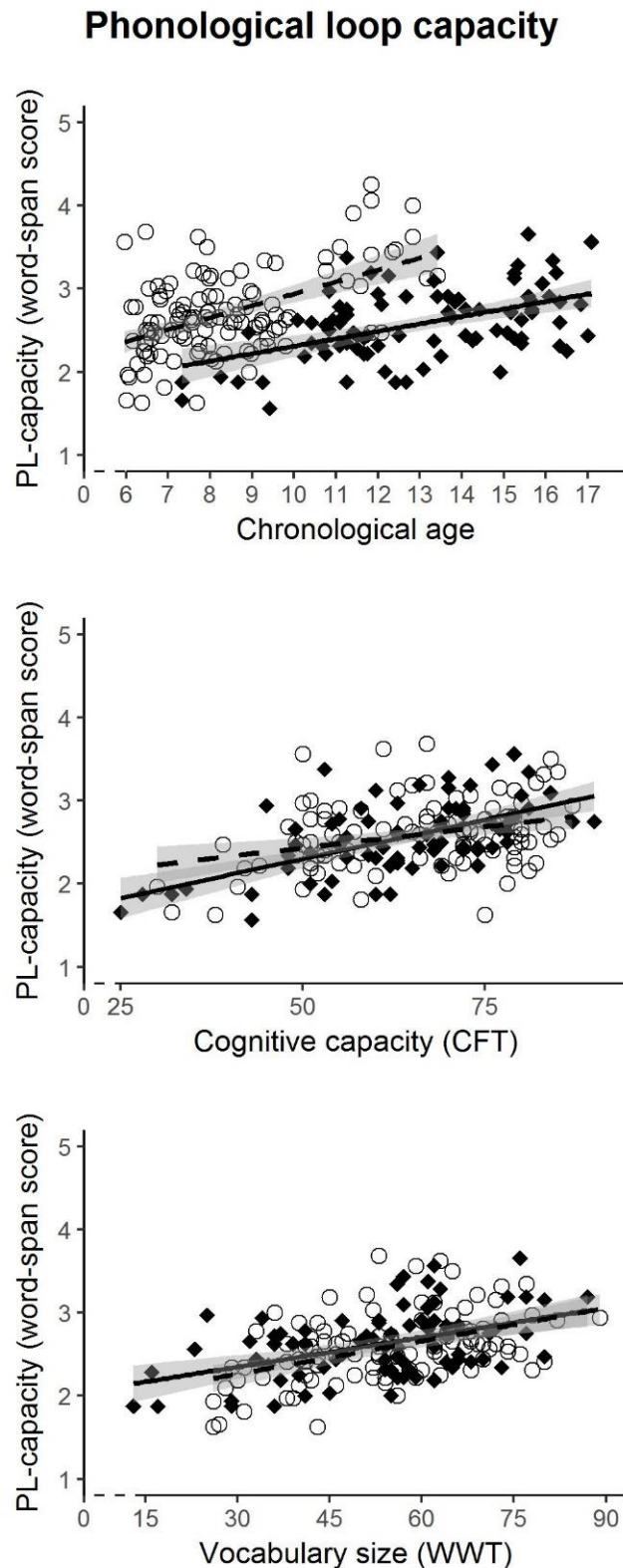


Abbildung 2. Developmental Trajectory Modelle der Kapazität der phonologischen Schleife bei Schüler\*innen mit Lernbehinderung (MBID, schwarze Rauten) und unbeeinträchtigten Schüler\*innen (TD, weiße Kreise).

lung der PL-Kapazität, aber keine signifikant verlangsamte Entwicklungsrate im Verhältnis zu ihrem chronologischen Alter aufwiesen.

Mit *kognitiver Kapazität* als EI zur Vorhersage der PL-Kapazität konnten wir sowohl einen signifikanten Intercept-Unterschied zwischen den Gruppen ( $p = .039$ ), als auch eine signifikante Interaktion von Gruppe  $\times$  KOG ( $p = .033$ ) feststellen, was auf einen Unterschied in den Slopes hinweist. Der Einsatz der Entwicklung erfolgte damit signifikant später und die Steigung für die Schüler\*innen mit Lernbehinderung war signifikant steiler. Dies deutet auf einen stärkeren Zusammenhang zwischen kognitiver Kapazität und der generellen PL-Kapazität als bei unbeeinträchtigten Schüler\*innen hin. Folglich zeigten Schüler\*innen mit Lernbehinderung einen verspäteten Entwicklungseinsatz in Verbindung mit einer differenziellen Entwicklungsrate ihrer PL-Kapazität im Verhältnis zu ihrer kognitiven Kapazität. Es sei jedoch darauf hingewiesen, dass eine Konfundierung der Werte für die kognitive Kapazität mit dem chronologischen Alter nicht ausgeschlossen werden kann; wurden die jüngeren Schüler\*innen mit niedrigen Werten in kognitiver Kapazität (CFT 1-R Rohwerte  $< 45$ ) aus der Analyse ausgeschlossen, so waren beide Effekte nicht mehr signifikant, d. h.  $p = .081$  für den Intercept-Unterschied und  $p = .073$  für den Slope-Unterschied.

Mit *Wortschatzumfang* als EI zur Vorhersage der PL-Kapazität wurde weder der Intercept-Unterschied zwischen den Gruppen signifikant ( $p = .331$ ), noch unterschieden sich die Slopes (Gruppe  $\times$  VOC;  $p = .595$ ). Dies deutet darauf hin, dass das Entwicklungsmuster der PL-Kapazität bei Kindern mit Lernbehinderung im Verhältnis zum Wortschatzumfang nicht von dem der unbeeinträchtigten Kinder abwich.

Zusammengefasst zeigten Schüler\*innen mit Lernbehinderung hinsichtlich der PL-Kapazität (gemessen anhand der durchschnittlichen Wortspannen-Werte über alle Aufgabenbedingungen hinweg) nur einen verspäteten Entwicklungseinsatz, aber die gleiche Entwicklungsrate über den gemessenen Bereich des chronologischen Alters. Kognitive Kapazität erwies sich als stärkerer Prädiktor für Schüler\*innen mit Lernbehinderung als für die Kontrollgruppe; dieser Unterschied blieb aber nicht bestehen, wenn Kinder mit niedrigeren Werten in kognitiver Kapazität aus der Analyse ausgeschlossen wurden. Es konnten keine differenziellen Effekte für den Zusammenhang von PL-Kapazität mit dem Wortschatzumfang festgestellt werden. Wortschatzumfang als EI führte zu keinem differenziellen Entwicklungsmuster [in der PL-Kapazität] zwischen der Experimental- und der Kontrollgruppe.

#### 5.4.2. Effektivität des Rehearsalprozesses

In der zweiten Analyse (Abbildung 3) wurde untersucht, ob Schüler\*innen mit Lernbehinderung differenzielle Entwicklungsmuster hinsichtlich der Effektivität des Rehearsalprozesses

aufwiesen. Wie bei der ersten Forschungsfrage wurde hiermit die Frage beantwortet, ob die Effektivität des Rehearsals bei Schüler\*innen mit Lernbehinderung auf einem ähnlichen Level einsetzte wie bei unbeeinträchtigten Kindern. Zudem ließ sich bestimmen, ob der Entwicklungsfortschritt (gemessen am Alter, der kognitiven Kapazität und am Wortschatzumfang) sich in einer ähnlichen Weise auf die Entwicklung des Rehearsalprozesses auswirkt. Da Rehearsal anhand des Wortlängeneffektes operationalisiert ist (d. h. am relativen Vorteil für kurze gegenüber langen Wörtern), wurde neben dem Gruppenfaktor und den EI auch die Aufgabenbedingung Wortlänge als Prädiktor in das Modell aufgenommen, jedoch ohne zwischen echten Wörtern und Kunstwörtern zu unterscheiden. Um die Interpretation zu erleichtern, wurde in Abbildung 3 die Größe des Wortlängeneffekts (die Differenz von kurzen und langen Wörtern) auf der y-Achse abgetragen, wohingegen im Regressionsmodell Wortlänge als Prädiktor aufgenommen wurde.

Mit *chronologischem Alter* als EI zur Vorhersage der Rehearsal-Effektivität ließ sich ein signifikanter Intercept-Effekt für Wortlänge  $\times$  Gruppe ( $p = .021$ ) nachweisen, wohingegen die Interaktion von Wortlänge  $\times$  Gruppe  $\times$  Alter ( $p = .210$ ), welche die Entwicklungsrate

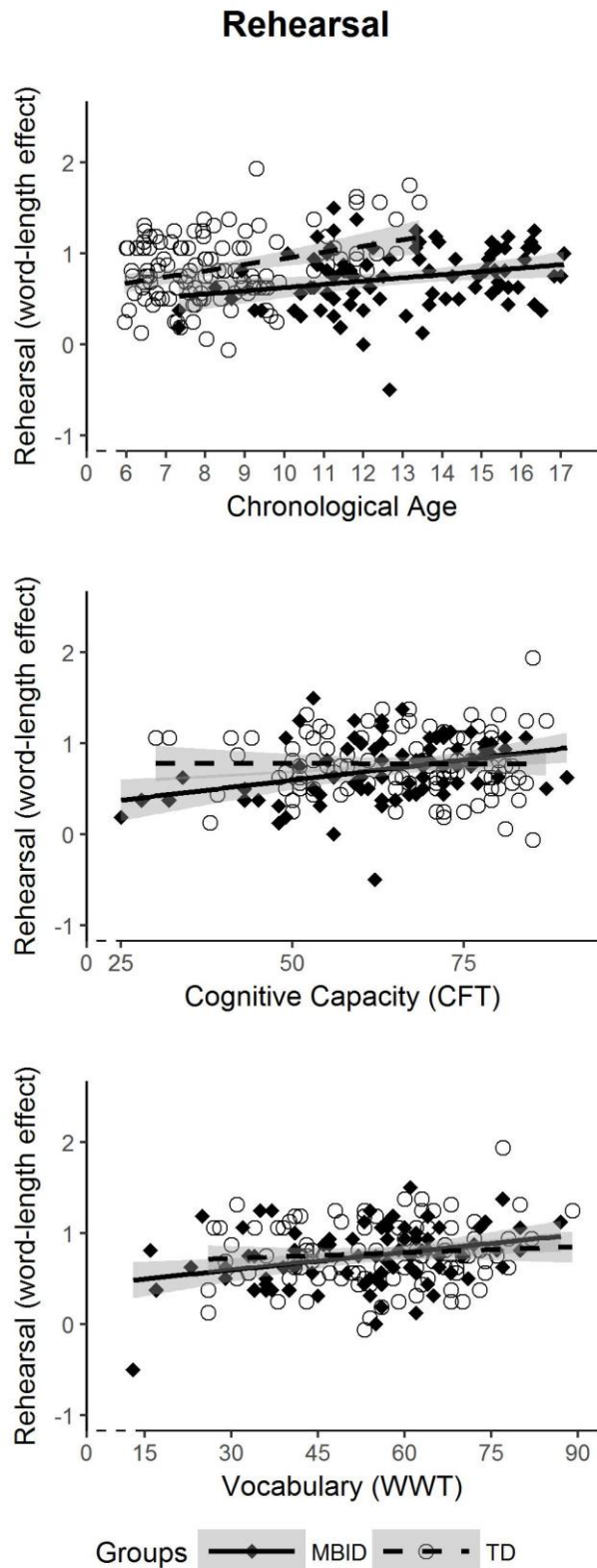


Abbildung 3. Developmental Trajectory Modelle des Rehearsalprozesses bei Schüler\*innen mit Lernbehinderung (MBID, schwarze Rauten) und unbeeinträchtigten Schüler\*innen (TD, weiße Kreise).



abbildet, nicht signifikant wurde. Das bedeutet, dass Schüler\*innen mit Lernbehinderung einen verspäteten Entwicklungseinsatz, aber keine verlangsamte Entwicklungsrate des Rehearsalprozesses im Verhältnis zu ihrem chronologischen Alter zeigten.

Mit *kognitiver Kapazität* als EI zur Vorhersage der Rehearsal-Effektivität ergab sich ein signifikanter Intercept-Effekt für Wortlänge  $\times$  Gruppe ( $p = .009$ ). Eine signifikante Interaktion von Wortlänge  $\times$  Gruppe  $\times$  KOG ( $p = .016$ ) wies auf einen Unterschied in den Slope-Koeffizienten hin. Die Steigung [der Regressionsgeraden] verlief für Schüler\*innen mit Lernbehinderung signifikant steiler als für die Kontrollgruppe. Folglich zeigten Schüler\*innen mit Lernbehinderung einen verspäteten Entwicklungseinsatz in Kombination mit einer beschleunigten Entwicklungsrate des Rehearsalprozesses im Verhältnis zur kognitiven Kapazität. Wiederum sorgte der Ausschluss der jüngeren Schüler\*innen mit niedrigen Werten in der kognitiven Kapazität (Rohwerte  $< 45$ ) dafür, dass beide Effekte nicht mehr signifikant wurden, d. h.  $p = .147$  für den Intercept-Unterschied und  $p = .192$  für den Unterschied der Slopes.

Mit *Wortschatzumfang* als EI konnten keine Unterschiede zwischen den Gruppen gefunden werden. Weder der Intercept-Unterschied (Wortlänge  $\times$  Gruppe;  $p = .212$ ) noch der Slope-Unterschied (Wortlänge  $\times$  Gruppe  $\times$  KOG;  $p = .192$ ) wurden signifikant. Dies zeigte, dass das Entwicklungsmuster der Rehearsal-Effektivität im Verhältnis zum Wortschatzumfang sich zwischen den Gruppen nicht unterschied.

Zusammenfassend zeigte sich für die Entwicklung der Rehearsal-Effektivität ein verzögerter Einsatz bei Schüler\*innen mit Lernbehinderung, wenn das chronologische Alter als Entwicklungsindikator herangezogen wurde. Der Befund einer differenziellen Entwicklungsrate des Rehearsals, wenn die kognitive Kapazität als EI eingesetzt wurde, sollte angesichts der nicht-signifikanten Intercept- und Slope-Effekte, wenn Kinder mit niedrigen KOG-Werten ausgeschlossen wurden, zurückhaltend interpretiert werden. Dies lässt eher darauf schließen, dass der Zusammenhang von Rehearsal und kognitiver Kapazität in beiden Gruppen ähnlich war. In keiner der beiden Gruppen konnte das lexikalische Wissen (Wortschatzumfang) die Effektivität des Rehearsalprozesses signifikant vorhersagen.

#### 5.4.3. Effektivität des Redintegrationprozesses

In der dritten Analyse (Abbildung 4) wurde untersucht, ob Schüler\*innen mit Lernbehinderung differenzielle Entwicklungsmuster hinsichtlich der Effektivität des Redintegrationprozesses aufwiesen. Wie vorher wurde hiermit die Frage beantwortet, ob die Effektivität der Redintegration bei Schüler\*innen mit Lernbehinderung auf einem ähnlichen Level einsetzte wie bei unbeeinträchtigten Kindern. Zudem ließ sich bestimmen, ob der Entwicklungsfortschritt (gemessen am Alter, der kognitiven Kapazität und am Wortschatzumfang) sich in einer ähnlichen

Weise auf die Entwicklung der Redintegration auswirkte. Da Redintegration anhand des Lexikalitätseffektes operationalisiert ist (d. h. am relativen Vorteil für echte Wörter gegenüber Kunstwörtern), wurde neben dem Gruppenfaktor und den EI auch die Aufgabenbedingung Lexikalität als Prädiktor in das Modell aufgenommen, jedoch ohne zwischen kurzen und langen Wörtern zu unterscheiden. Um die Interpretation zu erleichtern, wurde in Abbildung 4 die Größe des Lexikalitätseffekts (die Differenz von echten Wörtern und Kunstwörtern) auf der y-Achse abgetragen, wohingegen im Regressionsmodell Lexikalität als Prädiktor aufgenommen wurde.

Mit *chronologischem Alter* als EI zur Vorhersage der Redintegration-Effektivität ließ sich weder ein signifikanter Intercept-Effekt für Lexikalität  $\times$  Gruppe ( $p = .530$ ) nachweisen, noch eine signifikante Interaktion von Lexikalität  $\times$  Gruppe  $\times$  Alter ( $p = .617$ ). Dies bedeutet, dass Schüler\*innen mit Lernbehinderung im Verhältnis zum chronologischen Alter ein ähnliches Entwicklungsmuster für Redintegration zeigten wie die unbeeinträchtigte Kontrollgruppe.

Mit *kognitiver Kapazität* als EI zur Vorhersage der Redintegration-Effektivität zeigte sich kein signifikanter Intercept-Effekt für Lexikalität  $\times$  Gruppe ( $p = .718$ ). Eine nicht-signifikante Interaktion von Lexikalität  $\times$

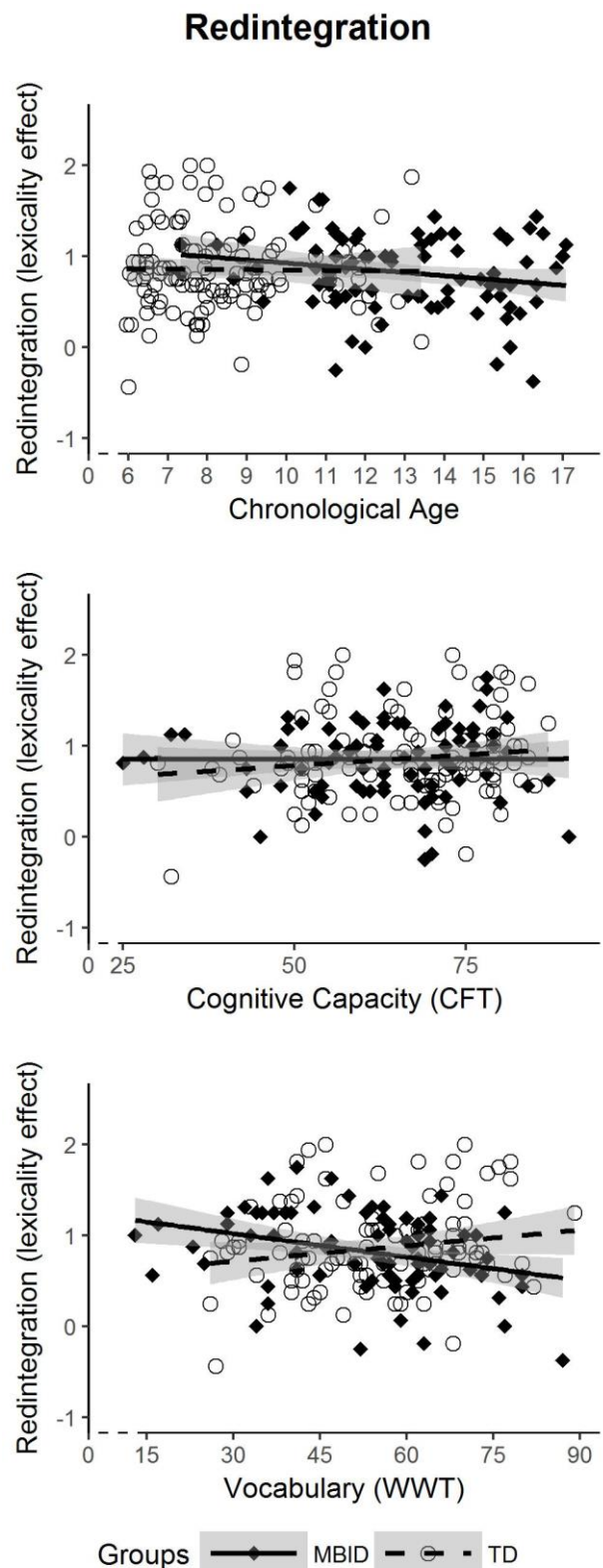


Abbildung 4. Developmental Trajectory Modelle des Redintegrationprozesses bei Schüler\*innen mit Lernbehinderung (MBID, schwarze Rauten) und unbeeinträchtigten Schüler\*innen (TD, weiße Kreise).

Gruppe  $\times$  KOG ( $p = .658$ ) wies darauf hin, dass es keinen Unterschied in den Slope-Koeffizienten gab. Folglich erschien die Entwicklung von Redintegration bei Schüler\*innen mit Lernbehinderung im Verhältnis zur kognitiven Kapazität unbeeinträchtigt zu sein.

Dagegen ergaben sich mit dem *Wortschatzumfang* als EI zur Vorhersage der Redintegration-Effektivität sowohl ein signifikanter Unterschied im Intercept (Lexikalität  $\times$  Gruppe;  $p = .024$ ) als auch in der Slope (Lexikalität  $\times$  Gruppe  $\times$  VOC;  $p = .005$ ). Der Intercept-Unterschied fiel zugunsten der Schüler\*innen mit Lernbehinderung aus, die bei einem geringeren Wortschatzumfang eine höhere Redintegration-Effektivität zeigten. Dagegen verringerte sich die Redintegration-Effektivität mit zunehmendem Wortschatzumfang. Dieser Befund stellt ein unerwartetes differenzielles Muster in der Redintegration-Entwicklung dar. Anschließende separate Analysen der DTs pro Gruppe ergaben, dass bei der unbeeinträchtigten Kontrollgruppe kein signifikanter Slope-Effekt feststellbar war,  $F(1, 100) = 2.062$ ;  $p = .154$ , wohingegen bei Schüler\*innen mit Lernbehinderung ein negativer Slope-Effekt signifikant wurde,  $F(1, 78) = 8.254$ ;  $p = .005$ . Dieser negative Slope-Effekt blieb auch signifikant, nachdem ein möglicher Ausreißer in der Untersuchungsgruppe mit hohem Wortschatz-Wert und niedriger Redintegration-Effektivität aus der Analyse ausgeschlossen wurde,  $F(1, 77) = 5.267$ ;  $p = .024$ .

Bezüglich der Effektivität von Redintegration waren die Befunde für Kinder mit Lernbehinderung hier folglich uneinheitlich. Einerseits schien es kein grundsätzlich differenzielles Entwicklungsmuster zu geben, da wir ähnliche Entwicklungseinsätze und -raten nachweisen konnten, wenn chronologisches Alter und kognitive Kapazität als EI herangezogen wurden. Andererseits war der Zusammenhang zwischen Redintegration und Wortschatzumfang invers und stellte damit einen differenziellen Effekt dar: Je größer der Wortschatzumfang von Schüler\*innen mit Lernbehinderung, desto schwächer fiel die Redintegration-Effektivität aus, selbst nach Ausschluss eines potenziellen Ausreißers. Dieser Befund ist unerwartet und muss weiter diskutiert werden.

## 5.5. Diskussion<sup>23</sup>

### 5.5.1. Entwicklung der Arbeitsgedächtnis-Prozesse

Das Ziel dieser Studie war, herauszufinden, ob Schüler\*innen mit Lernbehinderung differenzielle Entwicklungsmuster hinsichtlich dreier Aspekte des verbalen Arbeitsgedächtnisses aufwiesen: (a) Kapazität der PL, (b) Effektivität des Rehearsalprozesses und (c) Effektivität des Redintegrationprozesses.

---

<sup>23</sup> Der folgende Abschnitt ist eine Übersetzung des Abschnitts „Discussion“ (S. 10-13) aus Bruns, Ehl & Gro-sche (2019).

Wir konnten zeigen, dass Schüler\*innen mit Lernbehinderung verglichen mit einer unbeeinträchtigten Kontrollgruppe einen verspäteten Entwicklungseinsatz der PL-Kapazität im Verhältnis zum chronologischen Alter aufwiesen, was sich an einem signifikanten Intercept-Unterschied offenbarte. Schüler\*innen mit Lernbehinderung haben folglich von Beginn ihrer Entwicklung an eine geringere PL-Kapazität. Dieser Befund deckt sich mit weiteren Studien aus dem Feld (Mähler & Hasselhorn, 2003; Mahler, 2007; Schuchardt et al., 2010; Poloczek et al., 2014; 2016). Hinsichtlich des Wachstums der PL-Kapazität ließ sich keine verlangsamte Entwicklungsrate erkennen, wie der nicht-signifikante Slope-Effekt für die Interaktion mit dem chronologischen Alter anzeigte. Das bedeutet, dass sich die PL-Kapazität bei Schüler\*innen mit Lernbehinderung mit einer ähnlichen [parallelen] Rate zu entwickeln schien wie bei unbeeinträchtigten Kindern, aber sich der Abstand im Lauf der im Rahmen der Studie erhobenen Altersspanne auch nicht verringerte. Es lässt sich jedoch die Möglichkeit nicht ausschließen, dass sich diese Lücke im Laufe der weiteren Entwicklung schließen könnte.

Wenn die PL-Kapazität zu der kognitiven Kapazität ins Verhältnis gesetzt wurde, wiesen Schüler\*innen mit Lernbehinderung eine steilere Steigung auf als unbeeinträchtigte Kinder, was auf den ersten Blick den Erwartungen zuwiderzulaufen scheint. Eine solche beschleunigte Rate [bzw. stärkerer Zusammenhang] wurde auch von Henry (2001) und Schuchardt et al. (2011) berichtet. Eine mögliche Erklärung könnte darin liegen, dass die Entwicklung der PL-Kapazität von Schüler\*innen mit Lernbehinderung eher mit der mentalen Entwicklung einhergeht (siehe jedoch Colom et al., 2010), während die unbeeinträchtigte Kontrollgruppe bereits ein Plateau erreicht haben könnte und sich über dieses Niveau nicht weiter steigern kann. Die Befunde zum Zusammenhang mit der kognitiven Kapazität sollten insgesamt zurückhaltend interpretiert werden, da der Ausschluss von Kindern mit niedrigen Rohwerten dazu führte, dass die Unterschiede nicht mehr signifikant wurden. Daher erscheinen Replikationsstudien notwendig, bevor inhaltliche Schlüsse gezogen werden können.

Die Developmental Trajectories zur PL-Kapazität im Verhältnis zum Wortschatzumfang als EI waren für Kinder der Experimental- und Kontrollgruppe äquivalent, was als Hinweis gedeutet werden kann, dass die Entwicklung der allgemeinen Funktionsweise der PL im Verhältnis zum Langzeitgedächtnis-Wissen sich zwischen den Gruppen nicht unterschied. Daher scheint in beiden Gruppen gleichermaßen die Entwicklung der PL-Kapazität mit wachsendem Wortschatzumfang einherzugehen.

Die Prozesse Rehearsal und Redintegration wurden anhand der Effekte für Wortlänge bzw. Lexikalität untersucht. Beim Vergleich der Wortlängenbedingungen (kurze vs. lange Wörter, über beide Lexikalitätsbedingungen hinweg) ließ sich eine automatische Aktivierung des Rehearsals in beiden Gruppen nachweisen. Es sei angemerkt, dass diese Aussage vorrangig im

Rahmen der Konzeptionalisierung der Rehearsal-Aktivierung nach Hasselhorn et al. (2000) gilt. Aktuelle Studien haben die Rolle von Rehearsal bei der Entstehung des Wortlängeneffekts in Frage gestellt [Campoy, 2008], und Poloczek et al. (2014) haben Hinweise darauf gefunden, dass das Auftreten eines Wortlängeneffektes auch von Output-Verzögerungen abhängig sein kann.

Wir konnten einen grundsätzlichen signifikanten Wortlängeneffekt nachweisen, da die Erinnerungsleistung für kurze Wörter höher war als für lange Wörter. Schüler\*innen mit Lernbehinderung zeigten einen verspäteten Entwicklungseinsatz des Rehearsals, da bei ihnen der Vorteil für kurze Wörter weniger stark ausfiel als in der Kontrollgruppe. Die Effektivität des Rehearsals schien sich im Zuge des chronologischen Alters ähnlich zu entwickeln, was bedeutet, dass ältere Kinder in beiden Gruppen gleichermaßen ein stärkeres Rehearsal zeigten, denn es ließen sich keine Unterschiede in der Entwicklungsrate finden. Dieser Befund passt zu Mähler & Hasselhorn (2003), die eine Verzögerung, aber keine strukturelle Abweichung bei Jugendlichen mit Lernbehinderung zeigen konnten, und anderen Studien, die eine Entwicklungsverzögerung bei Schüler\*innen mit Lernbehinderung fanden (Henry, 2001; Hasselhorn & Mähler, 2007; Mähler, 2007). Es sei jedoch angemerkt, dass der Wortlängeneffekt auch von einem [sog.] proportionalen Skalierungs-Artefakt betroffen sein kann: Da der Wortlängeneffekt ein relatives Maß der Differenz zwischen kurzer und langer Bedingung ist, argumentieren Jarrold, Danielsson und Wang (2015), dass er ins Verhältnis zur generellen Leistung gesetzt werden sollte, d. h. in dem Fall zur Leistung bei kurzen echten Wörtern. Ein solches Artefakt lässt sich im vorliegenden Fall nicht ausschließen, würde jedoch über den Rahmen der aktuellen Studie hinausgehen.

Schüler\*innen mit Lernbehinderung schienen einen verzögerten Entwicklungseinsatz [des Rehearsalprozesses] aufzuweisen, wenn die kognitive Kapazität betrachtet wurde. Während in der vollständigen Stichprobe [ohne Ausschluss von möglichen Ausreißern nach unten] eine zunehmende kognitive Kapazität in der unbeeinträchtigten Kontrollgruppe die Rehearsal-Effektivität nicht vorhersagen konnte [d. h. es gab keinen reliablen Zusammenhang], war bei Schüler\*innen mit Lernbehinderung ein Zuwachs der Rehearsal-Effektivität [mit steigender kognitiver Kapazität] zu verzeichnen. Möglicherweise holen Schüler\*innen mit Lernbehinderung in ihrer Rehearsal-Entwicklung mit zunehmender kognitiver Reifung stärker auf. Wie bei der PL-Kapazität sollte dieser Befund mit Zurückhaltung interpretiert werden, da Kinder mit niedrigen Werten in der kognitiven Kapazität die Ergebnisse [übermäßig] zu beeinflussen schienen. Wenn sie aus den Analysen ausgeschlossen wurden, ließ sich kein Unterschied zwischen den Gruppen mehr finden, weder im Intercept- noch im Slope-Effekt. Dies deutet eher darauf hin, dass Schüler\*innen mit Lernbehinderung und unbeeinträchtigte Kinder hinsichtlich ihrer

entwicklungsbezogenen Relation zwischen Rehearsal und kognitiver Kapazität gleich sind. Dieser Effekt sollte in weiteren Studien eingehender untersucht werden, um aussagekräftige Schlussfolgerungen ziehen zu können.

In keiner der beiden Gruppen hing die Rehearsal-Effektivität mit dem Wortschatzumfang zusammen. Auch die Muster [der Gruppen] unterschieden sich nicht in Entwicklungseinsatz oder -rate. Dieser Befund war insofern plausibel, als dass der Wortschatzumfang, der auf die Messung des lexikalischen Wissens im Langzeitgedächtnis abzielt, nicht mit dem phonologisch basierten Rehearsalprozess zusammenhing.

Hinsichtlich der Redintegration konnte ein genereller Vorteil der echten Wörter gegenüber den Kunstwörtern gemessen werden, da echte Wörter im Langzeitgedächtnis verfügbar sind und damit deren Rekonstruktion leichter fällt. Insofern konnte der Effekt repliziert werden, dass Arbeitsgedächtnis Wortspannen-Aufgaben [auch] vom Langzeitgedächtnis abhängen. Hinsichtlich des *chronologischen Alters* waren Entwicklungseinsatz und -rate in beiden Gruppen ähnlich, was bedeutet, dass bei Schüler\*innen mit Lernbehinderung weder ein verspäteter Entwicklungseinsatz noch eine verzögerte Entwicklungsrate in der Redintegration-Effektivität konstatiert werden konnte. Die *kognitive Kapazität* konnte in keiner der beiden Gruppen die Redintegration-Effektivität vorhersagen. Dagegen rief der *Wortschatzumfang* als Entwicklungsindikator ein differenzielles Muster in der Vorhersage von Redintegration hervor. Schüler\*innen mit Lernbehinderung zeigten eine höhere Redintegration-Effektivität am unteren Ende des Wortschatzumfangs und wurden mit steigendem Wortschatzumfang weniger effektiv. Daraus schlussfolgern wir, dass Schüler\*innen mit Lernbehinderung Redintegration weniger effektiv nutzen konnten, als aufgrund ihres Wortschatzumfangs zu erwarten wäre.

Wenn man, unter Anwendung des traditionellen Gruppenmatching-Ansatzes, nur die Relation mit chronologischem Alter betrachtet hätte, dann wäre Redintegration von Schüler\*innen mit Lernbehinderung unbeeinträchtigt erschienen. Dies könnte [zunächst] als Stärke und Resource interpretiert werden, da gezeigt wurde, dass Schüler\*innen mit Lernbehinderung in der Lage waren, zur Rekonstruktion [von Arbeitsgedächtnis-Memoranda] auf das Langzeitgedächtnis zurückzugreifen, und dies in ähnlicher Art und Weise wie ihre unbeeinträchtigten Peers. Dieser Befund der unbeeinträchtigten Redintegration Effektivität wurde jedoch teilweise relativiert durch den negativen Zusammenhang zwischen Redintegration und Wortschatzumfang, was ein spezifisches differenzielles Entwicklungsmuster für Lernbehinderung darstellte, da sie nicht von einem größeren Umfang von Wörtern, die sie zur Redintegration nutzen könnten, zu profitieren schienen. Da Redintegration als Nutzung von existierendem Wissen im Langzeitgedächtnis zur Rekonstruktion von Spuren im Arbeitsgedächtnis konzipiert ist, wäre eigentlich zu erwarten, dass eine höhere Anzahl an verfügbaren Wörtern im Langzeitgedächtnis auch zu einer

effektiveren Redintegration führen sollte. Dies war bei Kindern mit Lernbehinderung jedoch nicht der Fall; vielmehr schien entgegen dieser Erwartung ein wachsender Wortschatzumfang einen eher abträglichen Effekt für die Nutzung von Langzeitgedächtnis-Inhalten zur Rekonstruktion von Spuren in der PL zu haben. Weitere Forschungsarbeiten sollten der Frage nachgehen, ob es sich hierbei eher um ein Problem ineffizienter Strategienutzung (z. B. Klauer & Lauth, 1997), um Schwierigkeiten in der Unterdrückung irrelevanter Informationen (z. B. Lilienthal, Rose, Tamez, Myerson & Hale, 2015), um eine suboptimale Organisation des mentalen Lexikons (z. B. Vitevitch, Chan & Roodenrys, 2012; Kenett, Gold et al., 2016), bei der ein größerer Wortschatzumfang für „Verwirrung“ sorgen könnte, oder um Probleme bei anderen Prozessen in der Abruf-Phase (z. B. Dell & O’Seaghdha, 1992) handelt.

In dieser Studie dienten Developmental Trajectories als Methode zur Aufdeckung von entwicklungsbezogenen Ähnlichkeiten und Unterschieden hinsichtlich der Arbeitsgedächtnis [Kapazität sowie der] Prozesse von Rehearsal und Redintegration. Die DT-Methode erwies sich in besonderem Maße als nützlich durch die Aufnahme von Entwicklungsindikatoren, um die Entwicklung über einen breiteren Altersbereich zu beschreiben und zu vergleichen. Außerdem ermöglichte sie die Untersuchung der Relation mit anderen Entwicklungsindikatoren (hier: kognitive Kapazität und Wortschatzumfang). Wir möchten an dieser Stelle auf drei Innovationen hinweisen: Erstens hätte die Differenzierung zwischen einem verspäteten Entwicklungseinsatz und einer verlangsamten Entwicklungsrate mit dem traditionellen Gruppenmatching-Ansatz im ANOVA-Design nicht gezeigt werden können. Zweitens ergaben sich durch die DTs differenzielle entwicklungsbezogene Relationen für Schüler\*innen mit Lernbehinderung: Kognitive Kapazität sagte in keiner der beiden Gruppen die Rehearsal-Effektivität vorher, wenn man mögliche Ausreißer beachtete (in der vollen Stichprobe der Schüler\*innen mit Lernbehinderung gab es einen positiven Zusammenhang von kognitiver Kapazität und Rehearsal), aber der Wortschatzumfang sagte Redintegration negativ vorher. Diese detaillierten Muster hätten mit dem ANOVA-Ansatz ebenfalls nicht aufgedeckt werden können. Drittens erwies sich die Nutzung eines größeren Altersbereiches als sinnvoll für die Betrachtung des breiteren Abbildes der Entwicklung, anstelle des Vergleichs von Clustern in ausgewählten, engen Altersabschnitten.

Auch wenn der DT-Ansatz zahlreiche Vorteile bietet, ist darauf hinzuweisen, dass das Design dieser Studie querschnittlich war. Daher sollte der Begriff „Entwicklung“ nicht wörtlich aufgefasst werden, da wir verschiedene Probanden [nur einmal] aus einem breiteren Spektrum von chronologischem Alter, kognitiver Kapazität und Wortschatzumfang getestet haben. Diese querschnittliche Art der Daten ist vergleichbar mit der im üblichen MA-Matching Ansatz. Eine wichtige Unterscheidung liegt jedoch in der Berücksichtigung eines breiteren (mental) Altersbereiches, was als explizite Stärke des DT-Ansatzes verstanden werden kann (Thomas et al.,

2009). Dagegen wäre[, wie in Kapitel 3.2 ausgeführt,] ein längsschnittliches Design, das im Rahmen von DTs gut umgesetzt werden kann, in der Lage, eine „echte“ Entwicklung abzubilden. Die Nutzung von korrelativen Daten zum Zusammenhang zwischen einem Ziel-Task und anderen Variablen, die als Entwicklungsindikatoren herangezogen werden, stellt einen ersten Schritt in der Auswertung dar, der dann durch weitere, längsschnittliche Untersuchungen überprüft und validiert werden muss. Thomas et al. (2009, S. 336) weisen darauf hin, dass ein „ursprünglich querschnittliches Design [idealerweise kombiniert wird] mit einem längsschnittlichen Follow-up“, was die Aufgabe von zukünftigen Studien darstellt.

Es sei darauf hingewiesen, dass die Ursachen für das Auftreten eines bestimmten Entwicklungsmusters nicht vollständig klar sind [und aus den Daten auch nicht erschlossen werden können]. DTs wurden bisher vor allem genutzt als „ein deskriptiv stärkeres empirisches Vokabular“ und vielseitiges Werkzeug zur Beschreibung von Entwicklungsverläufen (Thomas et al., 2009, S. 355), das auch als „Theorie-neutraler Marker für untypische Erscheinungsformen“ (Thomas et al., 2009, S. 340) verstanden werden kann. Einen allgemeineren Blick auf den zugrundeliegenden neurokonstruktivistischen Rahmen in einem allgemeinen theoretischen Kontext erhält man bei Karmiloff-Smith (1998), findet jedoch keine spezifischen Annahmen darüber, welche Faktoren die (versteckten) Ursachen für differenzielle Entwicklungsmuster sein könnten.

### 5.5.2. Limitationen

Es sind eine Reihe von methodischen und theoretischen Limitationen zu nennen. Die erste Limitation betrifft die Abdeckung des Altersbereiches. Thomas et al. (2009) empfehlen, dass die unbeeinträchtigte Kontrollgruppe den gesamten Altersbereich vom jüngsten mentalen Alter bis zum höchsten chronologischen Alter abdecken sollte. Die Kontrollgruppe in unserer Stichprobe reichte jedoch nur bis zu einem maximalen Alter von 13;5 Jahren (wohingegen in der Untersuchungsgruppe auch ältere Kinder bis zu 17;1 Jahren vertreten waren). Aus diesem Grund könnte das Design ein mögliches Plateau oder andere Veränderung in der Entwicklungsrate der Kontrollgruppe nicht aufdecken. Auch wenn das die Erklärungskraft für den Bereich der älteren Schüler\*innen begrenzt, sind die allgemeinen Befunde von dieser Einschränkung weniger betroffen: Die Intercepts werden am Entwicklungseinsatz verglichen, d. h. beim jüngsten gemessenen Alter der Untersuchungsgruppe, wo ausreichend Daten [für ein differenziertes Bild] vorliegen. Dies ist in ähnlicher Weise der Fall für die Interpretation des Slope-Vergleiches: Der ausschlaggebende Vergleich bezieht sich auf die früheren Stadien der Entwicklung, was durch die jüngeren Schüler\*innen der Kontrollgruppe abgedeckt wird. Es erscheint sinnvoll, eine breitere Altersspanne in der Untersuchungsgruppe abzudecken, da es möglicherweise län-



gere Zeit benötigt, bis eine Entwicklung oder Veränderung [reliabel] messbar ist. Diese Limitation betrifft jedoch nur die Analysen, in denen das chronologische Alter als Prädiktor eingesetzt wird. Für die beiden Entwicklungsindikatoren kognitive Kapazität und Wortschatzumfang wurde eine Teilstichprobe der Kontrollgruppe zum Vergleich herangezogen, für welche die Instrumente normiert sind. Damit deckt die Kontrollgruppe den gesamten Leistungsbereich der Untersuchungsgruppe ab.

Zweitens müssen auch methodische Einschränkungen berücksichtigt werden. Da die Schüler\*innen mit Lernbehinderung nur aus Förderschulen mit dem Förderschwerpunkt Lernen rekrutiert wurden, sind die Befunde auf diese spezielle Population begrenzt und sollten nicht ohne weiteres auf die gesamte Population der Schüler\*innen mit Lernbeeinträchtigung [z. B. im gemeinsamen Unterricht] generalisiert werden. Möglicherweise war die Messinvarianz über die beiden Gruppen nicht in allen Instrumenten gegeben, was einen Einfluss auf die Ergebnisse haben kann. Auch könnte es sein, dass in beiden Gruppen Bodeneffekte einen negativen Einfluss auf die Reliabilität der Messungen hatten, insbesondere für einzelne Kinder jüngeren (mental) Alters.

Drittens ist bekannt, dass neben Wortlänge und Lexikalität auch viele weitere Faktoren einen Einfluss auf die Leistung des verbalen Arbeitsgedächtnisses haben können. Dies könnte ein Problem für die Validität des Wortlängeneffektes als theoretisch fundierter [und eindeutiger] Indikator eines phonologischen, zeit-basierten Rehearsalprozesses darstellen. Insbesondere haben Jalbert, Neath, Bireta et al. (2011) und Jalbert, Neath und Surprenant (2011) gezeigt, dass in sämtlichen Studien, die einen Wortlängeneffekt gefunden haben, die Test Items mit der Neighborhood Size konfundiert waren [siehe auch Kapitel 2.1.2.3]. [...] Laut Roodenrys (2009) kann der Neighborhood Size Effekt auch durch die Wirkung des Redintegrationprozesses erklärt werden (siehe auch Clarkson et al., 2017; Derragh et al., 2017). Unter experimenteller Kontrolle der Neighborhood Size konnten Jalbert, Neath und Surprenant (2011) zeigen, dass der Wortlängeneffekt verschwand oder sogar umgekehrt wirkte [d. h. kurze Wörter mit einer niedrigen Neighborhood Size konnten schlechter behalten werden als lange Wörter mit einer hohen Neighborhood Size]. Dieser Punkt sollte in weiteren Studien berücksichtigt werden, um Stimuli zu konstruieren, in denen diese Konfundierung von Neighborhood Size und Wortlänge vermieden wird.

### **5.5.3. Implikationen**

In der vorliegenden Studie wurde bei Kindern mit Lernbehinderung die Gesamtkapazität der phonologischen Schleife sowie die Effektivität von zwei Prozessen im Rahmen der PL, nämlich Rehearsal und Redintegration, untersucht. Das Wissen und Verständnis darüber, welche kogni-

tiven Prozesse diesen Schüler\*innen zur Verfügung stehen und in welchen Bereichen spezifische Schwierigkeiten auftreten können, ist sinnvoll und notwendig, um Unterricht und Instruktionen besser auf sie und ihre schwerwiegenden schulischen Probleme ausrichten zu können. Dieses Wissen kann dazu genutzt werden, um den Unterricht entsprechend anzupassen, indem z. B. Instruktionen kürzer formuliert werden, sodass ihr Verständnis nicht so stark von einem funktionierenden Rehearsalprozess abhängig ist.

Da theoretische Modelle zu den Ursachen von Lernbehinderung bisher nicht genauer angeben, welche kognitiven Prozesse beeinträchtigt sein könnten, stellt diese Studie einen Versuch dar, Licht auf die Verfügbarkeit von Kapazität und Prozessen des verbalen Arbeitsgedächtnisses zu werfen. Indem wir Developmental Trajectories als Analysestrategie einsetzten, konnten wir zeigen, dass Schüler\*innen mit Lernbehinderung anscheinend einen verspäteten Entwicklungseinsatz hinsichtlich ihrer PL-Kapazität und des Rehearsalprozesses haben. Die Entwicklungsrate scheint jedoch nicht beeinträchtigt zu sein, was darauf hindeutet, dass das Defizit sich zumindest nicht verschärft. Der Redintegrationprozess wurde für diese Gruppe von Schüler\*innen erstmals empirisch untersucht. Der Befund, dass der Redintegrationprozess intakt erscheint, kann als mögliche Ressource interpretiert werden und einen möglichen Ansatzpunkt für erfolgsversprechende Interventionen bieten. Die differenziellen Befunde zur entwicklungsbezogenen Relation zwischen Redintegration und Wortschatzumfang könnten jedoch darauf hinweisen, dass es einen Unterschied in der kognitiven Struktur gibt bezüglich der Fähigkeit, Informationen im Langzeitgedächtnis zu nutzen; dies erfordert jedoch zunächst weitere replizierende Studien.

## 6. Struktur des Langzeitgedächtnisses (Studie 2)

### 6.1. Einleitung und Fragestellung

In der ersten Studie (Kapitel 5) wurde das Arbeitsgedächtnis als eine (von mehreren) kognitive Lernvoraussetzung für den Wissenserwerb an Kindern mit Lernbehinderung untersucht. In der nun anschließenden zweiten Studie soll eine weitere kognitive Lernvoraussetzung betrachtet werden, nämlich die Struktur des Langzeitgedächtnisses. Dazu werden die Binnenstruktur und die Außenstruktur des mentalen Lexikons von Kindern mit Lernbehinderung und unbeeinträchtigten Kindern des gleichen mentalen Alters verglichen.

Nach dem Modell der Informationsverarbeitung (Klauer & Leutner, 2012; Mietzel, 2017), das in dieser Arbeit zur Erklärung von Wissenserwerb herangezogen wird, sind sowohl Langzeit- als auch Arbeitsgedächtnis als zentrale Komponenten am Wissenserwerb beteiligt. Wissenserwerb wird in diesem Modell als die Verknüpfung und Einbettung von neuem Wissen in bestehende Wissensstrukturen verstanden (Büttner & Hasselhorn, 2007; Verhoeven, 2007).

Angesichts dessen, dass in Erklärungsmodellen zu Lernbehinderung (Grünke & Grosche, 2014; Klauer & Lauth, 1997; Verhoeven, 2007) dem (bereichsspezifischen) Vorwissen häufig eine bedeutsame Rolle zugesprochen wird, ohne dass dazu aktuelle belastbare empirische Untersuchungen vorliegen, ist es für das theoretische Verständnis des Phänomens Lernbehinderung relevant, die bestehende Wissensstruktur genauer daraufhin zu untersuchen, ob sie sich bei Kindern mit und ohne Lernbehinderung des gleichen mentalen Alters unterscheidet. Die Erforschung der Organisation des Langzeitgedächtnisses von Kindern mit Lernbehinderung hat aber auch praktische Relevanz, da Lehrkräfte anhand dieser Kenntnisse ihren Unterricht und Instruktionen besser auf die Bedürfnisse der Kinder anpassen können.

Darüber hinaus bieten die Erkenntnisse zur Langzeitgedächtnis-Struktur einen zusätzlichen Nutzen für die Interpretation eines Ergebnisses aus der ersten Studie, das zunächst überraschend erschien: Spezifisch bei Kindern mit Lernbehinderung wurde eine unerwartete Interaktion von Redintegrationprozess und Wortschatzumfang festgestellt. Der Befund zum Redintegrationprozess bei Kindern mit Lernbehinderung aus der ersten Studie verdient besondere Aufmerksamkeit aus folgendem Grund: Im Verhältnis zu chronologischem Alter und kognitiver Kapazität wirkte Redintegration bei den Kindern mit Lernbehinderung auf den ersten Blick unbeeinträchtigt. Jedoch zeigte sich in der Interaktion mit dem Langzeitgedächtnis überraschenderweise ein differenzielles Muster, nämlich eine Abnahme der Redintegration-Effektivität mit wachsendem Wortschatz. Wie könnte dieser Interaktionseffekt erklärt werden?

Einerseits muss einer inhaltlichen Erklärung nachgegangen werden, dass die Interaktion durch Gruppen-Unterschiede in der Struktur des Langzeitgedächtnisses bedingt sein könnte.

Für diesen inhaltlichen Aspekt können die Befunde aus dieser zweiten Studie zur Interpretation herangezogen werden. Andererseits könnte auch ein methodisches Artefakt aufgrund von Differential Item Functioning vorliegen, sodass die Interaktion lediglich auf einem Messfehler beruht. Diese Vermutung wird in der anschließenden dritten Studie überprüft und ist hier nur zur Orientierung bereits erwähnt.

Es steht also auf der inhaltlichen Ebene in Studie 2 die Klärung der Rolle des Langzeitgedächtnisses für den Redintegrationprozess an, ob Unterschiede zwischen den Gruppen in der Wissensstruktur für die Interaktion verantwortlich sein könnten. Da Redintegration die Nutzung von Langzeitgedächtnis-Inhalten für Arbeitsgedächtnis-Aufgaben bedeutet, ist es für das Verständnis der kognitiven Probleme von Schüler\*innen mit Lernbehinderung aufschlussreich, die Struktur der Inhalte des Langzeitgedächtnisses näher zu betrachten. Wenn die Wissensstruktur im Langzeitgedächtnis bei Schüler\*innen mit Lernbehinderung von derjenigen bei unbeeinträchtigten Kindern abweicht, dann könnte darin eine mögliche inhaltliche Erklärung für das Auftreten der Interaktion von Redintegration und Wortschatzumfang liegen.

Eine solche Erklärung könnte lauten, dass Kinder mit Lernbehinderung aufgrund einer abweichenden Wissensstruktur weniger systematisch oder zuverlässig auf ihr gespeichertes Wissen zurückgreifen können, wenn sie es in einer zeitkritischen Aufgabe (wie bei der Wiedergabe von Arbeitsgedächtnis-Spannungsaufgaben) benötigen. Die Ergebnisse haben für die Interpretation der Interaktion aus Studie 1 damit theoretische Relevanz, da sie die Rolle der Wissensstruktur für das Auftreten der Interaktion aufklären; darüber hinaus haben sie praktische Relevanz, da sich für Lehrkräfte ein besseres Verständnis der Probleme von Kindern mit Lernbehinderung ableiten lässt: Sind sie womöglich aufgrund ihrer Wissensstruktur weniger gut in der Lage, ihr Vorwissen für die Bewältigung von arbeitsgedächtnislastigen Aufgaben zu nutzen? Daraus könnten auch Interventionen abgeleitet werden, die die Organisation und Strukturierung des Vorwissens trainieren.

Um die Wissensstruktur im mentalen Lexikon zu erforschen, werden die Daten des Tests zur Messung des Wortschatzumfangs, des WWT 6-10 (Glück, 2011), genauer betrachtet. Es handelt sich hierbei also um eine vertiefende Analyse von Daten, die im Rahmen von Studie 1 bereits berichtet wurden. Allerdings wurde dort lediglich der Gesamtscore des globalen Wortschatzumfangs zur Schätzung der Größe des verfüg- und abrufbaren Wissens im Langzeitgedächtnis berücksichtigt. Die folgenden Analysen sollen hinsichtlich einer möglicherweise differenziellen Organisation des Langzeitgedächtnisses bei Kindern mit Lernbehinderung stärker in die Tiefe gehen.

Mit der bereits in Kapitel 2.2 dargelegten Aufteilung von Studie 2 in (1) Binnenstruktur und (2) äußere Struktur des Langzeitgedächtnisses werden zwei Ansätze zur Operationalisierung und den Gruppenvergleich verfolgt. Daraus ergeben sich die beiden folgenden Teilfragestellungen bzgl. des semantischen Langzeitgedächtnisses von Kindern mit Lernbehinderung:

Die erste Teilfragestellung betrifft die Binnenstruktur des mentalen Lexikons, d. h. das Verhältnis der WWT-Items untereinander, und lautet: „Unterscheidet sich die mentale Organisation des Wortschatzes zwischen Kindern mit Lernbehinderung und unbeeinträchtigten Kindern des gleichen mentalen Alters?“ Zur Beantwortung dieser Fragestellung werden die Korrelationsmuster der Items durch Prüfung der Messäquivalenz zwischen den Gruppen mithilfe einer MG-CFA untersucht.

Die zweite Teilfragestellung betrifft die Außenstruktur, d. h. das möglicherweise differenzielle Verhältnis zu weiteren personen- und itembezogenen Variablen, und lautet: „Gibt es Unterschiede zwischen den Gruppen dahingehend, ob personen- und itembezogene Variablen die Lösungswahrscheinlichkeit vorhersagen?“ Diese Frage nach dem differenziellen Einfluss wird anhand der Interaktionen dieser Variablen mit dem Gruppenfaktor im Rahmen eines generalisierten gemischten linearen Modells (GLMM) beantwortet.

In Studie 2 wird für beide Teilfragestellungen eine jüngere unbeeinträchtigte Gruppe von Kindern als Kontrollgruppe herangezogen, die hinsichtlich der mentalen Altersvariablen, nämlich kognitiver Kapazität und allgemeinem Wortschatzumfang, mit der Gruppe der Kindern mit Lernbehinderung vergleichbar ist. Dadurch lässt sich die Fragestellung beantworten, ob Kinder mit Lernbehinderung in der Organisation ihres Langzeitgedächtnisses eine abweichende Struktur aufweisen. Da bislang keine Forschungsergebnisse zu Binnen- und Außenstruktur des semantischen Gedächtnisses bei Kindern mit Lernbehinderung vorliegen, können keine gerichteten Hypothesen abgeleitet werden.

Gemäß der Development-Difference Kontroverse und der Logik des mental age Matching Ansatzes (siehe Kapitel 3.1.2) kann ein nicht-signifikantes Ergebnis als empirischer Hinweis auf eine Entwicklungsverzögerung gedeutet werden. Bei einer Entwicklungsverzögerung wäre das mentale Lexikon bei Kindern mit Lernbehinderung nur in dem Maße beeinträchtigt, wie es aufgrund des allgemein geringeren mentalen Alters zu erwarten wäre. Wenn Unterschiede zur MA-Gruppe zu beobachten sind, wäre dies ein Hinweis auf eine strukturell abweichende Entwicklung, da die Gedächtnis-Struktur von derjenigen der nach mentalem Alter gematchten Kontrollgruppe abweicht.

Allerdings muss an dieser Stelle die Aussagekraft des CFA-Ansatzes (erste Teilfragestellung) dahingehend eingeschränkt werden, dass sich nur ein nicht-signifikantes Ergebnis, d. h. erfüllte Messinvarianz, eindeutig interpretieren lässt (nämlich als Entwicklungsverzögerung). Wenn die

Messinvarianz nicht bestätigt werden kann, so ist dies lediglich ein empirischer Hinweis darauf, dass es zwischen den Gruppen (ungerichtete) Unterschiede in der Struktur des Langzeitgedächtnisses gibt – nicht jedoch, dass die Struktur des Langzeitgedächtnisses bei Kindern mit Lernbehinderung ‚schlechter‘ ausgeprägt ist als bei unbeeinträchtigten Kindern.

Stattdessen wäre es genauso gut möglich, dass die Struktur einer nach chronologischem Alter (CA) gematchten Kontrollgruppe viel eher vergleichbar ist. Die Erhebung einer nach chronologischem Alter vergleichbaren Gruppe war jedoch nicht möglich oder sinnvoll, da der WWT 6-10 nur für einen Altersbereich von ca. 10 Jahren sensitiv und normiert ist. Eine ältere Gruppe hätte hier starke Deckeneffekte erzielt. Insofern beschränkt sich der Vergleich auf die MA-Gruppe, sodass hier die Erklärungskraft des CFA-Ansatzes eingeschränkt werden muss. Diese Limitation gilt jedoch nicht für den zweiten Ansatz der Außenstruktur, denn hier lässt sich die Richtung der Abweichung bestimmen.

## 6.2. Methode

### 6.2.1. Stichprobe und Material

Die Stichprobe für diese Studie war im Wesentlichen identisch mit der aus Studie 1. Die Gruppe der Schüler\*innen mit Lernbehinderung ( $n = 93$ ) umfasste sechs zusätzliche Proband\*innen, die aus Studie 1 aufgrund fehlender Werte in den Arbeitsgedächtnis-Aufgaben ausgeschlossen werden mussten. Außerdem wurden in die jüngere, nach mentalem Alter vergleichbare unbeeinträchtigte Kontrollgruppe (insgesamt  $n = 114$ ) zwölf bilinguale Kinder zur Erhöhung der statistischen Power mit aufgenommen, die in Studie 1 aufgrund des monolingualen Status nicht berücksichtigt wurden. Für die bilingualen Kinder wurden dabei identische Einschlusskriterien angelegt, dass die Intelligenz und der Wortschatzumfang unbeeinträchtigt waren (CFT 1-R IQ > 85; WWT 6-10  $T$ -Wert > 40). Der Nachteil, dass die Kontrollgruppe nicht einheitlich monolingual ist, wird daher zugunsten einer größeren Stichprobe weniger stark gewichtet.

Tabelle 3 gibt die deskriptiven Angaben zur Stichprobenbeschreibung wieder. Es zeigten sich keine signifikanten Unterschiede auf Gruppenebene in den beiden Matching-Kriterien kognitive Kapazität (gemessen am Rohwert des CFT 1-R;  $p = .899$ ) und expressiver Wortschatzumfang (gemessen am Rohwert des WWT 6-10;  $p = .599$ ) und in der Geschlechterverteilung ( $p = .082$ ). Die Unterschiede zwischen den Gruppen im chronologischen Alter und dem Norm-Intelligenzwert waren erwartungsgemäß signifikant (beide  $p < .001$ ), da es sich um eine jüngere Kontrollgruppe handelt, die auf Ebene der Rohwerte von CFT und WWT gematcht ist. Die Rohwerte entsprechen in der Kontrollgruppe durchschnittlichen (d. h. unbeeinträchtigten)

Normwerten, in der älteren Untersuchungsgruppe der Kinder mit Lernbehinderung dagegen unterdurchschnittlichen Normwerten.

Tabelle 3

*Stichprobenmerkmale und deskriptive Angaben aller Variablen (M und SD) in Studie 2*

	LB (n = 93)		KG (n = 114)		Signifikanz
	M	SD	M	SD	
Geschlecht (männlich/weiblich)	53/40		50/64		$\chi^2(1) = 3.026$ ; $p = .082$ ; $\phi = .12$
Alter (Jahre; Monate)	13;2 <sup>a</sup>	2;3	8;9	1;1	$t(119.92) = 20.805$ ; $p < .001$ ; $d = 3.16$
Kognitive Kapazität (CFT 1-R Rohwerte)	65.8 <sup>b</sup>	11.8	65.6	12.9	$t(187.04) = 0.127$ $p = .899$ ; $d = 0.02$
Wortschatzumfang (WWT 6–10 Rohwerte)	53.4	16.8	54.6	14.1	$t(179.89) = 0.526$ ; $p = .599$ ; $d = 0.07$
Intelligenz (IQ Normwerte)	76.0 <sup>a</sup>	8.4	105.4	9.9	$t(200.92) = 22.869$ ; $p < .001$ ; $d = 3.16$
<u>Wortschatz nach Wortarten</u>					
Nomen	14.26	5.56	13.75	4.48	$t(175.3) = 0.707$ ; $p = .481$ ; $d = 0.10$
Verben	13.96	3.90	14.66	3.23	$t(178.1) = 1.389$ ; $p = .167$ ; $d = 0.20$
Adjektive	12.91	5.04	13.33	4.93	$t(194.97) = 0.601$ ; $p = .548$ ; $d = 0.08$
Kategorien	12.29	4.54	12.82	4.11	$t(187.87) = 0.879$ ; $p = .381$ ; $d = 0.12$

*Anmerkungen.* LB: Gruppe Schüler\*innen mit Lernbehinderung, KG: Kontrollgruppe. Die Rohwerte des Culture-Fair-Test 1 Revision (CFT 1-R; Weiß und Osterland, 2012) wurden zur Erfassung der kognitiven Kapazität verwendet; die Rohwerte des Wortschatz- und Wortfindungstest für 6-10 Jährige (WWT 6-10, Glück 2011) wurden zur Erfassung des Wortschatzumfangs verwendet; Intelligenznormwerte wurden aus verschiedenen Quellen ermittelt: für Schüler\*innen mit LB aus den offiziellen Schulakten, für jüngere KG-Schüler\*innen wurden die Normwerte des CFT 1-R herangezogen. Aufgrund ungleicher Varianzen werden korrigierte Freiheitsgrade in den *t*-Tests berichtet.

<sup>a</sup> n = 90. <sup>b</sup> n = 84

Die Testmaterialien entsprachen denen aus Studie 1 (siehe Kapitel 5.3.2) und umfassten den CFT 1-R (Weiß & Osterland, 2012) und die Wortspannen-Aufgaben mit echten Wörtern und Kunstwörtern jeweils mit kurzen und langen Wörtern, angelehnt an das Vorgehen in der AGTB 5-12 (Hasselhorn et al., 2012). Die IQ-Normwerte der Kinder mit Lernbehinderung entstammten den offiziellen Schulakten, für die Kinder der Kontrollgruppe wurden die Normwerte aus den Ergebnissen des CFT 1-R gebildet. Für den Wortschatz kam der WWT 6-10 (Glück, 2011) zum Einsatz.

Da in dieser Studie die Ergebnisse des WWT 6-10 an zentraler Stelle stehen, wird hier detaillierter darauf eingegangen. Der Wortschatz- und Wortfindungstest für 6- bis 10-Jährige (Glück, 2011) enthält einen rezeptiven und einen expressiven Subtest, jeweils als Kurz- und Langform. Im Rahmen der vorliegenden Studie wurde der expressive Subtest (d. h. Wortbenennung) in seiner Langform, bestehend aus insgesamt 95 Items, durchgeführt.

Zu jedem Item wird ein Bild gezeigt, das vom Kind benannt werden soll; dazu wird ein verbaler Hinweis vom PC abgespielt. Die Items lassen sich in vier Wortarten einteilen, zu denen in Tabelle 4 jeweils ein leichtes, mittleres und schwieriges Item als Beispiel angegeben wird, siehe auch Abbildung 5:

Tabelle 4  
*Beispiele für Items im WWT 6-10 nach Wortarten*

Wortart (Abkürzung)	Anzahl Items	Beispiele	verbaler Hinweisreiz
Nomen (N), Teile von Gegenständen	$k = 26$	„Geländer“, „Mantel“, „Riegel“	„Was ist das?“
Verben (V)	$k = 23$	„schälen“, „verbeugen“, „entgleisen“	„Was macht er/sie?“; „Was ist hier passiert?“
Kategorien (K)	$k = 23$	„Instrumente“, „Getreide“, „Gepäck“	„Wie nennt man das alles zusammen?“; „Was sind das alles?“
Adjektive (A)	$k = 23$	„weich“, „glatt“, „einfarbig“	„Was ist das Gegenteil von X?“

Die Testung beginnt mit einer allgemeinen Einführung und Instruktion, danach wird zunächst pro Wortart ein leicht zu lösendes Probeitem dargeboten. Durch die Bearbeitung dieser vier Übungsisems wird das Vorgehen für den Test geübt und das Verständnis der Instruktionen überprüft bzw. bei Bedarf wiederholt und erläutert. Außerdem erfüllen die Übungsisems die Funktion eines Abbruchkriteriums: wenn weniger als drei der vier Übungsisems richtig beantwortet werden, bricht der Test ab, da er für das Kind vermutlich zu schwierig ist (Glück, 2011).

Pro Item wird eine Zeitspanne von maximal 15 Sekunden gewährt, innerhalb derer eine Antwort zu erfolgen hat; bleibt diese aus, dann wird das Item automatisch als falsch gewertet und „keine Antwort“ festgehalten. Die Äußerungen des Kindes wurden in der Datenerhebung von der Testleitung direkt im PC dokumentiert. War die Antwort des Kindes eindeutig korrekt, genügte die Markierung als „richtig“ (Taste R); bei falschen oder nicht eindeutig korrekten (unsicheren) Antworten wurde die Antwort im Wortlaut eingetragen und vom Programm automatisch anhand einer hinterlegten Datenbank überprüft. Die Datenbank enthält Alternativantworten, die ebenfalls als korrekt gewertet werden; außerdem enthält sie eine Liste von häufig vorkommenden falschen Antworten. Für eine differenzierte Fehleranalyse kann die hinterlegte Fehlerart von falschen Antworten (z. B. „Übergeneralisierung“) ausgegeben werden.



Die übliche Dauer für den gesamten Durchlauf liegt bei ca. 20 Minuten; eine Pause ist laut Manual nicht vorgesehen (Glück, 2011).

Bei einigen Items gibt es skalenspezifische Besonderheiten. Zum einen weist in manchen Items zu Nomen und Verben ein Pfeil darauf hin, dass nur ein Teil (z. B. das Dach eines gezeigten Hauses) bzw. ein Aspekt des Bildes gemeint ist. Bei den Adjektiven erfolgt, wie im Theorieteil in Abschnitt 2.2.2.2 dargestellt, eine abweichende Präsentation des Stimulus. Der visuelle Stimulus ist kein spezifischer Hinweis auf das Zielwort, sondern es wird nur ein allgemeines Bild dargeboten (siehe Abbildung 5d). Dieses Bild ist für alle Adjektive gleich und zeigt ein Quadrat auf gelbem Grund, das diagonal in eine schwarze und eine weiße Hälfte geteilt ist. Dass aus dieser abweichenden Präsentation eine andersartige Aufgabenstellung erwächst, wurde in Abschnitt 2.2.2.2 gezeigt.

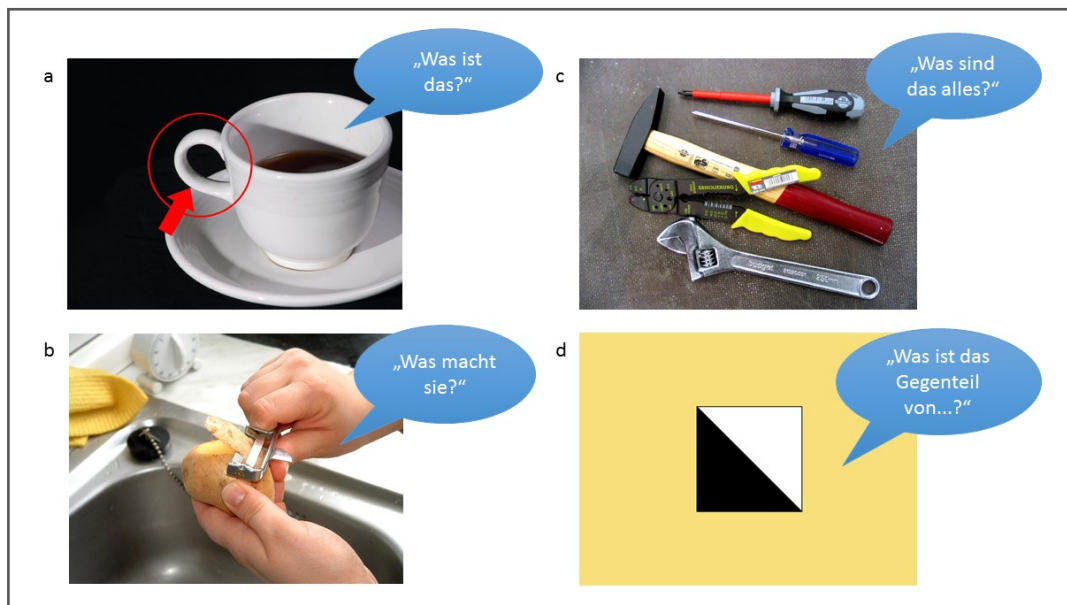


Abbildung 5. Beispiele für Items im WWT 6-10. a) Nomen, b) Verben, c) Kategorien, d) Adjektive. Aus: Glück (2011). Wortschatz- und Wortfindungstest für 6- bis 10-Jährige, 2. Auflage © Elsevier GmbH, Urban & Fischer, München. Abdruck erfolgt mit freundlicher Genehmigung des Verlages.

### 6.2.2. Auswertungsstrategien

Für die Auswertung der Daten zum Zweck des Gruppenvergleichs im Wortschatz wurden drei Ansätze in dieser Arbeit herangezogen. Zwei Ansätze (nämlich MG-CFA und GLMM) kamen in Studie 2 zur Anwendung, der dritte Ansatz (DIF) sei der Vollständigkeit halber an dieser Stelle bereits als Ausblick für Studie 3 (Kapitel 7) erwähnt.

- (1) Für die Bewertung der grundsätzlichen Messinvarianz auf Konstruktebene zum Vergleich des Messmodells („Misst der WWT für beide Gruppen ähnlich?“) wurde eine MG-CFA unter Nutzung der Item-Parceling Strategie herangezogen (Little et al., 2013).

- (2) Der Einfluss weiterer Variablen und deren Interaktion mit der Gruppenzugehörigkeit („Lässt sich die Lösungswahrscheinlichkeit eines Items durch weitere personen- und itembezogene Variablen und die Gruppenzugehörigkeit erklären?“) wurde in einem GLMM-Ansatz überprüft (Baayen et al., 2008).
- (3) Um Aussagen über die Vergleichbarkeit auf Itemebene zu ermöglichen („Gibt es Items, die für eine Gruppe systematisch zu leicht oder schwer sind?“), wurde in Studie 3 vor dem Hintergrund eines IRT-Frameworks das Differential Item Functioning (DIF) der Wörter betrachtet (Koller, Alexandrowicz & Hatzinger, 2012).

Die Auswertung im Rahmen des ersten Ansatzes steht vor der methodischen Herausforderung, dass die Stichprobe nur ca. 100 Personen pro Gruppe (Kinder mit Lernbehinderung und jüngere unbeeinträchtigte Kinder) umfasst. Für die Betrachtung des WWT 6-10 auf Itemebene mittels einer CFA wäre jedoch eine deutlich größere Stichprobe notwendig. Die (heuristischen) Empfehlungen in der Literatur für eine CFA reichen von einem liberalen Probanden-Item-Verhältnis von 3:1 bis hin zu einer strengeren Empfehlung von 10:1 (siehe MacCallum, Widaman, Zhang & Hong, 1999). Mit dem vorliegenden Datensatz von ca. 100 Personen pro Teilstichprobe und  $k = 95$  Items ist dieses Verhältnis deutlich unterschritten.

Dass die Stichprobe für die Untersuchung dennoch unverändert blieb, hatte verschiedene Gründe. Zunächst war die Gruppengröße ausreichend für den Vergleich auf Ebene von Mittelwerten, wie in Studie 1 erfolgt<sup>24</sup>. Bei Vergrößerung der Stichprobe auf mehrere Hundert Probanden pro Gruppe hätte der Zeitaufwand angesichts einer gesamten Testdauer von ca. 1,5 Stunden je Proband\*in ein unververtretbares Ausmaß angenommen. Zudem stellt es durchaus einen Vorteil dar, dass das Langzeitgedächtnis in der gleichen Stichprobe wie das Arbeitsgedächtnis (siehe Studie 1) untersucht werden kann, weil so Unterschiede in den beiden Gedächtnis-komponenten (Arbeits- und Langzeitgedächtnis) besser interpretiert werden können.

Um dieser methodischen Limitation zu begegnen und die Auswertung der Binnenstruktur mit MG-CFA auch ohne eine aufwändige Nachrekrutierung von Schüler\*innen zu ermöglichen, wurde auf die Item-Parceling Strategie zurückgegriffen, bei der die Anzahl der manifesten Variablen und damit der zu schätzenden Parameter reduziert wird. Das Vorgehen wird in Kapitel 6.2.3.3 detaillierter beschrieben. Bei den anderen beiden Ansätze (GLMM und IRT) wirkt sich die Stichprobengröße nicht so gravierend aus, da hier deutlich weniger Parameter geschätzt werden müssen als bei der MG-CFA; die Zusammenfassung zu Item-Parcels ist dort also nicht notwendig.

---

<sup>24</sup> Auch für einen kleinen bis mittleren Effekt von  $d = 0.35$  bei einem Signifikanzniveau von  $\alpha = .05$  für eine gerichtete Hypothese und den gegebenen Gruppengrößen ( $n_B = 93$  und  $n_{TD} = 114$ ) lag der Beta-Fehler bei 20% ( $\beta = 0.197$ ) und entsprechend wurde eine ausreichende Power von .803 erreicht.

Zwei weitere Varianten zum Umgang mit dem ungünstigen Probanden-Item-Verhältnis wurden verworfen: Ein Bootstrapping-Verfahren, in dem über mehrere hundert Iterationen pro Subskala eine festgelegte Anzahl von Items zufällig gezogen und über diese partiellen Subskalen die Messinvarianz-Überprüfung durchgeführt wird, wurde erprobt, erwies sich aber aufgrund häufig auftretender Konvergenzprobleme der Modelle als unbrauchbar. Ebenfalls im Rahmen dieser Arbeit nicht verwendet wurde der Ansatz der Exploratory Graph Analysis (Christensen, 2020; Golino & Epskamp, 2017). Dieser erscheint zwar vielversprechend, würde jedoch aufgrund der theoretischen Grundlagen der Netzwerk- und Graphentheorie sowie der technischen Umsetzung den Rahmen dieser Arbeit übersteigen. In den nächsten Abschnitten wird die Umsetzung der ersten beiden gewählten Strategien näher erläutert, die Betrachtung der dritten Strategie folgt in Studie 3 (Kapitel 7).

### 6.2.3. Methode zu Fragestellung 1: Binnenstruktur

Die Binnenstruktur des mentalen Lexikons wird anhand des Messmodells im Rahmen einer CFA operationalisiert, der Gruppenvergleich zwischen Kindern mit und ohne Lernbehinderung erfolgt mithilfe der Prozedur zur Überprüfung von Messinvarianz<sup>25</sup>. Im Messmodell stellen die Wortarten (Nomen, Verben, Kategorien, Adjektive) die latenten Variablen dar; als manifeste Variablen werden pro Wortart fünf Item-Parcels gebildet.

Die Messinvarianz (*measurement invariance*, z. B. Hirschfeld & von Brachel, 2014; Putnick & Bornstein, 2016) wird mit einer Multigroup-CFA überprüft, welche testet, ob ein vorab theoretisch festgelegtes Messmodell über mehrere Gruppen hinweg gleichermaßen gilt. Als Messmodell wird die Zuweisung von gemessenen, manifesten Variablen (auch Indikatorvariable genannt, Backhaus et al., 2013) zu nicht messbaren, latenten Variablen bezeichnet.

Da die Zuordnung der manifesten zu latenten Variablen und die Restriktion weiterer Parameter des Messmodells anhand theoretischer Vorannahmen erfolgen, ist die CFA ein strukturprüfendes Verfahren. Nicht-restringierte Parameter werden frei geschätzt, z. B. durch ein Maximum-Likelihood (ML) Verfahren (Backhaus et al., 2013). Das Ziel der Modellanpassung ist es, die Parameter so zu wählen bzw. zu schätzen, dass die Modellpassung zu den empirischen Daten

---

<sup>25</sup> Häufig wird die Messinvarianz für psychometrische Fragestellungen untersucht als Voraussetzung dafür, Vergleiche von Gruppenmittelwerten ziehen zu können (Putnick & Bornstein, 2016; Schwab & Helm, 2015). Mithilfe einer MG-CFA wird die (häufig nur implizit getroffene) Annahme geprüft, dass die Messergebnisse eines Testinstrumentes für verschiedene Gruppen tatsächlich die gleiche inhaltliche Bedeutung haben. Damit ist gemeint, dass sich gleiche latente Fähigkeiten von Personen aus verschiedenen Gruppen in einem gleichen oder sehr ähnlichen Rohwert widerspiegeln (Hirschfeld & von Brachel, 2014). In dieser Arbeit steht jedoch nicht die psychometrische Frage nach der Messgüte des Instrumentes im Vordergrund, sondern die inhaltliche Aussage, ob sich die Gruppen hinsichtlich der Binnenstruktur unterscheiden.

möglichst hoch ist. Eine hohe Modellpassung bemisst sich in einer möglichst geringen Diskrepanz zwischen der Varianz-Kovarianzmatrix, die aus den geschätzten Modell-Parametern resultiert, und derjenigen aus den empirischen Daten (Cheung & Rensvold, 2002).

In den folgenden Abschnitten werden das Konzept und die Berechnung der Multigroup-CFA erläutert. Dazu gibt der erste Abschnitt einen Überblick über die Indizes, die zur Beurteilung der allgemeinen und relativen Modellpassung herangezogen werden. Im zweiten Abschnitt wird die Überprüfung der Messinvarianz auf mehreren, zunehmend restringierten Stufen eingeführt. Der dritte Abschnitt erläutert die Item-Parceling Strategie, die das Problem der im Verhältnis zur Itemanzahl geringen Stichprobengröße abmildert. Darauf folgt im vierten Schritt die konkrete Spezifikation des Messmodells in dieser Studie. Abschließend wird als Ergänzung die parcel-allocation Methode vorgestellt, mit der abgesichert werden kann, dass es sich bei den Messinvarianz-Befunden nicht um zufällige Ergebnisse handelt, die lediglich aus der spezifischen Zusammensetzung der Parcels resultieren.

#### 6.2.3.1. Indizes für die Modellpassung

Zur Beurteilung der Modellpassung, auch Modellfit genannt, werden inferenzstatistische Indizes ( $\chi^2$ -Test und RMSEA; Backhaus et al., 2013; Cheung & Rensvold, 2002) sowie alternative deskriptive, sog. goodness-of-fit Indizes herangezogen (z. B. CFI, TLI; Cheung & Rensvold, 2002). Diese Indizes und entsprechende Konventionen, wann ein Modell als „gut“ anzunehmen oder als „schlecht“ abzulehnen ist, werden im Folgenden erläutert. Die Modellpassung kann (1) absolut für ein einzelnes Modell und (2) relativ in einem Modellvergleich beurteilt werden. Bei einem relativen Modellvergleich wird geprüft, ob die Aufnahme eines Parameters oder einer Restriktion zu einer signifikant besseren oder schlechteren Modellpassung führt.

Bei einem Likelihood Ratio- bzw.  $\chi^2$ -Test (Backhaus et al., 2013; Cheung & Rensvold, 2002) wird die Nullhypothese getestet, dass die aus den theoretischen Parametern resultierende Varianz-Kovarianzmatrix nicht von der empirisch beobachteten Varianz-Kovarianzmatrix abweicht (Backhaus et al., 2013; Cheung & Rensvold, 2002). Als zusätzliche Prüfgröße hat sich das Verhältnis des  $\chi^2$ -Wertes zu den Freiheitsgraden etabliert; ein Verhältnis von  $\chi^2/df < 2.5$  wird dabei als Indikator für einen angemessenen Modellfit betrachtet (Backhaus et al., 2013, S. 147).

Für den Root Mean Square Error of Approximation (RMSEA; Backhaus et al., 2013) deuten niedrigere Werte auf eine höhere Modellpassung hin. Brown und Cudeck (1993) schlagen (wohl-gemerkt basierend auf ihrer „subjektiven Erfahrung“, S. 146) folgende Richtwerte vor: der Modellfit gilt als „gut“ für Werte von  $RMSEA \leq .05$ ; als „akzeptabel“ gelten Werte bis  $RMSEA \leq .08$ ; und schließlich sind Werte von  $RMSEA \geq .10$  „inakzeptabel“. Für den RMSEA lässt sich die Irrtumswahrscheinlichkeit berechnen, dass ein  $RMSEA \leq .05$  zufällig zustande gekommen ist (Backhaus et al., 2013). Ein signifikantes Ergebnis bedeutet, dass der RMSEA mit

sehr niedriger Wahrscheinlichkeit zufällig kleiner als .05 ist, sodass auf eine gute Modellpassung geschlossen werden kann.

Neben den inferenzstatistischen Indizes stellen der comparative fit index (CFI; Bentler, 1990) und der Tucker-Lewis-Index (TLI; Tucker & Lewis, 1973) häufig gewählte goodness-of-fit Indizes (GFI) dar (Xia & Yang, 2019; für eine Übersicht und Performance-Vergleich von insgesamt 20 GFIs siehe Cheung & Rensvold, 2002). GFIs sind unabhängig von der Stichprobengröße (Cheung & Rensvold, 2002) und weniger anfällig bei nicht erfüllten Annahmen (Backhaus et al., 2013). Da für GFIs keine Verteilungsannahmen vorliegen, können sie jedoch nicht inferenzstatistisch ausgewertet werden, sondern es liegen lediglich Richtwerte und Konventionen vor. Die Indizes CFI und TLI sollten nach gängigen Konventionen (Hu & Bentler, 1999)  $CFI \geq .95$  und  $TLI \geq .95$  betragen, um für eine „gute“ Passung zu sprechen (auch wenn häufig eine liberalere Grenze von  $CFI \geq .90$  angelegt wird, z. B. Wu, Li & Zumbo, 2007; siehe auch Lance, Butts & Michels, 2006).

Bei der relativen Beurteilung der Modellpassung wird die Veränderung der Fit-Indizes betrachtet, wenn man ein restringiertes Modell mit einem weniger restringierten Modell vergleicht. Das restringierte Modell bildet mehr bzw. stärkere theoretische Annahmen dadurch ab, dass Parameter vorab fixiert werden, während diese im unrestringierten Modell durch die ML-Prozedur frei geschätzt werden können. Aufgrund der freien Schätzung der Parameter im unrestringierten Modell weist dieses eine höhere Passung zu den Daten auf. Die Frage bei einem relativen Modellvergleich ist, ob die Aufnahme von theoretischen Restriktionen zu einer deutlichen Verschlechterung der Modellpassung im Vergleich zum weniger restringierten Modell führt.

Der relative Modellvergleich wird zur Feststellung der Messinvarianz eingesetzt, d. h. zur Überprüfung, ob das Messmodell für verschiedene Gruppen ähnlich ist. Dazu wird für jede Gruppe ein getrenntes Messmodell geschätzt und ein gemeinsames Messmodell mit der Restriktion, dass bestimmte Parameter des Modells in beiden Gruppen äquivalent sind. Das Messmodell, in dem die Parameter für beide Gruppen gleichgesetzt werden, darf nicht deutlich schlechter zu den empirisch erfassten Daten passen als das Modell, in dem für jede Gruppe die Parameter eigenständig geschätzt werden (Cheung & Rensvold, 2002).

Der relative Modellvergleich erfolgt anhand der Differenzwerte der Fit-Indizes. Hierfür stehen wiederum inferenzstatistische Tests (für  $\chi^2$  der Likelihood Ratio Test, sowie RMSEA) für genestete, d. h. ineinander geschachtelte, Modelle sowie GFI-Indizes zur Verfügung.

Der Likelihood Ratio (LR) Test prüft, ob die Modellpassung des restringierten Modells im Vergleich zu einem weniger restringierten Modell signifikant schlechter ist. Dies geschieht an-

hand der Differenz der  $\chi^2$ -Werte, die zu den durch die Restriktion freiwerdenden Freiheitsgraden ins Verhältnis gesetzt werden. Die Modelle sind genestet, da sie die gleiche Struktur haben bis auf einen Parameter (bzw. eine Gruppe von Parametern), der im restringierten Modell (aufgrund theoretischer Annahmen) fixiert ist, während er im vollen Modell frei geschätzt wird. Ein signifikantes Ergebnis ( $\chi^2$ -Test der Differenz) im Modellvergleich bedeutet also eine deutlich schlechtere Modellpassung des restringierten Modells (Putnick & Bornstein, 2016), was so zu interpretieren ist, dass die zugrundeliegende theoretische Annahme mit den empirischen Daten nicht vereinbar ist.

Angewendet auf den Kontext der Messinvarianz-Prüfung bedeutet ein signifikantes Ergebnis im Modellvergleich, dass die Restriktion von zwischen den Gruppen äquivalenten Parametern zu einer deutlichen Verschlechterung der Modellpassung führt. Damit muss davon ausgegangen werden, dass das spezifizierte Messmodell nicht für beide Gruppen gleich gut passt, sondern dass unterschiedliche Messmodelle die Daten in den Gruppen besser beschreiben.

Neben der Signifikanzprüfung der  $\chi^2$ -Differenz kann auch das Verhältnis der  $\chi^2$ -Differenz zu den Freiheitsgraden ( $\chi^2/df < 3$ ) für den Modellvergleich herangezogen werden (z. B. Milfont & Fischer, 2010). Auch für die aus einem Modellvergleich resultierenden Differenzwerte der alternativen GFIs (hier CFI und TLI) liegen lediglich Konventionen und Richtgrößen als Empfehlungen vor. So empfehlen Cheung & Rensvold (2002) als Richtgröße, dass die Differenz der CFI- und TLI-Werte zweier Modelle den Wert von .01 nicht übersteigen sollte.

#### 6.2.3.2. Stufen der Messinvarianz

Die Messinvarianz wird mittels einer MG-CFA auf mehreren Stufen überprüft. In jeder Stufe werden schrittweise mehr Restriktionen (dass die Parameter für beide Gruppen äquivalent sein müssen) aufgenommen. So wird für ein Messmodell überprüft, ob sich die Struktur signifikant zwischen den Gruppen unterscheidet (Putnick & Bornstein, 2016), d. h., ob die jeweilige Restriktion zu einer signifikant schlechteren Modellpassung führt. Im weniger restringierten Modell ist es erlaubt, dass die Parameter innerhalb der Gruppen frei geschätzt werden, sodass die Modellpassung innerhalb jeder Gruppe möglichst groß ist. Im restringierten Modell werden die entsprechenden Parameter in beiden Gruppen gleichgesetzt. Nunmehr lautet die Anforderung an die Modellschätzung, Parameterwerte zu finden, die die empirischen Daten in beiden Gruppen gleichermaßen möglichst gut abbilden, ohne dass sie pro Gruppe frei variieren dürfen.

Die Restriktion führt zwar fast zwangsläufig zu einer etwas schlechteren Modellpassung, die Frage ist jedoch, ob diese Verschlechterung des Modellfits signifikant ausfällt (siehe relative Modellvergleiche im vorigen Abschnitt). Wenn eine signifikante Verschlechterung des Modellfits festgestellt wird, bedeutet das, dass in den Gruppen anscheinend die Parameter unterschiedlich geschätzt werden müssten, um die empirischen Daten angemessen abzubilden.

Für die Abfolge der Aufnahme von Parameter-Restriktionen wird in der Literatur folgende Reihenfolge vorgeschlagen (Hirschfeld & von Brachel, 2014; Putnick & Bornstein, 2016; Vandenberg & Lance, 2000):

- (1) Auf Ebene der *konfiguralen* Invarianz sind Faktorenanzahl und Ladungsmuster äquivalent.
- (2) Auf Ebene der *metrischen* (oder ‚schwachen‘) Invarianz ist die Stärke der Faktorladungen äquivalent.
- (3) Auf Ebene der *skalaren* (oder ‚starken‘) Invarianz sind die Item-Intercepts äquivalent.
- (4) Auf Ebene der *residualen* (oder ‚strikten‘) Invarianz sind die Residualvarianzen der Items äquivalent.

Im Folgenden werden die Stufen näher erläutert, danach wird am Beispiel eines WWT-Items „Kosmetika“ aus der Subskala Kategorien illustriert, wie es zu interpretieren wäre, wenn die jeweilige Stufe der Invarianz für die Gruppe der Kinder mit Lernbehinderung nicht erfüllt wäre (das fiktive Beispiel dient jedoch nur zur Veranschaulichung, da in der Auswertung Item-Parcels als manifeste Variablen eingesetzt wurden).

Im ersten Schritt wird das konfigurale Modell geschätzt und als Baseline- oder Null-Modell für den Vergleich mit dem metrischen Modell verwendet. Das konfigurale Modell legt als einzige Restriktion fest, dass die Form des Messmodells in beiden Gruppen gleich ist. Inhaltlich bedeutet dies, dass die Items in beiden Gruppen den gleichen Faktoren zugeordnet sind, die Items also in beiden Gruppen auf die gleichen latenten Variablen laden (Putnick & Bornstein, 2016). Falls im WWT keine konfigurale Invarianz vorläge, könnte dies bedeuten, dass „Kosmetika“ in der Kontrollgruppe auf den Faktor Kategorien, in der Gruppe der Kinder mit Lernbehinderung aber (stärker) auf den Faktor Nomen lädt. Wenn die konfigurale Invarianz erfüllt ist, kann die zweite Stufe, die metrische Invarianz, überprüft werden.

Gegen das konfigurale Modell wird in der zweiten Stufe, der metrischen Invarianz, ein Modell getestet, bei dem nicht nur die Zuordnung der Items zu den Faktoren gleich ist, sondern zusätzlich die Ladungsparameter der Items auf die Faktoren für die Gruppen gleichgesetzt sind. Inhaltlich bedeutet dies, dass die Items in beiden Gruppen gleich stark mit den jeweiligen latenten Konstrukten zusammenhängen, diese also in ähnlicher Weise repräsentieren (Cheung & Rensvold, 2002; Putnick & Bornstein, 2016). Schwab und Helm (2015, S. 178) interpretieren diese Stufe so, „dass in den Subpopulationen die Konstrukte die gleiche inhaltliche Bedeutung haben“. Falls metrische Invarianz im WWT nicht vorläge, könnte dies bedeuten, dass „Kosmetika“ bei Kindern mit Lernbehinderung weniger stark mit der latenten Fähigkeit, Kategorien zu benennen, zusammenhängt als in der anderen Gruppe. Wenn die metrische Invarianz erfüllt ist, kann die dritte Stufe, die skalare Invarianz, überprüft werden.

Die dritte Stufe, die skalare Invarianz, geht davon aus, dass nicht nur das Messmodell und die Ladungen gleich sind, sondern zusätzlich die Ankerpunkte und Skalierung der Items in beiden Gruppen gleich kalibriert sind (Wu et al., 2007). Die zusätzlich restringierten Parameter sind hier die Intercepts der Items. Das bedeutet, dass Mittelwertunterschiede in den Items auf Unterschiede im latenten Konstrukt zurückzuführen sind. Falls skalare Invarianz im WWT nicht vorläge, könnte dies bedeuten, dass „Kosmetika“ bei Kindern mit Lernbehinderung zwar häufiger als von der Kontrollgruppe korrekt beantwortet wird, dies jedoch nicht mit einer höheren latenten Fähigkeit, Kategorien zu benennen, einhergeht. Wenn die skalare Invarianz erfüllt ist, kann die vierte Stufe, die residuale Invarianz, überprüft werden.

Die vierte, residuale Stufe prüft schließlich, ob die itemspezifische Varianz und die Fehlervarianz (also diejenige Varianz in den Items, die nicht durch den latenten Faktor erklärt wird) in den Gruppen gleich ist. Falls residuale Invarianz im WWT nicht vorläge, könnte dies bedeuten, dass sich bei einem spezifischen Item wie „Kosmetika“ verzerrende Einflüsse oder Messfehler in den Gruppen unterschiedlich stark auf die Lösungswahrscheinlichkeit auswirken, sodass sich der Messfehler bzw. die Reliabilitäten der Items zwischen den Gruppen unterscheiden (Schwab & Helm, 2015). Falls residuale Invarianz im WWT nicht vorläge, könnte dies bedeuten, dass „Kosmetika“ bei Kindern mit Lernbehinderung eine niedrigere Reliabilität aufweist als bei der Kontrollgruppe.

Da Messinstrumente in der Praxis selten vollständig messinvariant sind, kann auch eine partielle Invarianz festgestellt werden (Putnick & Bornstein, 2016). Dazu werden die Restriktionen für einzelne manifeste Variablen gelockert, sodass deren Parameter für jede Gruppe einzeln frei geschätzt werden können und damit zwischen den Gruppen variieren dürfen. Anschließend wird die Modellpassung erneut überprüft. Wenn nur wenige Lockerungen der Restriktionen (mindestens die Hälfte der Parameter sollte fixiert bleiben) notwendig sind, sodass danach ein erneuter Modellvergleich nicht-signifikant ausfällt, so kann weiterhin von (partieller) Messinvarianz ausgegangen werden (Putnick & Bornstein, 2016).

#### 6.2.3.3. Item-Parceling Strategie

Um mit dem erwähnten problematischen Verhältnis von Itemanzahl zu Stichprobengröße umzugehen, wurde der Ansatz des Item-Parcelings (Little, Cunningham, Shahar & Widaman, 2002; Little et al., 2013) gewählt, bei dem mehrere einzelne Items zu Parcels zusammengefasst werden, die dann als gebündelte manifeste Variablen in das Messmodell einfließen. Ein Item-Parcel wird anhand des Mittelwertes aus mehreren Items berechnet, z. B. können aus sechs Items zwei Parcels erstellt werden, indem für jede Person der Mittelwert der ersten drei Items



für Parcel-1, und der anderen drei Items für Parcel-2 errechnet wird. Durch die Zusammenfassung von Items zu Parcels reduziert sich die Anzahl der manifesten Variablen und damit die Anzahl der zu schätzenden Modellparameter.

Über die Rechtfertigung von Item-parcels gegenüber manifesten Items zeichnen Little et al. (2013) eine lebhafte Kontroverse nach, bei der sich eine eher pragmatische und eine eher puristische Perspektive gegenüberstehen (Little et al. 2013; für Kritik siehe u. a. Meade & Kroustalis, 2006). Little et al. (2013) argumentieren, dass durch die Verwendung von Parcels die spezifische Varianz der manifesten Variablen minimiert wird. Dadurch ist eine höhere Reliabilität gewährleistet, und die Wahrscheinlichkeit von korrelierten Residuen und Cross-Ladungen (eine manifeste Variable lädt auf mehr als eine latente Variable) ist reduziert.

Die Parceling Strategie eignet sich insbesondere, wenn bei der Auswertung nicht die psychometrische Qualität einzelner Items, sondern das latente Konstrukt betrachtet werden soll (Little et al., 2013). Die Messinvarianz gilt dann jedoch explizit nur auf Ebene der Parcels; ein Rückschluss, dass Invarianz auf Itemebene vorläge, darf also nicht gezogen werden. In der vorliegenden Studie stand die Äquivalenz der Messmodelle zwischen den Gruppen auf der übergeordneten, latenten Konstruktebene im Fokus, um eine Aussage darüber zu treffen, ob die Binnenstruktur des mentalen Lexikons insgesamt zwischen den Gruppen vergleichbar ist. Die Fragestellung betrifft nicht die psychometrische Qualität einzelner Items, weswegen es angemessen erschien, die Parceling-Strategie für die Auswertung in dieser Arbeit zu nutzen.

#### 6.2.3.4. Modellspezifikation

Die Modelle wurden folgendermaßen spezifiziert: Für jede der vier Wortarten (Nomen, Verben, Kategorien, Adjektive) wurden fünf Item-Parcels erstellt, indem insgesamt 4-6 Items (stratifiziert) zufällig einem der Parcels zugeordnet wurde. Stratifiziert bedeutet, dass bei der zufälligen Zuordnung der Items zu Parcels die Itemschwierigkeit beachtet wurde, damit die Parcels nach Schwierigkeit parallel waren. Für die Parallelisierung wurden die Itemkennwerte aus der Normstichprobe (Glück, 2011) der Altersgruppe „7;6 – 7;11 Jahre“ (entspricht dem Altersdurchschnitt der MA-Kontrollgruppe) herangezogen.

In der Skala „Nomen“ mit insgesamt 26 Items enthielt das erste Parcel „N1“ sechs, alle anderen Parcels fünf Items. Für die anderen Skalen („Verben“, „Adjektive“ und „Oberkategorien“) mit je 23 Items enthielten die ersten drei Parcels fünf Items und die letzten beiden Parcels jeweils vier Items. Die Zusammensetzung der Parcels aus einzelnen Items finden sich in Tabelle A-3 im Anhang, die dazugehörigen deskriptiven Informationen zu den Parcels pro Gruppe in Tabelle A-4 im Anhang.

Mithilfe der `cfa()`-Funktion im R-Paket `lavaan` (Rosseel, 2012) wurden unter Anwendung der Maximum-Likelihood Schätzmethode die Messmodelle geschätzt. Die vier Wortarten

stellten die latenten Faktoren dar, denen die jeweiligen fünf Item-Parcels als manifeste Variablen zugeordnet wurden. Anhand der zweistufigen Variable Gruppe (Lernbehinderung vs. Kontrollgruppe) erfolgte die getrennte Schätzung für die Gruppen. Auf konfiguraler Ebene wurde nur die Faktorenstruktur festgelegt, alle weiteren Parameter konnten frei geschätzt werden. In den anschließenden Schritten wurden die Modelle gemäß den Ausführungen in Abschnitt 6.2.3.2 zunehmend restringiert: Das metrische Modell setzte die Faktorladungen zwischen den Gruppen gleich, das skalare Modell zusätzlich die Item-Intercepts, und schließlich das residuale Modell die Fehlervarianzen der manifesten Variablen.

Ausgegeben wurden die in Abschnitt 6.2.3.1 beschriebenen Werte zur Einschätzung der Modellgüte:  $\chi^2$ -Wert und Freiheitsgrade, RMSEA sowie die alternativen Fitmaße CFI und TLI. Für den relativen Vergleich der Messinvarianz-Modelle wurde zum einen der Likelihood Ratio Test herangezogen, der die Nullhypothese prüft, dass das restringiertere Modell nicht signifikant schlechter zu den Daten passte. Außerdem wurden die alternative Fit-Differenzen von  $\Delta$ CFI betrachtet. Ein  $\Delta$ CFI > .01 deutet darauf hin, dass die Restriktion zu einer deutlichen Verschlechterung des Modellfits führt (Cheung & Rensvold, 2002).

#### 6.2.3.5. Absicherung durch Parcel-Allocation

Ergänzend zu dem Parceling-Vorgehen, das im vorangehenden Abschnitt dargestellt wurde, kam zur Absicherung die Parcel-Allocation Methode zur Anwendung. Da die eher starken Voraussetzungen für eine zufällige Zuteilung der Items zu Parcels („großes  $N$ , viele Items, hohe Item-Kommunalität, geringe Item-Diversität“, Little et al., 2013, S. 295) im vorliegenden Fall nicht in allen Belangen ausreichend erfüllt sind, stellt die Parcel-Allocation Methode (Sterba & MacCallum, 2010) eine sinnvolle Weiterentwicklung der Parceling-Idee dar.

Bei der Methode `parcelAllocation` (Jorgensen, Pornprasertmanit, Schoemann & Rosseel, 2018) werden über eine größere Anzahl von Iterationen zufällige Item-zu-Parcel Zuordnungen gebildet. Im Prinzip kann das Verfahren als Bootstrapping verstanden werden. Für jedes Item wird die Faktorzugehörigkeit angegeben, und pro Faktor die Anzahl der zu bildenden Parcels festgelegt. Anschließend wird für jede Iteration eine neue zufällige Zuweisung der Items zu den festgelegten Faktoren vorgenommen und die Modellpassung ermittelt. In der vorliegenden Analyse wurden 100 Iterationen durchgeführt, und analog zu der ursprünglichen Analyse wurden fünf Parcels pro Subskala gebildet.

Für jede Messinvarianz-Stufe erhält man in der Parcel-Allocation Methode anstatt eines einzelnen Wertes eine Verteilung von Modellpassungswerten. Somit ist die Wahrscheinlichkeit geringer, dass die eine erfolgte spezifische Zuweisung für besonders hohe oder niedrige Modellfits

sorgt. Die Messinvarianz wird nun analog zu den oben dargelegten Stufen anhand von Modellvergleichen (konfigural vs. metrisch, metrisch vs. skalar, skalar vs. residual) überprüft. Die Fit-Indizes, anhand derer die Modellvergleiche vorgenommen werden, lassen sich aus dem arithmetischen Mittel über die aus den Iterationen resultierenden Fit-Indizes berechnen (Little et al., 2013). Durch die Verwendung der Parcel-Allocation Methode kann auch eingeschätzt werden, wie hoch die „Parcel-Zuordnungs-Variabilität“ (Little et al., 2013, S. 295) ist, sprich wie breit die Ergebnisse in Abhängigkeit davon streuen, welche Items zu Parcels zusammengefasst werden (Sterba & Rights, 2017).

Damit dient die Parcel-Allocation Methode zur Absicherung der Messinvarianz-Ergebnisse, um auszuschließen, dass diese lediglich aufgrund der zufälligen Zuteilung der Items zu Parcels zustande kamen.

#### **6.2.4. Methode zu Fragestellung 2: Außenstruktur**

##### 6.2.4.1. Generalisierte gemischte lineare Modelle

Die zweite Fragestellung betrifft die Außenstruktur des mentalen Lexikons. Die Wahrscheinlichkeit, ein WWT-Item korrekt zu lösen (Kriteriumsvariable), wird durch eine Reihe von weiteren Faktoren regressionsanalytisch vorhergesagt und auf differenzielle Einflüsse zwischen den Gruppen der Kinder mit und ohne Lernbehinderung überprüft.

Die empirische Datengrundlage weist dabei zwei Aspekte auf, die die Auswertung durch ein einfaches lineares Modell nicht erlauben, sondern die Erweiterung durch ein verallgemeinertes lineares gemischtes Modell (generalized linear mixed model, GLMM, siehe z. B. Baayen et al., 2008; Wong & Mason, 1985) notwendig machen: (1) da jedes Kind mehrere Items bearbeitete und jedes Item von mehreren Kindern beantwortet wurde, sind die Datenpunkte und auch die Residuen nicht unabhängig voneinander; (2) die Kriteriumsvariable hat als dichotome Variable nur die zwei Ausprägungen „1“ (korrekt beantwortet) und „0“ (nicht korrekt).

GLMMs als verallgemeinerte Form einer klassischen linearen Regression sind für solche Szenarien geeignet, denn hinsichtlich des oben genannten Aspekts (1) können Abhängigkeiten in den Daten durch die Aufnahme von personen- und itembezogenen Zufallseffekten berücksichtigt werden. Bezüglich des Aspekts (2) kann auch eine nicht-kontinuierliche Kriteriumsvariable durch das Modell vorhergesagt werden.

Für Aspekt (1) ist Folgendes festzustellen: Neben den sog. festen Effekten (die für die Forschungsfrage von Interesse sind), können Zufallseffekte modelliert werden, mit denen Abhängigkeiten in den Daten berücksichtigt werden und die Fehlervarianz minimiert wird. In Anlehnung an Baayen et al. (2008) werden personenbezogene Zufallseffekte als *by-subject random effects* und itembezogene als *by-item random effects* bezeichnet. Die folgenden Abschnitte befassen sich

mit Abhängigkeiten von Daten innerhalb von Personen, der Ansatz kann aber analog auf die Ebene der Items übertragen werden (Baayen et al., 2008).

Abhängigkeiten in Daten entstehen, wenn Personen mehrfach getestet werden, z. B. zu verschiedenen Messzeitpunkten oder, wie im vorliegenden Fall der WWT-Items, mit verschiedenen Items zu einem Messzeitpunkt. Das hat zur Folge, dass die Residuen nicht unabhängig oder unkorreliert sind (Casals, Girabent-Farrés & Carrasco, 2014). Die Messungen der Items sind also in den Personen genestet, sodass diese als „Kontexteinheit“ im Rahmen eines hierarchisch-linearen Modells (Richter & Naumann, 2002) verstanden werden müssen.

Die Abhängigkeiten illustrieren Bates, Mächler, Bolker und Walker (2015) am Beispiel der Auswirkungen von Schlafentzug auf die Reaktionsgeschwindigkeit: Bei allen Proband\*innen hat Schlafentzug insgesamt eine negative Auswirkung auf die Reaktionszeit, jedoch ist erstens diese Auswirkung des Entzugs unterschiedlich stark (jede Person hat ihre eigene Steigung, *random slope*) und zweitens haben die Personen unterschiedliche Ausgangsniveaus der Reaktionszeiten zu Beginn des Experiments (jede Person hat ihren eigenen Intercept, *random intercept*, wenn kein Schlafentzug gegeben ist).

Die resultierende komplexere Fehlerstruktur lässt sich im Rahmen von GLMMs durch die Aufnahme von Zufallseffekten (*random effects*) abbilden, sowohl für das generelle Niveau von Personen (*random intercept*) als auch für die Wechselwirkung mit einem oder mehreren festen Effekten (*random slope*). Auf diese Weise wird der Anteil der Varianz, die durch die festen Faktoren aufgeklärt wird, von dem Varianzanteil getrennt, der auf die Unterschiedlichkeit der Personen bzw. Items zurückzuführen ist. In einem Modell ohne Zufallseffekte würden die systematischen Unterschiede in die allgemeine Fehlervarianz einfließen, welche dadurch überschätzt würde. Durch die Spezifizierung von Zufallseffekten kann die komplexere Fehlerstruktur abgebildet und die restliche Fehlervarianz reduziert werden (Baayen et al., 2008).

Mit der Aufnahme eines *random intercepts* wird für jede Person geschätzt, um wieviel sie im Mittel vom globalen Intercept nach oben oder unten abweicht. Ein *random intercept* für Personen drückt also systematische (Leistungs-)Unterschiede zwischen den Personen aus. Wenn z. B. eine Person grundsätzlich Werte auf einem höheren Niveau erzielt als eine andere, drückt sich dies in einem positiven *random intercept* für diese Person aus: Der globale Intercept muss für diese Person nach oben korrigiert werden.

Wenn die Annahme besteht, dass sich eine Modellvariable stärker oder schwächer auf die verschiedenen Personen auswirkt, wird analog durch einen *random slope* für jede Person geschätzt, um wieviel der globale Slope-Effekt des Modells für diese Person korrigiert werden muss. Wenn z. B. eine Person grundsätzlich von Schlafentzug weniger stark beeinträchtigt wird, weil sie es beispielsweise aus ihrem beruflichen Umfeld gewohnt ist, dann drückt sich dies in

einem negativen random slope für diese Person aus: Der globale Slope Effekt des Modells muss für diese Person nach unten korrigiert werden.

Bei einem GLMM können nicht nur für Personen die genannten Zufallseffekte (*by-subject random effects*) spezifiziert werden, sondern auch für Items (*by-item random effects*), da diese ebenfalls mehrfach (von verschiedenen Personen) bearbeitet werden, sodass auch die Items Abhängigkeiten aufweisen. In einem GLMM können beide Arten von Zufallseffekten gleichzeitig berücksichtigt werden (Baayen et al., 2008).

Auch hinsichtlich der Kriteriumsvariable ist ein GLMM flexibel, was eine weitere Anforderung aufgrund der vorliegenden Daten ist. So ist es möglich, nicht nur kontinuierliche Variablen durch das Modell vorherzusagen, sondern auch ordinale und dichotome Variablen. Da in der vorliegenden Teilfragestellung die abhängige Variable dichotom ist („Wurde das Item korrekt beantwortet, ja oder nein?“), wird anstelle der Normalverteilung von einer Binomialverteilung ausgegangen (Kleinbaum & Klein, 2002). Infolgedessen wird für kategoriale Variablen eine logistische Regression (Kleinbaum & Klein, 2002; Peng, Lee & Ingersoll, 2002) herangezogen. Die Lösungswahrscheinlichkeit für ein Item wird durch S-förmige Kurve der logit-Funktion abgebildet, die als Link-Funktion dient und nur Werte zwischen 0 und 1 annehmen kann.

#### 6.2.4.2. Modellvariablen und -gleichung

Die Anpassung des generalisierten gemischten linearen Modells zur Beantwortung der zweiten Teilfragestellung nach der Außenstruktur der WWT-Items erfolgte mithilfe der Funktion `glmer()` aus dem R-Paket `lme4` (Bates et al., 2015). Dabei wurden insgesamt sieben Prädiktoren in das Modell aufgenommen, und zwar vier personen- und drei itembezogene Variablen, die im Folgenden angeführt sind.

Auf Ebene der Personen wurden die folgenden Prädiktoren aufgenommen:

- *Gruppe*: Die Gruppenzugehörigkeit ist eine dichotome, dummy-codierte Variable mit den Ausprägungen Lernbehinderung („LB“) und unbeeinträchtigte Kontrollgruppe („TD“), wobei letztere die Referenzkategorie darstellt.
- *Alter*: Das Alter ist eine kontinuierliche Variable und bemisst sich in Lebensmonaten zum Zeitpunkt der Testung mit dem WWT. Das Alter wurde zentriert.
- *AG-Kapazität*: Die Arbeitsgedächtnis-Kapazität ist eine kontinuierliche Leistungs-Variable, die anhand des Gesamtscores über alle Arbeitsgedächtnis-Wortspannungsaufgaben hinweg gemessen wurde (siehe Studie 1). Die AG-Kapazität wurde zentriert.
- *CFT*: Die kognitive Kapazität ist eine kontinuierliche Leistungsvariable zur Schätzung des mentalen Alters, die durch den Rohwert des CFT gemessen wurde (siehe Studie 1). Die kognitive Kapazität wurde zentriert.

Auf Ebene der Wörter (Items) flossen die folgenden drei Prädiktoren in das Modell ein:

- *Wortart*: Die Zugehörigkeit zur Wortart ist eine dummy-codierte kategoriale Variable mit den vier Ausprägungen „Nomen, Verb, Adjektiv, und Kategorie“, wobei die Wortart „Nomen“ die Referenzkategorie darstellt.
- *Frequency*: Die absolute Worthäufigkeit ist eine kontinuierliche Variable. Sie wurde aus dem ChildLex Corpus (Schroeder et al., 2015) ermittelt und für die Analyse logarithmiert und zentriert, damit eine Normalverteilung gegeben ist.
- *NBS*: Die Neighborhood-Size ist eine kontinuierliche, ebenfalls logarithmierte und zentrierte Variable.

Die Neighborhood-Size („NBS“) wurde mithilfe des R-Paketes `vwr` (Keuleers, 2013) unter Nutzung der Funktion `old20()` ermittelt (Yarkoni et al., 2008). Diese findet im hinterlegten `vwr`-Corpus diejenigen 20 Wörter, die den geringsten Levenshtein-Abstand (siehe Abschnitt 2.2.3.1) zum Zielwort aufwiesen. Der aus `old20()` resultierende Wert gibt den durchschnittlichen Levenshtein-Abstand an, den die 20 gefundenen Wörter zum Cue-Wort haben. Besitzt ein Cue-Wort einen niedrigen Wert (nahe 1), so bedeutet dies, dass es eine verhältnismäßig große Ähnlichkeit zu seinen Nachbarn aufweist, da insgesamt nur wenige Änderungen nötig sind, um zu den 20 nächsten Nachbarn zu gelangen. Die Variablen *Frequency* und *NBS* wurden zudem logarithmiert, damit sie einer Normalverteilung folgen.

Da für den Gruppenvergleich insbesondere die Interaktionen mit dem Faktor *Gruppe* von Bedeutung waren, wurden die zweifach-Interaktionen von *Gruppe* mit den Variablen *Wortart*, *AG-Kapazität*, *kognitiver Kapazität*, sowie *Frequency* und *Neighborhood-Size* ebenfalls mit in das Modell aufgenommen. Dadurch kann die Frage beantwortet werden, ob die genannten Prädiktoren sich in ihren Effekten auf die Lösungswahrscheinlichkeit im WWT zwischen den Gruppen unterschieden. Weitere Interaktionen zwischen den anderen Prädiktoren wurden zwecks Komplexitätsreduktion nicht in das Modell aufgenommen, da keine theoretischen Annahmen hierzu vorlagen. Ebenfalls nicht berücksichtigt wurden an dieser Stelle Interaktionen mit dem Faktor *Alter*, da die Gruppen hinsichtlich der Altersbereiche sehr unterschiedlich ausgeprägt waren.

Um der innerhalb von Personen sowie Items genesteten Datenstruktur Rechnung zu tragen, wurden Zufallseffekte sowohl auf Personen- wie auch auf Item-Ebene angenommen. Für die Personen wurde deren jeweiliges Niveau berücksichtigt (d. h. ein random intercept modelliert) sowie die Auswirkung des Wortart-Effektes für jede Person angepasst (d. h. ein random slope für Wortart modelliert). Technisch drückt sich dies im R-Code durch `(1 + Wortart | ID)` aus. Pro Person („ID“) wird eine Anpassung des Intercepts („1“) und eine Anpassung des Effektes für Wortart („+ Wortart“) vorgenommen. Auf Seiten der Items wurde für jedes Item

bzw. Wort eine individuelle Schwierigkeit in Form eines random intercepts modelliert. Im R-Code wird dies spezifiziert durch `(1|Wort)`. Der Code für das vollständige Modell (mit allen hier beschriebenen Effekten und Interaktionen) lautet wie folgt:

```
> glmer(Punkt ~ Gruppe + Alter + Wortart + ag_ges + cft +
> log(freq.abs) + log(nbs) +
> Wortart:Gruppe + ag_ges:Gruppe + cft:Gruppe +
> log(freq.abs):Gruppe + log(nbs):Gruppe +
> (1 + Wortart|ID) + (1|Wort)
```

Die Datenaufbereitung und -auswertung erfolgte in R Studio (R Core Team, 2017; R Studio Team, 2018), für die Auswertung der Multigroup-CFA kamen insbesondere die Pakete `lavaan` (Rosseel, 2012) sowie `semTools` (Jorgensen et al., 2020) zum Einsatz. Für die GLMMs wurde auf das Paket `lme4` (Bates et al., 2015) zurückgegriffen, und zur Berechnung von Signifikanzwerten für einzelne Prädiktoren im Modell wurde das Paket `lmerTest` genutzt (Kuznetsova, Brockhoff & Christensen, 2017).

### 6.3. Ergebnisse

#### 6.3.1. Ergebnisse zu Fragestellung 1: Binnenstruktur

Die Ergebnisdarstellung erfolgt in drei Schritten. Zunächst wird allgemein überprüft, ob ein vierfaktorielles Modell, das die vier Wortarten differenziert, tatsächlich wie erwartet besser zu den Daten passt als ein einfaktorielles Modell. Als nächstes werden die Ergebnisse zur Messinvarianz-Analyse präsentiert. Abschließend wird mit der `parcelAllocation`-Methode abgesichert, dass die gefundenen Ergebnisse nicht auf der zufälligen Zuordnung der Items zu den Parcels beruhen.

Im ersten Schritt wurde die Gültigkeit der postulierten vierfaktoriellen Lösung (mit den vier Wortarten als Faktoren) unabhängig von der Gruppenzugehörigkeit überprüft: Führt die Aufnahme von vier Faktoren gegenüber einem globalen Faktor zu einer signifikant besseren Modellpassung? Dies wurde für die gesamte Stichprobe über beide Gruppen (LB und TD) hinweg ermittelt. Die Ergebnisse der Modellfits für das einfaktorielle und das vierfaktorielle Modell sind in Tabelle 5 dargestellt.

Tabelle 5  
*Globaler Modellvergleich ein- vs. vier-faktorielles Modell über beide Gruppen hinweg*

	$\chi^2$	Df	$\Delta\chi^2$	$\Delta Df$	<i>p</i> -Wert
Ein-Faktor-Modell	379.31	170	---	---	---
Vier-Faktoren-Modell (Wortarten)	186.86	164	192.45	6	< .001

*Anmerkung.* Df: Freiheitsgrade des Modells.

Das vierfaktorielle Modell wies eine signifikant bessere Passung zu den Daten auf als die einfaktorielle Lösung,  $\Delta\chi^2(6) = 192.45$ ,  $p < .001$ . Es kann grundsätzlich davon ausgegangen werden, dass die Aufteilung der Item-Parcels nach Wortarten und die damit verbundene Bildung von Subskalen mit den empirischen Daten vereinbar ist. Diese Aussage bezieht sich auf die gesamte Stichprobe. Im nächsten Schritt ist zu prüfen, ob die Modellparameter für beide Gruppen gleichermaßen gelten.

Die Ergebnisse der Modellvergleiche zur Messinvarianz können Tabelle 6 entnommen werden. In den Zeilen sind die Invarianz-Stufen abgetragen; in den Spalten finden sich die Angaben zur Modellpassung und die Fit-Indizes:  $\chi^2$  mit Freiheitsgraden (df),  $\Delta\chi^2$  mit df und  $p$ -Werten, CFI und  $\Delta$ CFI, sowie TLI und RMSEA mit  $p$ -Werten für RMSEA.

Es zeigte sich zunächst, dass das konfigurale Modell gut zu den Daten passte (CFI = .956). Es kann also angenommen werden, dass die grundsätzliche Struktur des Messmodells mit den vier Wortarten als latenten Variablen in beiden Gruppen gleichermaßen galt. Die zweite Stufe, die der metrischen Invarianz, führte zu keinem signifikant schlechteren Modellfit,  $\Delta\chi^2(16) = 16.66$ ;  $p = .408$ ;  $\Delta$ CFI = .002. Folglich kann auch angenommen werden, dass die Ladungen der Items auf die latenten Faktoren für beide Gruppen vergleichbar waren. Dagegen wies bei der dritten Stufe, der skalaren Invarianz, eine signifikant schlechtere Modellpassung,  $\Delta\chi^2(16) = 29.95$ ;  $p = .018$ , darauf hin, dass die Annahme von vergleichbaren Intercepts zwischen den Gruppen nicht gerechtfertigt ist, auch wenn der  $\Delta$ CFI = .008 den Konventionsrichtwert von  $\Delta$ CFI = .01 nicht überschritt.

Vor der Überprüfung der vierten Stufe, der residualen Invarianz, wurde an dieser Stelle eine punktuelle Lockerung der Restriktion von äquivalenten Parcel-Intercepts zugelassen, um auf partielle Messinvarianz auf der skalaren Stufe zu testen (Putnick & Bornstein, 2016). Basierend auf dem Ergebnis des Lagrange Multiplier Tests (Rosseel, 2012) wurde für das erste Parcel in der Verben-Subskala („Verb-Parcel1“) eine getrennte Schätzung der Intercepts zugelassen. Die Lockerung dieses einen Intercept-Parameters genügte bereits, dass der Modellfit des partiellen skalaren Modells nicht mehr signifikant schlechter war als das metrische Modell,  $\Delta\chi^2(15) = 23.8$ ;  $p = .069$ ;  $\Delta$ CFI = .005. Damit kann auf der skalaren Stufe partielle Messinvarianz angenommen werden.

Abschließend wurde als vierte Stufe die residuale Invarianz überprüft. Die Aufnahme der Restriktion, dass die Residuen der Item-Parcels in beiden Gruppen äquivalent sind, führte zu keinem signifikant schlechteren Modellfit im Vergleich zur vorigen Stufe der partiellen skalaren Invarianz,  $\Delta\chi^2(20) = 12.93$ ;  $p = .88$ ;  $\Delta$ CFI = -.004. Insofern kann auch für die residuale Stufe partielle Invarianz angenommen werden.



Tabelle 6  
*Modellvergleiche zur Überprüfung der Messinvarianz*

	$\chi^2$	Df	$\Delta\chi^2$	$\Delta$ Df	p-Wert	CFI	$\Delta$ CFI	TLI	RMSEA	p-Wert RMSEA
Konstante Faktorenstruktur (konfigural)	410.90	328	---	---	---	.956	---	.948	.0495	.510
Konstante Ladungsstruktur (metrisch)	427.56	344	16.66	16	.408	.954	.002	.949	.0486	.552
Konstante Achsenabschnitte (skalar) <sup>a</sup>	451.36	359	23.80	15	.069	.949	.005	.946	.0500	.492
Konstante Messfehlervarianzen (residual) <sup>a</sup>	464.29	379	12.93	20	.880	.950	-.004	.950	.0481	.578

*Anmerkungen.* Df: Freiheitsgrade des Modells, CFI: Comparative Fit Index, TLI: Tucker-Lewis Index, RMSEA: Root-Mean-Square-Error-of-Approximation, p-Wert RMSEA: Test der Null-Hypothese, dass RMSEA größer als .05 ist.

<sup>a</sup> partielle Invarianz: Lockerung eines Parameters (V1-Parcel wurde frei geschätzt).

Bei der Anwendung der multiplen Parcel-Allocation Methode, in der die Modellvergleiche in einhundert Iterationen von Item-zu-Parcel Zuweisungen durchgeführt wurden, zeigten sich ähnliche Ergebnisse. Allerdings konnte hier die partielle Invarianz nicht überprüft werden, da durch die Prozedur die Parcels in jeder Iteration neu zusammengestellt werden. Dadurch war es nicht möglich, einzelne Parameter zu ermitteln, für die die Restriktionen gelockert werden könnten. Die Ergebnisse der multiplen Parcel-Allocation finden sich in Tabelle 7. Diese berichtet analog zur oben beschriebenen Darstellung die Angaben zur Modellpassung pro Messinvarianz-Stufe. Bei den Werten handelt es sich in dieser Tabelle jedoch um die Mittelwerte über einhundert Iterationen. In Klammern sind die Standardabweichungen und darunter die minimalen und maximalen Werte angegeben, welche eine Abschätzung ermöglichen, wie stark die Modellpassungswerte in Abhängigkeit von der Parcel-Zusammensetzung schwanken (Sterba & Rights, 2017).

Die Modellvergleiche wurden anhand der durchschnittlichen Fit-Indizes vorgenommen. So zeigte sich auch hier, dass das konfigurale Modell zu den Daten passte ( $CFI = .954$ ). Die Modellpassung der metrischen Invarianz war im LR-Test mit  $\Delta\chi^2(16) = 28.23, p = .03$ , zwar signifikant schlechter als bei der konfiguralen Invarianz, allerdings sprechen sowohl das Verhältnis von  $\Delta\chi^2/df = 1.76$  als auch ein  $\Delta CFI = .007$  dafür, dass Messinvarianz vorlag. Auf der nächsten Stufe der skalaren Invarianz dagegen wurde zum einen  $\Delta\chi^2(16) = 55.16, p < .001$  signifikant, zum anderen überstieg das Verhältnis von  $\Delta\chi^2/df = 3.46$  den Richtwert von 2 deutlich, und auch  $\Delta CFI = .023$  lag über dem empfohlenen Richtwert. Es kann daher keine skalare Invarianz angenommen werden.

Da partielle Invarianz im Rahmen der Parcel-Allocation Methode nicht überprüft werden kann, entfiel an dieser Stelle die Überprüfung der partiellen skalaren und residualen Invarianz (vgl. Putnick & Bornstein, 2016). Folglich konnten die obigen Ergebnisse durch die Parcel-Allocation Methode bis zum Zwischenschritt der metrischen Invarianz bekräftigt werden.

Aus der nachgelagerten Analyse von einhundert randomisierten Item-zu-Parcel Zuweisungen kann geschlussfolgert werden, dass es sich bei den gefundenen Ergebnissen der metrischen Messinvarianz mit großer Wahrscheinlichkeit nicht um zufällige Befunde handelte, die auf die spezifische Kombination von Items zu einem Parcel zurückzuführen wären. Die darüber hinausgehenden Befunde der partiellen skalaren und residualen Invarianz konnten durch die Parcel-Allocation Methode aufgrund der oben beschriebenen technischen Einschränkung nicht bestätigt (aber auch nicht verworfen) werden.

Tabelle 7  
*Modellergleiche mit Parcel-Allocation Methode zur Überprüfung der Messinvarianz anhand zufälliger Item- $\xi$ -Parcel Zuweisungen*

	$\chi^2$	Df	$\Delta\chi^2$	$\Delta$ Df	p-Wert	CFI	$\Delta$ CFI	TLI	RMSEA
Konstante Faktorenstruktur (konfigural)	408.279	328	---	---	---	.954	---	.947	.0478
Konstante Ladungsstruktur (metrisch)	436.511	344	28.232	16	.030	.947	.007	.941	.0504
Konstante Achsenabschnitte (skalar) <sup>a</sup>	491.826	360	55.315	16	< .001	.924	.022	.920	.0591
Konstante Messfehler-varianzen (residual) <sup>a</sup>	519.241	380	27.415	20	.124	.920	.004	.920	.0590

*Anmerkungen.* Df: Freiheitsgrade des Modells, CFI: Comparative Fit Index, TLI: Tucker-Lewis Index, RMSEA: Root-Mean-Square-Error-of-Approximation. Die Angaben entsprechen den durchschnittlichen Fit-Werten aus 100 iterierten Modelle.

### 6.3.2. Ergebnisse zu Fragestellung 2: Außenstruktur

Die zweite Teilfragestellung zum differenziellen Einfluss von weiteren Prädiktoren auf die Lösungswahrscheinlichkeit wurde mithilfe eines GLMMs beantwortet. Hier stellte die dichotome Variable „Punkt“ die abhängige Variable dar: Wurde das Item von dem Kind korrekt benannt („1“) oder nicht („0“)? Der Datensatz umfasste insgesamt 19665 ( $n = 207 \times k = 95$ ) Antworten. Im GLMM-Ansatz wirkte sich das Problem der Stichprobengröße weniger stark aus, da hier deutlich weniger Modellparameter geschätzt werden mussten. Deswegen erfolgten die Auswertungen auf Ebene der einzelnen Items, ohne Parcels zu bilden.

Tabelle 8

*Parameter des verallgemeinerten gemischten linearen Modells (GLMM) zur Vorhersage der Lösungswahrscheinlichkeit der korrekten Bild-Benennung im WWT 6-10*

Fester Effekt	Estimate	S.E.	z-Wert	p-Wert
Intercept	0.870	0.294	2.957	<b>.003</b>
GruppeLB	-0.958	0.239	-4.011	<b>.000</b>
Chronologisches Alter	0.576	0.118	4.893	<b>.000</b>
AG-Kapazität	0.255	0.075	3.398	<b>.001</b>
Kognitive Kapazität	0.330	0.077	4.289	<b>.000</b>
Wortart Adjektive	-0.274	0.450	-0.608	.543
Wortart Kategorien	0.021	0.399	0.052	.958
Wortart Verben	0.768	0.408	1.882	.060
Wort-Frequency	0.581	0.195	2.979	<b>.003</b>
Wort-Neighborhoodsize	0.330	0.192	1.719	.086
GruppeLB × AG-Kapazität	-0.027	0.121	-0.226	.821
GruppeLB × Wortart Kognitive Kapazität	-0.005	0.119	-0.046	.963
GruppeLB × Wortart Adjektive	-0.227	0.185	-1.224	.221
GruppeLB × Wortart Kategorien	-0.191	0.126	-1.517	.129
GruppeLB × Wortart Verben	-0.172	0.123	-1.399	.162
GruppeLB × Wort-Frequency	0.039	0.054	0.721	.471
GruppeLB × Wort-Neighborhoodsize	0.091	0.053	1.730	.084

*Anmerkungen.* Die Variablen Gruppe sowie Wortart sind dummy-codiert, für Gruppe sind unbeeinträchtigte Kinder die Referenzgruppe, bei Wortart ist die Referenzbedingung Nomen. Chronologisches Alter (in Monaten), Kognitive Kapazität (CFT 1-R) und AG-Kapazität (mittlere Wortspannen-Leistung, siehe Studie 1) sind personenbezogene Variablen; Wortart, Wort-Frequency und Wort-Neighborhoodsize sind itembezogene Variablen. Signifikante Effekte ( $p < .05$ ) sind fett markiert.

Das gemischte lineare Modell konvergierte mit den in Abschnitt 6.2.4.2 angegebenen Zufallseffekten: by-subject random intercept sowie random slope für Wortart; außerdem ein by-

item random intercept. Das volle Modell, bei dem ein by-subject random slope-Effekt berücksichtigt wurde, passte signifikant besser zu den Daten als ein Modell, das nur random intercept Effekte annahm,  $\Delta\chi^2(9) = 173.04$ ;  $p < .001$ . Als feste Effekte enthielt das volle Modell die Haupteffekte Gruppenzugehörigkeit, Alter, AG-Kapazität, CFT-Score, Wortart, Frequency, Neighborhood-Size, sowie alle zweifach-Interaktionen dieser Variablen mit dem Gruppenfaktor (außer der Interaktion von Gruppe  $\times$  Alter, s.o.). Die Werte der einzelnen Parameter dieses Modells sind in Tabelle 8 aufgeführt, unter Angabe der standardisierten Schätzer für die Beta-Gewichte, SE,  $z$ -Werte und dazugehörigen  $p$ -Werte.

Für die Interpretation ist zu beachten, dass die Faktoren Gruppe und Wortart kategorial und dummy-codiert sind. Der Intercept-Effekt gibt somit an, ob sich die Werte der Referenzgruppe (TD) in der Referenzbedingung (Nomen) bei durchschnittlicher Ausprägung aller kontinuierlichen Kovariaten (da diese  $z$ -standardisiert und damit zentriert wurden) signifikant von Null unterscheiden. Die einzelnen Faktorstufen in den Zeilen (z. B. „GruppeLB“) geben wieder, inwiefern sich der Wert dieser Gruppe in der Referenzbedingung (Nomen) von der TD-Gruppe unterscheidet. Eine Interaktion, beispielsweise gekennzeichnet als „GruppeLB:WortartV“, gibt an, ob und wie stark sich der Effekt für die Wortart „Verb“ für die LB-Gruppe vom Wortart-Effekt in der Referenzgruppe unterscheidet. Ein nicht-signifikanter Interaktionseffekt lässt jedoch nicht automatisch darauf schließen, dass sich der Effekt für die Wortart für die LB-Gruppe auch statistisch von Null unterscheidet.

Aus den Ergebnissen in Tabelle 8 lässt sich ein signifikanter Haupteffekt für „GruppeLB“ ablesen, der besagt, dass Kinder mit Lernbehinderung in der Referenzbedingung (Nomen) signifikant niedrigere Lösungswahrscheinlichkeiten aufwiesen als Kinder der Kontrollgruppe ( $z = -4.011$ ;  $p < .001$ ). In der Kontrollgruppe schien statistisch keine der Wortarten signifikant leichter oder schwerer zu lösen sein als die Referenzbedingung (Nomen); die Wortart „Verben“ verfehlte dabei allerdings nur knapp die Signifikanzgrenze ( $z = 1.882$ ;  $p = .06$ ). In der Kontrollgruppe zeigten sich signifikante Effekte für die Kovariaten AG-Kapazität ( $z = 3.398$ ;  $p < .001$ ) und kognitive Kapazität ( $z = 4.289$ ;  $p < .001$ ). Eine höhere AG-Kapazität und eine höhere kognitive Kapazität gingen folglich in der Kontrollgruppe mit höheren Lösungswahrscheinlichkeiten im Wortschatz einher. Auf Seiten der Items bzw. Wörter hatte die Frequency einen signifikant positiven Einfluss auf die Lösungswahrscheinlichkeit ( $z = 2.979$ ;  $p = .003$ ), während die Neighborhood-Size nicht signifikant wurde ( $z = 1.719$ ;  $p = .086$ ). Diese Ergebnisse beziehen sich auf die Kontrollgruppe.

Kinder mit Lernbehinderung wichen von diesen Befunden nicht ab, da sich keine signifikanten Interaktionen mit dem Faktor Gruppe einstellten. So waren die  $p$ -Werte der Interaktionen mit den einzelnen Stufen der kategorialen Variable Wortart allesamt nicht signifikant (alle

$p \geq .129$ ). Auch für die kontinuierlichen Kovariaten AG-Kapazität, kognitive Kapazität, Wort-Frequency und Neighborhood-Size ergaben sich keine signifikanten Gruppeninteraktionen (alle  $p \geq .084$ ). Kinder mit Lernbehinderung wichen also nicht von der Kontrollgruppe ab, was die Auswirkung der genannten Effekte auf die Lösungswahrscheinlichkeit der WWT-Items betraf.

Auch wenn also keine der Gruppeninteraktionen mit den einzelnen Wortart-Stufen im paarweisen Vergleich mit der Referenz-Wortart Nomen signifikant wurde, könnte insgesamt ein globaler Interaktionseffekt für die Variable Wortart auftreten. Daher wurde in einem weiteren Modellvergleich geprüft, ob die Aufnahme einer Interaktion von Gruppe  $\times$  Wortart die Modellpassung insgesamt verbesserte, siehe Tabelle 9.

In diesem Modellvergleich wurden die Interaktionen mit dem Gruppenfaktor per Likelihood Ratio Test schrittweise auf eine Verbesserung der Modellpassung geprüft. Das baseline-Modell enthielt nur die Haupteffekte der oben angeführten Variablen und keine Interaktionen; ein Zwischen-Modell enthielt alle Gruppen-Interaktionen außer derjenigen mit den Variablen Wortart und Alter; das volle Modell enthielt alle fünf Gruppen-Interaktionen mit den Kovariaten außer Alter.

Das volle Modell zeigte weder im Vergleich zum Zwischen-Modell,  $\Delta\chi^2(3) = 3.17; p = .367$ , noch im Vergleich zum Baseline-Modell eine bessere Passung,  $\Delta\chi^2(7) = 7.32; p = .397$ . Von einer Wechselwirkung mit dem Faktor Gruppe kann also nicht ausgegangen werden.

Tabelle 9  
*Modellvergleich bei Berücksichtigung der Gruppen-Interaktion*

	logLik	Df	$\Delta\chi^2$	$\Delta$ Df	$p$ -Wert
Reduziertes Modell 1 (ohne Interaktionen)	-8925.2	21	---	---	---
Reduziertes Modell 2 (ohne Wortart-Interaktion)	-8923.1	25	4.1474	4	.386
Volles Modell	-8921.5	28	3.1702	3	.366

*Anmerkungen.* logLik: logarithmierte Likelihood, bei diesem Index stehen geringere Werte für eine bessere Modellpassung, Df: Freiheitsgrade des Modells. Reduziertes Modell 1 ohne sämtliche zweifach-Interaktionen; reduziertes Modell 2 mit Interaktionen zwischen Gruppe und allen personen- und itembezogenen Variablen außer Wortart; volles Modell mit allen zweifach-Interaktionen zwischen Gruppe und den personen- und itembezogenen Variablen

Insgesamt zeigten sich folglich bei Kindern mit Lernbehinderung im Modellvergleich, abgesehen von einer niedrigeren Lösungswahrscheinlichkeit in der Referenzbedingung (Nomen), keine Abweichungen in der Untersuchungsgruppe von der Kontrollgruppe. Dies kann als em-

pirischer Hinweis darauf interpretiert werden, dass bei Lernbehinderung eine Entwicklungsverzögerung vorliegen könnte, da die Zusammenhangsmuster vergleichbar waren mit denen einer jüngeren unbeeinträchtigten Kontrollgruppe des gleichen mentalen Alters.

## 6.4. Diskussion

### 6.4.1. Zusammenfassung und Interpretation der Ergebnisse

In Studie 2 wurde der Abruf aus dem mentalen Lexikon anhand der Bildbenennung als Teil des Langzeitgedächtnisses näher betrachtet. Die Fragestellung betraf die Vergleichbarkeit auf Gruppenebene zwischen Kindern mit und ohne Lernbehinderung hinsichtlich der Struktur des Abrufs aus dem mentalen Lexikon, der anhand der expressiven Testbedingung des WWT 6-10 gemessen wurde. Dazu wurde in der ersten Teilfragestellung untersucht, ob die Binnenstruktur des Instrumentes über beide Gruppen hinweg messinvariant, also vergleichbar ist. In der zweiten Teilfragestellung wurden zur Untersuchung der Außenstruktur personen- und itembezogene Variablen auf differenzielle Prädiktion der Lösungswahrscheinlichkeit getestet.

Die erste Teilfragestellung zur Invarianz der Faktorenstruktur zwischen den Gruppen wurde mithilfe einer konfirmatorischen Faktorenanalyse im Multigruppen-Vergleich (Hirschfeld & von Brachel, 2014; Putnick & Bornstein, 2016) beantwortet. Damit wurde die Annahme überprüft, ob die Messmodelle der latenten Variablen in beiden Gruppen gleichermaßen gelten. Aufgrund des ungünstigen Verhältnisses von Stichprobe zur Anzahl an Items wurde beim CFA-Ansatz auf die Item-Parceling Strategie zurückgegriffen. Dabei werden mehrere Items zu einem Parcel zusammengefasst, wodurch die itemspezifische Fehlervarianz minimiert wird (Little et al., 2013). Die Ergebnisse lassen sich infolgedessen nicht auf Ebene der einzelnen Items interpretieren, sondern auf Ebene der latenten Konstrukte.

Gemäß dem üblichen Vorgehen zur Überprüfung von Messinvarianz (Putnick & Bornstein, 2016) wurde in vier Schritten geprüft, ob das Messmodell zunehmenden Restriktionen der Gruppen-Äquivalenz standhielt. Auf der konfiguralen und metrischen Stufe erwiesen sich die Daten zum WWT 6-10 der Untersuchungsgruppe (Kinder mit Lernbehinderung) und der MA-Kontrollgruppe als vollständig messinvariant. Dies bedeutet, dass die grundsätzliche Faktorenstruktur, also die Zuordnung von Parcels zu den latenten Faktoren zwischen den Gruppen ähnlich ist (‚konfigural‘) und dass zudem die Ladungen der Parcels auf die latenten Faktoren zwischen den Gruppen ähnlich ausfallen (‚metrisch‘). Diese Ergebnisse wurden durch die anschließende Parcel-Allocation Methode bekräftigt, womit ausgeschlossen werden konnte, dass die Ergebnisse auf der spezifischen Item-zu-Parcel Konstellation beruhten.

Für die darauf folgende Stufe der skalaren Messinvarianz war dies jedoch nicht mehr eindeutig und nur partiell der Fall: Auch wenn nach dem Kriterium der alternativen Fit-Indizes der

konventionsgemäße Wert von  $\Delta CFI < .01$  nicht unterschritten wurde, wies ein signifikanter  $\chi^2$ -Wert für den Modellvergleich darauf hin, dass die Annahme gleicher Intercepts zwischen den Gruppen zu einer deutlichen Verschlechterung der Modellpassung führte. Dies war nach punktueller Lockerung eines einzelnen Intercept-Parameters nicht mehr der Fall. Daher kann von partieller skalarer Messinvarianz ausgegangen werden, dass folglich (mit der Ausnahme des einen Item-Parcels in der Subskala Verben) die Unterschiede in den manifesten Variablen bei beiden Gruppen auf Unterschiede in den latenten Variablen zurückzuführen sind. Auch auf der anschließenden vierten Stufe der partiellen residualen Messinvarianz konnte gezeigt werden, dass die Residuen der Items für beide Gruppen äquivalent sind. Da die Überprüfung von partieller Invarianz mit der `parcelAllocation`-Methode nicht möglich ist, können diese Ergebnisse nicht validiert werden.

Zusammenfassend kann daraus geschlossen werden, dass sich keine grundsätzlichen Unterschiede in der Struktur des Langzeitgedächtnisses zwischen Kindern mit und ohne Lernbehinderung finden ließen. Dies zeigt sich darin, dass die Messinvarianz zwischen den Gruppen auf konfiguraler und metrischer Ebene vollständig und auf der skalaren und residualen Ebene partiell erfüllt war. Damit kann eine für beide Gruppen vergleichbare Faktoren- und Ladungsstruktur angenommen werden, hinsichtlich der Binnenstruktur der Zuordnung von Item-Parcels zu den vier latenten Fähigkeiten (Benennung von Nomen, Verben, Kategorien, und Gegensätzen).

Somit gab es keine systematischen Unterschiede in den Messmodellen der Gruppen, bis auf die punktuelle Lockerung der Restriktion im Rahmen der partiellen skalaren Invarianz. Die Vergleichbarkeit der Binnenstruktur des Langzeitgedächtnisses von Kindern mit bzw. ohne Lernbehinderung kann als empirischer Hinweis interpretiert werden, dass der Abruf aus dem mentalen Lexikon von Kindern mit Lernbehinderung kein strukturelles Defizit aufweist, sondern höchstens entwicklungsverzögert ist.

In Abschnitt 6.2.26.1 wurde auf das Problem hingewiesen, dass ein signifikantes Ergebnis keine eindeutige Interpretation zuließe, da die Richtung der Abweichung nicht feststellbar ist. So könnten Kinder mit Lernbehinderung theoretisch auch ‚nach oben‘ abweichen und eine mit der CA-Gruppe vergleichbare Struktur aufweisen. Angesichts des vorliegenden nicht-signifikanten Ergebnisses muss diesem Hinweis allerdings nicht weiter nachgegangen werden.

Die Beantwortung der zweiten Teilfragestellung zur differenziellen Vorhersage der Lösungswahrscheinlichkeit erfolgte mithilfe eines generalisierten gemischten linearen Modells. Die Antwort im dichotomen Format (korrekte oder falsche Benennung) einer einzelnen Person auf ein einzelnes Item stellte die abhängige Variable dar, pro Person lagen also 95 Messpunkte (WWT-Items) vor. Die Lösungswahrscheinlichkeit für diese dichotome Variable wurde in dem Modell durch eine Reihe von festen und zufälligen Faktoren, die entweder Charakteristika der Personen



oder der Items darstellten, regressionsanalytisch vorhergesagt. Da jede Person mehrere Items beantwortete und jedes Item von mehreren Personen bearbeitet wurde, waren die Messpunkte und damit die Residuen nicht unkorreliert. Um dieser Datenstruktur Rechnung zu tragen und zufällige personen- und itembezogene Effekte (z. B. eine Person ist systematisch besser als eine andere, oder ein Item ist systematisch schwieriger als ein anderes) von der allgemeinen Residualvarianz zu trennen, wurden ein by-subject random intercept und random slope (für Wortart) sowie ein by-item random intercept angenommen.

Bei den festen Effekten der personenbezogenen Variablen war vor allem die Gruppenzugehörigkeit von Interesse, wobei auch eine Reihe von Kovariaten berücksichtigt wurde, um deren möglichen Einfluss zu testen. So gingen in die Analysen das chronologische Alter, die kognitive Kapazität (CFT-Rohwert) und die AG-Kapazität ein. Auch die Items bringen diverse lexikalische Charakteristika mit sich, die ebenfalls die Lösungswahrscheinlichkeit systematisch beeinflussen könnten. Hier wurden die Wortart (d. h. die Zugehörigkeit des Items zu einer der vier WWT-Wortarten), die Frequency und die Neighborhood-Size in das Modell aufgenommen.

Die Fragestellung, inwiefern die Gruppen vergleichbar waren, wurde durch Interaktionen der genannten Faktoren mit der Gruppenvariable beantwortet. In einem allgemeinen Modellvergleich zur Vorhersagekraft der Interaktionen führte die Aufnahme der Interaktionen zu keiner signifikanten Verbesserung der Modellpassung ( $p = .397$ ). Im vollen Modell erwies sich lediglich der Haupteffekt für „Gruppe“ als signifikant, da für Kinder mit Lernbehinderung eine niedrigere Lösungswahrscheinlichkeit in der Referenzbedingung (Nomen) festzustellen war. Jedoch wurde keine der Interaktionen zwischen den Kovariaten und dem Gruppenfaktor signifikant. Dies deutet darauf hin, dass bei Kindern mit Lernbehinderung der Abruf aus dem mentalen Lexikon mit denselben Faktoren und in einem ähnlichen Ausmaß zusammenhängt wie bei unbeeinträchtigten Kindern des gleichen mentalen Alters. Zum Beispiel unterschieden sich die Gruppen nicht darin, dass eine höhere AG-Kapazität mit einer höheren Wahrscheinlichkeit, die Items im WWT 6-10 korrekt zu benennen, einhergeht, und die Auftretenshäufigkeit eines Wortes hatte in beiden Gruppen den gleichen Einfluss auf die Lösungswahrscheinlichkeit. Es ließen sich somit keine gruppenspezifischen differenziellen Zusammenhänge des Gedächtnisabrufs mit personen- oder itembezogenen Variablen feststellen.

Insgesamt lässt sich für die kognitive Lernvoraussetzung des Abrufs aus dem mentalen Lexikon schlussfolgern, dass Schüler\*innen mit Lernbehinderung eine ähnliche kognitive Struktur (Thomas et al., 2009) wie die unbeeinträchtigte jüngere Kontrollgruppe aufwiesen. Die Kontrollgruppe war zwar nach chronologischem Alter jünger, dafür aber nach mentalem Alter (gemessen am Rohwert des CFT 1-R) vergleichbar. Gemäß dem MA-Matching Ansatz (siehe Kapitel 3.1) ist die Langzeitgedächtnis-Struktur damit als entwicklungsverzögert im Rahmen des

allgemeinen niedrigeren mentalen Alters zu interpretieren; eine spezifische und qualitativ andersartige Beeinträchtigung im Langzeitgedächtnis, die über die Maße des jüngeren Entwicklungsalters hinausgeht (siehe auch Mähler, 2007; Zigler & Balla, 1982) scheint nicht vorzuliegen.

Für diese Interpretation sprechen zusammenfassend die Befunde aus zwei untersuchten Teilfragestellungen, die mit unterschiedlichen methodischen Herangehensweisen beantwortet wurden: Der Vergleich der Messmodelle durch eine Multigroup-CFA deutete darauf hin, dass die Binnenstruktur des Langzeitgedächtnisses als äquivalent angesehen werden kann, und im Rahmen der regressionsanalytischen Vorhersage der Lösungswahrscheinlichkeiten zeigten sich keine signifikanten Interaktionen mit dem Faktor Gruppe, sodass auf eine ähnliche Außenstruktur geschlossen werden kann.

#### **6.4.2. Limitationen und Ausblick**

Die Studie unterliegt sowohl methodischen als auch theoretischen Limitationen. Die geringe Stichprobengröße ist eine methodische Einschränkung, die insbesondere beim Ansatz der MG-CFA in der ersten Teilfragestellung die Identifikation eines Modells angesichts der Vielzahl der zu schätzenden Parameter erschwert. So traten auch in einem zunächst angestrebten Versuch eines partiellen Bootstrappings immer wieder Konvergenzprobleme auf. Durch die Strategie des Item-Parcelings konnte diese Limitation teilweise umgangen werden, jedoch birgt auch die zufällige Zuweisung von Items zu Parcels, wie sie im ersten Auswertungsschritt vorgenommen wurde, Probleme. Zu diesen Problemen gehört, dass Effekte auch auf die Zusammensetzung der zufällig erstellten Parcels zurückzuführen sein könnten (Sterba & Rights, 2017). Daher wurde in einem nachgelagerten Schritt die Item-zu-Parcel Zuweisung über einhundert Iterationen permutiert und für die jeweils entstehenden Parcels die Überprüfung der Messinvarianz durchgeführt. Dies führte zu weitestgehend ähnlichen Ergebnissen und legitimiert somit die eingesetzte Strategie des Itemparcelings.

Weiterhin lässt das Parceling keine Rückschlüsse darauf zu, ob und welche spezifischen Items sich für die Modellpassung als problematisch erweisen, sondern liefert lediglich Hinweise auf die globale Faktorenstruktur auf Ebene der Parcels als manifeste Variablen und des zu Grunde liegenden latenten Konstruktes. Dies sollte zwar bei der Interpretation berücksichtigt werden, schränkt aber die Aussagekraft zur übergeordneten Binnenstruktur des mentalen Lexikons auf der latenten Konstruktebene nicht ein. Für die Identifikation kritischer, nicht äquivalent messender Items (Differential Item Functioning) wird in Studie 3 (Kapitel 7) der IRT-Ansatz genutzt.

Auf theoretischer Ebene ist kritisch anzumerken, dass der hier gemessene Abruf von Wörtern aus dem mentalen Lexikon nur einen kleinen Ausschnitt des Langzeitgedächtnisses dar-

stellt. Damit ergeben sich mindestens drei theoretische Limitationen hinsichtlich der Konzeption des Langzeitgedächtnisses, auf die in den folgenden Abschnitten genauer eingegangen wird: (1) Einspeicherungs- und Konsolidierungsprozesse bleiben unberücksichtigt; (2) es wird nicht zwischen Alltags- und Bildungssprache differenziert; (3) die Netzwerkstruktur des mentalen Lexikons wird nicht abgebildet.

Insofern muss die Gültigkeit der Aussagen in dieser Studie dementsprechend eingeschränkt werden, denn alle drei Limitationen lassen sich nur durch weitere Forschungsarbeiten mit neuen und eigens zu diesem Zweck erhobenen Daten befriedigend adressieren. Hinsichtlich der Limitationen (2) und (3) ist es jedoch anhand der verfügbaren Daten zumindest tentativ möglich, erste Hinweise als Ausblick zur zukünftigen Erforschung zu erhalten. Diese Möglichkeiten werden im Anschluss an die jeweilige Limitation dargestellt.

Die erste Limitation betrifft den Fokus auf dem *Abruf* von Wörtern (Lexemen) aus dem mentalen Lexikon. Aussagen zu den vorgeschalteten Teilprozessen des Gedächtnismodells (Anderson, 2013), nämlich Einspeicherung und Konsolidierung, können somit nicht getroffen werden. Es wäre durchaus plausibel, hier differenzielle Befunde für Schüler\*innen mit Lernbehinderung zu erwarten. Möglicherweise werden schon bei der Einspeicherung weniger oder undifferenziertere Merkmale abgespeichert (Büttner, 1998), oder es ist eine größere Anzahl an Wiederholungen für eine adäquate Einspeicherung nötig. Hinsichtlich der Konsolidierung könnte im Rahmen der Strategie-Hypothese nach Klauer und Lauth (1997) vermutet werden, dass weniger oder ineffektivere Strategien zum Memorieren bereitstehen bzw. diese inadäquat genutzt werden. Um diese Aspekte zu untersuchen, wären geeignete experimentelle Interventionsstudien zu planen, die z. B. testen, inwiefern die Unterrichtung einer expliziten Strategie in beiden Gruppen zu vergleichbaren Leistungssteigerungen führt.

Weiterhin führt die Reduktion auf den Wortabruf dazu, dass die Inhalte des semantischen Gedächtnisses nur unzureichend in ihrer Komplexität repräsentiert werden. So gehört zu jedem Wort ein Konzept (Lemma), welches wiederum mit anderen Konzepten vernetzt ist (Glück, 2011). Inwiefern sich solche propositionalen Netze zur Wissensrepräsentation auf konzeptueller Ebene zwischen Kindern mit und ohne Lernbehinderung unterscheiden, lässt sich nicht einfach quantitativ erfassen. Hier erscheinen explorative, qualitative Verfahren zur Hypothesengenerierung nötig.

Die zweite Limitation betrifft die fehlende Differenzierung des Sprachregisters nach Alltags- und Bildungssprache (Morek & Heller, 2012) in der vorliegenden Studie. Die Limitation wird hier zunächst erläutert und daraufhin ein erster tentativer Ausblick zur Auswertung der vorhandenen Daten vorgeschlagen. Die Unterscheidung in Alltags- und Bildungssprache könnte eine

weitere plausible Variable darstellen, mit der Unterschiede zwischen den Gruppen in der Lösungswahrscheinlichkeit von Items erklärt werden könnten. Ein Grund könnte sein, dass Kinder mit Lernbehinderung möglicherweise in einem geringeren oder abweichenden Maß mit bildungssprachlichen Äußerungen in Berührung kommen und somit bei diesen Stimuli systematisch benachteiligt wären. Jedoch ist eine klare Einteilung in bildungssprachlich oder alltagssprachlich gerade für einzelne lexikalische Einheiten nicht eindeutig möglich, da sich die Einteilung häufiger auf ganze Sätze oder Textabschnitte bezieht (Morek & Heller, 2012). In die Einschätzung fließen neben lexikalischen auch syntaktische und grammatische Ebenen mit ein (z. B. wie lang sind Wörter, Sätze; wie viele und welche Konjunktionen werden verwendet; Stahns, 2016). Insofern kann eine eindeutige und trennscharfe Zuordnung einzelner Wörter des WWT 6-10 zu entweder Alltags- oder Bildungssprache nicht vorgenommen werden.

Im Bewusstsein dieser Einschränkungen wurde dennoch ein Versuch der Auswertung unternommen, um einen ersten Hinweis zu erhalten, ob möglicherweise die Zugehörigkeit zu einem Sprachregister zusätzlich die Lösungswahrscheinlichkeit vorhersagen kann. Dazu wurde als approximative Variable eine heuristische Zuteilung zu einem Register vorgenommen, orientiert an der Operationalisierung nach Stahns (2016). Als Entscheidungsregel wurde dabei eine einfache Heuristik angewandt: Wie wahrscheinlich erschien es, dass ein Kind das entsprechende Wort in einer Alltagssituation im Gespräch mit seinen Eltern verwendet? So wurden z. B. „langweilig“ oder „Möbel“ als alltagssprachlich eingeschätzt, während „demonstrieren“ oder „dirigieren“ der Bildungssprache zugeordnet wurden. Da die Variable einer theoretischen Fundierung entbehrt und es zudem viele Zweifelsfälle gab, die ähnlich wahrscheinlich einem alltags- wie bildungssprachlichen Kontext zugeordnet werden können (z. B. „Fackel“ oder „Krücke“), darf dieses tentative Ergebnis nur mit Zurückhaltung interpretiert werden und soll hier lediglich als Ausblick für weitere Untersuchungen dargestellt werden.

Für das Modell mit „Register“ als weiterem Prädiktor ergab sich eine signifikant bessere Modellpassung,  $\Delta\chi^2(1) = 7.73$ ;  $p = .005$ . Die Aufnahme der Interaktion von „Register“ mit dem Faktor Gruppe führte dagegen zu keinem signifikant besseren Modellfit,  $\Delta\chi^2(1) = 0.65$ ;  $p = .419$ . Dieses vorläufige Ergebnis darf wie gesagt nur sehr zurückhaltend interpretiert werden. Es könnte allerdings einen ersten Hinweis darauf geben, dass Kinder diejenigen Wörter, die gemäß einer Spontaneinschätzung eher in Alltagssituationen Verwendung finden, leichter

aus dem mentalen Lexikon abrufen können. Dies zeigt sich ebenfalls in einer signifikanten mittleren bis hohen punktbiserialen Korrelation von  $r = -.46$  ( $p < .001$ )<sup>26</sup> mit der im Manual berichteten Itemschwierigkeit: Wörter, die als bildungssprachlich eingeschätzt wurden, wurden in der Normstichprobe seltener korrekt beantwortet.

Der positive Zusammenhang der Zugehörigkeit zum Alltagsregister mit der Lösungswahrscheinlichkeit scheint jedoch für beide Gruppen gleichermaßen zu gelten, denn das Ausbleiben einer signifikanten Interaktion deutet darauf hin, dass als bildungssprachlich eingeschätzte Wortschatzelemente für Schüler\*innen mit Lernbehinderung nicht schwieriger zu lösen waren als für die MA-Kontrollgruppe. Es könnte sich somit in zukünftigen Forschungsarbeiten lohnen, das Konstrukt von Alltags- und Bildungssprache mit einzubeziehen. Dies setzt jedoch eine intensivere theoretische Auseinandersetzung und eine klarere Operationalisierung des Konstruktes voraus, um die empirische Erfassung reliabel und valide zu gewährleisten. Bis dahin muss die Interpretation spekulativ bleiben.

Die dritte Limitation schließlich betrifft die fehlende Berücksichtigung der Netzwerkstruktur in den berichteten Auswertungen. Die plausible Annahme, dass Wissen im semantischen Gedächtnis als Netzwerk organisiert ist (z. B. Collins & Loftus, 1975; zusammenfassend Anderson, 2013) ist erst in jüngster Vergangenheit empirisch überprüfbar (z. B. Bassett & Sporns, 2017; Beckage & Colunga, 2016). Der Multigroup-CFA Ansatz betrachtet zwar die Zusammenhangsmuster innerhalb der Items bzw. Parcels, ist aber nicht in der Lage, eine Netzwerkstruktur abzubilden. Das GLMM betrachtet nur den Einfluss von anderen personenorientierten und lexikalischen Prädiktoren, nicht jedoch die Verhältnisse der Items untereinander.

Insofern erscheint es sicherlich aufschlussreich, die Netzwerkstruktur im Langzeitgedächtnis der beiden Gruppen zu vergleichen. Eine quantitative Beschreibung von Netzwerken ist mittlerweile mit Methoden aus der Graphentheorie möglich (Boccaletti, Latora, Moreno, Chavez & Hwang, 2006; Borge-Holthoefer & Arenas, 2010). Für vielfältige Fragestellungen in den Kognitionswissenschaften haben sich diese Methoden bereits etabliert (Baronchelli et al., 2013; Castro & Siew, 2019) und sind für Forschungsfragen im Bildungskontext ebenfalls ein vielversprechender Ansatz (Siew, 2020). So scheint die mittels der Netzwerk-Theorie abbildbare Struktur des mentalen Lexikons für den Abruf eine wichtige Rolle zu spielen (Chan & Vitevitch, 2010).

Der Netzwerk-Ansatz kann auch dazu genutzt werden, Gruppen mit und ohne Beeinträchtigungen zu vergleichen, z. B. Erwachsene mit und ohne Störungen im Autismus-Spektrum (Kenett, Gold et al., 2016). Auch für Personenmerkmale wie Kreativität und Intelligenz (Kenett,

---

<sup>26</sup> Alltagssprache war als „0“ kodiert, Bildungssprache als „1“. Die negative Korrelation bedeutet also, dass bildungssprachlich eingeschätzte Items geringere Werte in der Itemschwierigkeit (zu interpretieren als Lösungswahrscheinlichkeit) aufwiesen.

Anaki & Faust, 2014; Kenett, Beaty et al., 2016; Benedek et al., 2017) ließen sich Zusammenhänge mit differenzieller assoziativer Netzwerkstruktur nachweisen. Diese methodischen Entwicklungen stellen folglich einen aussichtsreichen Ansatz dar, mit dem die Netzwerkstruktur von Wissensinhalten im semantischen Gedächtnis modelliert und zwischen Gruppen verglichen werden könnte.

Die Netzwerk-Analyse verspricht einen hohen Informationsmehrwert, weil mit ihr die Frage beantwortet werden kann, ob das mentale Lexikon bei Kindern mit Lernbehinderung tatsächlich schlechter vernetzt und/oder „chaotischer strukturiert“ ist (Bruns & Grosche, 2019, Feb), wie häufig postuliert wird (z. B. Büttner, 1998; Lauth, Brunstein & Grünke, 2014; Schröder, 2000). Im Rahmen der vorliegenden Arbeit wäre eine Umsetzung aber erstens wegen der theoretischen Komplexität (Boccaletti et al., 2006) und zweitens wegen der notwendigen, aber sehr zeit- und ressourcenaufwendigen Erhebung von Wort-Assoziationen als Datengrundlage nicht möglich gewesen. Dennoch soll im Rahmen der Gesamtdiskussion (Kapitel 8.3) eine Netzwerk-Auswertung der vorliegenden WWT-Daten als Ausblick berichtet werden, um den inhaltlichen Mehrwert zu demonstrieren.

#### **6.4.3. Fazit Studie 2**

Im Rahmen dieser Studie wurde die globale Struktur im Langzeitgedächtnis, gemessen am expressiven Wortschatz bzw. dem Abruf aus dem mentalen Lexikon, für Schüler\*innen mit Lernbehinderung untersucht. Es konnten dabei keine wesentlichen Unterschiede zur Langzeitgedächtnis-Struktur von jüngeren unbeeinträchtigten Schüler\*innen nachgewiesen werden. Dies zeigte sich in Teilfragestellung 1 am Befund der metrischen Messinvarianz. Diese besagt, dass die Faktoren- und Ladungsstruktur zwischen den Gruppen vergleichbar sind. Inhaltlich bedeutet das, dass die einzelnen Items bzw. Parcels in beiden Gruppen ähnlich stark mit den latenten Konstrukten zusammenhängen; die Zusammensetzung der latenten Fähigkeiten (die Benennung von Nomen, Verben, Kategorien und Gegensätzen) ist also in beiden Gruppen gleich, sodass nicht von einem systematischen Unterschied in der Repräsentation der verschiedenen Wortarten ausgegangen werden kann. Dass die skalare Messinvarianz nicht eindeutig erfüllt war, deutete auf einen möglichen differenziellen Einfluss anderer Variablen hin. Diesbezüglich konnten im Rahmen des GLMMs in Teilfragestellung 2 keine Belege für einen differenziellen Einfluss von AG-Kapazität, kognitiver Kapazität sowie von lexikalischen Variablen (Häufigkeit und Neighborhood-Size) nachgewiesen werden.

Es wird daher die Schlussfolgerung gezogen, dass Kinder mit Lernbehinderung eine ähnliche Langzeitgedächtnis-Struktur<sup>27</sup> aufweisen wie unbeeinträchtigte Kinder, die chronologisch jünger sind und ein vergleichbares mentales Alter haben. Im Sinne der Frage nach dem Entwicklungsmuster von delay oder deviance (Mähler, 2007; Zigler & Balla, 1982) liegt hiermit also ein Beleg gegen eine strukturelle Abweichung vor. Stattdessen sprechen die Ergebnisse dafür, dass Kinder mit Lernbehinderung im Langzeitgedächtnis lediglich eine graduelle Entwicklungsverzögerung aufweisen, die mit der generell verlangsamten mentalen Entwicklung konform ist. Als Ausblick sei hier auf zusätzliche Auswertungen von Netzwerkanalysen in der Gesamtdiskussion (Kapitel 8.3) verwiesen, da sie tiefergehende Einblicke in die Struktur des mentalen Lexikons von Kindern mit Lernbehinderung versprechen und möglicherweise in der Lage sind, Unterschiede in der Netzwerkstruktur aufzudecken.

---

<sup>27</sup> Insoweit die Binnenstruktur mit den hier verwendeten Methoden abbildbar ist.

## 7. Absicherung des Redintegration-Effekts (Studie 3)

In diesem Kapitel wird die dritte Teilstudie vorgestellt, die als ein ‚short-paper‘ und Folgestudie zu Studie 1 zu verstehen ist. Das Ziel besteht in der Absicherung, dass ein in Studie 1 gefundener differenzieller Effekt (siehe Abschnitt 5.4.3) kein durch Differential Item Functioning (DIF) verursachtes Artefakt darstellt.

Bei besagtem Effekt aus der ersten Studie handelte es sich um eine unerwartete Interaktion zwischen Redintegration und Wortschatzumfang bei Kindern mit Lernbehinderung: Im Verhältnis zu chronologischem Alter und kognitiver Kapazität wirkte Redintegration bei den Kindern mit Lernbehinderung auf den ersten Blick unbeeinträchtigt. Im Verhältnis mit dem Wortschatz (als Proxy für das Langzeitgedächtnis) zeigte sich jedoch überraschend ein differenzielles Muster, nämlich eine Abnahme der Redintegration-Effektivität mit wachsendem Wortschatz bei Kindern mit Lernbehinderung. Mit anderen Worten: Je mehr Wörter ein Kind mit Lernbehinderung in seinem Wortschatz besaß, desto geringer fiel im Durchschnitt die Redintegration-Effektivität aus, d. h. desto schlechter konnte es im Durchschnitt die Wörter in den Arbeitsgedächtnis-Aufgaben aus dem Langzeitgedächtnis rekonstruieren.

Einerseits könnte die Interaktion inhaltlich erklärt werden, wenn Kinder mit Lernbehinderung eine von unbeeinträchtigten Kindern abweichende Struktur im Langzeitgedächtnis aufweisen; dieser Frage wurde im Rahmen der zweiten Studie nachgegangen. In der nun folgenden dritten Studie wird die methodische Erklärung untersucht, ob die beobachtete Interaktion von Redintegrationprozess und Wortschatzumfang auch auf Differential Item Functioning bei der Messung des Wortschatzes zurückzuführen sein könnte. Denn bei der Besprechung der unerwartet aufgetretenen Interaktion (siehe Diskussionsteil Studie 1, Kapitel 5.5; und Bruns & Grosche, 2018, Juni) erschien u. a. eine fehlende Item-Äquivalenz des Instrumentes für den Wortschatzumfang (WWT 6-10) als eine mögliche und plausible Erklärung.

Bevor eine inhaltliche Interpretation der Interaktion dahingehend zulässig ist, dass Schüler\*innen mit Lernbehinderung hinsichtlich der Nutzung ihres Wortschatzwissens für Redintegrationprozesse tatsächlich eine differenzielle entwicklungsbezogene Relation (siehe Kapitel 3.2.4) aufweisen, muss ein methodisch bedingtes Artefakt als Ursache ausgeschlossen werden.

Ein solches Artefakt könnte möglicherweise darin bestehen, dass im WWT 6-10 zwischen den Gruppen keine vollständige Messinvarianz gegeben ist. Die im Rahmen von Studie 2 erfolgte Überprüfung der Messinvarianz mithilfe der Parceling-Strategie erlaubte allerdings nur Aussagen über die globale Äquivalenz auf der Konstruktebene der latenten Faktoren, nicht jedoch eine Aussage auf Itemebene. Denn aufgrund der Zusammenfassung der Items zu Parcels können keine einzelnen Items auf ihre Messäquivalenz hin überprüft werden. Daher soll in dieser Folgestudie die Messäquivalenz auf Itemebene überprüft werden, um diejenigen Items zu



identifizieren, die DIF aufweisen. Daraufhin können diese problematischen Items aus dem WWT-Score ausgeschlossen und die Interaktion des Redintegration mit dem bereinigten Wortschatzumfang erneut überprüft werden. Dies erfolgt in zwei Schritten:

Als erstes werden die WWT-Items pro Subskala einzeln analysiert, ob sie DIF aufweisen. Als DIF wird im Kontext der Item-Response-Theorie (IRT) der Fall bezeichnet, wenn ein Item nicht für zwei oder mehrere Gruppen nicht gleich ‚funktioniert‘, d. h. für eine Personengruppe (z. B. Altersgruppen, Geschlechter oder Kinder mit vs. ohne Lernbehinderung) systematisch schwerer oder leichter zu lösen ist.

Im zweiten Schritt werden die so identifizierten DIF-Items aus der Berechnung des Wortschatzumfangs ausgeschlossen, um einen bereinigten Wortschatz-Wert zu erhalten, bei dem DIF keinen Einfluss hat. Sollte sich in der erneuten Developmental Trajectory Analyse der Interaktion von Redintegrationprozess und bereinigtem Wortschatz herausstellen, dass diese nicht mehr signifikant wird, so muss DIF als (mit) ursächlich für das Auftreten in Studie 1 angenommen werden und darf inhaltlich nicht weiter interpretiert werden. Hat die Interaktion jedoch auch mit dem bereinigten WWT-Score als Prädiktor weiterhin Bestand, so kann DIF als Ursache ausgeschlossen werden. Dann erscheint es plausibel, eine inhaltliche Erklärung für das Auftreten der unerwarteten Interaktion anzunehmen. Eine solche inhaltliche Erklärung wäre z. B. die Annahme, dass Kinder mit Lernbehinderung bei der Rekonstruktion von Arbeitsgedächtnis-Inhalten weniger effektiv auf ihr im Langzeitgedächtnis gespeichertes Wissen zugreifen können.

## 7.1. Einleitung und Fragestellung

Wie bei Leistungstests üblich ist das Antwortformat der Items im WWT 6-10 dichotom, da nur zwei Möglichkeiten eintreten können: das Item wurde korrekt beantwortet („1“) oder nicht („0“). Zur Untersuchung solcher Variablen bietet sich die Modellierung mithilfe des Raschmodells aus dem Framework der Item Response Theory (IRT) an, die auch probabilistische Testtheorie genannt wird (Eid & Schmidt, 2014). Auch wenn IRT häufig als gegenläufige Alternative zur klassischen Testtheorie aufgefasst wird, plädiert Moosbrugger (2012) dafür, sie als eine Ergänzung zur klassischen Testtheorie zu bezeichnen.

Eine umfassende Darlegung der mathematischen Grundlagen des Raschmodells würde den Rahmen der Arbeit sprengen, daher sei an dieser Stelle auf die einschlägigen Publikationen verwiesen (Fischer & Molenaar, 1995; Moosbrugger, 2012). Im Folgenden werden die wichtigsten Aspekte knapp zusammengefasst: (1) die Begriffe der Personen- und Itemparameter und deren Schätzung; (2) Annahmen des Raschmodells, insbesondere die lokale stochastische Unabhängigkeit sowie die Subgruppeninvarianz; und (3) Verfahren zur Modellprüfung, insbesondere quasi-exakte Prozeduren bei kleinen Stichproben.

Im Rahmen einer IRT-Modellschätzung wird aus dem Antwortverhalten (*responses*) in manifesten Variablen auf die dahinterliegende latente Merkmalsausprägung (den Personenparameter  $\xi$ ) geschlossen. Auch die Items besitzen einen Parameter (den Itemparameter  $\sigma$ ). Üblicherweise wird im bildungswissenschaftlichen und psychometrischen Kontext der Personenparameter als die Fähigkeit einer Person und der Itemparameter als die Schwierigkeit eines Items interpretiert (Debelak, 2019).

Durch eine sog. itemcharakteristische Funktion (IC-Funktion) wird „die Beziehung zwischen dem manifesten Antwortverhalten [...] und der Ausprägung auf dem latenten Trait [Merkmal]“ (Moosbrugger, 2012, S. 233) mathematisch abgebildet. Diese IC-Funktion ist bei einem Raschmodell eine logistische Funktion (Fischer, 1995) und drückt das Verhältnis von Personen- zu Itemparameter als Wahrscheinlichkeit dafür aus, dass eine Person (mit ihrem gegebenen Personenparameter) ein Item (mit dessen gegebenem Itemparameter) lösen wird. Diese beiden Parameter werden vom Modell so geschätzt, dass sie die empirisch erhobenen Daten möglichst gut abbilden (Molenaar, 1995). Die Modellpassung der Parameter zu den empirischen Daten wird durch die Likelihood angegeben. Mithilfe von iterativen Verfahren werden die Parameter schrittweise angepasst, sodass diese Likelihood maximiert wird. In dieser Studie wird die im R-Paket `eRm` (Mair, Hatzinger & Maier, 2015) implementierte conditional maximum likelihood (CML)-Methode zur Parameterschätzung genutzt, die bei Mair & Hatzinger (2007) und Molenaar (1995) erläutert wird.

Personen- und Itemparameter können auf einer gemeinsamen Skala angegeben werden: Liegt die Personenfähigkeit über der Itemschwierigkeit, ist es wahrscheinlicher, dass die Person das Item löst, umgekehrt entsprechend. Sind Personenfähigkeit und Itemschwierigkeit gleich, liegt die Wahrscheinlichkeit, dass die Person das Item löst, bei  $P = .50$  (Moosbrugger, 2012).

Zwei zentrale Voraussetzungen des Raschmodells sind die lokale stochastische Unabhängigkeit und die Subgruppeninvarianz (Koller et al., 2012). Die lokale stochastische Unabhängigkeit bedeutet, dass die Skala eindimensional ist, also nur eine latente Fähigkeit gemessen wird. Die Eindimensionalität zeigt sich daran, dass es bei einem gegebenen („lokalen“) Wert des Personenparameters keine Korrelationen zwischen den Items geben darf. Systematische Unterschiede in den Items sollen also allein auf variierende Personenfähigkeiten zurückgeführt werden (Moosbrugger, 2012).

Die Subgruppeninvarianz bedeutet, dass die Schätzung der Itemparameter unabhängig von der Stichprobe ist, also davon, welche Personen den Test bearbeitet haben. Unabhängig von z. B. Geschlecht oder Alter der Probanden sollten die Itemparameter (d. h. Schwierigkeit der Items) gleich sein. Wenn die Invarianzannahme verletzt ist, misst der Test für die Gruppen

systematisch unterschiedlich. Dies wird als Differential Item Functioning (DIF) bezeichnet, wenn die Items systematisch zu schwierig oder zu leicht für eine spezifische Subgruppe sind.

Die Annahme der Subgruppeninvarianz lässt sich überprüfen, indem die Stichprobe in zwei Hälften (z. B. Schüler\*innen-Gruppen) geteilt wird und das Modell für jede Gruppe separat geschätzt wird. Wenn diese separat geschätzten Itemparameter der Gruppen nicht voneinander abweichen, ist die Subgruppeninvarianz erfüllt (Moosbrugger, 2012). Ein Grund dafür, dass die Subgruppeninvarianz nicht erfüllt ist, also ein Item systematisch schwerer oder leichter ist, könnte z. B. sein, dass spezifisches Erfahrungswissen nötig ist, welches nur eine Gruppe besitzt, das der anderen Gruppe hingegen fehlt.

Die Überprüfung der Modellkonformität, insbesondere zur Identifikation von Items mit DIF, kann mit Verfahren auf Skalenebene und auf Ebene einzelner Items vollzogen werden. Ein häufig gewähltes Verfahren zur Überprüfung der Subgruppeninvarianz für die gesamte Skala ist der Likelihood Ratio Test (LR-Test) nach Andersen (1973). Ähnlich wie in Studie 2 wird die Passung des Modells zu den Daten, ausgedrückt durch die Likelihood, eines gemeinsamen Modells mit derjenigen Modellpassung verglichen, die aus einer nach Subgruppen getrennten Parameterschätzung resultiert. Der LR-Test prüft damit die Invarianz auf Ebene der gesamten Skala. Soll für jedes einzelne Item geprüft werden, ob sich bei einer nach Gruppen getrennten Schätzung unterschiedliche Itemparameter ergeben, kommt der Wald-Test zum Einsatz.

Für geringe Stichprobengrößen, bei denen die Annahme einer  $\chi^2$ -Verteilung der Likelihood nicht erfüllt ist, schlagen Koller et al. (2012) und Koller & Hatzinger (2013) non-parametrische, quasi-exakte Tests vor, die auf zufällig simulierten Matrizen beruhen (Ponocny, 2001; Verhelst, 2008). Ausgehend von der Annahme der Suffizienz der Randsummen (Fischer, 1995) werden binäre Matrizen (d. h. mit Einträgen von „0“ oder „1“) mit ebenso vielen Personen und Items wie im ursprünglichen Datensatz simuliert, sodass deren Randsummen den empirisch gefundenen Randsummen entsprechen (Koller et al., 2012). Aus der sehr großen Anzahl möglicher Matrizen wird über einen Permutationsalgorithmus eine zufällige Auswahl getroffen. Nach der Logik von quasi-exakten Tests wird ausgezählt, wie häufig in den simulierten Matrizen das in der Ausgangsmatrix gefundene Ergebnis oder ein extremeres Ergebnis produziert wurde. Dieser Häufigkeitswert, geteilt durch die Anzahl der simulierten Matrizen, kann als  $p$ -Wert interpretiert werden (Ponocny, 2001).

Das non-parametrische Pendant zum LR-Test für die globale Überprüfung der Subgruppeninvarianz ist die Teststatistik  $T_{10}$ . (Koller et al., 2012). Koller (2010, zit. n. Koller et al., 2012) berichtet, dass diese non-parametrischen Testvarianten bereits ab Stichprobengrößen von 50-100 Probanden pro Gruppe bei maximal zwei Gruppen angewendet werden können. Basierend

auf einer neueren Power-Analyse empfehlen Koller, Maier & Hatzinger (2015) eine Stichprobengröße von  $n \geq 100$  Probanden, mit der auch moderate Verletzungen der Subgruppeninvarianz durch die Teststatistik  $T_{10}$  aufgedeckt werden können; hinsichtlich der alpha-Fehlerrate (Type-I-error) erwies sich  $T_{10}$  bereits bei sehr kleinen Stichproben von  $n = 30$  als robust (siehe auch Debelak, 2019). Da die Stichprobe in der vorliegenden Studie in beiden Gruppen (nahezu) 100 Probanden umfasste, erschien die Verwendung der  $T_{10}$ -Statistik zur Absicherung der Subgruppeninvarianz für diese Studie angemessen. Zur zusätzlichen Überprüfung der Eindimensionalität der Skala wird die non-parametrische Teststatistik  $T_{11}$  herangezogen, die sich bei Debelak (2019) als sensitiv gegenüber Verletzungen der Modellannahmen bewährt hat.

Die Fragestellung für diese Studie 3 lässt sich in zwei aufeinander aufbauende Schritten formulieren: (1) Gibt es in den Subskalen des WWT 6-10 Items, die für die Kinder mit Lernbehinderung und die Kinder der Kontrollgruppe systematisch unterschiedlich messen, d. h. DIF aufweisen? (2) Daran schließt sich die Fragestellung der Re-Analyse an, nachdem die DIF-Items aus dem Wortschatz-Wert ausgeschlossen wurden: Hat die Interaktion von Redintegration und Wortschatzumfang Bestand, auch wenn DIF-Items aus dem Prädiktor ausgeschlossen werden? Auch wenn in Studie 2 anhand des Strukturmerkmals Wortarten weitgehend gezeigt werden konnte, dass auf Konstruktebene (mit Item-Parcels) Messinvarianz vorlag, wird vermutet, dass einzelne Items des WWT 6-10 für die beiden Gruppen unterschiedlich funktionieren könnten (also DIF vorliegt). Gleichzeitig wird aber erwartet, dass auch bei Ausschluss dieser DIF-behafteten Items die Interaktion zwischen Redintegrationprozess und Wortschatzumfang signifikant bleibt.

## 7.2. Methode

Die Stichprobe ( $N = 207$ ) ist identisch zu derjenigen aus Studie 2 (siehe Abschnitt 6.2). In die Auswertung flossen die Daten aller Personen ein, die den WWT 6-10 vollständig bearbeitet hatten und bei denen auf Seite der Kontrollgruppe keine Ausschlusskriterien vorlagen, die auf eine Beeinträchtigung schließen ließen. Das Material bestand ebenfalls aus dem bereits in Kapitel 5.3 und Kapitel 6.2 ausführlicher geschilderten Verfahren des WWT 6-10 (Glück, 2011).

Die Effektivität des Redintegrationprozesses, die abschließend re-analysiert wird, wurde im Rahmen der Theorie (Kapitel 2.1.1.2) und in Studie 1 (Kapitel 5.4.3) ebenfalls bereits eingehend erläutert: operationalisiert wird der Redintegrationprozess als Lexikalitätseffekt in Arbeitsgedächtnis-Wortspannenaufgaben. Wortspannenaufgaben erfordern die unmittelbare serielle Wiedergabe einer Sequenz von Stimuli. Der Lexikalitätseffekt besagt, dass Wortspannen mit echten Wörtern besser behalten werden können als Wortspannen mit Kunstwörtern.

Für alle Analysen und Modellschätzungen kam das R-Paket `eRm` (Mair et al., 2015) zum Einsatz. Um Items mit DIF zu identifizieren, wurde zunächst pro Wortart-Subskala des WWT 6-10 (Nomen, Verben, Adjektive, Oberbegriffe) ein Raschmodell angepasst. Die Überprüfung der Messäquivalenz dieses gefitteten Modells erfolgt zum einen mit dem Andersen LR-Test (Andersen, 1973), welcher Auskunft über die globale Vergleichbarkeit der gesamten Skala gibt. Zum anderen wird als analoges non-parametrisches Verfahren für kleine Stichproben die quasi-exakte Teststatistik  $T_{10}$  (Koller et al., 2012; Ponocny, 2001) eingesetzt. Auch  $T_{10}$  informiert über die Vergleichbarkeit der gesamten Skala, ist aber robuster gegen Voraussetzungsverletzungen (Koller et al., 2015). Der Vergleich auf Ebene der einzelnen Items zum Zweck der Identifikation von problematischen Items erfolgt anhand des Wald-Tests.

### 7.2.1. Ermittlung von Items mit Differential Item Functioning

Die Ermittlung von DIF in dieser Studie erfolgte nach Wortarten getrennt (Nomen, Verben, Adjektive, und Kategorien). Im Folgenden wird das Vorgehen anhand der Subskala für Nomen exemplarisch beschrieben, bei der Auswertung der weiteren Subskalen wird analog verfahren. Die Identifikation umfasste vier Schritte: 1) Schätzung eines gemeinsamen Raschmodells über beide Gruppen und erste globale DIF-Analyse (Schätzung von getrennten Raschmodellen pro Gruppe und Vergleich der Anpassungsgüte). 2) Identifikation von einzelnen DIF-Items. 3) weitere globale DIF-Analyse mit der resultierenden reduzierten Subskala.

Im ersten Schritt wurde für alle 26 Items der Subskala Nomen ein Raschmodell für beide Gruppen gemeinsam mithilfe des R-Befehls `RM()` angepasst. Dieses wird über den globalen Andersen Likelihood Ratio Test auf Subgruppeninvarianz überprüft. Dabei wird die Likelihood als Maß der Anpassungsgüte von Daten und Modell für zwei Varianten verglichen: die Likelihood des gemeinsamen Modells ohne Differenzierung nach Gruppen, und die Likelihood eines Modells, in dem die Parameter nach Gruppen getrennt geschätzt werden<sup>28</sup>. Die Nullhypothese besagt, dass es keinen Unterschied zwischen den Likelihood-Werten für das gemeinsame und das getrennte Modell gibt. Ein signifikantes Ergebnis besagt, dass eine nach Gruppen getrennte Schätzung der Parameter zu einer deutlich höheren Modellpassung führen würde. In dem Fall muss davon ausgegangen werden, dass die Skala insgesamt nicht invariant misst.

Der zweite Schritt umfasst die Identifikation derjenigen Items, die in den Gruppen systematisch unterschiedliche Schwierigkeitsparameter aufweisen. Dies erfolgt über einen iterativen Prozess, der über die R-Funktion `stepwiseIt()` aufgerufen wird. In jedem Schritt wird dasjenige Item ausgeschlossen, dessen Itemparameter die größte Diskrepanz zwischen den

---

<sup>28</sup> Die Schätzung der getrennten Parameter erfolgt mithilfe der R-Funktion `LR-Test()` automatisch, wenn ein Teilungskriterium angegeben wird.

Gruppen aufweist. Als Teststatistik für die Diskrepanz wird der Wald-Test herangezogen, welcher die Differenz der Itemparameterschätzwerte ins Verhältnis zu deren geschätzten Standardfehlern setzt (Koller et al., 2012). Damit ist die Teststatistik standardnormalverteilt und kann als z-Wert interpretiert werden. Nach jedem Ausschluss eines Items wird ein neues Raschmodell mit den verbleibenden Items geschätzt, so lange bis ein Stop-Kriterium erreicht ist. Das Stopkriterium kann entweder die globale Messinvarianz der Skala (LR-Test) sein, oder auf dem Wald-Test basieren: es werden so lange Items ausgeschlossen, bis kein Item mehr einen signifikanten z-Wert aufweist. In dieser Studie wurde der Wald-Test als Stopkriterium gewählt. Sobald sich bei keinem Item mehr die Itemparameter signifikant zwischen den Gruppen unterscheiden, endet die Eliminierung von Items.

Abschließend werden im dritten Schritt die verbliebenen Items der reduzierten Skala erneut dem Andersen-LR-Test und der non-parametrischen Variante, der quasi-exakten Teststatistik  $T_{10}$ , als globale Tests der Subgruppeninvarianz der gesamten Subskala unterzogen. Damit wird sichergestellt, dass die resultierende reduzierte Subskala messinvariant für die Gruppen ist. Außerdem gibt die quasi-exakte Teststatistik  $T_{11}$  die lokale stochastische Unabhängigkeit der Subskala an, um die Eindimensionalität der reduzierten Skala zu überprüfen.

Der Vergleich von Itemparametern zwischen Gruppen kann auch graphisch anhand eines GOF-Plots verdeutlicht werden, bei dem die Itemparameter der einen Gruppe auf der x-Achse, die der anderen Gruppe auf der y-Achse abgetragen werden. Wenn die Konfidenz-Ellipsen die Diagonale einschließen, kann davon ausgegangen werden, dass die Itemschwierigkeitsparameter in beiden Gruppen ähnlich sind. Liegt ein Item oberhalb der Diagonalen, bedeutet dies, dass es für die LB-Gruppe systematisch schwieriger ist und vice versa.

### 7.2.2. Developmental Trajectory Re-Analyse

Die mittels des oben beschriebenen Ausschlussverfahrens identifizierten Items wurden aus ihren jeweiligen Subskalen entfernt, sodass nur diejenigen Items in die Berechnung eines neuen, bereinigten WWT-Gesamtscores einfließen, bei denen keine systematische Gruppenunterschiede feststellbar waren.

Analog zum Vorgehen in Studie 1 wurde ein Developmental Trajectory für Redintegration erstellt: die Leistung einer Person in einer Arbeitsgedächtnis-Spannenaufgabe (über beide Wortlängen-Bedingungen [kurze vs. lange Wörter] hinweg) wird im Rahmen eines multivariaten allgemeinen linearen Modells mit Messwiederholung anhand der Faktoren Lexikalität, Gruppenzugehörigkeit und dem Entwicklungsindikator vorhergesagt. Der Entwicklungsindikator war hier der reduzierte WWT-Score. Da die Redintegration-Effektivität anhand des Lexikalitätseffekts operationalisiert wird, erfolgt die Auswertung über die beiden Wortlängen-Bedingungen hinweg.

Für die einfachere Interpretation ist in der grafischen Darstellung (Abbildung 10) auf der y-Achse bereits der Lexikalitätseffekt als Differenz von echten Wörtern und Kunstwörtern abgetragen. Die Auswertungen des Developmental Trajectorys erfolgte analog zum Vorgehen in Studie 1 mit SPSS (IBM Corp., 2017; siehe auch Abschnitt 3.2.2; Thomas et al., 2009).

### 7.3. Ergebnisse

#### 7.3.1. Ausschluss von Items mit DIF

Aus den vollständigen Skalen wurden insgesamt  $k = 29$  Items ausgeschlossen, die anhand ihrer zwischen den Gruppen divergierenden Fit-Werte als DIF-behaftet identifiziert wurden. Die Skala mit den wenigsten ausgeschlossenen Items waren die Adjektive mit 5 Items, während bei den Nomen mit 12 Items die meisten Items entfernt wurden. Die deskriptiven Angaben finden sich in Tabelle 10.

Die Ergebnisse werden in den nächsten Abschnitten pro Subskala getrennt nach dem folgenden Schema berichtet: Zunächst erfolgt die Beschreibung des Raschmodells für die ungekürzte Skala; hier ist v.a. der Andersen-LR-Test zur globalen Bewertung der Messinvarianz über die beiden Gruppen von Interesse. Es folgt eine Auflistung derjenigen Items, die aufgrund ihrer unzureichenden Subgruppeninvarianz ausgeschlossen wurden. An dieser Stelle die Fit-Werte anzugeben, ist jedoch nicht sinnvoll, da in jedem Schritt das Raschmodell neu geschätzt wird, wodurch sich auch die Itemparameter mit jedem Schritt verändern. Daher werden nur die Itemstatistiken für die verbleibenden Items der resultierenden Subskalen berichtet. Der Vollständigkeit halber sind für die finalen reduzierten Subskalen die Schwierigkeitsparameter und deren Konfidenzintervallen pro Gruppe und die Ergebnissen des Waldtests in Tabelle A-5 im Anhang angefügt.

Pro Subskala werden zwei GOF-Plots gezeigt: der erste beinhaltet die gesamte Skala, der zweite nur die verbliebenen Items nach der Item-Reduktion. Auch hier zeigt sich das Problem, dass durch den iterativen Prozess des Itemausschlusses jedes Mal ein neues Raschmodell geschätzt wird, wodurch sich auch die Itemparameter ändern. So müsste im vollständigen Modell z. B. das Item „Wappen“ nicht ausgeschlossen werden, denn im GOF-Plot schließt die Konfidenzellipse die Diagonale mit ein, auch der Waldtest ( $\chi = 1.166; p = .244$ ) ist nicht signifikant. Dies ändert sich jedoch nach der Entfernung von insgesamt 11 Items davor, sodass im 12. Iterationsschritt „Wappen“ mit  $\chi = 1.991; p = .046$  signifikant wird und daher ebenfalls aus der Skala ausgeschlossen wird.

Tabelle 10

Deskriptive Angaben zur vollen und reduzierten Skala des Wortschatzumfangs in Studie 3

	LB (n = 93)		KG (n = 114)		Signifikanz
	M	SD	M	SD	
Gesamtskala (k = 95)	53.42 (56.3%)	16.79 (17.8%)	54.57 (57.4%)	14.11 (14.9%)	$t(179.89) = 0.526$ ; $p = .599$ ; $d = 0.07$
<u>Vollständige Subskalen</u>					
Nomen (k = 26)	14.26 (54.8%)	5.56 (21.4%)	13.75 (52.9%)	4.48 (17.2%)	$t(175.3) = 0.707$ ; $p = .481$ ; $d = 0.10$
Verben (k = 23)	13.96 (60.7%)	3.90 (17.9%)	14.66 (63.7%)	3.23 (14.0%)	$t(178.1) = 1.389$ ; $p = .167$ ; $d = 0.20$
Adjektive (k = 23)	12.91 (56.1%)	5.04 (21.9%)	13.33 (58.0%)	4.93 (21.4%)	$t(194.97) = 0.601$ ; $p = .548$ ; $d = 0.08$
Kategorien (k = 23)	12.29 (53.4%)	4.54 (19.7%)	12.82 (55.8%)	4.11 (17.9%)	$t(187.87) = 0.879$ ; $p = .381$ ; $d = 0.12$
reduzierte Gesamtskala (k = 66)	36.30 (55.0%)	11.75 (17.8%)	38.20 (57.9%)	9.39 (14.2%)	$t(172.13) = 1.258$ ; $p = .210$ ; $d = 0.18$
<u>Reduzierte Subskalen</u>					
Nomen (k = 14)	6.73 (48.1%)	3.24 (23.1%)	7.11 (50.8%)	2.41 (17.2%)	$t(164.06) = 0.950$ ; $p = .344$ ; $d = 0.14$
Verben (k = 17)	10.71 (63.0%)	3.03 (17.9%)	11.18 (65.8%)	2.25 (13.2%)	$t(165.78) = 1.253$ ; $p = .212$ ; $d = 0.18$
Adjektive (k = 18)	9.48 (52.7%)	3.99 (22.1%)	10.36 (57.6%)	4.10 (22.8%)	$t(198.77) = 1.553$ ; $p = .122$ ; $d = 0.22$
Kategorien (k = 17)	9.40 (55.3%)	3.43 (20.2%)	9.54 (56.1%)	2.95 (17.4%)	$t(182.61) = 0.324$ ; $p = .746$ ; $d = 0.05$

*Anmerkungen.* Summenwerte der Punkte (obere Zahl) sowie Prozentwerte der korrekt gelösten Items (in Klammern), zur besseren Vergleichbarkeit aufgrund unterschiedlicher Item-Anzahlen. Aufgrund ungleicher Varianzen werden korrigierte Freiheitsgrade in den  $t$ -Tests berichtet.

Für die vollständige Subskala der Nomen mit  $k = 26$  Items wurde der Andersen LR-Test signifikant,  $LR(25) = 118.471$ ;  $p < .001$ . Die log-Likelihood für die LB-Gruppe betrug  $L = -919.54$ , während bei der TD-Gruppe der Wert  $L = -1301.75$  betrug. Im Zuge der Itementfernung wurden die folgenden 12 Items in der angegebenen Reihenfolge ausgeschlossen: *Armreif, Geländer, Kriechke, Mantel, Container, Absatz, Ventilator, Kompass, Automat, Klinge, Federball, Wappenstein*. Es verblieben damit die folgenden 14 Items in der Subskala „Nomen“: *Schubkarre, Hocker, Pyramide, Ellenbogen, Satellit, Orchester, Knospe, Ferse, Schleier, Riegel, Fackel, Schnalle, Henkel, Gipfel*.



Für die verbleibende Subskala ergab sich weder im Andersen LR-Test,  $LR(13) = 17.152$ ;  $p = .192$ , noch in der quasi-exakten Teststatistik zur Subgruppeninvarianz ( $T_{10}$ ,  $p = .082$ ) ein signifikanter Unterschied zwischen den Gruppen. Die lokale stochastische Unabhängigkeit  $T_{11}$  zeigte jedoch an, dass auch bei den verbliebenen Items noch Items in ihrer Lösungswahrscheinlichkeit nicht unabhängig voneinander zu sein schienen ( $p = .032$ ). Das bedeutet, dass andere, nicht erfasste Faktoren bei der Lösung der Items möglicherweise eine Rolle spielen. Abbildung 6a zeigt den GOF-Plot für die vollständige Subskala Nomen, Abbildung 6b die verbliebenen Items des finalen Raschmodells der Subskala Nomen.

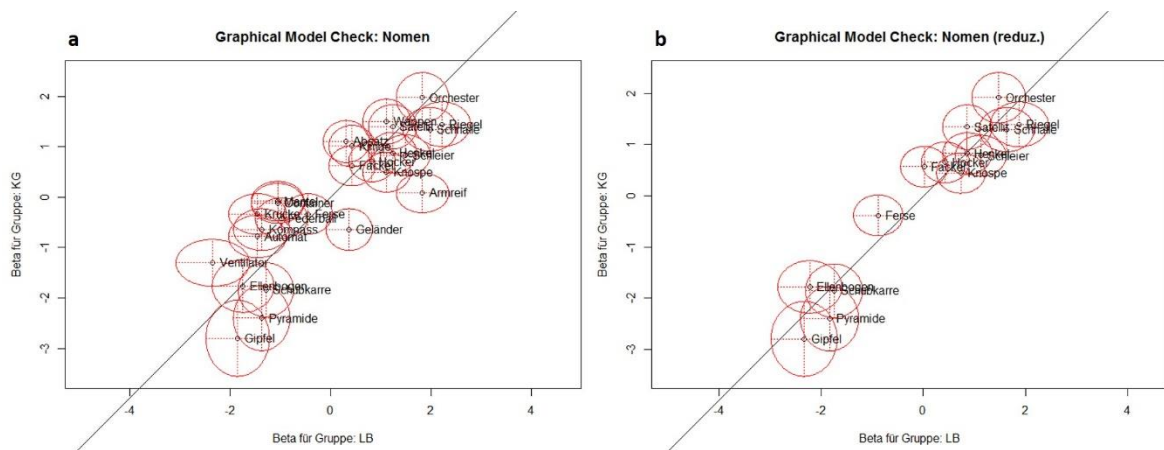


Abbildung 6. GOF-Plot für Subskala Nomen. a) Vollständige Subskala, b) reduzierte Subskala; Itemparameter für die LB-Gruppe auf der x-Achse, für die KG auf der y-Achse

Für die vollständige Subskala der Verben mit  $k = 23$  Items wurde der Andersen LR-Test signifikant,  $LR(22) = 69.32$ ;  $p < .001$ . Die log-Likelihood für die LB-Gruppe betrug  $L = -753.048$ , während bei der TD-Gruppe der Wert  $L = -1003.62$  betrug. Im Zuge der Itementfernung wurden die folgenden 6 Items in der angegebenen Reihenfolge ausgeschlossen: *fressen*, *wiegen*, *demonstrieren*, *warten*, *verbeugen*, *wehen*. Es verblieben damit die folgenden 17 Items in der

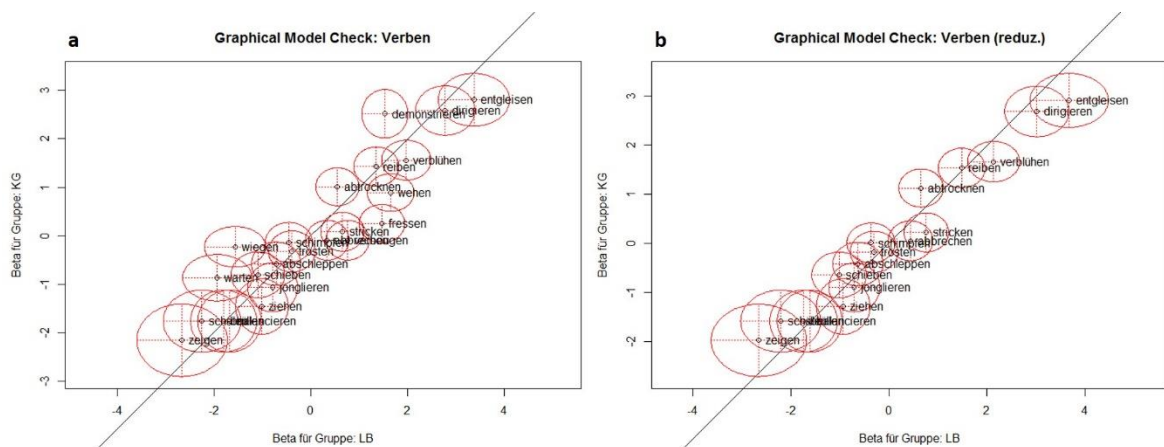


Abbildung 7. GOF-Plot für Subskala Verben. a) Vollständige Subskala, b) reduzierte Subskala; Itemparameter für die LB-Gruppe auf der x-Achse, für die KG auf der y-Achse

Subskala „Verben“: *zeigen, entgleisen, jonglieren, schälen, schieben, balancieren, abtrocknen, abbrechen, verblühen, dirigieren, trösten, reiben, ziehen, schimpfen, abschleppen, stricken, brüllen*. Für die verbleibende Subskala ergab sich weder im Andersen LR-Test,  $LR(16) = 17.424, p = .359$ , noch in der quasi-exakten Teststatistik zur Subgruppeninvarianz ( $T_{10}, p = .132$ ) ein signifikanter Unterschied zwischen den Gruppen. Bei den Verben (als einziger Subskala) schien auch die Bedingung der lokalen stochastischen Unabhängigkeit erfüllt zu sein ( $T_{11}, p = .165$ ). Abbildung 7a zeigt den GOF-Plot für die vollständige Subskala Verben, Abbildung 7b die verbliebenen Items des finalen Raschmodells der Subskala Verben.

Für die vollständige Subskala der Adjektive (Gegensätze) mit  $k = 23$  Items wurde der Andersen LR-Test signifikant,  $LR(22) = 57.057; p < .001$ . Die log-Likelihood für die LB-Gruppe betrug  $L = -763.621$ , während bei der TD-Gruppe der Wert  $L = -887.96$  betrug. Im Zuge der Itementfernung wurden die folgenden 5 Items in der angegebenen Reihenfolge ausgeschlossen: *spitz, glatt, hoch, hässlich, lieb*. Es verblieben damit die folgenden 18 Items in der Subskala „Adjektive“: *alt, früh, weich, wild, wolkenlos, einfach, einfarbig, mutig, vertraut, traurig, nah, langweilig, innen, sauer, unzufrieden, altmodisch, trocken, ungefährlich*. Für die verbleibende Subskala ergab sich weder im Andersen LR-Test,  $LR(17) = 10.657, p = .874$ , noch in der quasi-exakten Teststatistik zur Subgruppeninvarianz ( $T_{10}, p = .786$ ) ein signifikanter Unterschied zwischen den Gruppen. Die lokale stochastische Unabhängigkeit  $T_{11}$  zeigte jedoch an, dass auch bei den verbliebenen Items noch Items in ihrer Lösungswahrscheinlichkeit nicht unabhängig voneinander zu sein schienen ( $p < .001$ ). Das bedeutet, dass andere, nicht erfasste Faktoren bei der Lösung der Items möglicherweise eine Rolle spielen. Abbildung 8a zeigt den GOF-Plot für die vollständige Subskala Adjektive, Abbildung 8b die verbliebenen Items des finalen Raschmodells der Subskala Adjektive.

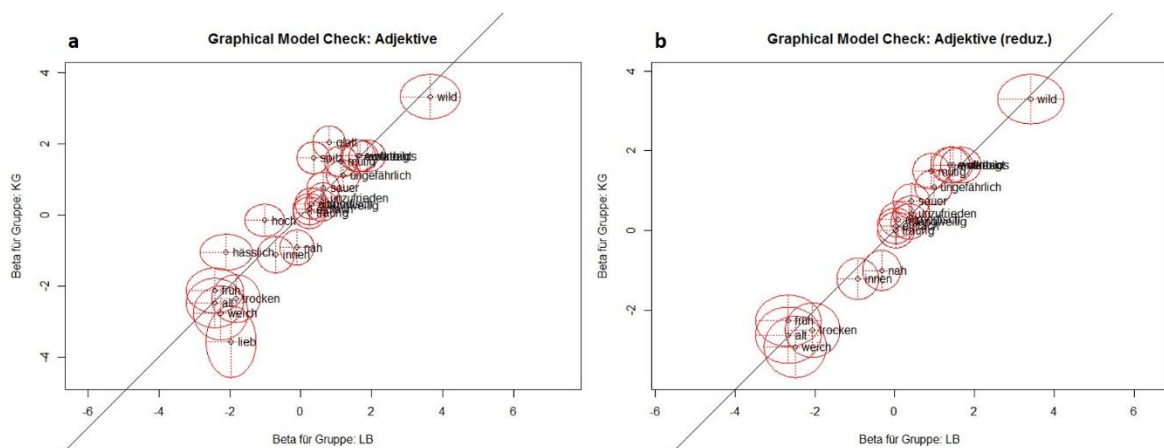


Abbildung 8. GOF-Plot für Subskala Adjektive. a) Vollständige Subskala, b) reduzierte Subskala; Itemparameter für die LB-Gruppe auf der x-Achse, für die KG auf der y-Achse

Schließlich wurde für die vollständige Subskala der Kategorien mit  $k = 23$  Items der Andersen LR-Test signifikant,  $LR(22) = 67.472$ ;  $p < .001$ . Die log-Likelihood für die LB-Gruppe betrug  $L = -806.57$ , während bei der TD-Gruppe der Wert  $L = -1125.29$  betrug. Im Zuge der Itementfernung wurden die folgenden 6 Items in der angegebenen Reihenfolge ausgeschlossen: *Jahreszeiten, Kosmetika, Gewürze, Gebäude, Pflanzen, Baufahrzeuge*. Es verblieben damit die folgenden 17 Items in der Subskala „Kategorien“: *Gemüse, Getreide, Gepäck, Instrumente, Kopfbedeckungen, Geschirr, Sportarten, Küchengeräte, Feste, Schmuck, Werkzeuge, Besteck, Lebensmittel, Insekten, Möbel, Zeichen, Berufe*. Für die verbleibende Subskala ergab sich weder im Andersen LR-Test,  $LR(16) = 14.872$ ,  $p = .534$ , noch in der quasi-exakten Teststatistik zur Subgruppeninvarianz ( $T_{10}$ ,  $p = .522$ ) ein signifikanter Unterschied zwischen den Gruppen. Die lokale stochastische Unabhängigkeit  $T_{11}$  zeigte jedoch an, dass auch bei den verbliebenen Items noch Items in ihrer Lösungswahrscheinlichkeit nicht unabhängig voneinander zu sein schienen ( $p < .001$ ). Das bedeutet, dass andere, nicht erfasste Faktoren bei der Lösung der Items möglicherweise eine Rolle spielen. Abbildung 9a zeigt den GOF-Plot für die vollständige Subskala Kategorien, Abbildung 9b die verbliebenen Items des finalen Raschmodells der Subskala Kategorien.

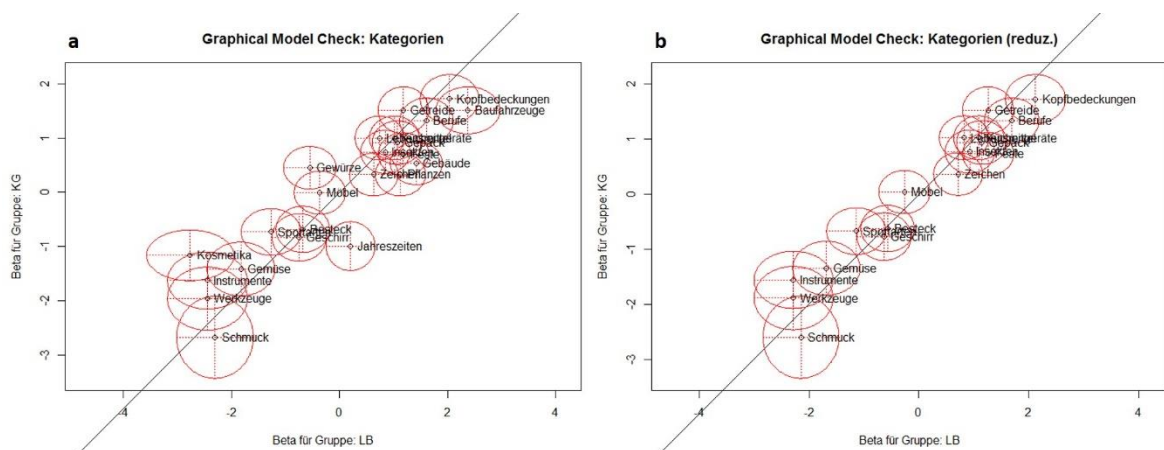


Abbildung 9. GOF-Plot für Subskala Kategorien. a) Vollständige Subskala, b) reduzierte Subskala; Itemparameter für die LB-Gruppe auf der x-Achse, für die KG auf der y-Achse

Die erste Fragestellung kann folgendermaßen beantwortet werden: Insgesamt wurden  $k = 29$  Items wegen DIF ausgeschlossen (mit einem Range von 5-12 Items je Subskala). Daraus ergab sich ein reduzierter und damit um DIF bereinigter WWT-Gesamtscore ( $WWT_{red}$ ) aus insgesamt  $k_{red} = 66$  Items. Die deskriptiven Statistiken für die reduzierte Skala ( $WWT_{red}$ ) im direkten Vergleich zur vollständigen Skala ( $WWT_{ges}$ ) sind in Tabelle 10 (Seite 153) nach Gruppen und Subskalen getrennt dargestellt. Die Gruppen (LB und TD) unterschieden sich auch im bereinigten Score auf Ebene der Mittelwerte nicht signifikant voneinander ( $p = .210$ ); ebenso gab es in keiner der Subskalen signifikante Unterschiede zwischen den Gruppen (alle  $p > .122$ ). Die

Korrelation zwischen der vollständigen und der reduzierten Skala erwies sich mit  $r = .982$  ( $p < .001$ ) als äußerst hoch.

### 7.3.2. Re-Analyse des Redintegration-Effekts

In der Developmental Trajectory Reanalyse mit dem reduzierten WWT-Gesamtscore als Entwicklungsindikator zur Vorhersage der Redintegration-Effektivität zeigte sich ein signifikanter Unterschied im Intercept (am niedrigsten WWT<sub>red</sub>-Score der LB-Gruppe). Dies wurde anhand der Interaktion von Lexikalität  $\times$  Gruppe,  $F(1, 180) = 5.638$ ;  $p = .019$ ;  $\eta_p^2 = .03$ , gemessen. Wie im Developmental Trajectory mit dem vollständigen Wortschatz-Score (Studie 1) fiel der Intercept-Effekt zugunsten der LB-Gruppe aus, die bei einem geringeren Wortschatzumfang eine höhere Redintegration-Effektivität aufwies. Über beide Gruppen hinweg zeigte sich (im Durchschnitt) kein Zusammenhang zwischen dem Wortschatzumfang und der Redintegration-Effektivität (Lexikalität  $\times$  WWT<sub>red</sub>  $p = .768$ ). Dies wird jedoch qualifiziert durch die signifikante Interaktion mit dem Gruppenfaktor (Gruppe  $\times$  Lexikalität  $\times$  WWT<sub>red</sub>), was zeigt, dass sich die Gruppen weiterhin in den Steigungskoeffizienten der Regressionsgeraden unterscheiden,  $F(1, 180) = 6.967$ ;  $p = .009$ ;  $\eta_p^2 = .039$ .

Damit kann die zweite Fragestellung dahingehend beantwortet werden, dass die in Studie 1 gefundene Interaktion zwischen Redintegration-Effektivität und Wortschatzumfang bei Kindern mit Lernbehinderung erhalten bleibt, auch wenn DIF als mögliche Ursache ausgeschlossen wird. Es scheint sich also um ein differenzielles Entwicklungsmuster im Zusammenhang mit dem Wortschatzumfang zu handeln.

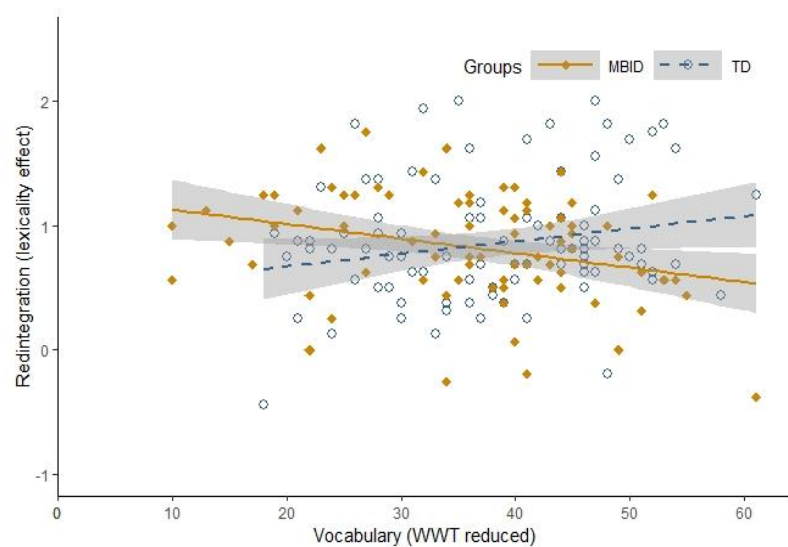


Abbildung 10. Developmental Trajectory Reanalyse für die Effektivität des Redintegrationprozesses im Verhältnis zum Wortschatzumfang (bereinigt) für Schüler\*innen mit Lernbehinderung (MBID, orange Rauten) und unbeeinträchtigte Schüler\*innen (TD, blaue Kreise).

## 7.4. Diskussion

### 7.4.1. Zusammenfassung und Interpretation der Ergebnisse

In dieser Studie wurden mittels DIF-Analysen im IRT-Framework einzelne Items identifiziert, die für eine unzureichende Messinvarianz zwischen den Gruppen im WWT 6-10 verantwortlich waren. Diese itembasierte Analyse war über die vorangegangenen Analysen mit Item-Parceling (Studie 2) nicht möglich gewesen. Mithilfe eines iterativen Prozesses wurden getrennt nach Subskala insgesamt  $k = 29$  Items ausgeschlossen. Das entspricht knapp einem Drittel aller Items des Instrumentes, die für Schüler\*innen mit Lernbehinderung systematisch zu schwierig oder zu leicht zu beantworten waren.

Die Bereinigung der WWT-Skala geschah vorrangig zu dem Zweck, eine in Studie 1 (Kapitel 5.4.3) gefundene Interaktion mit der Redintegration-Effektivität abzusichern und damit auszuschließen, dass dieser Effekt auch durch DIF zu erklären ist. Es zeigte sich, dass die Interaktion erhalten blieb, auch wenn der reduzierte WWT-Score als Entwicklungsindikator eingesetzt wurde. Dies kann als ein starker empirischer Hinweis interpretiert werden, dass DIF als Ursache für besagte Interaktion ausgeschlossen werden kann.

Da somit eine methodische Ursache als Erklärung ausgeschlossen werden kann, sollten nun wiederum inhaltliche Interpretationen stärker in den Fokus rücken, die die Befunde als differenzielles Entwicklungsmuster bei Kindern mit Lernbehinderung deuten. Wenn Redintegration als Prozess zur Rekonstruktion von bereits teilweise zerfallenen Gedächtnisspuren im Arbeitsgedächtnis anhand des im Langzeitgedächtnis verfügbaren Wissens betrachtet wird (Schweickert, 1993; Roodenrys & Miller, 2008), dann wäre zu erwarten, dass mit einem größeren Wortschatzumfang auch eine höhere Effektivität des Redintegrationprozesses einhergeht. Denn je mehr Wörter im mentalen Lexikon verfügbar sind, desto eher sollte dieses für die Rekonstruktion genutzt werden können.

Während schon in der Kontrollgruppe dieser Effekt nicht eintritt, sich also weder ein positiver noch ein negativer Zusammenhang beobachten lässt, ist bei Schüler\*innen mit Lernbehinderung ein stabiler negativer Zusammenhang feststellbar: Je größer der Wortschatzumfang von Kindern mit Lernbehinderung ist, desto geringer fällt die Effektivität ihres Redintegrationprozesses im Durchschnitt aus. Dies stellt eine Abweichung vom aufgrund des mentalen Alters zu erwartenden Muster dar, was in der Argumentation von Thomas et al. (2009, S. 353) eine Form von „Atypikalität“ darstellt, die über einen *delay* hinausgeht. So ergibt sich durch diesen Befund der Hinweis auf eine strukturell abweichende Nutzung von Langzeitgedächtnis-Inhalten bei der Rekonstruktion von Arbeitsgedächtnis-Memoranda, da bei Kindern mit Lernbehinderung ein größerer Wortschatzumfang offenbar einen abträglichen Effekt auf die Redintegration hat.

Vermutungen über mögliche Ursachen wurden im Rahmen von Studie 1 (Kapitel 5.5) bereits angestellt. Da es bei Studie 3 lediglich um eine methodische Absicherung ging, werden diese hier nicht weiter vertieft. Im Rahmen der Gesamtdiskussion wird dann detaillierter aufgegriffen, inwiefern die Ergebnisse vor dem Hintergrund zweier möglicher Ansätze erklärt werden können: (1) Könnten die Probleme bei der Nutzung von Vorwissen für Redintegration auf Ebene der Informationsverarbeitung liegen; (2) ließen sie sich möglicherweise auch durch Probleme auf Ebene der Strategienutzung erklären.

#### 7.4.2. Limitationen

Es ist angezeigt, auf vier Limitationen dieser Studie hinzuweisen. Erstens erfolgte der Ausschluss von Items nicht theoriebasiert oder aus inhaltlichen Gründen, sondern nur auf Basis der Fit-Werte. Insbesondere aus der Subskala der Nomen wurden viele Items ausgeschlossen, deren Itemparameter im mittleren Schwierigkeitsbereich lagen. Dies ist insofern ungünstig, als sich viele Personenparameter ebenfalls in einem mittleren Bereich befanden, wodurch die Messung für diese Personen weniger differenziert war.

Zweitens schienen die Subskalen, bis auf „Verben“, nicht vollständig eindimensional zu sein, da die Annahme der lokalen stochastischen Unabhängigkeit (gemessen an der quasi-exakten non-parametrischen Teststatistik  $T_{11}$ ) verletzt wurde. Dies deutet darauf hin, dass weitere Faktoren einen Einfluss auf die Lösungswahrscheinlichkeit haben könnten. Hier stehen weitere Untersuchungen aus, jedoch können einige Variablen (Alter, CFT, AG-Kapazität, Frequency, NBS) bereits begründet ausgeschlossen werden, da sie im Rahmen des GLMM in Studie 2 keine signifikanten Interaktionen hervorriefen.

Drittens erwies sich die Korrelation zwischen dem vollständigen WWT-Score und dem reduzierten Score als äußerst hoch. Dass im Developmental Trajectory die Interaktion zwischen Redintegration und dem reduzierten WWT-Score ebenfalls signifikant wurde, ist daher nicht überraschend.

Viertens wäre sicherlich auch die Messinvarianz der Arbeitsgedächtnis-Spannenaufgaben zu überprüfen, zumal gerade für Wortspannen mit Kunstwörtern aufgrund von Bodeneffekten auch Reliabilitätsprobleme auftreten können. Eine solche Analyse wäre aber verglichen mit der WWT-Analyse ungleich komplexer, denn es muss berücksichtigt werden, dass wegen des adaptiven Vorgehens nicht alle Personen die identischen Spannen-Items bearbeiteten. Somit wäre die Verteilung der bearbeiteten Items sehr unbalanciert, da nur wenige Personen die längeren Items bearbeiteten. Dies methodisch angemessen abzubilden, wäre über den Rahmen des short-Papers hinausgegangen.

Wenn ein grundsätzliches, übergreifendes Problem der Messqualität für die Arbeitsgedächtnis-Aufgaben vorgelegen hätte, so wäre zu erwarten, dass sich dies ebenfalls in den Interaktionen mit den beiden anderen Entwicklungsindikatoren (chronologisches Alter und CFT-Rohwert) zeigte. Die Interaktion war aber ausschließlich bei der spezifischen Kombination von Redintegrationprozess und Wortschatzumfang zu beobachten, was dafür spricht, dass es sich hier tatsächlich um ein differenzielles Entwicklungsmuster handelt.

### 7.4.3. Fazit Studie 3

Mit dieser Follow-up-Studie konnte gezeigt werden, dass der differenzielle Zusammenhang zwischen Redintegration und Wortschatzumfang, der in Studie 1 für Kinder mit Lernbehinderung festgestellt worden war, nicht auf ein methodisches Artefakt zurückzuführen ist. Der vollständige Gesamtscore des WWT 6-10 beinhaltete zwar Items mit DIF, aber nach deren Ausschluss blieb die Interaktion zwischen Redintegration und dem Wortschatzumfang erhalten.

Insofern sind nunmehr inhaltliche Interpretationen angemessen, die den Befund als differenzielles Entwicklungsmuster deuten: Bei Kindern mit Lernbehinderung fällt die Effektivität des Redintegrationprozesses bei einem größeren Wortschatz signifikant schwächer aus. Bemerkenswert ist, dass diese Interaktion ausschließlich und spezifisch bei der Kombination von Redintegration und Wortschatzumfang auftrat. In keinem der beiden anderen Entwicklungsindikatoren (chronologisches Alter, CFT) gab es einen (differenziellen) Zusammenhang mit Redintegration, und der Wortschatzumfang sorgte bei der AG-Kapazität und dem Rehearsalprozess ebenfalls nicht für differenzielle Muster. Schüler\*innen mit Lernbehinderung scheinen also spezifisch bei der Rekonstruktion von Arbeitsgedächtnis-Memoranda aus dem Langzeitgedächtnis nicht von einem größeren Wortschatz zu profitieren. Der Frage, ob eine Processing- oder eine Strategie-Hypothese besser geeignet, ist diese Ergebnisse zu erklären, wird im Rahmen der nun folgenden Gesamtdiskussion (Kapitel 8) vertieft nachgegangen.

## 8. Gesamtdiskussion

### 8.1. Zusammenfassung und Interpretation der Ergebnisse

Die in dieser Arbeit aufgeworfenen Fragestellungen zur Verfügbarkeit und Funktionalität von Arbeits- und Langzeitgedächtnis bei Kindern mit Lernbehinderung wurden in drei Teilstudien untersucht. Die erste Studie betrachtete die Kapazität des Arbeitsgedächtnisses und zwei Aufrechterhaltungs-Prozesse; die zweite Studie zum Langzeitgedächtnis ging auf die Binnen- und Außenstruktur des mentalen Lexikons ein; schließlich sicherte die dritte Studie einen Interaktionseffekt aus Studie 1 gegen ein methodisches Artefakt ab.

In der nun folgenden Gesamtdiskussion der Arbeit werden zunächst die Ergebnisse der Studien zusammengefasst und zueinander in Beziehung gesetzt (Kapitel 8.1). Anschließend werden theoretische und methodische Limitationen der Arbeit aufgegriffen und die Aussagekraft der Arbeit dementsprechend eingeschränkt und Desiderate formuliert (Kapitel 8.2). Um speziell auf die Limitation einzugehen, dass die Netzwerkstruktur des Langzeitgedächtnisses nicht ausreichend berücksichtigt werden kann, und eine methodische Möglichkeit aufzuzeigen, diese in zukünftigen Arbeiten abzubilden, wird in Kapitel 8.3 als Ausblick eine tentative Netzwerkanalyse berichtet. Sodann wird in Kapitel 8.4 auf die theoretischen und praktischen Implikationen eingegangen, wobei besonderes Augenmerk auf den differenziellen Befund zum Redintegrationprozess und dessen Erklärung durch die Strategie-Hypothese gelegt wird. Abschließend fasst die Schlussbetrachtung (Kapitel 8.5) die zentralen Erkenntnisse der Arbeit zusammen.

#### 8.1.1. Studie 1 – Arbeitsgedächtnis

In Studie 1 wurde das Arbeitsgedächtnis untersucht. Die dazu verwendeten Wortspannungsaufgaben lassen Rückschlüsse auf die Gesamtkapazität der phonologischen Schleife und die Funktionsweise von zwei Prozessen zur Aufrechterhaltung der Gedächtnisspur zu (Hasselhorn et al., 2012): (1) Der phonologisch basierte Rehearsalprozess lässt sich als inneres Nachsprechen innerhalb der phonologischen Schleife verstehen und zeigt sich empirisch im Wortlängeneffekt, d. h., kurze Wörter werden besser behalten als lange Wörter. (2) Der lexikalisch basierte Redintegrationprozess, als Rekonstruktion der Gedächtnisspur aus dem Langzeitgedächtnis vorstellbar (Schweickert, 1993), lässt sich am Lexikalitätseffekt messen, d. h., echte Wörter können besser als Kunstwörter behalten werden. Während eine Reihe von Studien bereits die Kapazität der phonologischen Schleife und den Rehearsalprozess untersucht hat, wurde der Redintegrationprozess in der vorliegenden Arbeit zum ersten Mal bei Kindern mit Lernbehinderung erfasst.

Die erste Fragestellung dieser Arbeit (Seite 85) kann anhand der Ergebnisse folgendermaßen beantwortet werden: Als Entwicklungsmuster im verbalen Arbeitsgedächtnis bei Kindern mit Lernbehinderung lässt sich für AG-Kapazität, Rehearsal- und Redintegrationprozess vorrangig



ein verspäteter Entwicklungseinsatz nachweisen („delayed onset“, Thomas et al., 2009). Das bedeutet, dass die Kapazität der phonologischen Schleife und die Effektivität der beiden Prozesse zu Beginn der Entwicklung niedriger ausgeprägt sind als bei unbeeinträchtigten Kindern, dann aber mit zunehmendem Alter in einer ähnlichen Rate ansteigen. Dies gilt nicht nur für das Verhältnis zum chronologischen Alter, sondern auch, wenn das Verhältnis zum mentalen Alter (gemessen an kognitiver Kapazität und Wortschatzumfang) betrachtet wird.

Die Befunde dieser Arbeit zum Arbeitsgedächtnis sprechen insgesamt für das Muster eines verzögerten Entwicklungseinsatzes. Damit lassen sie sich in den Kontext bisheriger Forschungen zum Arbeitsgedächtnis einreihen, die ebenfalls eine Entwicklungsverzögerung feststellen (siehe Kapitel 2.1.3; Danielsson et al., 2012; Hasselhorn & Mähler, 2010; Henry & MacLean, 2002; Mähler, 2007; Mähler & Hasselhorn, 2003; Poloczek et al., 2016; van der Molen et al., 2009). Eine Einordnung der Ergebnisse wurde bereits in Kapitel 5.5 vorgenommen und wird daher hier nicht weiter ausgeführt. Für ein strukturell abweichendes Entwicklungsmuster, das auch in einigen Studien nachgewiesen wurde (Hasselhorn & Mähler, 2007; Schuchardt et al., 2011), werden in der vorliegenden Arbeit bis auf eine Ausnahme fast keine Hinweise gefunden.

Für den Redintegrationprozess ergibt sich allerdings ein spezifisches differenzielles Muster, das nähere Betrachtung verdient. Zunächst können im Verhältnis zu chronologischem Alter und kognitiver Kapazität keinerlei Unterschiede zwischen Kindern mit und ohne Lernbehinderung im Redintegrationprozess festgestellt werden. Im Verhältnis zum Wortschatzumfang zeigt sich jedoch eine spezifische Interaktion: Bei Schüler\*innen mit Lernbehinderung geht ein größerer Wortschatzumfang mit einer geringeren Effektivität des Redintegrationprozesses einher. Das bedeutet, der Rückgriff auf Informationen aus dem Langzeitgedächtnis zur Rekonstruktion von Gedächtnisspuren im Arbeitsgedächtnis fällt Kindern mit Lernbehinderung umso schwerer, je größer ihr Wortschatzumfang ist. Da dies bei unbeeinträchtigten Kindern nicht der Fall ist, scheint hier ein atypisches Muster vorzuliegen.

An dieser Stelle ist die methodische Stärke des eingesetzten Developmental Trajectory Ansatzes hervorzuheben, denn der klassische Gruppenmatching-Ansatz wäre nicht in der Lage gewesen, den Effekt zu entdecken. Nur durch den Einsatz verschiedener MA-Variablen (hier kognitive Kapazität und Wortschatzumfang) konnte dieser differenzielle Zusammenhang aufgedeckt werden.

Der Interaktions-Befund macht eine nähere Untersuchung notwendig. Zum einen stellt sich die Frage, inwiefern das Langzeitgedächtnis bei Kindern mit Lernbehinderung strukturell andersartig sein könnte, was den Effekt zumindest teilweise inhaltlich erklären könnte. Für die Beantwortung dieser Frage werden die Ergebnisse von Studie 2 herangezogen (siehe nächster Abschnitt). Zum anderen muss ausgeschlossen werden können, dass es sich lediglich um ein

methodisch bedingtes Artefakt handelt, das durch Differential Item Functioning im Wortschatz-Testverfahren verursacht sein könnte; diese Frage überprüft Studie 3. Eine theoretische Einordnung und Bewertung der Strategie-Hypothese nach Klauer und Lauth (1997) hinsichtlich ihrer Erklärungskraft erfolgt in Kapitel 8.4.

### 8.1.2. Studie 2 – Langzeitgedächtnis

Zum semantischen Langzeitgedächtnis und dessen Struktur bei Schüler\*innen mit Lernbehinderung liegen keine aktuellen empirischen Studien vor. Daher wird im Rahmen von Studie 2 das semantische Gedächtnis, eingegrenzt auf den Abruf aus dem mentalen Lexikon, anhand eines expressiven Wortschatztests erfasst (WWT 6-10, Glück, 2011). Da der Struktur und dem Vernetzungsgrad für die Organisation des semantischen Gedächtnisses eine bedeutsame Rolle zugesprochen wird (Renkl, 2015; Vernooij, 2007), verfolgt diese Arbeit zwei Ansätze, mit denen die Binnen- und die Außenstruktur abgebildet werden können. Ein Vergleich der Binnenstruktur zwischen den Gruppen erfolgt anhand der Überprüfung der Messinvarianz mithilfe einer Multigroup-CFA, bezogen auf die Wortart als strukturierendes Merkmal. Hier lässt sich für die Binnenstruktur volle metrische und der partielle residuale Messinvarianz zwischen den Gruppen feststellen. Für die Außenstruktur wird gefragt, ob in einem gemischten linearen Modell differenzielle Zusammenhänge zwischen personen- und itembezogenen Variablen auftreten. Hier treten keine signifikanten Interaktionen zwischen dem Gruppenfaktor und den personen- und itembezogenen Variablen auf.

Die zweite Fragestellung der Arbeit (Kapitel 6.1) lässt sich also folgendermaßen beantworten: Weder in der Binnenstruktur noch in der Außenstruktur lassen sich Unterschiede zwischen Schüler\*innen mit Lernbehinderung und der unbeeinträchtigten jüngeren Kontrollgruppe (nach mentalem Alter vergleichbar) feststellen. Die gefundene Messinvarianz kann als empirischer Hinweis darauf gedeutet werden, dass die Gruppen eine sehr ähnliche Binnenstruktur aufweisen hinsichtlich des Verhältnisses der Einträge des mentalen Lexikons untereinander. Die ausbleibenden Gruppen-Interaktionen sind ein entsprechender Hinweis darauf, dass in keinem der geprüften Bereiche (kognitive Kapazität, Arbeitsgedächtnis-Kapazität, Eigenschaften der Items) ein differenzielles Zusammenhangs- und Entwicklungsmuster vorliegt („developmental relation“, Thomas et al., 2009, S. 339). Die Zusammenhänge mit anderen Domänen und die Verarbeitung von sprachlichen Eigenschaften scheinen bei Kindern mit Lernbehinderung daher weitgehend analog zu verlaufen wie bei unbeeinträchtigten Kindern des gleichen mentalen Alters. Man kann also davon ausgehen, dass hinsichtlich der Struktur des semantischen Gedächtnisses allenfalls eine Entwicklungsverzögerung vorliegt, wohingegen ein strukturelles Defizit ausgeschlossen werden kann.

Neben der Interpretation des Ergebnisses für die Langzeitgedächtnisstruktur als eigenständige kognitive Lernvoraussetzung können diese Befunde auch dazu beitragen, die in Studie 1 gefundene Interaktion zwischen Redintegrationprozess und Wortschatzumfang zu interpretieren. Da die Struktur des Langzeitgedächtnisses offenbar nicht spezifisch beeinträchtigt ist, scheidet sie als mögliche inhaltliche Erklärung für das Auftreten besagter Interaktion aus.

### **8.1.3. Studie 3 – Absicherung der Redintegration-Interaktion**

Schließlich dient die als Folgestudie zu Studie 1 konzipierte Studie 3 dazu, eine methodische Ursache für das Auftreten der Interaktion zwischen Redintegrationprozess und Wortschatzumfang auszuschließen. Als eine solche Ursache kommt die Möglichkeit in Betracht, dass Differential Item Functioning die Erfassung des Wortschatzes beeinflusst. Das bedeutet, dass manche Items für die Gruppen systematisch unterschiedlich schwierig zu beantworten sein könnten. Die Erklärung, dass es sich bei der in Studie 1 gefundenen Interaktion um ein durch DIF bedingtes Artefakt handelt, kann aber ausgeschlossen werden, wenn man WWT-Items mit DIF mittels einer IRT-Analyse identifiziert und darauf basierend einen bereinigten WWT-Score berechnet. Dies geschieht im Rahmen von Studie 3.

Bezogen auf die dritte Fragestellung können durch die IRT-Analyse insgesamt 29 Items mit DIF identifiziert und aus der Skala für den Wortschatzumfang ausgeschlossen werden. Anschließend wird dieser bereinigte Wortschatzumfang anstelle des WWT-Gesamtscores in die Developmental Trajectory Analyse, analog zu Studie 1, eingesetzt, um den Zusammenhang mit der Effektivität des Redintegrationprozesses zwischen den Gruppen zu vergleichen. Hierbei zeigt sich, dass die Interaktion weiterhin signifikant ist. Das differenzielle Entwicklungsmuster bei Schüler\*innen mit Lernbehinderung hinsichtlich der Nutzung des Langzeitgedächtnisses für die Rekonstruktion von Arbeitsgedächtnis-Memoranda bleibt also auch bei Kontrolle von DIF bestehen.

Hinsichtlich des Wortschatzumfangs kann also DIF als Ursache für die Interaktion ausgeschlossen werden. Auch bei den Arbeitsgedächtnisaufgaben wäre eine Überprüfung von DIF nötig; dies ist jedoch aufgrund der adaptiven Items methodisch deutlich aufwändiger, vor allem aber inhaltlich nicht plausibel, da die Interaktion spezifisch bei der Kombination von Wortschatzumfang und Redintegrationprozess auftritt. Daher rückt nun eine inhaltliche Erklärung der Interaktion in den Vordergrund, die im Rahmen der theoretischen Implikationen (Abschnitt 8.4) diskutiert werden soll; insbesondere wird dabei auf die Erklärungskraft durch die Strategie-Hypothese (Klauer & Lauth, 1997) eingegangen.

## 8.2. Limitationen

Zunächst sind mindestens vier Aspekte als Limitationen kritisch anzumerken: (1) Lassen sich keine Aussagen über Ursachen tätigen; (2) wird keine Abgrenzung von Lernbehinderung zu anderen Formen von Lernbeeinträchtigungen und keine Identifikation von Subgruppen vorgenommen. Die weiteren beiden Limitationen beziehen sich auf (3) Desiderata bei der Erfassung des Arbeitsgedächtnisses und (4) Einschränkungen bei der Abbildung des semantischen Langzeitgedächtnisses.

Bezogen auf die erste Limitation bleibt offen, ob nachweisbare Defizite in einzelnen Bereichen den Ausgangspunkt oder die Folge einer Lernbehinderung darstellen (Büttner & Hasselhorn, 2007; Schröder, 2000). Die Daten der vorliegenden Arbeit sind, wie bereits erwähnt, nicht längsschnittlich erhoben und können daher eine „echte“ Entwicklung nicht abbilden; insbesondere müssten auch Daten zu einem Zeitpunkt vor Feststellung einer Lernbehinderung vorliegen, um die Ergebnisse abzusichern und kausale Aussagen treffen zu können (Thomas et al., 2009). Es handelt sich bei den berichteten Befunden um (differenzielle) Zusammenhänge mit Indikatoren, die als Maß für den Stand der mentalen Entwicklung herangezogen wurden. Wie bereits zu Beginn der Arbeit angemerkt, wird mit der rein deskriptiven Beschreibung auch kein (monokausales) Erklärungsmodell oder eine Ableitung von Präskriptionen zur Intervention angestrebt.

Für die Ableitung eines Erklärungsmodelles reichen die der Untersuchung zugrundeliegenden Daten also nicht aus. Denn die Feststellung einer Entwicklungsverzögerung bzw. eines verspäteten Entwicklungseinsatzes ermöglicht zwar eine präzisere Beschreibung und Einordnung des Entwicklungsmusters, die Ursachen für diese verzögerte Entwicklung bleiben aber weiterhin offen. Wohl aber liefern die Ergebnisse deskriptive Hinweise darauf, ob bestimmte Strukturen und Prozesse von Arbeits- und Langzeitgedächtnis zur Verfügung stehen. Ebenfalls muss offen bleiben, welche Entwicklungsmuster bei älteren Personen mit Lernbehinderung auftreten (Bruns & Grosche, 2017, Mai; Mähler & Hasselhorn, 2003), weil die Befunde nicht über die erhobene Altersspanne hinaus extrapoliert werden können.

Bezogen auf die zweite Limitation wurden in der Stichprobe dieser Arbeit nur Kinder mit Lernbehinderung und unbeeinträchtigte Kinder des gleichen mentalen (und in Studie 1 chronologischen Alters) verglichen. Daher können keine Aussagen darüber gemacht werden, ob die gefundenen Entwicklungsmuster bei Arbeits- und Langzeitgedächtnis spezifisch für Lernbehinderung, oder nicht auch bei anderen Beeinträchtigungen wie z. B. spezifischer Lernstörung, Lernschwäche oder auch Sprachentwicklungsstörungen zu beobachten sind (Alt, 2011; Brandenburg et al., 2013; 2015; Hasselhorn et al., 2010). Zukünftige Untersuchungen sollten also auch die Population der Schüler\*innen mit Lernbehinderung mit anderen Formen von Lern-

und Sprachbeeinträchtigungen kontrastieren, um zum einen festzustellen, inwiefern die Befunde spezifisch für das Phänomen Lernbehinderung gelten, oder ob sie allgemeiner bei Problemen von Aneignungsprozessen auftreten, und damit zum anderen spezifische oder übergreifende Charakteristika herauszuarbeiten.

Neben diesem angesprochenen Verhältnis nach außen ist auch das Verhältnis der Gruppe der Schüler\*innen mit Lernbehinderung nach innen zu betrachten: Auch wenn in der vorliegenden Arbeit Schüler\*innen mit Lernbehinderung als eine Gruppe betrachtet wurden, ist deutlich darauf hinzuweisen, dass innerhalb der Gruppe mit einer großen Heterogenität zu rechnen ist (Heimlich, 2009). Eine Lernbehinderung kann vielerlei Ursachen haben, sich aber am Ende auf ähnliche Art und Weise äußern („Äquifinalität“; Piquart & Silbereisen, 2007). Dies würde für die vorliegende Arbeit bedeuten, dass mögliche Beeinträchtigungen in den untersuchten kognitiven Domänen nicht für alle Kinder mit Lernbehinderung gleichermaßen ein Problem darstellen, oder dass sie sich unterschiedlich auswirken könnten. Dadurch könnten Effekte verdeckt bleiben oder Befunde nicht für alle Subgruppen gleichermaßen generalisierbar sein. Es existiert jedoch keine Annahme zu einer solchen Aufteilung in Subgruppen, die sowohl theoretisch begründbar wie auch empirisch eindeutig operationalisierbar wäre. Daher bleibt der Anspruch der Arbeit bei der Beschreibung der in Kapitel 1.1 definierten, breiter gefassten Gruppe in eben ihrer Heterogenität, wie sie in der Schulwirklichkeit angetroffen wird. Perspektivisch könnte anhand der erhobenen Daten z. B. mittels einer Clusteranalyse eine empirische Unterteilung in Subgruppen erfolgen, was jedoch über den Rahmen der vorliegenden Arbeit hinausgeht.

Bezogen auf die dritte Limitation lässt sich mit den Daten aus Studie 1 nicht genauer lokalisieren, wo Schüler\*innen mit Lernbehinderung Schwierigkeiten haben; außerdem wirkt sich die Neighborhood-Size konfundierend auf den Wortlängeneffekt aus. Für eine präzisere Untersuchung des Arbeitsgedächtnisses in zukünftigen Studien ergeben sich daher mindestens zwei Punkte: zum einen erscheint es sinnvoll, zwischen Item- und Order-Information differenzieren zu können (siehe Kapitel 2.1.2), ob also ein Element einer Sequenz vollständig nicht mehr erinnert, oder nur die Reihenfolge nicht zuverlässig rekonstruiert werden konnte. Hierzu ist die Erfassung nicht nur auf Ebene der gesamten Sequenz, sondern auf Ebene einzelner Sequenzelemente nötig. Zum anderen ist in zukünftigen Studien bei der Auswahl der Stimuli zur Konstruktion der Sequenzen zu beachten, dass Wortlänge häufig mit Neighborhood-Size konfundiert ist und sich daher in bisherigen Untersuchungen der Wortlängeneffekt nicht eindeutig vom Neighborhood-Size-Effekt trennen lässt (Derragh et al., 2017; Jalbert, Neath, Bireta et al., 2011).

Für Aussagen über den Umgang mit Distraktoren im Arbeitsgedächtnis wäre es nötig, detailliertere Daten zu nutzen als sie in den hier berichteten Studien vorlagen. So müssten die

Antworten auf Item-Ebene dokumentiert werden um festzustellen, ob mehr Intrusionen stattfinden (Oberauer, Awh & Sutterer, 2017). Aber auch neuartige Untersuchungsdesigns wären denkbar, die die Zusammensetzung der Stimuli variieren (Giofrè, Carretti & Belacchi, 2016), oder Distraktoren systematisch variieren und über Eyetracking deren ‚Anziehungskraft‘ messen (z. B. Thibaut & French, 2016; Ivanova & Hallowell, 2012).

Bezogen auf die vierte Limitation wurde bereits im Rahmen der Diskussion von Studie 2 zum Langzeitgedächtnis kritisch angemerkt (Kapitel 6.4.2), dass die Kontrollgruppe dort nur nach mentalem Alter vergleichbar ist. Auf eine nach chronologischem Alter vergleichbare Kontrollgruppe musste jedoch verzichtet werden, da das eingesetzte Maß zur Erfassung des expressiven Wortschatzes (WWT 6-10) für Kinder älter als 10 Jahre nicht normiert ist und Deckeneffekte hervorgebracht hätte. Desweiteren scheinen die Subskalen des WWT 6-10 auch nach Ausschluss von möglichen DIF-Items nicht vollständig eindimensional zu sein, worauf die  $T_{11}$ -Statistik im Rahmen von Studie 3 hinweist. Ein möglicher Grund hierfür mag u. a. in der Unterscheidung im Register von Bildungs- und Alltagssprache liegen, wofür es erste, aufgrund deutlicher methodischer Einschränkungen bei der Zuteilung zum Register jedoch sehr vorsichtig zu interpretierende, Hinweise geben könnte (siehe Kapitel 6.4.2). Daher wäre es für zukünftige Studien wünschenswert, den Wortschatz breiter zu erfassen und das Register zu berücksichtigen, um dessen Anteil an der Varianzaufklärung zu testen. Ein angemessenes Testinstrument zu finden, das gleichzeitig ein breites Vokabular und einen breiten Altersbereich abdeckt, erweist sich jedoch als schwierig.

Weiterhin kann die in Studie 2 als Auswertungsverfahren gewählte MG-CFA nur ein Strukturmerkmal berücksichtigen; hier wurde die Zugehörigkeit zu einer der vier Wortarten Nomen, Verben, Adjektive und Kategorien als strukturierende latente Variable herangezogen. Die Methodik ist aber nicht dazu geeignet, komplexere Annahmen zur Netzwerkstruktur des semantischen Gedächtnisses angemessen abzubilden. Die Konstruktion eines solchen Netzwerkes beruht in der Regel auf Daten aus funktionalen bildgebenden Verfahren (Bassett & Sporns, 2017) oder aus Assoziationen (z. B. Kenett, Gold et al., 2016), die beide sehr zeitaufwändig zu erfassen sind. Aus einer Pilotierungsphase liegen von Kindern mit und ohne Lernbehinderung bereits einige Daten in einem Assoziations-Paradigma vor, die jedoch für die vorliegende Arbeit nicht ausreichend aussagekräftig sind (Bruns & Grosche, 2019, Feb). Mit den vorliegenden Leistungsdaten des WWT 6-10 aus Studie 2 kann der Netzwerk-Ansatz jedoch veranschaulicht werden, wie im folgenden Abschnitt dargestellt wird.

### 8.3. Ausblick auf Netzwerkanalyse

Um der Limitation der mangelnden Abbildung einer Netzwerkstruktur zu begegnen, soll als Ausblick eine Netzwerkanalyse der Daten des WWT 6-10 vorgestellt werden. Dazu wird zunächst die Konstruktion eines Netzwerkes erläutert, und es werden Indizes zur Beschreibung des Vernetzungsgrades und der Netzwerkstruktur genannt. Anschließend werden die aus den WWT-Daten generierten Netzwerke der Gruppen dargestellt und verglichen.

Ein Netzwerk besteht aus Knoten (*nodes*) und deren Verbindungen (*edges*). Bei semantischen Netzwerken repräsentieren die Knoten Wörter; die Verbindungen ergeben sich aus der Stärke des Zusammenhangs zwischen diesen Wörtern. Bei sog. assoziativen Netzwerken werden die Korrelationen aus der Überlappung von Assoziationen auf vorgegebene Cue-Wörter ermittelt (z. B. Kenett, Gold et al., 2016). Hierbei wird den Proband\*innen eine Reihe von Cue-Wörtern vorgegeben, zu denen sie innerhalb einer Minute alles frei assoziieren sollen, was ihnen zum Cue-Wort einfällt. Aus den Überlappungen der Assoziationen (auch über verschiedene Probanden hinweg) ergeben sich korrelative Zusammenhänge zwischen den Cue-Wörtern: mehr gemeinsame Assoziationen sorgen für eine höhere Überlappung zwischen den Cue-Wörtern und damit für eine höhere Korrelation.<sup>29</sup> Mit einem anschließenden Filterverfahren (Massara, di Matteo & Aste, 2016; Tumminello, Aste, di Matteo & Mantegna, 2005) werden geringfügige Verbindungen entfernt. Alle verbleibenden Verbindungen werden gleich stark gewichtet, indem sie in einer *Adjacency*-Matrix auf den Wert 1 festgesetzt werden. Daraus wird ein ungewichtetes und ungerichtetes Netzwerk modelliert, das graphisch dargestellt und mithilfe von quantitativen Methoden der Graphentheorie analysiert werden kann.

Die Graphentheorie liefert eine Reihe von Indizes zur quantitativen Schätzung der Strukturbeschaffenheit eines Netzwerkes. Dabei gibt der *degree* die Anzahl der Verbindungen pro Knoten wieder und beschreibt damit die generelle Vernetzungsdichte. Die Strukturierung und Organisation des Netzwerkes kann u. a. durch die folgenden Indizes beschrieben werden: Die Ausbreitung des Netzwerkes als durchschnittliche kürzeste Entfernung zwischen sämtlichen Knoten-Paaren ist die *average shortest path length* (*ASPL*). Der *clustering coefficient* (*CC*) gibt die Dichte der Vernetzung in Clustern an, d. h., wie sehr benachbarte Knoten wiederum selbst miteinander verbunden sind. Die Zusammensetzung in einzelne Module wird durch den *Modularitätsindex* (*Q*; Newman, 2005) dargestellt. Schließlich drückt der sog. *„small-world-ness“-Index* (*S*) aus, wie „chaotisch“ ein Netzwerk organisiert ist (Kenett, Beaty et al., 2016). Für eine tiefergehende Einführung und Erläuterung dieser Indizes wird hier auf die einschlägigen Publikationen zur

---

<sup>29</sup> Im vorliegenden Fall der WWT-Daten werden die Korrelationen zwischen den Items herangezogen, wie im übernächsten Absatz beschrieben wird.

Netzwerktheorie verwiesen (z. B. Baronchelli et al., 2013; Boccaletti et al., 2006; Borge-Holthofer & Arenas, 2010).

Mit den Daten der WWT-Items wurde nach dem oben geschilderten Vorgehen (siehe Kenett, Gold et al., 2016; Bruns & Grosche, 2019, Feb) für jede Gruppe ein Netzwerk aus der Item-Korrelationsmatrix konstruiert. Die Netzwerkparameter wurden mit dem R-Paket `NetworkToolbox` (Christensen, 2018) berechnet. Abbildung 11 zeigt die resultierenden Netzwerke der beiden Gruppen. Um einen inferenzstatistischen Vergleich der Parameter zwischen den Gruppen zu ermöglichen, wurden in einem partiellen Bootstrapping Verfahren (Kenett, Gold et al., 2016) pro Gruppe  $k = 100$  Netzwerke aus jeweils 49 zufällig gezogenen Items erstellt. Die Verteilungen der oben vorgestellten Indizes der Subnetzwerke konnten daraufhin mit  $t$ -Tests auf signifikante Unterschiede hin überprüft werden.

Tabelle 11

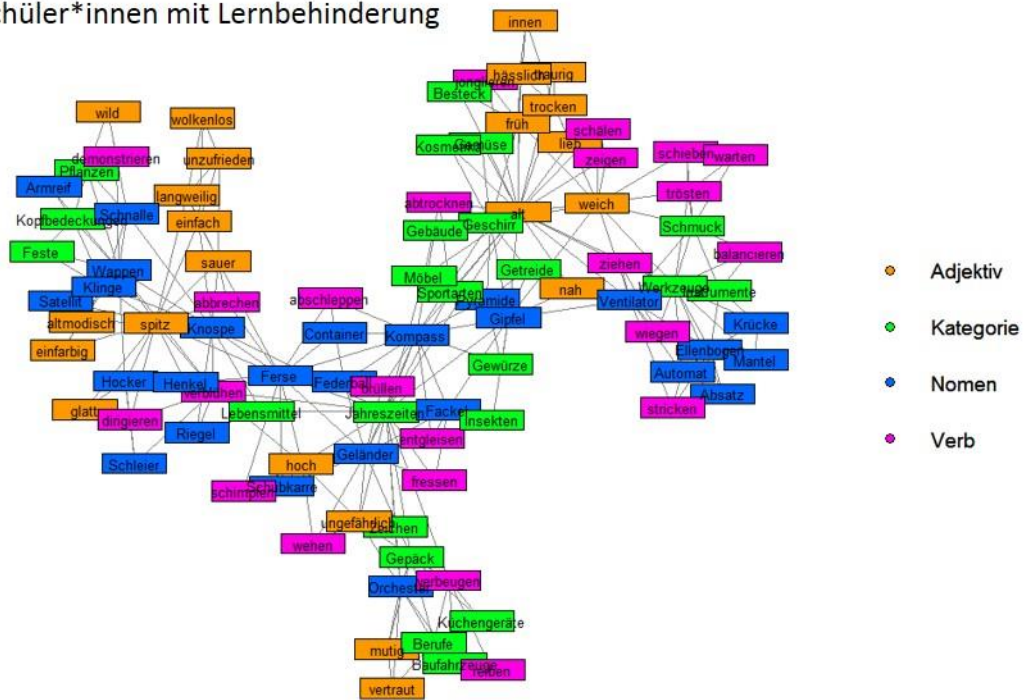
*Ergebnisse der Netzwerkanalyse der Daten des WWT 6-10*

Netzwerk-Maß	LB	KG	Signifikanz (partial bootstrap)
ASPL	3.346	3.470	
ASPL <sub>part</sub>	2.601	2.817	$t(172.48) = 414.67; p < .001$
D	7	7	
D <sub>part</sub>	5	5.99	$t(99) = 99; p < .001$
CC	.741	.738	
CC <sub>part</sub>	.725	.724	$t(101.21) = 2.161; p = .033$
S	7.971	7.500	
S <sub>part</sub>	5.569	5.208	$t(141.99) = 129.44; p < .001$
Q	.626	.611	
Q <sub>part</sub>	.494	.542	$t(112.71) = 163.18; p < .001$

*Anmerkungen.* ASPL: average shortest pathlength, Ausbreitung des Netzes. D: diameter, längste shortest pathlength im Netzwerk. CC: clustering coefficient, Nachbarn von Knoten sind selbst benachbart. S: small-world-ness Index, Verhältnis von CC zu ASPL, Indikator für ‚chaotische‘ Organisation (Kenett, Beaty et al., 2016). Q: modularity index, Untergliederung des Netzwerkes in kleinere Submodule. Die mit ‚part‘ indizierten Maße beziehen sich auf die Ergebnisse des partiellen bootstrappings ( $i = 100$  Iterationen) wieder, bei dem pro Iteration  $k = 49$  Items zufällig gezogen und daraus ein Netzwerk gebildet wurde, zum Zwecke der Signifikanzprüfung (siehe letzte Spalte).



a) WWT-Netzwerk  
Schüler\*innen mit Lernbehinderung



b) WWT-Netzwerk  
Kontrollgruppe

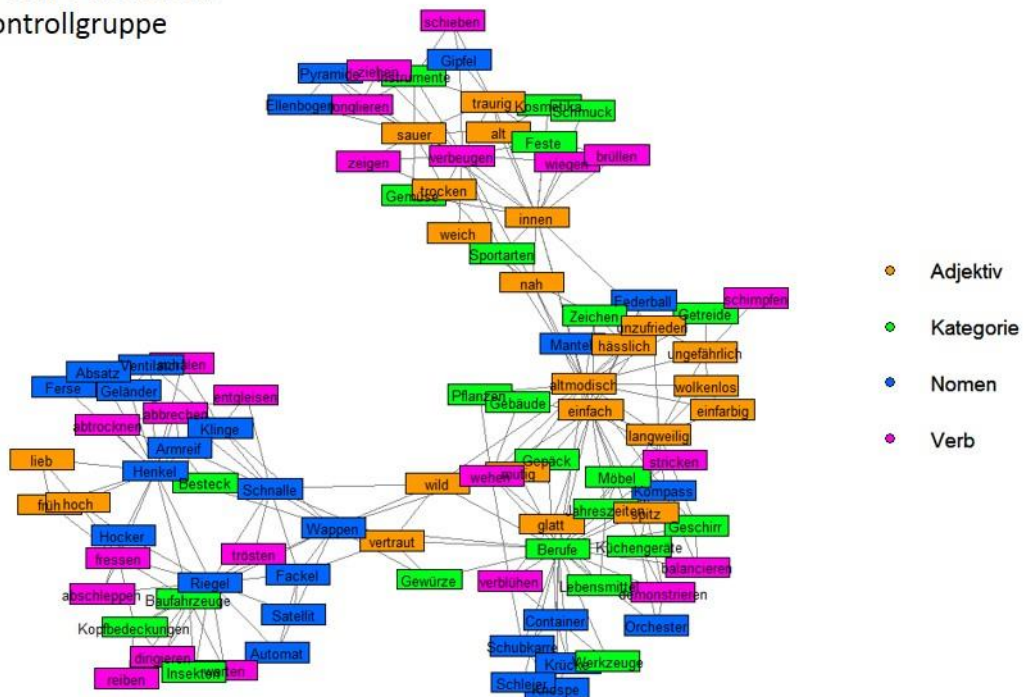


Abbildung 11. Semantische Netzwerke anhand der WWT-Daten für a) Schüler\*innen mit Lernbehinderung und b) unbeeinträchtigte Kontrollgruppe.

Die Ergebnisse weisen darauf hin, dass sich die Vernetzungsstruktur zwischen den Gruppen unterscheiden könnte. Inhaltlich passen diese Ergebnisse zu den in Bruns und Grosche (2019, Feb.) formulierten Hypothesen, insofern als das mentale Lexikon von Kindern mit Lernbehinderung chaotischer vernetzt erscheint. Das zeigt sich in den ermittelten Parametern (siehe Tabelle 11) folgendermaßen: die Ausbreitung der Netzwerke war niedriger (ASPL:  $p < .001$ ), die Cluster-Dichte jedoch höher (CC:  $p = .033$ ). Der ‚Chaos-Indikator‘ der small-world-ness war ebenfalls höher ausgeprägt (S:  $p < .001$ ), während die Modularität geringer ausfiel (Q:  $p < .001$ ). Diese Ergebnisse decken sich mit den explorativen Befunden zu Netzwerken, die aus Assoziationsdaten einer Pilotierung erstellt wurden (Bruns & Grosche, 2019, Feb.).

Mit diesen ersten Auswertungen kann gezeigt werden, dass die Abbildung von Netzwerkstrukturen mithilfe neuerer Netzwerk-Tools möglich ist und für die Fragestellung zu Unterschieden im mentalen Lexikon von Kindern mit und ohne Lernbehinderung einen Erkenntnisgewinn verspricht. Für belastbarere Aussagen wäre eine umfangreiche Erfassung von Assoziationsdaten notwendig (Kenett, Gold et al., 2016).

Auf diese Weise könnte die Fragestellung untersucht werden, ob eventuell auftretende Unterschiede eher quantitativer oder qualitativ-struktureller Art sind. Da bereits ein Zusammenhang zwischen Intelligenz und ‚strukturierenden‘ Merkmalen von Netzwerken nachgewiesen werden konnte (Barbey, 2018; Kenett, Beaty et al., 2016), ließe sich daraus ableiten, dass bei Kindern mit Lernbehinderung eher unstrukturierte und chaotische Netzwerke zu erwarten sind.

Weiterhin könnten aus den Analysen Informationen zur (differenziellen) Vernetztheit einzelner Wörter gewonnen werden, die wiederum als mögliche Prädiktoren für eine Erweiterung des in Studie 2 beschriebenen GLMMs eingesetzt werden können. So stellten Bassett und Mattar (2017) fest, dass interindividuelle Unterschiede im Arbeitsgedächtnis mit der Modularität funktionaler Aktivierungsmuster im Gehirn zusammenhängen, und Hills et al. (2009) konnten bei Kleinkindern den Erwerbszeitpunkt von Wörtern anhand deren Vernetztheit im sprachlichen Input vorhersagen. Insofern erscheint es aussichtsreich, diesen methodischen Zugang weiter zu verfolgen, um häufig postulierte, aber bisher kaum überprüfbare Annahmen über den allgemeinen Wissenserwerb (Bjorklund, 1987) und die Vernetzung des Vorwissens bei Kindern mit Lernbehinderung (Lauth, Brunstein et al., 2014) empirisch zu fundieren.

#### **8.4. Theoretische und praktische Implikationen**

Die Ergebnisse der Arbeit lassen sich in eine allgemeine und eine spezifische Aussage differenzieren: zum einen lässt sich insgesamt betrachtet das Muster einer Entwicklungsverzögerung (*delay* bzw. *delayed onset*) nachweisen, massive Defizite auf breiter Ebene in Arbeits- und Lang-

zeitgedächtnis sind demnach nicht zu vermuten. Zum anderen sollte der spezifische differenzielle Befund genauer betrachtet werden: Bei der Interaktion mit dem Langzeitgedächtnis scheinen größere Schwierigkeiten vorzuliegen, die über das mentale Alter hinausgehen. Dieser Befund ist zum einen auf theoretischer Ebene interessant, da den kognitiven Lernvoraussetzungen lange eine untergeordnete Rolle zugesprochen wurde – dies wird im folgenden Abschnitt besprochen. Zum anderen ist der Befund auf praktischer Ebene bedeutsam, weil er belegt, dass Maßnahmen zur Entlastung des Arbeitsgedächtnisses notwendig und sinnvoll sind.

#### **8.4.1. Theoretische Implikation: Strategie- vs. Processing-Hypothese**

Im Folgenden soll die Frage erörtert werden, inwiefern der konstatierte Interaktionseffekt mit der eingangs der Arbeit erwähnten Strategie-Hypothese (Klauer & Lauth, 1997) erklärt werden könnte. Wie in Kapitel 1.3 dargelegt, besagt die Strategie-Hypothese, dass Schüler\*innen mit Lernbehinderung nicht grundsätzlich in ihren kognitiven Funktionen beeinträchtigt seien, sondern lediglich ungünstige und ineffektive Lernaktivitäten an den Tag legten. Sie hätten nicht so viele Lernstrategien und Herangehensweisen an Aufgaben zur Verfügung, bzw. wählten diese oft nicht aufgabenadäquat aus und gingen insgesamt weniger planvoll vor. Empirisch festgestellte (Aktivierungs-)Unterschiede in basalen kognitiven Leistungsvariablen führen Klauer und Lauth (1997) vor allem auf unterschiedliche Anstrengungsbereitschaft und Metakognition zurück, da diese sich auch auf die Aufgabenbearbeitung in wissenschaftlichen Testsituationen auswirkten. Wenn die Strategie-Hypothese den Effekt nicht plausibel erklären kann, sollte deren Aussage über die niedrige Relevanz von kognitiven Lernvoraussetzungen (zumindest hinsichtlich des Zusammenspiels von Arbeits- und Langzeitgedächtnis) hinterfragt werden.

Die generelle Gültigkeit der Strategie-Hypothese soll hier nicht in Frage gestellt werden, denn es lässt sich auch für diese Arbeit nicht völlig ausschließen, dass auch motivationale, metakognitive und strategische Variablen einen Einfluss auf die Bearbeitung der Arbeitsgedächtnisaufgaben gehabt haben könnten. Es soll aber im Folgenden gezeigt werden, dass die Erklärung der Interaktion mithilfe der Strategie-Hypothese aus den folgenden zwei Gründen weniger plausibel ist als die Annahme von beeinträchtigten kognitiven Lernvoraussetzungen.

Zunächst kann durch den Ansatz des mentalen Altersvergleichs die Effizienz von Strategien und deren Einsatz zu einem Teil kontrolliert werden. So zeigen die Gruppen vergleichbare Leistungen in den Variablen des CFT 1-R und WWT 6-10, die zur Schätzung des mentalen Alters herangezogen werden. Die allgemeine mentale Entwicklung, die sich auch auf die Strategien und Metakognition auswirken dürfte, ist also zwischen den Gruppen parallelisiert. Da es plausibel ist anzunehmen, dass sich die postulierten Strategiedefizite auf mehrere Domänen und Aufgabenformate auswirken, müsste folglich eine sehr spezifische Hypothese dafür vorliegen,

warum ausgerechnet die Bearbeitung von Arbeitsgedächtnis-Aufgaben andersartige Anforderungen an Strategien, Motivation und Metakognition stellen sollte. Eine solche Hypothese liegt nicht vor, aber es könnte immerhin sein, dass die AG-Aufgaben eine größere inhaltliche Nähe zu misserfolgsbehafteten Aufgabenformaten aufweisen. Folglich könnte es sein, dass Kinder mit Lernbehinderung eher Strategien und Herangehensweisen zeigen, die einer erfolgreichen Bewältigung in besonderem Maße abträglich sind.

Doch selbst wenn es eine solche Annahme gäbe, müsste sie nochmals spezifiziert werden, da keine allgemeine Minderleistung im Arbeitsgedächtnis vorliegt, sondern der fragliche Effekt nur bei der spezifischen Kombination von Redintegration und Wortschatz in Form einer disordinalen Interaktion auftritt. Dieser Effekt steht in deutlichem Kontrast zu anderen gleichzeitig gemessenen kognitiven Variablen, auf die sich ein Strategiedefizit ebenfalls (zumindest graduell) auswirken müsste. Wenn ein allgemeines Strategiedefizit vorläge, dann wäre zu erwarten, dass sich dieses auch auf die anderen Maße des Arbeitsgedächtnisses und deren Zusammenhänge mit anderen Entwicklungsindikatoren ebenfalls auswirkte. Das ist aber nicht der Fall, denn der Redintegrationprozess wird weder von den anderen beiden Entwicklungsindikatoren (chronologisches Alter und kognitive Kapazität) differenziell vorhergesagt; noch ist bei den beiden anderen abhängigen Variablen (AG-Kapazität und Rehearsalprozess) im Verhältnis zum Wortschatz ein differenzielles Muster zu beobachten.

Indem also der allgemeine Einfluss von Strategien, Motivation und Metakognition über den Ansatz des mental age Matchings zumindest teilweise kontrolliert ist, und da die fragliche Interaktion ausschließlich bei einer spezifischen Kombination von Variablen eintritt, erscheint es weniger plausibel, dass die Strategie-Hypothese diesen Befund angemessen erklären kann. Insofern kann der differenzielle Zusammenhang im Vergleich zur MA-Kontrollgruppe durchaus als Hinweis auf Besonderheiten in der Informationsverarbeitung unabhängig von der Strategienutzung gedeutet werden. Auf eine solche Interpretation wird im Folgenden eingegangen.

Es wäre nun also eine alternative Erklärung zu überlegen, wie der Befund gedeutet werden kann, dass ein umfangreicherer Wortschatz zu einer geringeren Effektivität des Redintegrationprozesses führt. Hier erscheint es plausibel, sich den Zugriff auf das mentale Lexikon als Suchprozess vorzustellen (Masuda, Porter & Lambiotte, 2017), der bei einer wohlorganisierten Struktur eher und schneller erfolgreich ist (Vitevitch et al., 2012). Beim Redintegrationprozess wird auf teilweise zerfallene Gedächtnisspuren zurückgegriffen, die aus dem LZG rekonstruiert werden müssen; hierbei können Verwechslungen entstehen. Sind nun mehr Wörter im mentalen Lexikon enthalten und dann noch möglicherweise schlechter und chaotischer vernetzt, so könnte dies die Suche und Auswahl der korrekten Kandidaten beeinträchtigen und für mehr Verwechslungen sorgen. Dass die Verwechslungsgefahr steigt, zeigt sich z. B. daran, dass in

AG-Aufgaben bei Wörtern mit vielen ähnlichen Nachbarn mehr Intrusionen zu verzeichnen sind, auch wenn sie insgesamt besser behalten werden können (Roodenrys, 2009).

Die Erklärung für die mit wachsendem Wortschatzumfang reduzierte Redintegration-Effektivität durch die Verwechslungs-Hypothese ließe sich auch als Mangel an Inhibition auffassen, da bei Selektionsprozessen die Unterdrückung konkurrierender Kandidaten eine wichtige Rolle spielt (Klimesch, 2012; Mostofsky & Simmonds, 2008). Passolunghi, Cornoldi und de Liberto (1999) konnten zeigen, dass Schüler\*innen, die schwache Leistungen in Problemlöse-Aufgaben zeigten, mehr Intrusionsfehler machten und größere Schwierigkeiten hatten, sich nur relevante Informationen zu merken. Folglich erscheint die Annahme plausibel, dass es Kindern mit Lernbehinderung bei wachsendem Wortschatzumfang zunehmend schlechter gelingt, die Aktivierung von semantisch oder phonologisch ähnlichen, jedoch inkorrekten, Wörtern zu unterdrücken (de Beni & Palladino, 2000). Da mit einem größeren Wortschatzumfang die Anzahl der möglichen Kandidaten ansteigt, ist zu vermuten, dass die Auswahl des korrekten Wortes Kindern mit Lernbehinderung zunehmend schwerer fällt.

Wie relevant Inhibition für kognitive Prozesse sein kann, wird sowohl generell (Dempster & Corkill, 1999; Gorfein & MacLeod, 2007), als auch für andere Formen psychischer Beeinträchtigungen diskutiert: z. B. für ADHS (Willcutt et al., 2005), Autismus (Rubenstein & Merzenich, 2003), sowie für Lernbeeinträchtigungen im weiteren Sinne (*poor comprehenders*, Borella, Carretti & Pelegrina, 2010). Auch gibt es Hinweise auf einen Zusammenhang von Inhibitionsleistungen und Arbeitsgedächtnis (Chiappe, Hasher & Siegel, 2000), indem Inhibition dafür sorgt, dass der Speicher von Distraktoren und alten Inhalten bereinigt wird (Dagry, Vergauwe & Barrouillet, 2017). Für Dyslexie formulieren Hancock, Pugh und Hoeft (2017) eine „neural noise hypothesis“, in der sie ein Ungleichgewicht von exzitatorischen und inhibitorischen Signalen als eine mögliche Ursache ansehen. Die reduzierte Inhibition könnte erklären, dass das sog. neuronale Rauschen (*perceptual neural noise*) schlechter unterdrückt werden kann, womit wiederum die bereits lange festgestellten Defizite in der phonologischen Bewusstheit bei Personen mit Dyslexie erklärbar wären. Vor dem Hintergrund dieser Debatte und aufgrund der inhaltlichen Nähe der Beeinträchtigung (Probleme in der Aneignung von sprachlichem Material) dürfte es vielversprechend sein, die Inhibitions-Hypothese auch für Kinder mit Lernbehinderung weiter zu verfolgen.

Daher und weil die Strategie-Hypothese, wie oben gezeigt wurde, den Befund der differenziellen Redintegration-Interaktion nicht befriedigend zu erklären vermag, erscheint es sinnvoll, für Kinder mit Lernbehinderung die alternative Hypothese zu überprüfen, inwiefern sie Einschränkungen in Inhibitionsfunktionen zeigen (analog zu de Beni & Palladino, 2000; Passolunghi et al., 1999), und ob die Einschränkungen möglicherweise auch mit strukturellen und

funktionalen Merkmalen der Organisation des Langzeitgedächtnisses zusammenhängen. Im Rahmen einer Meta-Analyse zu exekutiven Funktionen bei Schülerinnen und Schülern mit intellektuellen Beeinträchtigungen (welche eine große Überschneidung mit Lernbehinderung aufweisen) stellten Danielsson et al. (2012) Defizite bei Inhibitionsaufgaben fest, die signifikant über das mentale Alter hinausgingen.

Erste Hinweise auf eine eventuell ‚chaotischere‘ Vernetzungsqualität des semantischen Gedächtnisses bei assoziationsbasierten Netzwerken liegen vor (Bruns & Grosche, 2019, Feb.). Die oben angeführten tentativen Ergebnisse zur alternativen Netzwerkanalyse der Daten des WWT 6-10 aus Studie 2 weisen ebenfalls in diese Richtung (siehe Abschnitt 8.3). Insofern scheint die genauere Untersuchung des semantischen Langzeitgedächtnisses ein vielversprechender Ansatz für weitere Forschungen zu sein, der das theoretische Verständnis des Konstruktes Lernbehinderung weiter vorantreiben dürfte.

#### **8.4.2. Praktische Implikation**

Auf praktischer Ebene tragen die Ergebnisse dazu bei, im schulischen Alltag besser und zielgerichteter auf die Bedürfnisse von Schüler\*innen mit Lernbehinderung reagieren zu können. Präskriptive Aussagen lassen sich aus dem vorliegenden deskriptiven Wissen zur präziseren Beschreibung der Verfügbarkeit und Funktionalität von kognitiven Lernvoraussetzungen jedoch nicht unmittelbar ableiten (Kuhl, 2019). Ansätze zur Förderung und Wege, diese Erkenntnisse in den Unterricht zu integrieren (z. B. Schulze & Kuhl, 2019; Schulze, Kuhl & Breucker, 2020), müssen ihrerseits einer empirischen Prüfung unterzogen werden. Dennoch können die Hinweise aus dieser Arbeit auf die Verfügbarkeit von kognitiven Lernvoraussetzungen auch für die schulische Praxis von Nutzen sein. Auf diese Weise können Lehrkräfte besser einschätzen, auf welche kognitiven Lernvoraussetzungen ihre Schüler\*innen in welchem Maß zurückgreifen können, und wo ggf. eine Adaptation des Unterrichts notwendig ist.

Da diese Arbeit zeigt, dass Einschränkungen im Arbeits- und Langzeitgedächtnis eher gradueller Art sind, können Lehrkräfte also zuversichtlich sein, dass in diesen Bereichen mit der Zeit, und entsprechender zusätzlicher Förderung, eine positive und mindestens im Rahmen des mentalen Entwicklungsalters liegende Veränderung zu erwarten ist. Ob jedoch Rückstände vollständig aufgeholt werden, kann aus den Daten dieser Arbeit nicht geschlossen werden. Da die Befunde im Allgemeinen auf eine Entwicklungsverzögerung hindeuten, erscheint es zielführend, Unterricht und Interventionen vor allem am mentalen Alter auszurichten. So könnte durchaus eine Instruktion von Strategien, die dem mentalen Alter angemessen sind, als eine sinnvolle und aussichtsreiche Maßnahme betrachtet werden.

Gleichzeitig muss der spezifische differenzielle Befund zur Nutzung von Langzeitgedächtnisinhalten berücksichtigt werden, da sich beim Redintegrationprozess ein Hinweis auf eine

strukturelle Abweichung nachweisen lässt. Während also Gedächtnisprozesse und -strukturen allgemein als kognitive Lernvoraussetzungen nicht über die Maße beeinträchtigt erscheinen, sollten spezifische Teilaspekte wie die Nutzung von Inhalten im Langzeitgedächtnis von Lehrkräften berücksichtigt werden. Es kann im Unterricht also hilfreich sein, das Arbeitsgedächtnis und insbesondere den Redintegrationprozess (zunächst) zu entlasten, indem möglichst wenige Informationen aus dem Langzeitgedächtnis rekonstruiert werden müssen, denn insbesondere in diesem Aspekt zeigten sich größere Schwierigkeiten im Zusammenspiel von Arbeits- und Langzeitgedächtnis.

### 8.5. Schlussbetrachtung

In dieser Arbeit wurde die Frage behandelt, ob Schüler\*innen mit Lernbehinderung abweichende Entwicklungsmuster in den kognitiven Lernvoraussetzungen Arbeits- und Langzeitgedächtnis aufwiesen. Im wesentlichen ergaben sich hinsichtlich der AG-Kapazität und des Rehearsalprozesses sowie der Struktur des Langzeitgedächtnisses Hinweise auf eine Entwicklungsverzögerung, welche im Rahmen des mentalen Alters lag. Als differenzieller Effekt wurde gezeigt, dass sich dagegen eine größere Wissensbasis negativ auf die Rekonstruktion von Arbeitsgedächtnis-Inhalten aus dem Langzeitgedächtnis auszuwirken schien. Da die Strategie-Hypothese diesen Effekt nicht vollständig erklären kann, ist die Vernachlässigung der kognitiven Lernvoraussetzungen nicht gerechtfertigt. Auf praktischer Ebene wecken die Ergebnisse Zuversicht, dass Kinder mit Lernbehinderung in ihren kognitiven Lernvoraussetzungen in vielen Bereichen nicht über ihr mentales Entwicklungsalter hinaus beeinträchtigt sind, sie also in zunehmendem Maße auf diese Ressourcen zugreifen können. Bei der Rekonstruktion aus dem Langzeitgedächtnis benötigen sie jedoch möglicherweise mehr gezielte Unterstützung.

Die Fragestellung wurde in drei Studien an einer Stichprobe von  $n = 93$  Kindern mit Lernbehinderung im Alter der Sekundarstufe 1 und einer Kontrollgruppe von insgesamt  $n = 120$  unbeeinträchtigten Kindern (bzw.  $n = 102$  für Studien 2 und 3) untersucht. Aufgrund der Stichproben-Einschlusskriterien für Lernbehinderung können die Ergebnisse nur auf Kinder generalisiert werden, bei denen ein sonderpädagogischer Förderbedarf und eine leichte bis grenzwertige Intelligenzminderung vorliegt (IQ 55-85 Punkte). Aufgrund der querschnittlich erhobenen Daten können nur Zusammenhänge mit Entwicklungsindikatoren (des mentalen Alters) betrachtet werden. Aussagen über die Entwicklungsmuster bei jüngeren oder älteren Personen mit Lernbehinderung (Mähler & Hasselhorn, 2003) sind ebenfalls nicht möglich.

Die erste Studie prüfte die Fragestellung, ob Kinder mit Lernbehinderung bei der Bearbeitung von Wortspannungsaufgaben differenzielle Entwicklungsmuster im Developmental Trajectory Ansatz hinsichtlich der AG-Kapazität sowie dem Rehearsal- und Redintegrationprozess

zur Aufrechterhaltung aufwiesen. Die zweite Studie betrachtete die Fragestellung, ob sich die Binnen- und Außenstruktur des mentalen Lexikons, gemessen am expressiven Wortschatz, zwischen Kindern mit und ohne Lernbehinderung unterschied. Studie 3 diente schließlich dazu, DIF im WWT 6-10 als methodische Ursache für die Interaktion von Redintegrationprozess und Wortschatzumfang auszuschließen.

Die Ergebnisse können folgendermaßen zusammengefasst werden:

- In Arbeitsgedächtnis-Kapazität und Rehearsalprozess wiesen Schüler\*innen mit Lernbehinderung bezogen auf das chronologische Alter einen verspäteten Entwicklungseinsatz auf; bezogen auf die mentalen Altersvariablen (kognitive Kapazität und Wortschatz) zeigten sich keine reliablen Unterschiede in den Entwicklungsmustern (Studie 1).
- Der Redintegrationprozess wirkte im Verhältnis zu chronologischem Alter und kognitiver Kapazität unbeeinträchtigt; es zeigte sich aber ein differenzielles Verhältnis zum Wortschatz. Je größer der Wortschatz, desto schwerer fiel Kindern mit Lernbehinderung die Rekonstruktion von Arbeitsgedächtnis-Inhalten (Studie 1). Dieses Muster hatte auch Bestand, wenn DIF im WWT 6-10 als methodische Ursache ausgeschlossen wurde (Studie 3).
- In der Binnen- und Außenstruktur des mentalen Lexikons konnten keine Unterschiede zwischen Kindern mit Lernbehinderung und einer nach mentalem Alter vergleichbaren Kontrollgruppe festgestellt werden. Wenn überhaupt, so liegt lediglich eine Entwicklungsverzögerung vor.

Allgemein lassen sich die Ergebnisse so interpretieren, dass kognitive Lernvoraussetzungen bei Kindern mit Lernbehinderung tatsächlich einer genaueren Betrachtung wert sind. So wurden zum einen großenteils verzögerte Entwicklungsmuster in Arbeits- und Langzeitgedächtnis gefunden; zum anderen aber auch ein differenzielles Muster bei der Nutzung des Langzeitgedächtnisses für die Rekonstruktion. Mit der Strategie-Hypothese kann dieser Effekt nicht hinreichend erklärt werden, sie sollte also in der Stärke ihrer Aussage gegen die Relevanz kognitiver Lernvoraussetzungen relativiert werden.

Zukünftige Studien sollten das semantische Gedächtnis bei Kindern mit Lernbehinderung auf Unterschiede in der Netzwerk-Struktur prüfen. Für die schulische Praxis können zwar keine unmittelbaren Implikationen für Unterricht und Intervention abgeleitet werden; das präzisere Beschreibungswissen über die Verfügbarkeit von Arbeits- und Langzeitgedächtnis-Komponenten kann für Lehrkräfte aber nützlich sein, Anforderungen besser an die Bedürfnisse von Kindern mit Lernbehinderung anzupassen.



## Literaturverzeichnis

- Acheson, D. J., Postle, B. R. & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36* (1), 17–36. <https://doi.org/10.1037/a0017679>
- Adams, E. J., Nguyen, A. T. & Cowan, N. (2018). Theories of working memory. Differences in definition, degree of modularity, role of attention, and purpose. *Language, Speech, and Hearing Services in Schools*, *49* (3), 340–355. [https://doi.org/10.1044/2018\\_LSHSS-17-0114](https://doi.org/10.1044/2018_LSHSS-17-0114)
- Allen, R. J. & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55* (1), 64–88. <https://doi.org/10.1016/j.jml.2006.02.002>
- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, *25* (2), 92–98. <https://doi.org/10.1027/1015-5759.25.2.92>
- Alloway, T. P. (2010). Working memory and executive function profiles of individuals with borderline intellectual functioning. *Journal of Intellectual Disability Research*, *54* (5), 448–456. <https://doi.org/10.1111/j.1365-2788.2010.01281.x>
- Alloway, T. P. & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106* (1), 20–29. <https://doi.org/10.1016/j.jecp.2009.11.003>
- Alloway, T. P. & Archibald, L. M. D. (2008). Working memory and learning in children with developmental coordination disorder and specific language impairment. *Journal of Learning Disabilities*, *41* (3), 251–262. <https://doi.org/10.1177/0022219408315815>
- Almeida, J., Knobel, M., Finkbeiner, M. & Caramazza, A. (2007). The locus of the frequency effect in picture naming. When recognizing is not enough. *Psychonomic Bulletin & Review*, *14* (6), 1177–1182. <https://doi.org/10.3758/BF03193109>
- Alt, M. (2011). Phonological working memory impairments in children with specific language impairment. Where does the problem lie? *Journal of Communication Disorders*, *44* (2), 173–185. <https://doi.org/10.1016/j.jcomdis.2010.09.003>
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, *38* (1), 123–140.
- Anderson, J. R. (2000). *Learning and memory. An integrated approach* (2nd ed.). New York, NY: Wiley.
- Anderson, J. R. (2013). *Kognitive Psychologie* (Lehrbuch, 7. Aufl.). Berlin: Springer.
- Andrews, S. (1989). Frequency and neighborhood effects on lexical access. Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15* (5), 802–814.
- Annaz, D., Karmiloff-Smith, A., Johnson, M. H. & Thomas, M. S. C. (2009). A cross-syndrome study of the development of holistic face recognition in children with autism, Down syndrome, and Williams syndrome. *Journal of Experimental Child Psychology*, *102* (4), 456–486. <https://doi.org/10.1016/j.jecp.2008.11.005>

- Arango, P. S., Aparicio, A. & Tenorio, M. (2018). Developmental trajectories of children with Down syndrome by socio-economic status. The case of Latin America. *Journal of Intellectual Disability Research*, 62 (9), 759–774. <https://doi.org/10.1111/jir.12516>
- Arciuli, J. & Simpson, I. C. (2011). Statistical learning in typically developing children. The role of age and speed of stimulus presentation. *Developmental Science*, 14 (3), 464–473. <https://doi.org/10.1111/j.1467-7687.2009.00937.x>
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory. A proposed system and its control processes. In K. W. Spence & J. T. Spence (Hrsg.), *The psychology of learning and motivation. Advances in research and theory* (S. 89–195). New York: Academic Press.
- Atwell, J. A., Conners, F. A. & Merrill, E. C. (2003). Implicit and explicit learning in young adults with mental retardation. *American Journal on Mental Retardation*, 108 (1), 56. [https://doi.org/10.1352/0895-8017\(2003\)108<0056:IAELIY>2.0.CO;2](https://doi.org/10.1352/0895-8017(2003)108<0056:IAELIY>2.0.CO;2)
- Baayen, R. H. (2010). Demythologizing the word frequency effect. A discriminative learning perspective. *The Mental Lexicon*, 5, 436–461.
- Baayen, R. H., Davidson, D. J. & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59 (4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Backhaus, K., Erichson, B. & Weiber, R. (2013). *Fortgeschrittene Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (Springer-Lehrbuch, 2., Aufl.). Berlin: Springer.
- Backman, J. E., Mamen, M. & Ferguson, H. B. (1984). Reading level design. Conceptual and methodological issues in reading research. *Psychological Bulletin*, 96 (3), 560–568.
- Baddeley, A. D. (1986). *Working memory* (Oxford Psychology Series, Bd. 11). Oxford: Clarendon Press.
- Baddeley, A. D. (2000). The episodic buffer. A new component of working memory? *Trends in Cognitive Sciences*, 4 (11), 417–423.
- Baddeley, A. D. (2012). Working memory. Theories, models, and controversies. *Annual Review of Psychology*, 63 (1), 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A. D., Gathercole, S. E. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105 (1), 158–173. <https://doi.org/10.1037/0033-295X.105.1.158>
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. H. Bower (Hrsg.), *Advances in research and theory* (The psychology of learning and motivation, Bd. 8, Bd. 8, S. 47–89). New York, NY: Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baddeley, A. D., Lewis, V. & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, 36 (A), 233–252. <https://doi.org/10.1080/14640748408402157>
- Baddeley, A. D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14 (6), 575–589. [https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/10.1016/S0022-5371(75)80045-4)

- Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, 1 (6), 589–595.
- Barbey, A. K. (2018). Network neuroscience theory of human intelligence. *Trends in Cognitive Sciences*, 22 (1), 8–20. <https://doi.org/10.1016/j.tics.2017.10.001>
- Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N. & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, 17 (7), 348–360. <https://doi.org/10.1016/j.tics.2013.04.010>
- Barry, C., Hirsh, K. W., Johnston, R. A. & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language*, 44 (3), 350–375. <https://doi.org/10.1006/jmla.2000.2743>
- Barry, C., Morrison, C. M. & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures. Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology Section A*, 50 (3), 560–585. <https://doi.org/10.1080/783663595>
- Bartsch, L. M., Singmann, H. & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition*, 46 (5), 796–808. <https://doi.org/10.3758/s13421-018-0805-9>
- Bassett, D. S. & Mattar, M. G. (2017). A network neuroscience of human learning. Potential to inform quantitative theories of brain and behavior. *Trends in Cognitive Sciences*, 21 (4), 250–264. <https://doi.org/10.1016/j.tics.2017.01.010>
- Bassett, D. S. & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, 20 (3), 353–364. <https://doi.org/10.1038/nn.4502>
- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M. & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108 (18), 7641–7646. <https://doi.org/10.1073/pnas.1018985108>
- Bates, D., Mächler, M., Bolker, B. M. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 49 (1), 1–48.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S. et al. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21 (1), 85–123.
- Bayliss, D. M., Bogdanovs, J. & Jarrold, C. (2015). Consolidating working memory. Distinguishing the effects of consolidation, rehearsal and attentional refreshing in a working memory span task. *Journal of Memory and Language*, 81, 34–50. <https://doi.org/10.1016/j.jml.2014.12.004>
- Beaujean, A. A., Benson, N., McGill, R. & Dombrowski, S. (2018). A misuse of IQ scores. Using the dual discrepancy/consistency model for identifying specific learning disabilities. *Journal of Intelligence*, 6 (3), 36. <https://doi.org/10.3390/jintelligence6030036>
- Beckage, N. M. & Colunga, E. (2016). Language networks as models of cognition. Understanding cognition through language. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard & B. Job

- (Hrsg.), *Towards a theoretical framework for analyzing complex linguistic networks* (Understanding Complex Systems, S. 3–28). Berlin: Springer.
- Beelmann, A. & Raabe, T. (2007). *Dissoziales Verhalten von Kindern und Jugendlichen. Erscheinungsformen, Entwicklung, Prävention und Intervention* (Klinische Kinderpsychologie, Bd. 10). Göttingen: Hogrefe.
- Belmont, J. M. & Mitchell, D. W. (1987). The general strategies hypothesis as applied to cognitive theory in mental retardation. *Intelligence*, *11* (1), 91–105. [https://doi.org/10.1016/0160-2896\(87\)90029-8](https://doi.org/10.1016/0160-2896(87)90029-8)
- Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M. & Neubauer, A. C. (2017). How semantic memory structure and intelligence contribute to creative thought. A network science approach. *Thinking & Reasoning*, *23* (2), 158–183. <https://doi.org/10.1080/13546783.2016.1278034>
- Beni, R. de & Palladino, P. (2000). Intrusion errors in working memory tasks. Are they related to reading comprehension? *Learning and Individual Differences*, *12*, 131–143.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107* (2), 238–246.
- Berman, M. G., Jonides, J. & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35* (2), 317–333. <https://doi.org/10.1037/a0014873>
- Bjorklund, D. F. (1987). How age changes in knowledge base contribute to the development of children's memory. An interpretative review. *Developmental Review*, *7*, 93–130.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. (2006). Complex networks. Structure and dynamics. *Physics Reports*, *424* (4-5), 175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>
- Bodenmann, G. (2006). Editorial. Lerntheorien und Kommunikation. *Sprache Stimme Gehör*, *30* (1), 1. <https://doi.org/10.1055/s-2006-931523>
- Borella, E., Carretti, B. & Pelegrina, S. (2010). The specific role of inhibition in Reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, *43* (6), 541–552. <https://doi.org/10.1177/0022219410371676>
- Borge-Holthoefer, J. & Arenas, A. (2010). Semantic networks. Structure and dynamics. *Entropy*, *12* (5), 1264–1302. <https://doi.org/10.3390/e12051264>
- Borkowski, J. G., Schneider, W. & Pressley, M. (1989). The challenges of teaching good information processing to learning disabled students. *International Journal of Disability, Development and Education*, *36* (3), 169–185. <https://doi.org/10.1080/0156655893603002>
- Bormann, T. (2011). The role of lexical-semantic neighborhood in object naming. Implications for models of lexical access. *Frontiers in Psychology*, *2*, 127. <https://doi.org/10.3389/fpsyg.2011.00127>
- Brandenburg, J., Kluszczewski, J., Fischbach, A., Büttner, G., Grube, D., Mähler, C. et al. (2013). Arbeitsgedächtnisfunktionen von Kindern mit Minderleistungen in der Schriftsprache. Zur

- Dissoziation von Lese- und Rechtschreibfertigkeiten und zur Relevanz des IQ-Diskrepanzkriteriums. *Lernen und Lernstörungen*, 2 (3), 147–159. <https://doi.org/10.1024/2235-0977/a000037>
- Brandenburg, J., Kleszczewski, J., Fischbach, A., Schuchardt, K., Büttner, G. & Hasselhorn, M. (2015). Working memory in children with learning disabilities in reading versus spelling. Searching for overlapping and specific cognitive factors. *Journal of Learning Disabilities*, 48 (6), 622–634. <https://doi.org/10.1177/0022219414521665>
- Brown, K. A., Parikh, S. & Patel, D. R. (2020). Understanding basic concepts of developmental diagnosis in children. *Translational Pediatrics*, 9 (Suppl 1), S9-S22. <https://doi.org/10.21037/tp.2019.11.04>
- Browne, M. & Cudeck, R. (1993). Alternative ways of assessing equation model fit. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 136–162). Newbury Park: Sage Publications, Inc.
- Bruns, G. & Grosche, M. (2017, Mai). *Arbeitsgedächtnis und Wortlängeneffekt bei Erwachsenen mit Lernbeeinträchtigungen*. Posterpräsentation. Arbeitsgruppe Empirische Sonderpädagogische Forschung (AESF), Rostock, Deutschland.
- Bruns, G. & Grosche, M. (2018, Juni). *Unerwartete Ergebnisse im Arbeitsgedächtnis bei Kindern mit Lernbehinderung*. Posterpräsentation. Arbeitsgruppe Empirische Sonderpädagogische Forschung (AESF), Genf, Schweiz.
- Bruns, G. & Grosche, M. (2019, Februar). *Modellierung von semantischen Netzen bei Schülerinnen und Schülern mit und ohne Lernbeeinträchtigung*. Posterpräsentation. 7. GEBF Tagung, Köln, Deutschland.
- Bryant, P. & Goswami, U. (1986). Strengths and weaknesses of the reading level design. A comment on Backman, Mamen, and Ferguson. *Psychological Bulletin*, 100 (1), 101–103.
- Burgess, N. & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55 (4), 627–652. <https://doi.org/10.1016/j.jml.2006.08.005>
- Büttner, G. (1998). Gedächtnisleistungen von Lernbehinderten. Kompetenzen und Defizite. In M. Greisbach, U. Kullik & E. Souvignier (Hrsg.), *Von der Lernbehindertenpädagogik zur Praxis schulischer Lernförderung* (S. 45–53). Lengerich: Pabst.
- Büttner, G. (2003). Gedächtnisentwicklung im Kindes- und Jugendalter. *Sprache Stimme Gehör*, 27 (1), 24–30. <https://doi.org/10.1055/s-2003-37876>
- Büttner, G. (2007). Kognitive Bedingungsfaktoren von Lern- und Verhaltensstörungen. In F. Linderkamp & M. Grünke (Hrsg.), *Lern- und Verhaltensstörungen. Genese, Diagnostik, Intervention* (S. 44–54). Weinheim: Beltz.
- Büttner, G. & Hasselhorn, M. (2007). Förderung von Lern- und Gedächtnisleistungen. In J. Walter & F. B. Wember (Hrsg.), *Sonderpädagogik des Lernens* (Handbuch Sonderpädagogik, Bd. 2, S. 281–292). Göttingen: Hogrefe.

- Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science*, 24 (6), 440–445. <https://doi.org/10.1177/0963721415606630>
- Campoy, G. (2008). The effect of word length in short-term memory. Is rehearsal necessary? *Quarterly Journal of Experimental Psychology*, 61 (5), 724–734. <https://doi.org/10.1080/17470210701402364>
- Carney, D. P. J., Henry, L. A., Messer, D. J., Danielsson, H., Brown, J. H. & Rönnerberg, J. (2013). Using developmental trajectories to examine verbal and visuospatial short-term memory development in children and adolescents with Williams and Down syndromes. *Research in Developmental Disabilities*, 34 (10), 3421–3432. <https://doi.org/10.1016/j.ridd.2013.07.012>
- Carroll, J. B. & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 25 (1), 85–95. <https://doi.org/10.1080/14640747308400325>
- Casals, M., Girabent-Farrés, M. & Carrasco, J. L. (2014). Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000–2012). A systematic review. *PLOS ONE*, 9 (11), 1–10.
- Caselli, C., Casadio, P. & Bates, E. (1999). Comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, 26, 69–111.
- Castro, N. & Siew, C. S. Q. (2019). Contributions of modern network science to the cognitive sciences. Revisiting research spirals of representation and process. *Preprint*. <https://doi.org/10.31234/osf.io/gkmb8>
- Ceci, S. J. (1984). A developmental study of learning disabilities and memory. *Journal of Experimental Child Psychology*, 38 (2), 352–371. [https://doi.org/10.1016/0022-0965\(84\)90131-0](https://doi.org/10.1016/0022-0965(84)90131-0)
- Chan, K. Y. & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, 34 (4), 685–697. <https://doi.org/10.1111/j.1551-6709.2010.01100.x>
- Chapman, R. S., Hesketh, L. J. & Kistler, D. J. (2002). Predicting longitudinal change in language production and comprehension in individuals with down syndrome. Hierarchical linear modeling. *Journal of Speech, Language, and Hearing Research*, 45 (5), 902–915. [https://doi.org/10.1044/1092-4388\(2002/073\)](https://doi.org/10.1044/1092-4388(2002/073))
- Chen, Q. & Mirman, D. (2012). Competition and cooperation among similar representations. Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, 119 (2), 417–430. <https://doi.org/10.1037/a0027175>
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9 (2), 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Chiappe, P., Hasher, L. & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition*, 28 (1), 8–17.
- Christensen, A. P. (2018). NetworkToolbox. Methods and Measures for Brain, Cognitive, and Psychometric Network Analysis in R. *R Journal*. <https://doi.org/10.17605/OSF.IO/6KMAV>

- Christensen, A. P. (2020). *Towards a network psychometrics approach to assessment. Simulations for redundancy, dimensionality, and loadings (Unpublished doctoral dissertation)*. University of North Carolina at Greensboro, Greensboro, NC, USA. <https://doi.org/10.31234/osf.io/84kgd>
- Clarkson, L., Roodenrys, S., Miller, L. M. & Hulme, C. (2017). The phonological neighbourhood effect on short-term memory for order. *Memory*, 25 (3), 391–402. <https://doi.org/10.1080/09658211.2016.1179330>
- Clerc, J. & Courbois, Y. (2017). Evidence of a phonological similarity effect after rehearsal training in adolescents with intellectual disability. *Journal of Cognitive Education and Psychology*, 16 (2), 127–143. <https://doi.org/10.1891/1945-8959.16.2.127>
- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82 (6), 407–428.
- Colom, R., Quiroga, M. Á., Shih, P. C., Martínez, K., Burgaleta, M., Martínez-Molina, A. et al. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence*, 38 (5), 497–505. <https://doi.org/10.1016/j.intell.2010.06.008>
- Conway, A. R.A., Kane, M. J. & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7 (12), 547–552. <https://doi.org/10.1016/j.tics.2003.10.005>
- Cornoldi, C. & Giofrè, D. (2014). The crucial role of working memory in intellectual functioning. *European Psychologist*, 19 (4), 260–268. <https://doi.org/10.1027/1016-9040/a000183>
- Craig, S. & Lewandowsky, S. (2013). Working memory supports inference learning just like classification learning. *Quarterly Journal of Experimental Psychology*, 66 (8), 1493–1503. <https://doi.org/10.1080/17470218.2013.818703>
- Dagry, I., Vergauwe, E. & Barrouillet, P. (2017). Cleaning working memory. The fate of distractors. *Journal of Memory and Language*, 92, 327–342. <https://doi.org/10.1016/j.jml.2016.08.002>
- Danielsson, H., Henry, L. A., Messer, D. J. & Rönnerberg, J. (2012). Strengths and weaknesses in executive functioning in children with intellectual disability. *Research in Developmental Disabilities*, 33 (2), 600–607. <https://doi.org/10.1016/j.ridd.2011.11.004>
- Danielsson, H., Zottarel, V., Palmqvist, L. & Lanfranchi, S. (2015). The effectiveness of working memory training with individuals with intellectual disabilities. A meta-analytic review. *Frontiers in Psychology*, 6, 1230. <https://doi.org/10.3389/fpsyg.2015.01230>
- David, A. & Wei, L. (2005). The composition of the bilingual lexicon. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (Hrsg.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (S. 594–607).
- Debelak, R. (2019). An Evaluation of Overall Goodness-of-Fit Tests for the Rasch Model. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02710>
- Dell, G. S. & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42, 287–314.

- Dempster, F. N. & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior. Unifying themes for educational psychology. *Educational Psychology Review*, 11 (1), 1–88. <https://doi.org/10.1023/A:1021992632168>
- Derraugh, L. S., Neath, I., Surprenant, A. M., Beaudry, O. & Saint-Aubin, J. (2017). The effect of lexical factors on recall from working memory. Generalizing the neighborhood size effect. *Canadian Journal of Experimental Psychology*, 71 (1), 23–31. <https://doi.org/10.1037/cep0000098>
- Deutscher Bildungsrat. (1976). *Empfehlungen der Bildungskommission zur pädagogischen Förderung behinderter und von Behinderung bedrohter Kinder und Jugendlicher*. Verabschiedet auf d. 34. Sitzung d. Bildungskomm. am 12./13. Okt. 1973 in Bonn. Stuttgart: Klett-Cotta.
- Deutsches Institut für Medizinische Dokumentation und Information (Hrsg.) (2020). *Systematisches Verzeichnis. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision. German Modification*. Verfügbar unter <https://www.icd-code.de/icd/code/ICD-10-GM.html>
- Deyne, S. de & Storms, G. (2008). Word associations. Network and semantic properties. *Behavior Research Methods*, 40 (1), 213–231.
- Deyne, S. de, Verheyen, S. & Storms, G. (2016). Structure and organization of the mental lexicon. A network approach derived from syntactic dependency relations and word associations. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard & B. Job (Hrsg.), *Towards a theoretical framework for analyzing complex linguistic networks* (Understanding Complex Systems, S. 47–79). Berlin: Springer. [https://doi.org/10.1007/978-3-662-47238-5\\_3](https://doi.org/10.1007/978-3-662-47238-5_3)
- Dilkina, K., McClelland, J. L. & Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Research*, 1365, 66–81.
- Dosher, B. A. & Ma, J.-J. (1998). Output loss or rehearsal loop. Output-time versus pronunciation-time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24 (2), 316–335.
- Duff, M. C., Covington, N. V., Hilverman, C. & Cohen, N. J. (2020). Semantic memory and the hippocampus. Revisiting, reaffirming, and extending the reach of their critical relationship. *Frontiers in Human Neuroscience*, 13, 1–17. <https://doi.org/10.3389/fnhum.2019.00471>
- Ehlert, A., Schroeders, U. & Fritz-Stratmann, A. (2012). Kritik am Diskrepanzkriterium in der Diagnostik von Legasthenie und Dyskalkulie. *Lernen und Lernstörungen*, 1 (3), 169–184. <https://doi.org/10.1024/2235-0977/a000018>
- Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion* (Bachelorstudium Psychologie). Göttingen: Hogrefe.
- Eser, K.-H. (2006). Lernbehinderung, die Behinderung "auf den zweiten Blick". Sind (junge) Menschen mit Lernbehinderung überhaupt behindert? In C. Bölke, K.-H. Eser & A. Gronewold (Hrsg.), *Junge Menschen mit Behinderung in der Berufsausbildungsvorbereitung* (Berichte und Materialien, Bd. 16, S. 45–65). Offenbach am Main: INBAS.



- Farrell, S., Oberauer, K., Greaves, M., Pasiecznik, K., Lewandowsky, S. & Jarrold, C. (2016). A test of interference versus decay in working memory. Varying distraction within lists in a complex span task. *Journal of Memory and Language*, *90*, 66–87. <https://doi.org/10.1016/j.jml.2016.03.010>
- Fenson, L., Vella, D. & Kennedy, M. (1989). Children's knowledge of thematic and taxonomic relations at two years of age. *Child Development*, *60* (4), 911–919. <https://doi.org/10.2307/1131032>
- Ferrari, M. (2009). Borderline intellectual functioning and the intellectual disability construct. *Intellectual and Developmental Disabilities*, *47* (5), 386–389. <https://doi.org/10.1352/1934-9556-47.5.386>
- Fischer, G. H. (1995). Derivations of the Rasch model. In G. H. Fischer & I. W. Molenaar (Hrsg.), *Rasch models. Foundations, recent developments, and applications* (S. 15–38). New York, NY: Springer. [https://doi.org/10.1007/978-1-4612-4230-7\\_2](https://doi.org/10.1007/978-1-4612-4230-7_2)
- Fischer, G. H. & Molenaar, I. W. (Hrsg.). (1995). *Rasch models. Foundations, recent developments, and applications*. New York, NY: Springer. <https://doi.org/10.1007/978-1-4612-4230-7>
- Fletcher, J., Maybery, M. T. & Bennett, S. (2000). Implicit learning differences. A question of developmental level? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26* (1), 246–252. <https://doi.org/10.1037//0278-7393.26.1.246>
- Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23* (1), 83–94.
- Gathercole, S. E., Alloway, T. P., Willis, C. S. & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93* (3), 265–281. <https://doi.org/10.1016/j.jecp.2005.08.003>
- Gathercole, S. E., Hitch, G. J., Service, E. & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33* (6), 966–979. <https://doi.org/10.1037//0012-1649.33.6.966>
- Gathercole, S. E., Pickering, S. J., Hall, M. & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A*, *54* (1), 1–30. <https://doi.org/10.1080/02724980042000002>
- Gebauer, G. F. & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33* (1), 34–54. <https://doi.org/10.1037/0278-7393.33.1.34>
- Gilbert, J. K., Compton, D. L. & Kearns, D. M. (2011). Word and person effects on decoding accuracy. A new look at an old question. *Journal of Educational Psychology*, *103* (2), 489–507. <https://doi.org/10.1037/a0023001>
- Giofrè, D., Carretti, B. & Belacchi, C. (2016). How semantic organisation influences primary school children's working memory. *Journal of Cognitive Psychology*, *29* (3), 327–336. <https://doi.org/10.1080/20445911.2016.1270950>

- Glidden, L. M. & Mar, H. H. (1978). Availability and accessibility of information in the semantic memory of retarded and nonretarded adolescents. *Journal of Experimental Child Psychology*, 25 (1), 33–40. [https://doi.org/10.1016/0022-0965\(78\)90036-X](https://doi.org/10.1016/0022-0965(78)90036-X)
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige. WWT 6-10*. München: Elsevier.
- Gold, A. (2011). *Lernschwierigkeiten*. Stuttgart: Kohlhammer.
- Gold, A. (2015). Lernschwierigkeiten. Wie man einen pädagogisch-psychologischen Dauerbrenner immer wieder aufs Neue befeuern kann. *Zeitschrift für Pädagogische Psychologie*, 29 (3-4), 123–132. <https://doi.org/10.1024/1010-0652/a000155>
- Goldhahn, D., Eckart, T. & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection. From 100 to 200 languages. *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 759–765.
- Golino, H. F. & Epskamp, S. (2017). Exploratory graph analysis. A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, 12 (6), e0174035. <https://doi.org/10.1371/journal.pone.0174035>
- Gorfein, D. S. & MacLeod, C. M. (Hrsg.). (2007). *Inhibition in cognition*. Washington: American Psychological Association. <https://doi.org/10.1037/11587-000>
- Goswami, U. & Bryant, P. (1989). The interpretation of studies using the reading level design. *Journal of Reading Behavior*, 21 (4), 413-424.
- Graves, W. W., Grabowski, T. J., Mehta, S. & Gordon, J. K. (2007). A neural signature of phonological access. Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience*, 19 (4), 617–631. <https://doi.org/10.1162/jocn.2007.19.4.617>
- Greenspan, S. (2017). Borderline intellectual functioning. An update. *Current Opinion in Psychiatry*, 30 (2), 113–122. <https://doi.org/10.1097/YCO.0000000000000317>
- Grosche, M., Bruns, G., Pulst, L., Grünke, M., Scanlon, D. & Sideridis, G. (2019, Juni). *Developing a coding system for a typology of "learning disability" definitions*. Posterpräsentation. 43rd IARLD Conference, Kreta, Griechenland.
- Grosche, M., Grünke, M., Scanlon, D. & Sideridis, G. (2018, Juli). *Towards an international understanding of learning disabilities. How are learning disabilities defined and operationalized in your country/culture?* Roundtable. 42nd IARLD Conference, Ghent, Belgien.
- Grube, D., Lingen, M. & Hasselhorn, M. (2008). Entwicklung des phonologischen Arbeitsgedächtnisses. Zur Rolle von Rehearsal und Lexikalität für den Ähnlichkeitseffekt. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40 (4), 200–207. <https://doi.org/10.1026/0049-8637.40.4.200>
- Grünke, M. & Cavendish, W. M. (2016). Learning disabilities around the globe. Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal*, 14 (1), 1–8.

- Grünke, M. & Grosche, M. (2014). Lernbehinderung. In G. W. Lauth, M. Grünke & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (2. Aufl., S. 76–89). Göttingen: Hogrefe.
- Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M. & Neath, I. (2018). Word length, set size, and lexical factors. Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44* (11), 1824–1844.
- Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory*, *13* (3-4), 318–324. <https://doi.org/10.1080/09658210344000350>
- Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall and nonword learning. In A. S. C. Thorn & M. P. A. Page (Hrsg.), *Interactions between short-term and long-term memory in the verbal domain* (S. 108–135). Hove: Psychology Press.
- Haeberlin, U. (1991). Die Integration von leistungsschwachen Schülern. Ein Überblick über empirische Forschungsergebnisse zu Wirkungen von Regelklassen, Integrationsklassen und Sonderklassen auf "Lernbehinderte". *Zeitschrift für Pädagogik*, *37* (2), 167–189.
- Hancock, R., Pugh, K. R. & Hoeft, F. (2017). Neural noise hypothesis of developmental dyslexia. *Trends in Cognitive Sciences*, *21* (6), 434–448. <https://doi.org/10.1016/j.tics.2017.03.008>
- Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M. et al. (2016). How concepts are encoded in the human brain. A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage*, *135*, 232–242. <https://doi.org/10.1016/j.neuroimage.2016.04.063>
- Hartley, R. T. & Barnden, J. A. (1997). Semantic networks. Visualizations of knowledge. *Trends in Cognitive Sciences*, *1* (5), 169–175.
- Hasenäcker, J., Schröter, P. & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43* (7), 1093–1108. <https://doi.org/10.1037/xlm0000353>
- Hashimoto, N., McGregor, K. K. & Graham, A. (2007). Conceptual organization at 6 and 8 years of age. Evidence from the semantic priming of object decisions. *Journal of Speech, Language, and Hearing Research*, *50* (1), 161–176. [https://doi.org/10.1044/1092-4388\(2007/014\)](https://doi.org/10.1044/1092-4388(2007/014))
- Hasselhorn, M. (1987). Kognitive Bedingungen der Leistungsdefizite lernschwacher Schüler bei Gedächtnisanforderungen. *Zeitschrift für Pädagogische Psychologie*, *1* (2), 91–98.
- Hasselhorn, M. (2005). Lernen im Altersbereich zwischen 4 und 8 Jahren. Individuelle Voraussetzungen, Entwicklung, Diagnostik und Förderung. In T. Guldemann & B. Hauser (Hrsg.), *Bildung 4- bis 8-jähriger Kinder*. Münster: Waxmann.
- Hasselhorn, M. & Gold, A. (2013). *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren* (3. Aufl.). Stuttgart: Kohlhammer.
- Hasselhorn, M., Grube, D. & Mähler, C. (2000). Theoretisches Rahmenmodell für ein Diagnostikum zur differentiellen Funktionsanalyse des phonologischen Arbeitsgedächtnisses. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Lese- Rechtschreibschwierigkeiten*

- (Jahrbuch der pädagogisch-psychologischen Diagnostik. Tests und Trends, Bd. 1, S. 167–181). Göttingen: Hogrefe.
- Hasselhorn, M. & Mähler, C. (2007). Phonological working memory of children in two German special schools. *International Journal of Disability, Development and Education*, 54 (2), 225–244. <https://doi.org/10.1080/10349120701330545>
- Hasselhorn, M., Schuchardt, K. & Mähler, C. (2010). Phonologisches Arbeitsgedächtnis bei Kindern mit diagnostizierter Lese- und/oder Rechtschreibstörung. Zum Einfluss von Wortlänge und Lexikalität auf die Gedächtnisspanne. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42 (4), 211–216. <https://doi.org/10.1026/0049-8637/a000024>
- Hasselhorn, M., Schumann-Hengstler, R., Gronauer, J., Grube, D., Mähler, C., Schmid, I. et al. (2012). *Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren*. (AGTB 5-12). Göttingen: Hogrefe.
- Hassiotis, A. (2015). Borderline intellectual functioning and neurodevelopmental disorders. Prevalence, comorbidities and treatment approaches. *Advances in Mental Health and Intellectual Disabilities*, 9 (5), 275–283. <https://doi.org/10.1108/AMHID-06-2015-0028>
- Heimlich, U. (2009). *Lernschwierigkeiten. Sonderpädagogische Förderung im Förderschwerpunkt Lernen* (UTB, Bd. 3192). Bad Heilbrunn: Julius Klinkhardt.
- Heimlich, U. & Wember, F. B. (2014). Lernschwierigkeiten, Lernbeeinträchtigung, Lernbehinderung. In F. B. Wember, R. Stein & U. Heimlich (Hrsg.), *Handlexikon Lernschwierigkeiten und Verhaltensstörungen* (51-54). Stuttgart: Kohlhammer.
- Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, 9 (4-6), 233–247. <https://doi.org/10.1080/09658210042000085>
- Henry, L. A. (2010). The episodic buffer in children with intellectual disabilities. An exploratory study. *Research in Developmental Disabilities*, 31 (6), 1609–1614. <https://doi.org/10.1016/j.ridd.2010.04.025>
- Henry, L. A. & MacLean, M. (2002). Working memory performance in children with and without intellectual disabilities. *American Journal on Mental Retardation*, 107 (6), 421–432. [https://doi.org/10.1352/0895-8017\(2002\)107<0421:WMPICW>2.0.CO;2](https://doi.org/10.1352/0895-8017(2002)107<0421:WMPICW>2.0.CO;2)
- Henry, L. A. & MacLean, M. (2003). Relationships between working memory, expressive vocabulary and arithmetical reasoning in children with and without intellectual disabilities. *Educational and Child Psychology*, 20 (3), 51–64.
- Henry, L. A. & Winfield, J. (2010). Working memory and educational achievement in children with intellectual disabilities. *Journal of Intellectual Disability Research*, 54 (4), 354–365. <https://doi.org/10.1111/j.1365-2788.2010.01264.x>
- Hills, T. T., Maouene, M., Maouene, J., Sheya, A. & Smith, L. (2009). Longitudinal analysis of early semantic networks. Preferential attachment or preferential acquisition? *Psychological Science*, 20 (6), 729–739.

- Hirschfeld, G. & Brachel, R. von (2014). Multiple-group confirmatory factor analysis in R. A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation*, 19 (7). Verfügbar unter <http://pare-online.net/getvn.asp?v=19&n=7>
- Holmes, J., Gathercole, S. E. & Dunning, D. L. (2010). Poor working memory. Impact and interventions. In J. Holmes (Ed.), *Developmental disorders and interventions* (Advances in child development and behavior, vol. 39, vol. 39, pp. 1–43). London: Academic Press.
- Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D.A., Martin, S. & Stuart, G. (1997). Word-frequency effects on short-term memory tasks. Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (5), 1217–1232.
- Ibernon, L., Touchet, C. & Pochon, R. (2018). Emotion recognition as a real strength in Williams syndrome. Evidence from a dynamic non-verbal task. *Frontiers in Psychology*, 9, 463. <https://doi.org/10.3389/fpsyg.2018.00463>
- Ise, E. & Schulte-Körne, G. (2012). Implizites Lernen und LRS. Spielen Defizite im impliziten Lernen eine Rolle bei der Entstehung von Schwierigkeiten im Lesen und Rechtschreiben? *Lernen und Lernstörungen*, 1 (2), 79–97. <https://doi.org/10.1024/2235-0977/a000011>
- Isiordia, M. & Ferrer, E. (2016). Curve of factors model. A latent growth modeling approach for educational research. *Educational and Psychological Measurement*, 78 (2), 203–231. <https://doi.org/10.1177/0013164416677143>
- Ivanova, M. V. & Hallowell, B. (2012). Validity of an eye-tracking method to index working memory in people with and without aphasia. *Aphasiology*, 26 (3-4), 556–578. <https://doi.org/10.1080/02687038.2011.618219>
- Jackson, E., Leitao, S. & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language & Communication Disorders*, 51 (1), 61–73. <https://doi.org/10.1111/1460-6984.12185>
- Jackson, N. E. & Butterfield, E. C. (1989). Reading-level-match designs. Myths and realities. *Journal of Reading Behavior*, 21 (4), 387–412.
- Jacquemot, C., Dupoux, E. & Bachoud-Lévi, A.-C. (2011). Is the word-length effect linked to subvocal rehearsal? *Cortex*, 47 (4), 484–493. <https://doi.org/10.1016/j.cortex.2010.07.007>
- Jalbert, A., Neath, I., Bireta, T. J. & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37 (2), 338–353. <https://doi.org/10.1037/a0021804>

- Jalbert, A., Neath, I. & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, *39* (7), 1198–1210. <https://doi.org/10.3758/s13421-011-0094-z>
- Janse, E. & Newman, R. S. (2013). Identifying nonwords. Effects of lexical neighborhoods, phonotactic probability, and listener characteristics. *Language and Speech*, *56* (4), 421–441. <https://doi.org/10.1177/0023830912447914>
- Jarrold, C., Danielsson, H. & Wang, X. (2015). Absolute and proportional measures of potential markers of rehearsal, and their implications for accounts of its development. *Frontiers in Psychology*, *6*, 299. <https://doi.org/10.3389/fpsyg.2015.00299>
- Johnson, M. H. (2011). Interactive specialization. A domain-general framework for human functional brain development? *Developmental Cognitive Neuroscience*, *1* (1), 7–21. <https://doi.org/10.1016/j.dcn.2010.07.003>
- Jones, G. & Witherstone, H. L. (2011). Lexical and sublexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory & Cognition*, *39* (4), 588–599. <https://doi.org/10.3758/s13421-010-0045-0>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M. & Rosseel, Y. (2020). semTools. Useful tools for structural equation modeling (Version 0.5-3.) [Computer software].
- Kaiser, M. (2017). Mechanisms of connectome development. *Trends in Cognitive Sciences*, *21* (9), 703–717. <https://doi.org/10.1016/j.tics.2017.05.010>
- Kanter, G. O. (1980). Lernbehindertenpädagogik. Gegenstandsbestimmung, Begriffsklärung. In G. O. Kanter & O. Speck (Hrsg.), *Pädagogik der Lernbehinderten* (Handbuch der Sonderpädagogik, Bd. 4, 2. Aufl., S. 7–33). Berlin: Marhold.
- Kanter, G. O. (2007). Gegenstand und Aufgaben einer Pädagogik und Psychologie bei Beeinträchtigungen des Lernens. In J. Walter & F. B. Wember (Hrsg.), *Sonderpädagogik des Lernens* (Handbuch Sonderpädagogik, Bd. 2, S. 33–59). Göttingen: Hogrefe.
- Kanter, G. O. (2007). Reizwort "Lernbehinderung". In K. Salzberg-Ludwig & E. Grüning (Hrsg.), *Pädagogik für Kinder- und Jugendliche in schwierigen Lern- und Lebenssituationen* (Heil- und Sonderpädagogik, S. 15–26). Stuttgart: Kohlhammer.
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2* (10), 389–398. [https://doi.org/10.1016/S1364-6613\(98\)01230-3](https://doi.org/10.1016/S1364-6613(98)01230-3)
- Karmiloff-Smith, A., Thomas, M. S. C., Annaz, D., Humphreys, K., Ewing, S., Brace, N. et al. (2004). Exploring the Williams syndrome face-processing debate. The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry*, *45* (7), 1258–1274. <https://doi.org/10.1111/j.1469-7610.2004.00322.x>
- Kassambara, A. (2017). ggpubr. 'ggplot2' based publication ready plots (Version 0.1.6) [Computer software]. Verfügbar unter <https://CRAN.R-project.org/package=ggpubr>

- Kauschke, C. (2003). Entwicklung, Störungen und Diagnostik lexikalischer Prozesse - Wortverständnis und Wortproduktion. *Sprache Stimme Gehör*, 27 (3), 110–118. <https://doi.org/10.1055/s-2003-42535>
- Kenett, Y. N., Anaki, D. & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience*, 8, 407. <https://doi.org/10.3389/fnhum.2014.00407>
- Kenett, Y. N., Beaty, R. E., Silvia, P. J., Anaki, D. & Faust, M. (2016). Structure and flexibility. Investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement. *Psychology of Aesthetics, Creativity, and the Arts*, 10 (4), 377–388. <https://doi.org/10.1037/aca0000056>
- Kenett, Y. N., Gold, R. & Faust, M. (2016). The hyper-modular associative mind. A computational analysis of associative responses of persons with Asperger syndrome. *Language and Speech*, 59 (3), 297–317. <https://doi.org/10.1177/0023830915589397>
- Keuleers, E. vwr. Useful functions for visual word recognition research (Version 0.3.0) [Computer software]. Verfügbar unter <https://CRAN.R-project.org/package=vwr>
- Klauer, K. J. (1977). *Lernbehindertenpädagogik* (5. Aufl.). Berlin: Marhold.
- Klauer, K. J. (1980). Erkenntnismethoden der Lernbehindertenpädagogik. In G. O. Kanter & O. Speck (Hrsg.), *Pädagogik der Lernbehinderten* (Handbuch der Sonderpädagogik, Bd. 4, 2. Aufl., S. 76–89). Berlin: Marhold.
- Klauer, K. J. & Lauth, G. W. (1997). Lernbehinderungen und Leistungsschwierigkeiten bei Schülern. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. (Themenbereich D, Serie I Pädagogische Psychologie, Psychologie des Unterrichts und der Schule)* (S. 701–738). Göttingen: Hogrefe.
- Klauer, K. J. & Leutner, D. (2012). *Lehren und Lernen. Einführung in die Instruktionspsychologie* (2. Aufl.). Weinheim: Beltz.
- Kleber, E. W. (1980). *Grundkonzeption einer Lernbehindertenpädagogik*. München: E. Reinhardt.
- Kleinbaum, D. G. & Klein, M. (2002). *Logistic regression. A self-learning text* (Statistics for biology and health, 2nd ed.). New York, NY: Springer.
- Kluszczewski, J., Brandenburg, J., Fischbach, A., Grube, D., Hasselhorn, M. & Büttner, G. (2015). Working memory functioning in children with poor mathematical skills. Relationships to IQ-achievement discrepancy and additional reading and spelling difficulties. *Zeitschrift für Psychologie*, 223 (2), 83–92. <https://doi.org/10.1027/2151-2604/a000206>
- Kluszczewski, J., Brandenburg, J., Fischbach, A., Schuchardt, K., Grube, D., Hasselhorn, M. et al. (2018). Development of working memory from grade 3 to 5. Differences between children with and without mathematical learning difficulties. *International Journal of Disability, Development and Education*, 20, 1–17. <https://doi.org/10.1080/1034912X.2017.1419555>
- Kliegl, R., Grabner, E., Rolfs, M. & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16 (1-2), 262–284. <https://doi.org/10.1080/09541440340000213>

- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16 (12), 606–617. <https://doi.org/10.1016/j.tics.2012.10.007>
- Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis Eine Einführung in eRm. Eine Einführung in eRm* (UTB Psychologie, Bd. 3786). Wien: Facultas.wuv.
- Koller, I. & Hatzinger, R. (2013). Nonparametric tests for the Rasch model. Explanation, development, and application of quasi-exact tests for small samples. *InterStat* (11). Verfügbar unter <http://interstat.statjournals.net/YEAR/2013/abstracts/1311002.php>
- Koller, I., Maier, M. J. & Hatzinger, R. (2015). An empirical power analysis of quasi-exact tests for the Rasch model. *Methodology*, 11 (2), 45–54. <https://doi.org/10.1027/1614-2241/a000090>
- Kölm, J., Gresch, C. & Kuhl, P. (2019). Zuwanderungsbezogene Disparitäten bei der Diagnose eines sonderpädagogischen Förderbedarfs Lernen und der besuchten Schulart. *Zeitschrift für Erziehungswissenschaft*, 22 (4), 771–789. <https://doi.org/10.1007/s11618-019-00896-y>
- Koßmann, R. (2019). *Schule und „Lernbehinderung“. Wechselseitige Erschließungen*. Bad Heilbrunn: Klinkhardt.
- Kraut, M. A., Pitcock, J. & Hart, J. (2004). Neural mechanisms of semantic memory. *Current Neurology and Neuroscience Reports*, 4 (6), 461–465. <https://doi.org/10.1007/s11910-004-0069-6>
- Kretschmann, R. (2007). Lernschwierigkeiten, Lernstörung und Lernbehinderung. In J. Walter & F. B. Wember (Hrsg.), *Sonderpädagogik des Lernens* (Handbuch Sonderpädagogik, Bd. 2, S. 4–32). Göttingen: Hogrefe.
- Kretschmar, F., Schlesewsky, M. & Staub, A. (2015). Dissociating word frequency and predictability effects in reading. Evidence from coregistration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41 (6), 1648–1662. <https://doi.org/10.1037/xlm0000128>
- Kuhn, J.-T., Raddatz, J., Holling, H. & Dobel, C. (2013). Dyskalkulie vs. Rechenschwäche. Basisnumerische Verarbeitung in der Grundschule. *Lernen und Lernstörungen*, 2 (4), 229–247. <https://doi.org/10.1024/2235-0977/a000044>
- Kultusministerkonferenz. (1960). Gutachten zur Ordnung des Sonderschulwesens.
- Kultusministerkonferenz. (1972). Empfehlung zur Ordnung des Sonderschulwesens.
- Kultusministerkonferenz. (1999). Empfehlungen zum Förderschwerpunkt Lernen. Beschluss der Kultusministerkonferenz vom 01.10.1999.
- Kultusministerkonferenz (2016). *Sonderpädagogische Förderung in Schulen 2005 bis 2014. Statistische Veröffentlichungen der Kultusministerkonferenz*. (Dokumentation Nr. 210 - Februar 2016).
- Kultusministerkonferenz. (2019). Empfehlungen zur schulischen Bildung, Beratung und Unterstützung von Kindern und Jugendlichen im sonderpädagogischen Schwerpunkt LERNEN.
- Kultusministerkonferenz (2020). *Sonderpädagogische Förderung in Schulen 2009 bis 2018. Statistische Veröffentlichungen der Kultusministerkonferenz*. (Dokumentation Nr. 223 - Februar 2020).



- Künkler, T. (2008). >Lernen im Zwischen<. Zum Zusammenhang von Lerntheorien, Subjekt-konzeptionen und dem Vollzug des Lernens. In K. Mitgutsch, E. Sattler, K. Westphal & I. M. Breinbauer (Hrsg.), *Dem Lernen auf der Spur. Die pädagogische Perspektive* (S. 33–50). Stuttgart: Klett-Cotta.
- Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package. Tests in linear mixed effects models. *Journal of Statistical Software*, *82* (13). <https://doi.org/10.18637/jss.v082.i13>
- Lance, C. E., Butts, M. M. & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods*, *9* (2), 202–220. <https://doi.org/10.1177/1094428105284919>
- Lauth, G. W., Brunstein, J. C. & Grünke, M. (2014). Lernstörungen im Überblick. Arten, Klassifikation, Verbreitung und Erklärungsperspektiven. In G. W. Lauth, M. Grünke & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (2. Aufl., S. 17–31). Göttingen: Hogrefe.
- Lauth, G. W., Hammes-Schmitz, E. & Lebens, M. (2014). Eine empirische Bedingungsanalyse von Lernstörungen. *Empirische Sonderpädagogik*, *6* (4), 350–364.
- Lauth, G. W. & Schlottke, P. F. (2005). Lernbehinderte Kinder und Jugendliche. In P. F. Schlottke (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich D: Praxisgebiete: Störungen im Kindes- und Jugendalter: Verhaltensauffälligkeiten* (S. 327–348). Göttingen: Hogrefe.
- Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y. et al. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *Journal of Child Psychology and Psychiatry*, *52* (2), 212–220. <https://doi.org/10.1111/j.1469-7610.2010.02311.x>
- Levy, Y. (2018). 'Developmental Delay' reconsidered. The critical role of age-dependent, co-variant development. *Frontiers in Psychology*, *9*, 503. <https://doi.org/10.3389/fpsyg.2018.00503>
- Lewandowsky, S. & Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research*, *63* (2), 163–173. <https://doi.org/10.1007/PL00008175>
- Lewandowsky, S., Yang, L.-X., Newell, B. R. & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38* (4), 881–904. <https://doi.org/10.1037/a0027298>
- Lifshitz, H., Kilberg, E. & Vakil, E. (2016). Working memory studies among individuals with intellectual disability. An integrative research review. *Research in Developmental Disabilities*, *59*, 147–165. <https://doi.org/10.1016/j.ridd.2016.08.001>
- Lilienthal, L., Rose, N. S., Tamez, E., Myerson, J. & Hale, S. (2015). Individuals with low working memory spans show greater interference from irrelevant information because of poor source monitoring, not greater activation. *Memory & Cognition*, *43* (3), 357–366. <https://doi.org/10.3758/s13421-014-0465-3>

- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel. Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9 (2), 151–173. [https://doi.org/10.1207/S15328007SEM0902\\_1](https://doi.org/10.1207/S15328007SEM0902_1)
- Little, T. D., Rhemtulla, M. & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, 18 (3), 285–300.
- Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words. The neighborhood activation model. *Ear and Hearing*, 19 (1), 1–36.
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4 (1), 84–99.
- Mähler, C. (2007). Arbeitsgedächtnisfunktionen bei lernbehinderten Kindern und Jugendlichen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39 (2), 97–106. <https://doi.org/10.1026/0049-8637.39.2.97>
- Mähler, C. & Hasselhorn, M. (1990). Gedächtnisdefizite bei lernbehinderten Kindern. Entwicklungsverzögerung oder Strukturdifferenz? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22 (4), 354–366.
- Mähler, C. & Hasselhorn, M. (2003). Automatische Aktivierung des Rehearsalprozesses im phonologischen Arbeitsgedächtnis bei lernbehinderten Kindern und Erwachsenen. *Zeitschrift für Pädagogische Psychologie*, 17 (3/4), 255–260. <https://doi.org/10.1024//1010-0652.17.3.255>
- Mähler, C., Jörns, C., Radtke, E. & Schuchardt, K. (2015). Chancen und Grenzen eines Trainings des Arbeitsgedächtnisses bei Kindern mit und ohne Lese-/Rechtschreibschwierigkeiten. *Zeitschrift für Erziehungswissenschaft*, 18 (3), 453–471. <https://doi.org/10.1007/s11618-015-0643-5>
- Mähler, C., Jörns, C. & Schuchardt, K. (2019). Training working memory of children with and without dyslexia. *Children*, 6 (47). <https://doi.org/10.3390/children6030047>
- Mähler, C. & Schuchardt, K. (2009). Working memory functioning in children with learning disabilities. Does intelligence make a difference? *Journal of Intellectual Disability Research*, 53 (1), 3–10. <https://doi.org/10.1111/j.1365-2788.2008.01105.x>
- Mähler, C. & Schuchardt, K. (2016). The importance of working memory for school achievement in primary school children with intellectual or learning disabilities. *Research in Developmental Disabilities*, 58, 1–8. <https://doi.org/10.1016/j.ridd.2016.08.007>
- Mair, P. & Hatzinger, R. (2007). CML based estimation of extended Rasch models with the eRm package in R. *Psychology Science*, 49 (1), 26–43.
- Mair, P., Hatzinger, R. & Maier, M. J. (2015). Package 'eRm'. Extended Rasch modeling [Computer software]. Verfügbar unter <https://mran.microsoft.com/snapshot/2016-10-12/web/packages/eRm/eRm.pdf>
- Majerus, S. (2009). Verbal short-term memory and temporary activation of language representation. The importance of distinguishing item and order information. In A. S. C. Thorn & M. P. A. Page (Hrsg.), *Interactions between short-term and long-term memory in the verbal domain* (S. 244–276). Hove: Psychology Press.

- Mamen, M., Ferguson, H. B. & Backman, J. E. (1986). No difference represents a significant finding. The logic of the reading level design. A response to Bryant and Goswami. *Psychological Bulletin*, 100 (1), 104–106.
- Marton, K. & Eichorn, N. (2014). Interaction between working memory and long-term memory. A study in children with and without language impairment. *Zeitschrift für Psychologie*, 222 (2), 90–99. <https://doi.org/10.1027/2151-2604/a000170>
- Massara, G. P., Matteo, T. di & Aste, T. (2016). Network filtering for big data. Triangulated maximally filtered graph. *Journal of Complex Networks*, 5 (2), 161–178. <https://doi.org/10.1093/comnet/cnw015>
- Masuda, N., Porter, M. A. & Lambiotte, R. (2017). Random walks and diffusion on networks. *Physics Reports*, 716–717, 1–58. <https://doi.org/10.1016/j.physrep.2017.07.007>
- McClelland, J. L. & Rumelhart, D. E. (1981). Interactive activation model of context effects in letter perception. Part 1. An account of basic findings. *Psychological Review*, 88 (5), 375–407.
- Meade, A. W. & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods*, 9 (3), 369–403. <https://doi.org/10.1177/1094428105283384>
- Meibauer, J., Demske, U., Geilfuß-Wolfgang, J., Pafel, J., Ramers, K. H., Rothweiler, M. et al. (Hrsg.). (2015). *Einführung in die germanistische Linguistik* (3. Aufl.). Stuttgart: J. B. Metzler.
- Melby-Lervåg, M. & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49 (2), 270–291. <https://doi.org/10.1037/a0028228>
- Melby-Lervåg, M. & Hulme, C. (2016). There is no convincing evidence that working memory training is effective. A reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychonomic Bulletin & Review*, 23 (1), 324–330. <https://doi.org/10.3758/s13423-015-0862-z>
- Meltzer, J. A., Rose, N. S., Deschamps, T., Leigh, R. C., Panamsky, L., Silberberg, A. et al. (2016). Semantic and phonological contributions to short-term repetition and long-term cued sentence recall. *Memory & Cognition*, 44 (2), 307–329. <https://doi.org/10.3758/s13421-015-0554-y>
- Merrill, E. C., Sperber, R., McCauly, C., Littlefield, J., Rider, E. A. & Shapiro, D. (1987). Picture encoding speed and mental retardation. *Intelligence*, 11, 169–191.
- Mietzel, G. (2017). *Pädagogische Psychologie des Lernens und Lehrens* (9. Aufl.). Göttingen: Hogrefe.
- Milfont, T. L. & Fischer, R. (2010). Testing measurement invariance across groups. Applications in cross-cultural research. *International Journal of Psychological Research*, 3 (1), 111–121.
- Miller, L. M. & Roodenrys, S. (2012). Serial recall, word frequency, and mixed lists. The influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38 (6), 1731–1740. <https://doi.org/10.1037/a0028470>
- Ministerium für Schule und Weiterbildung NRW. (2016). Ausbildungsordnung sonderpädagogische Förderung. AO-SF. Verfügbar unter <https://bass.schul-welt.de/6225.htm>
- Mirman, D., Landrigan, J.-F. & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, 143 (5), 499–520. <https://doi.org/10.1037/bul0000092>

- Molen, M. J. van der (2010). Working memory structure in 10- and 15-year old children with mild to borderline intellectual disabilities. *Research in Developmental Disabilities, 31* (6), 1258–1263. <https://doi.org/10.1016/j.ridd.2010.07.019>
- Molen, M. J. van der, Henry, L. A. & Luit, J. E. H. van (2014). Working memory development in children with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 58* (7), 637–650. <https://doi.org/10.1111/jir.12061>
- Molen, M. J. van der, Luit, J. E. H. van, Jongmans, M. J. & Molen, M. W. van der (2007). Verbal working memory in children with mild intellectual disabilities. *Journal of Intellectual Disability Research, 51* (2), 162–169. <https://doi.org/10.1111/j.1365-2788.2006.00863.x>
- Molen, M. J. van der, Luit, J. E. H. van, Jongmans, M. J. & Molen, M. W. van der (2009). Memory profiles in children with mild intellectual disabilities. Strengths and weaknesses. *Research in Developmental Disabilities, 30* (6), 1237–1247. <https://doi.org/10.1016/j.ridd.2009.04.005>
- Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Hrsg.), *Rasch models. Foundations, recent developments, and applications* (S. 39–51). New York, NY: Springer. [https://doi.org/10.1007/978-1-4612-4230-7\\_3](https://doi.org/10.1007/978-1-4612-4230-7_3)
- Monnier, C. & Syssau, A. (2008). Semantic contribution to verbal short-term memory. Are pleasant words easier to remember than neutral words in serial recall and serial recognition? *Memory & Cognition, 36* (1), 35–42. <https://doi.org/10.3758/MC.36.1.35>
- Moosbrugger, H. (2012). Item-Response-Theory (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., 228-274). Berlin: Springer.
- Morek, M. & Heller, V. (2012). Bildungssprache – Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik, 57* (1). <https://doi.org/10.1515/zfal-2012-0011>
- Mostofsky, S. H. & Simmonds, D. J. (2008). Response inhibition and response selection. Two sides of the same coin. *Journal of Cognitive Neuroscience, 20* (5), 751–761.
- Norris, D. G., Hall, J. & Gathercole, S. E. (2019). Can short-term memory be trained? *Memory & Cognition, 47* (5), 1012–1023. <https://doi.org/10.3758/s13421-019-00901-z>
- Oberauer, K., Awh, E. & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory. Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43* (1), 1–22. <https://doi.org/10.1037/xlm0000302>
- Oldfield, R. C. & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology, 17* (4), 273–281. <https://doi.org/10.1080/17470216508416445>
- O'Shaughnessy, T. E. & Swanson, H. L. (1998). Do immediate memory deficits in students with learning disabilities in reading reflect a developmental lag or deficit? A selective meta-analysis of the literature. *Learning Disability Quarterly, 21* (1), 123–148.

- Passolunghi, M. C., Cornoldi, C. & Liberto, S. de (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Memory & Cognition*, 27 (5), 779–790. <https://doi.org/10.3758/BF03198531>
- Peereman, R. & Content, A. (1995). Neighborhood size effect in naming. Lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (2), 409–421.
- Peng, C.-Y. J., Lee, K. L. & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96 (1), 3–14.
- Peng, P. & Fuchs, D. (2016). A meta-analysis of working memory deficits in children with learning difficulties. Is there a difference between verbal domain and numerical domain? *Journal of Learning Disabilities*, 49 (1), 3–20. <https://doi.org/10.1177/0022219414521667>
- Peng, P. & Fuchs, D. (2017). A randomized control trial of working memory training with and without strategy instruction. Effects on young children's working memory and comprehension. *Journal of Learning Disabilities*, 50 (1), 62–80. <https://doi.org/10.1177/0022219415594609>
- Peugh, J. L. & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, 65 (5), 717–741. <https://doi.org/10.1177/0013164405278558>
- Pickering, S. J. & Gathercole, S. E. (2004). Distinctive working memory profiles in children with special educational needs. *Educational Psychology*, 24 (3), 393–408. <https://doi.org/10.1080/0144341042000211715>
- Pinquardt, M. & Silbereisen, R. K. (2007). Biopsychosoziale Erklärungsansätze. In F. Linderkamp & M. Grünke (Hrsg.), *Lern- und Verhaltensstörungen. Genese, Diagnostik, Intervention* (S. 30–43). Weinheim: Beltz.
- Pochon, R., Touchet, C. & Ibernou, L. (2017). Emotion recognition in adolescents with Down syndrome. A nonverbal approach. *Brain Sciences*, 7 (6). <https://doi.org/10.3390/brainsci7060055>
- Poloczek, S., Büttner, G. & Hasselhorn, M. (2012). Relationships between working memory and academic skills. Are there differences between children with intellectual disabilities and typically developing children? *Journal of Cognitive Education and Psychology*, 11 (1), 20–38. <https://doi.org/10.1891/1945-8959.11.1.20>
- Poloczek, S., Büttner, G. & Hasselhorn, M. (2014). Phonological short-term memory impairment and the word length effect in children with intellectual disabilities. *Research in Developmental Disabilities*, 35 (2), 455–462. <https://doi.org/10.1016/j.ridd.2013.11.025>
- Poloczek, S., Henry, L. A., Danielsson, H., Büttner, G., Mähler, C., Messer, D. J. et al. (2016). Strategic verbal rehearsal in adolescents with mild intellectual disabilities. A multi-centre European study. *Research in Developmental Disabilities*, 58, 83–93. <https://doi.org/10.1016/j.ridd.2016.08.014>

- Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the rasch model. *Psychometrika*, *66* (3), 437–459. <https://doi.org/10.1007/BF02294444>
- Powell, J. J.W. & Wagner, Sandra, J. (2014). An der Schnittstelle Ethnie und Behinderung benachteiligt. Jugendliche mit Migrationshintergrund an deutschen Sonderschulen weiterhin überrepräsentiert. In G. Wansing & M. Westphal (Hrsg.), *Behinderung und Migration. Inklusion, Diversität, Intersektionalität* (S. 177–199). Wiesbaden: Springer.
- Pressley, M., Borkowski, J. G. & Schneider, W. (1989). Good information processing. What it is and how education can promote it. *International Journal of Educational Research*, *13* (8), 857–867. [https://doi.org/10.1016/0883-0355\(89\)90069-4](https://doi.org/10.1016/0883-0355(89)90069-4)
- Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L. & Woldorff, M. G. (2013). *Principles of cognitive neuroscience* (2. ed.). Sunderland, Mass.: Sinauer Associates.
- Putnick, D. L. & Bornstein, M. H. (2016). Measurement invariance conventions and reporting. The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- R Core Team. (2018). R. A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Verfügbar unter <https://www.R-project.org>
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J. et al. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28* (3), 164–171. <https://doi.org/10.1027/1015-5759/a000123>
- Renkl, A. (2015). Wissenserwerb. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Springer-Lehrbuch, S. 3–24). Berlin: Springer. [https://doi.org/10.1007/978-3-642-41291-2\\_1](https://doi.org/10.1007/978-3-642-41291-2_1)
- Richter, T. & Naumann, J. (2002). Mehrebenenanalysen mit hierarchisch-linearen Modellen. *Zeitschrift für Medienpsychologie*, *14* (4), 155–159.
- Ritchie, G., Tolan, G. A. & Tehan, G. (2015). Redintegration, task difficulty, and immediate serial recall tasks. *Canadian Journal of Experimental Psychology*, *69* (1), 54–63. <https://doi.org/10.1037/cep0000031>
- Röhm, A. (2020). Sprache, Arbeitsgedächtnis und mathematische Kompetenz von Schulkindern mit SES. *Lernen und Lernstörungen*, *9* (2), 1–12. <https://doi.org/10.1024/2235-0977/a000292>
- Röhm, A., Starke, A. & Ritterfeld, U. (2017). Die Rolle von Arbeitsgedächtnis und Sprachkompetenz für den Erwerb mathematischer Basiskompetenzen im Vorschulalter. *Psychologie in Erziehung und Unterricht*, *64* (2), 1–13. <https://doi.org/10.2378/peu2016.art26d>
- Roodenrys, S. & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, *8* (2), 71–78. <https://doi.org/10.1080/096582100387623>

- Roodenrys, S. (2009). Explaining phonological neighbourhood effects in short-term memory. In A. S. C. Thorn & M. P. A. Page (Hrsg.), *Interactions between short-term and long-term memory in the verbal domain* (S. 177–197). Hove: Psychology Press.
- Roodenrys, S. & Dunn, N. (2008). Unimpaired implicit learning in children with developmental dyslexia. *Dyslexia*, *14* (1), 1–15. <https://doi.org/10.1002/dys.340>
- Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M. & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28* (6), 1019–1034. <https://doi.org/10.1037//0278-7393.28.6.1019>
- Roodenrys, S. & Miller, L. M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Memory & Cognition*, *36* (3), 578–587. <https://doi.org/10.3758/MC.36.3.578>
- Rosenquist, C., Conners, F. A. & Roskos-Ewoldsen, B. (2003). Phonological and visuo-spatial working memory in individuals with intellectual disability. *American Journal on Mental Retardation*, *108* (6), 403–413. [https://doi.org/10.1352/0895-8017\(2003\)108<403:PAVWMI>2.0.CO;2](https://doi.org/10.1352/0895-8017(2003)108<403:PAVWMI>2.0.CO;2)
- Rosseel, Y. (2012). lavaan. An R package for structural equation modeling. *Journal of Statistical Software*, *48* (2), 1–36.
- Rothweiler, M. & Kauschke, C. (2007). Lexikalischer Erwerb. In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache* (Handbuch Sonderpädagogik, Bd. 1, S. 42–57). Göttingen: Hogrefe.
- RStudio Team. (2016). RStudio. Integrated development environment for R [Computer software]. Boston, MA: RStudio, Inc. Verfügbar unter <http://www.rstudio.com/>
- Rubenstein, J. L. R. & Merzenich, M. M. (2003). Model of autism. Increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2* (5), 255–267.
- Sachs, O., Weis, S., Zellagui, N., Sass, K., Huber, W., Zvyagintsev, M. et al. (2011). How different types of conceptual relations modulate brain activation during semantic priming. *Journal of Cognitive Neuroscience*, *23* (5), 1263–1273. <https://doi.org/10.1162/jocn.2010.21483>
- Sala, G. & Gobet, F. (2017). Working memory training in typically developing children. A meta-analysis of the available evidence. *Developmental Psychology*, *53* (4), 671–685. <https://doi.org/10.1037/dev0000265>
- Sala, G. & Gobet, F. (2020). Working memory training in typically developing children. A multilevel meta-analysis. *Psychonomic Bulletin & Review*, *27* (3), 423–434. <https://doi.org/10.3758/s13423-019-01681-y>
- Schermer, F. J. (2013). *Lernen und Gedächtnis* (5. Aufl.). Stuttgart: Kohlhammer.
- Schmetz, D. (2004). Förderschwerpunkt Lernen. *Zeitschrift für Heilpädagogik* (3), 113–128.
- Schmidt, C., Brandenburg, J., Busch, J., Büttner, G., Grube, D., Mähler, C. et al. (2020). Developmental trajectories of phonological information processing in upper elementary students with reading or spelling disabilities. *Reading Research Quarterly*, *1* (4), 134. <https://doi.org/10.1002/rrq.299>

- Schneider, W., Körkel, J. & Weinert, F. E. (1989). Domain-specific knowledge and memory performance. A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, 81 (3), 306–312.
- Schnotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Bild- und Textverstehen. *Zeitschrift für experimentelle Psychologie*, 46 (3), 216–235.
- Schoot, R. van de, Schmidt, P., Beuckelaer, A. de, Lek, K. & Zondervan-Zwijenburg, M. (2015). Editorial. Measurement invariance. *Frontiers in Psychology*, 6, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Schröder, U. J. (2000). *Lernbehindertenpädagogik. Grundlagen und Perspektiven sonderpädagogischer Lernhilfe*. Stuttgart: Kohlhammer.
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A. & Kliegl, R. (2015). childLex—Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen. *Psychologische Rundschau*, 66 (3), 155–165. <https://doi.org/10.1026/0033-3042/a000275>
- Schuchardt, K., Gebhardt, M. & Mähler, C. (2010). Working memory functions in children with different degrees of intellectual disability. *Journal of Intellectual Disability Research*, 54 (4), 346–353. <https://doi.org/10.1111/j.1365-2788.2010.01265.x>
- Schuchardt, K., Mähler, C. & Hasselhorn, M. (2008). Working memory deficits in children with specific learning disorders. *Journal of Learning Disabilities*, 41 (6), 514–523. <https://doi.org/10.1177/0022219408317856>
- Schuchardt, K., Mähler, C. & Hasselhorn, M. (2011). Functional deficits in phonological working memory in children with intellectual disabilities. *Research in Developmental Disabilities*, 32 (5), 1934–1940. <https://doi.org/10.1016/j.ridd.2011.03.022>
- Schuchardt, K., Roick, T., Mähler, C. & Hasselhorn, M. (2008). Unterscheidet sich die Struktur des Arbeitsgedächtnisses bei Schulkindern mit und ohne Lernstörung? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40 (3), 147–151. <https://doi.org/10.1026/0049-8637.40.3.147>
- Schulze, S. & Kuhl, J. (2019). Integration von Arbeitsgedächtnistrainings in die mathematische Lernförderung. *Lernen und Lernstörungen*, 8 (1), 47–59. <https://doi.org/10.1024/2235-0977/a000229>
- Schulze, S., Kuhl, J. & Breucker, T. (2020). Entwicklung und Pilotierung einer arbeitsgedächtnissensiblen Mathematikförderung. Ergebnisse einer Einzelfallstudie. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 89, 1–22. <https://doi.org/10.2378/vhn2020.art10d>
- Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M. & Richlan, F. (2016). Words in context. The effects of length, frequency, and predictability on brain responses during natural reading. *Cerebral Cortex*, 26 (10), 3889–3904. <https://doi.org/10.1093/cercor/bhw184>
- Schwab, S. & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik*, 7 (3), 175–193.



- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, 21 (2), 168–175. <https://doi.org/10.3758/BF03202729>
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology Section A*, 51 (2), 283–304. <https://doi.org/10.1080/027249898391639>
- Shipstead, Z., Harrison, T. L. & Engle, R. W. (2016). Working memory capacity and fluid intelligence. Maintenance and disengagement. *Perspectives on Psychological Science*, 11 (6), 771–799. <https://doi.org/10.1177/1745691616650647>
- Siew, C. S. Q. (2020). Applications of network science to education research. Quantifying knowledge and the development of expertise through network analysis. *Education Sciences*, 10 (4), 101–116. <https://doi.org/10.3390/educsci10040101>
- Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L. & Schneider, M. (2018). Domain-specific prior knowledge and learning. A meta-analysis. *Preprint*.
- Skowronek, M., Schuchardt, K. & Mähler, C. (2018). Die Entwicklung von Kindern mit umfassenden Lernschwierigkeiten im Verlauf der Grundschuljahre. Schulleistungen, Arbeitsgedächtnis, phonologische Informationsverarbeitung und Selbstkonzept. *Zeitschrift für Pädagogische Psychologie*, 32 (4), 223–236. <https://doi.org/10.1024/1010-0652/a000228>
- Smiley, S. S. & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations. A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, 28 (2), 249–257. [https://doi.org/10.1016/0022-0965\(79\)90087-0](https://doi.org/10.1016/0022-0965(79)90087-0)
- Sommers, M. S. & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in adults. Interaction of lexical competition and semantic context. *Psychology and Aging*, 14 (3), 458–472.
- Souvignier, E. (2008). Lernbehinderung. In W. Schneider (Hrsg.), *Handbuch der Pädagogischen Psychologie* (Handbuch der Psychologie, Bd. 10, S. 663–671). Göttingen: Hogrefe.
- Souza, A. S. & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology*, 107, 1–21. <https://doi.org/10.1016/j.cogpsych.2018.09.002>
- Spanoudis, G. C. & Natsopoulos, D. (2011). Memory functioning and mental verbs acquisition in children with specific language impairment. *Research in Developmental Disabilities*, 32, 2916–2926.
- Stahns, R. (2016). Bildungssprachliche Merkmale von Texten und Items. Zur Operationalisierung des Konstrukts "Bildungssprache". *Didaktik Deutsch*, 21 (41), 44–55.
- Stanovich, K. E. (2005). The future of a mistake. Will discrepancy measurement continue to make the learning disabilities field a pseudoscience? *Learning Disability Quarterly*, 28 (1), 103–106.
- Sterba, S. K. & MacCallum, R. C. (2010). Variability in parameter estimates and model fit across repeated allocations of items to parcels. *Multivariate Behavioral Research*, 45 (2), 322–358. <https://doi.org/10.1080/00273171003680302>

- Sterba, S. K. & Rights, J. D. (2017). Effects of parceling on model selection. Parcel-allocation variability in model ranking. *Psychological Methods*, 22 (1), 47–68. <https://doi.org/10.1037/met0000067>
- Stuart, G. & Hulme, C. (2009). Lexical and semantic influences on immediate serial recall. A role for redintegration. In A. S. C. Thorn & M. P. A. Page (Hrsg.), *Interactions between short-term and long-term memory in the verbal domain* (S. 157–176). Hove: Psychology Press.
- Sullivan, A. L., Kohli, N., Farnsworth, E. M., Sadeh, S. & Jones, L. (2016). Longitudinal models of reading achievement of students with learning disabilities and without disabilities. *School Psychology Quarterly*. <https://doi.org/10.1037/spq0000170>
- Swanson, H. L. (2008). Working memory and intelligence in children. What develops? *Journal of Educational Psychology*, 100 (3), 581–602. <https://doi.org/10.1037/0022-0663.100.3.581>
- Thibaut, J.-P. & French, R. M. (2016). Analogical reasoning, control and executive functions. A developmental investigation with eye-tracking. *Cognitive Development*, 38, 10–26. <https://doi.org/10.1016/j.cogdev.2015.12.002>
- Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C. & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research*, 52 (2), 336–358. [https://doi.org/10.1044/1092-4388\(2009/07-0144\)](https://doi.org/10.1044/1092-4388(2009/07-0144))
- Thorn, A. S. C. & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of non-words are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31 (4), 729–735. <https://doi.org/10.1037/0278-7393.31.4.729>
- Thorn, A. S. C., Frankish, C. R. & Gathercole, S. E. (2009). The influence of long-term knowledge on short-term memory. Evidence for multiple mechanisms. In A. S. C. Thorn & M. P. A. Page (Hrsg.), *Interactions between short-term and long-term memory in the verbal domain* (S. 198–219). Hove: Psychology Press.
- Thorn, A. S. C., Gathercole, S. E. & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory. An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology*, 50 (2), 133–158. <https://doi.org/10.1016/j.cogpsych.2004.07.001>
- Thorn, A. S. C. & Page, M. P. A. (2009). Current issues in understanding interactions between short-term and long-term memory. In A. S. C. Thorn & M. P. A. Page (Hrsg.), *Interactions between short-term and long-term memory in the verbal domain* (S. 1–15). Hove: Psychology Press.
- Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38 (1), 1–10.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Hrsg.), *Organization of memory* (S. 381–403). New York, NY: Academic Press.
- Tumminello, M., Aste, T., Matteo, T. di & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102 (30), 10421–10426.

- Turner, J. E., Henry, L. A., Smith, P. T. & Brown, P. A. (2004). Redintegration and lexicality effects in children. Do they depend upon the demands of the memory task? *Memory & Cognition*, 32 (3), 501–510. <https://doi.org/10.3758/BF03195842>
- Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature. Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3 (1), 4–70.
- Verhelst, N. D. (2008). An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika*, 73 (4), 705–728. <https://doi.org/10.1007/S11336-008-9062-3>
- Vernooij, M. (2007). Aspekte des Lernens bei Kindern mit Lernbeeinträchtigungen. In K. Salzberg-Ludwig & E. Grüning (Hrsg.), *Pädagogik für Kinder- und Jugendliche in schwierigen Lern- und Lebenssituationen* (Heil- und Sonderpädagogik, S. 47–64). Stuttgart: Kohlhammer.
- Vinter, A. & Detable, C. (2003). Implicit learning in children and adolescents with mental retardation. *American Journal on Mental Retardation*, 108 (2), 94. [https://doi.org/10.1352/0895-8017\(2003\)108<0094:ILICAA>2.0.CO;2](https://doi.org/10.1352/0895-8017(2003)108<0094:ILICAA>2.0.CO;2)
- Vitevitch, M. S., Chan, K. Y. & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, 67 (1), 30–44. <https://doi.org/10.1016/j.jml.2012.02.008>
- Vitevitch, M. S. & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, 31 (4), 491–504.
- Walker, I. & Hulme, C. (1999). Concrete words are easier to recall than abstract words. Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25 (5), 1256–1271.
- Weiß, R. H. (2008). *CFT 20-R mit WS/ZF-R. Grundintelligenztest Skala 2 - Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R)*. Göttingen: Hogrefe.
- Weiß, R. H. & Osterland, J. (2012). *CFT 1-R. Grundintelligenztest Skala 1 - Revision*. Göttingen: Hogrefe.
- Wember, F. B. (1987). Empirische Befunde zum konkret-operatorischen Denken und schulischen Lernen bei Schülerinnen und Schülern der Schule für Lernbehinderte. *Heilpädagogische Forschung*, 13 (2), 103-113.
- Wember, F. B. (2007). Differenzierung des Unterrichts. In J. Walter & F. B. Wember (Hrsg.), *Sonderpädagogik des Lernens* (Handbuch Sonderpädagogik, Bd. 2, S. 393–420). Göttingen: Hogrefe.
- Wember, F. B. & Heimlich, U. (2014). Lernen. In F. B. Wember, R. Stein & U. Heimlich (Hrsg.), *Handlexikon Lernschwierigkeiten und Verhaltensstörungen* (S. 54–58). Stuttgart: Kohlhammer.
- Werning, R. & Lütje-Klose, B. (2012). *Einführung in die Pädagogik bei Lernbeeinträchtigungen* (UTB Sonderpädagogik, Bd. 2391, 3. Aufl.). München: UTB.
- Wickham, H. (2011). *ggplot2. Elegant graphics for data analysis* [Computer software]: Springer New York. Verfügbar unter <http://ggplot2.org>

- Wilbert, J. & Grosche, M. (2012). A dissociation of implicit and explicit spatial sequence learning in a group of students with learning difficulties. *Journal of Cognitive Education and Psychology*, 11 (3), 301–316. <https://doi.org/10.1891/1945-8959.11.3.301>
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V. & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder. A meta-analytic review. *Biological Psychiatry*, 57 (11), 1336–1346. <https://doi.org/10.1016/j.biopsych.2005.02.006>
- Willett, J. B., Singer, J. D. & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context. Statistical models and methodological recommendations. *Development and Psychopathology*, 10 (2), 395–426.
- Williams, J. L., Miciak, J., McFarland, L. & Wexler, J. (2016). Learning disability identification criteria and reporting in empirical research. A review of 2001-2013. *Learning Disabilities Research & Practice*, 31 (4), 221–229.
- Winters, J. J. & Brzoska, M. A. (1975). Development of lexicon in normal and retarded persons. *Psychological Reports*, 37 (2), 391–402. <https://doi.org/10.2466/pr0.1975.37.2.391>
- Wittchen, H.-U. (2011). *Klinische Psychologie & Psychotherapie* (Springer-Lehrbuch). Berlin: Springer.
- Wong, G. Y. & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80 (391), 513–524.
- Wu, A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis. A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12 (12), 1–26.
- Xia, Y. & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data. The story they tell depends on the estimation methods. *Behavior Research Methods*, 51 (1), 409–428. <https://doi.org/10.3758/s13428-018-1055-2>
- Yarkoni, T., Balota, D. & Yap, M. (2008). Moving beyond Coltheart's N. A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15 (5), 971–979. <https://doi.org/10.3758/PBR.15.5.971>
- Yates, T. M., Egeland, B. & Sroufe, L. A. (2003). Rethinking resilience. A developmental process perspective. In S. S. Luthar (Hrsg.), *Resilience and vulnerability. Adaptation in the context of childhood adversities* (S. 243–266). Cambridge, U.K.: Cambridge University Press.
- Zielinski, W. (1980). *Lernschwierigkeiten. Verursachungsbedingungen, Diagnose, Behandlungsansätze* (Kohlhammer-Standards Psychologie). Stuttgart: Kohlhammer.
- Zielinski, W. (1995). *Lernschwierigkeiten. Ursachen, Diagnostik, Intervention* (Kohlhammer-Standards Psychologie, 2., überarbeitete und erweiterte Auflage). Stuttgart: Kohlhammer.
- Zigler, E. & Balla, D. A. (Hrsg.). (1982). *Mental retardation. The developmental-difference controversy*. Hillsdale, NJ: Lawrence Erlbaum.

## Tabellenverzeichnis

Tabelle 1 <i>Stichprobenmerkmale und deskriptive Angaben aller Variablen (M und SD) in Studie 1</i>	89
Tabelle 2 <i>Korrelationen zwischen Entwicklungsindikatoren in Studie 1</i>	91
Tabelle 3 <i>Stichprobenmerkmale und deskriptive Angaben aller Variablen (M und SD) in Studie 2</i>	112
Tabelle 4 <i>Beispiele für Items im WWT 6-10 nach Wortarten</i>	113
Tabelle 5 <i>Globaler Modellvergleich ein- vs. vier-faktorielles Modell über beide Gruppen hinweg</i>	128
Tabelle 6 <i>Modellvergleiche zur Überprüfung der Messinvarianz</i>	130
Tabelle 7 <i>Modellvergleiche mit Parcel-Allocation Methode zur Überprüfung der Messinvarianz</i>	132
Tabelle 8 <i>Parameter des verallgemeinerten gemischten linearen Modells (GLMM) zur Vorhersage der Lösungswahrscheinlichkeit der korrekten Bild-Benennung im WWT 6-10</i>	133
Tabelle 9 <i>Modellvergleich bei Berücksichtigung der Gruppen-Interaktion</i>	135
Tabelle 10 <i>Deskriptive Angaben zur vollen und reduzierten Skala des Wortschatzumfangs in Studie 3</i>	153
Tabelle 11 <i>Ergebnisse der Netzwerkanalyse der Daten des WWT 6-10</i>	169
Tabelle A-1 <i>Liste der Wortspannen-Stimuli pro Bedingung in Studie 1</i> .....	208
Tabelle A-2 <i>Parameter der Developmental Trajectory Modelle in Studie 1</i> .....	209
Tabelle A-3 <i>Zuordnung der Items zu Parcels nach Wortarten in Studie 2</i> .....	211
Tabelle A-4 <i>Deskriptive Angaben der einzelnen Parcels nach Wortarten</i> .....	212
Tabelle A-5 <i>Kennwerte der Items des WWT 6-10 und Ergebnisse des Wald-Tests in Studie 3</i> .....	213

## Abbildungsverzeichnis

<i>Abbildung 1.</i> Szenarien im Developmental Trajectory Design (Thomas et al., 2009, S. 346)	76
<i>Abbildung 2.</i> Developmental Trajectory Modelle der Kapazität der phonologischen Schleife	95
<i>Abbildung 3.</i> Developmental Trajectory Modelle des Rehearsalprozesses	97
<i>Abbildung 4.</i> Developmental Trajectory Modelle des Redintegrationprozesses	99
<i>Abbildung 5.</i> Beispiele für Items im WWT 6-10. a) Nomen, b) Verben, c) Kategorien, d) Adjektive	114
<i>Abbildung 6.</i> GOF-Plot für Subskala Nomen. a) Vollständige Subskala, b) reduzierte Subskala	154
<i>Abbildung 7.</i> GOF-Plot für Subskala Verben. a) Vollständige Subskala, b) reduzierte Subskala	154
<i>Abbildung 8.</i> GOF-Plot für Subskala Adjektive. a) Vollständige Subskala, b) reduzierte Subskala	155
<i>Abbildung 9.</i> GOF-Plot für Subskala Kategorien. a) Vollständige Subskala, b) reduzierte Subskala	156
<i>Abbildung 10.</i> Developmental Trajectory Reanalyse für die Effektivität des Redintegrationprozesses im Verhältnis zum Wortschatzumfang (bereinigt)	157
<i>Abbildung 11.</i> Semantische Netzwerke anhand der WWT-Daten für a) Schüler*innen mit Lernbehinderung und b) unbeeinträchtigte Kontrollgruppe.	170

## Anhang A – Tabellen

Tabelle A-1

Liste der Wortspannen-Stimuli pro Bedingung in Studie 1

Echte Wörter		Kunstwörter	
Kurz	Lang	Kurz	Lang
Ball	Briefkasten	Bork	Bestrugeln
Baum	Eisenbahn	Deil	Franulich
Eis	Erdbeere	Fen	Karflumen <sup>a</sup>
Fisch	Fernseher	Grul <sup>b</sup>	Külinge
Haus	Kneifzange	Laum	Laujossung
Pilz	Lichtschalter	Lurf <sup>b</sup>	Mindinnen
Schuh	Luftballon	Natt	Reseubelt
Stern	Schaukelpferd	Sim	Schleibunder
Topf	Zahnbürste	Tirk	Wuralten <sup>a</sup>

*Anmerkungen.* Die Stimuli für echte Wörter wurden der deutschen Arbeitsgedächtnis-Testbatterie entnommen (AGTB 5-12, Hasselhorn et al. 2012); Kunstwörter: für jede Länge wurden sieben Stimuli für Pseudowörter aus Hasselhorn et al. (2010) entnommen, wobei für jede Liste zwei Wörter hinzugefügt wurden, um die Anzahl der Wörter pro Wortpool anzupassen.

<sup>a</sup> Zwei zusätzliche dreisilbige Kunstwörter wurden aus der AGTB 5-12 hinzugefügt, Kunstwort-Wiederholungs-Aufgabe. <sup>b</sup> Zwei zusätzliche einsilbige Kunstwörter wurden hinzugefügt.

Tabelle A-2  
 Parameter der Developmental Trajectory Modelle in Studie 1

AV	Gruppe	Parameter	Entwicklungsindikatoren (EI)		
			CA	KOG	VOC
PL-Kapazität	Gesamt	Intercept <sup>a</sup>	$F(1, 203) = 16.48;$ $p < .001; \eta^2 = .075$	$F(1, 178) = 4.32;$ $p = .039; \eta^2 = .024$	$F(1, 178) = 0.95;$ $p = .331; \eta^2 = .005$
		Slope <sup>a</sup>	$F(1, 203) = 3.36;$ $p = .068; \eta^2 = .016$	$F(1, 178) = 4.63;$ $p = .033; \eta^2 = .025$	$F(1, 178) = 0.28;$ $p = .595; \eta^2 = .002$
	LB	Intercept <sup>b</sup>	$F(1, 85) = 355.87;$ $p < .001; \eta^2 = .798$	$F(1, 78) = 216.06;$ $p < .001; \eta^2 = .735$	$F(1, 78) = 374.19;$ $p < .001; \eta^2 = .828$
		Slope <sup>b</sup>	$F(1, 85) = 32.96;$ $p < .001; \eta^2 = .279$	$F(1, 78) = 46.41;$ $p < .001; \eta^2 = .373$	$F(1, 78) = 21.02;$ $p < .001; \eta^2 = .212$
	KG	Intercept <sup>b</sup>	$F(1, 118) = 2511.2;$ $p < .001; \eta^2 = .955$	$F(1, 100) = 291.99;$ $p < .001; \eta^2 = .745$	$F(1, 100) = 319.03;$ $p < .001; \eta^2 = .761$
		Slope <sup>b</sup>	$F(1, 118) = 54.57;$ $p < .001; \eta^2 = .279$	$F(1, 100) = 14.48;$ $p < .001; \eta^2 = .126$	$F(1, 100) = 29.08;$ $p < .001; \eta^2 = .225$
Rehearsal	Gesamt	Intercept <sup>a</sup>	$F(1, 203) = 5.39;$ $p = .021; \eta^2 = .026$	$F(1, 178) = 7.00;$ $p = .009; \eta^2 = .038$	$F(1, 178) = 1.57;$ $p = .212; \eta^2 = .009$
		Slope <sup>a</sup>	$F(1, 203) = 1.58;$ $p = .210; \eta^2 = .008$	$F(1, 178) = 5.94;$ $p = .016; \eta^2 = .032$	$F(1, 178) = 1.72;$ $p = .192; \eta^2 = .010$
	LB	Intercept <sup>c</sup>	$F(1, 85) = 26.53;$ $p < .001; \eta^2 = .238$	$F(1, 78) = 10.50;$ $p = .002; \eta^2 = .119$	$F(1, 78) = 56.72;$ $p < .001; \eta^2 = .421$
		Slope <sup>c</sup>	$F(1, 85) = 6.99;$ $p = .010; \eta^2 = .076$	$F(1, 78) = 11.62;$ $p = .001; \eta^2 = .130$	$F(1, 78) = 7.55;$ $p = .007; \eta^2 = .088$
	KG	Intercept <sup>c</sup>	$F(1, 118) = 351.63;$ $p < .001; \eta^2 = .749$	$F(1, 100) = 50.22;$ $p < .001; \eta^2 = .334$	$F(1, 100) = 88.65;$ $p < .001; \eta^2 = .470$
		Slope <sup>c</sup>	$F(1, 118) = 15.70;$ $p < .001; \eta^2 = .117$	$F(1, 100) = 0.003;$ $p = 0.96; \eta^2 = .000$	$F(1, 100) = 0.93;$ $p = .337; \eta^2 = .009$
Redintegration	Gesamt	Intercept <sup>a</sup>	$F(1, 203) = 0.40;$ $p = .530; \eta^2 = .002$	$F(1, 178) = 0.13;$ $p = .718; \eta^2 = .001$	$F(1, 178) = 5.19;$ $p = .024; \eta^2 = .028$
		Slope <sup>a</sup>	$F(1, 203) = 0.25;$ $p = .617; \eta^2 = .001$	$F(1, 178) = 0.20;$ $p = .658; \eta^2 = .001$	$F(1, 178) = 8.21;$ $p = .005; \eta^2 = .044$
	LB	Intercept <sup>c</sup>	$F(1, 85) = 66.06;$ $p < .001; \eta^2 = .437$	$F(1, 78) = 30.99;$ $p < .001; \eta^2 = .284$	$F(1, 78) = 132.71;$ $p < .001; \eta^2 = .63$
		Slope <sup>c</sup>	$F(1, 85) = 1.631;$ $p = .205; \eta^2 = .019$	$F(1, 78) = 0.17;$ $p = .678; \eta^2 = .002$	$F(1, 78) = 8.25;$ $p = .005; \eta^2 = .096$
	KG	Intercept <sup>c</sup>	$F(1, 118) = 212.29;$ $p < .001; \eta^2 = .643$	$F(1, 100) = 17.69;$ $p < .001; \eta^2 = .150$	$F(1, 78) = 37.54;$ $p < .001; \eta^2 = .273$
		Slope <sup>c</sup>	$F(1, 118) = 0.08;$ $p = .776; \eta^2 = .001$	$F(1, 100) = 0.97;$ $p = .328; \eta^2 = .010$	$F(1, 100) = 2.06;$ $p = .154; \eta^2 = .020$

*Anmerkungen.* Die drei abhängigen Variablen (AV) sind: Die Kapazität der phonologischen Schleife (PL) gemessen an der durchschnittlichen Leistung über alle vier Aufgabenbedingungen hinweg; der Rehearsalprozess gemessen am Wortlängen-Effekt, d.h. Wortlänge wird als Faktor in das Modell einbezogen; Redintegrationprozess gemessen am Lexikalitätseffekt, d.h. Lexikalität wird als Faktor in das Modell einbezogen. Für jede AV wird ein Gesamtmodell geschätzt, um Intercept und Slope zwischen den Gruppen zu vergleichen; zusätzliche Modelle werden separat pro Gruppe geschätzt, um festzustellen, ob die Effekte in jeder Gruppe signifikant von Null abweichen. Intercept bezeichnet einen Niveauunterschied zum jüngsten gemessenen Alter in der LB-



---

Gruppe. Slope bezeichnet einen Unterschied in den Steigungskoeffizienten zwischen den Gruppen. Drei Entwicklungsindikatoren (EI) werden als Prädiktoren verwendet: das chronologische Alter (CA), gemessen in Jahren und Monaten, die kognitive Kapazität (KOG), gemessen in CFT 1-R Rohwerten, und der Wortschatzumfang (VOC), gemessen in WWT 6-10 Rohwerten.

<sup>a</sup>In den Gesamtmodellen beziehen sich die Parameter auf die Unterschiede zwischen den Gruppen. <sup>b</sup>In den Gruppenmodellen für die PL-Kapazität beziehen sich die Parameter auf den Unterschied von Null. <sup>c</sup>In den Gruppenmodellen für Rehearsal- und Redintegrationprozess beziehen sich die Parameter auf Unterschiede zwischen den jeweiligen Bedingungen (d.h. Länge und Lexikalität) innerhalb jeder Gruppe.

---

Tabelle A-3

*Zuordnung der Items zu Parcels nach Wortarten in Studie 2*

<b>Nomen:</b>	<b>Oberkategorien:</b>
<i>N1:</i> Absatz, Schnalle, Mantel, Geländer, Schleier, Gipfel	<i>K1:</i> Gepäck, Zeichen, Feste, Geschirr, Schmuck
<i>N2:</i> Riegel, Armreif, Container, Krücke, Ellenbogen	<i>K2:</i> Gebäude, Insekten, Gewürze, Kosmetika, Instrumente
<i>N3:</i> Satellit, Wappen, Knospe, Ferse, Schubkarre	<i>K3:</i> Küchengeräte, Berufe, Pflanzen, Möbel, Werkzeuge
<i>N4:</i> Orchester, Hocker, Fackel, Henkel, Federball	<i>K4:</i> Lebensmittel, Baufahrzeuge, Jahreszeiten, Sportarten
<i>N5:</i> Klinge, Ventilator, Automat, Kompass, Pyramide	<i>K5:</i> Gemüse, Kopfbedeckungen, Getreide, Besteck
<b>Verben:</b>	<b>Gegensätze (Adjektive):</b>
<i>V1:</i> abtrocknen, fressen, verbeugen, brüllen, schälen	<i>A1:</i> einfarbig, langweilig, spitz, traurig, innen
<i>V2:</i> demonstrieren, wehen, schieben, trösten, ziehen	<i>A2:</i> wolkenlos, unzufrieden, hoch, trocken, lieb
<i>V3:</i> reiben, verblühen, abschleppen, wiegen, balancieren	<i>A3:</i> wild, einfach, nah, früh, weich
<i>V4:</i> entgleisen, jonglieren, abrechnen, warten	<i>A4:</i> mutig, ungefährlich, glatt, sauer
<i>V5:</i> zeigen, dirigieren, stricken, schimpfen	<i>A5:</i> alt, vertraut, altmodisch, hässlich

Tabelle A-4

*Deskriptive Angaben der einzelnen Parcels für Nomen (N), Verben (V), Adjektive (A) und Oberkategorien (K), getrennt nach Gruppen für Schüler\*innen mit Lernbehinderung (LB) und unbeeinträchtigte Schüler\*innen (KG) in Studie 2*

Parcel	Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
N1	LB	92	0.518	0.253	-0.313	-0.651
	KG	114	0.522	0.223	0.221	-0.728
N2	LB	93	0.557	0.232	-0.616	0.026
	KG	114	0.563	0.246	0.067	-0.772
N3	LB	93	0.492	0.299	0.082	-0.909
	KG	114	0.484	0.251	0.004	-0.733
N4	LB	93	0.439	0.280	0.292	-0.851
	KG	114	0.395	0.254	0.228	-0.804
N5	LB	93	0.746	0.280	-1.097	0.527
	KG	114	0.682	0.221	-0.426	-0.511
V1	LB	93	0.634	0.222	-0.291	-0.352
	KG	114	0.714	0.208	-0.350	-0.566
V2	LB	93	0.596	0.245	-0.241	-0.752
	KG	114	0.614	0.209	-0.231	-0.113
V3	LB	93	0.628	0.239	-0.115	-0.547
	KG	114	0.618	0.222	-0.094	-0.531
V4	LB	93	0.658	0.190	-0.447	-0.240
	KG	114	0.684	0.184	-0.579	0.092
V5	LB	93	0.459	0.273	-0.073	-0.661
	KG	114	0.503	0.227	0.135	-0.279
A1	LB	93	0.512	0.276	-0.269	-0.818
	KG	114	0.523	0.290	-0.182	-1.009
A2	LB	93	0.639	0.244	-0.761	0.537
	KG	114	0.674	0.225	-0.491	-0.482
A3	LB	93	0.596	0.248	-0.835	0.075
	KG	114	0.653	0.218	-0.488	-0.275
A4	LB	93	0.508	0.281	-0.088	-0.884
	KG	114	0.486	0.252	0.157	-0.731
A5	LB	93	0.548	0.293	-0.110	-0.747
	KG	114	0.553	0.331	-0.249	-0.997
K1	LB	93	0.542	0.263	-0.256	-0.417
	KG	114	0.600	0.250	-0.217	-0.715
K2	LB	93	0.632	0.231	-0.480	-0.056
	KG	114	0.602	0.269	-0.037	-1.077
K3	LB	93	0.499	0.261	0.026	-0.597
	KG	114	0.530	0.262	0.192	-0.821
K4	LB	93	0.542	0.253	-0.534	-0.432
	KG	114	0.589	0.224	-0.198	-0.517
K5	LB	93	0.405	0.282	0.119	-0.769
	KG	114	0.406	0.246	0.291	-0.120

Tabelle A-5

*Kenntnisse der Items des WWT 6-10 und Ergebnisse des Wald-Tests für den Gruppenvergleich zwischen Schüler\*innen mit Lernbehinderung (LB) und unbeeinträchtigte Schüler\*innen (KG) in Studie 3*

Wortart	Zielwort	Gruppe LB		Gruppe KG		Wald-Test vor Ausschluss		Wald-Test nach Ausschluss	
		<i>Beta</i>	<i>S.E.</i>	<i>Beta</i>	<i>S.E.</i>	<i>z</i> -Statistik	<i>p</i> -Wert	<i>z</i> -Statistik	<i>p</i> -Wert
Nomen	Schubkarre	1.287	0.282	1.837	0.274	-1.398	.162	-0.277	.782
Nomen	Hocker	-0.824	0.237	-0.708	0.204	-0.370	.711	0.706	.480
Nomen	Krücke	1.458	0.293	0.331	0.204	3.156	.002		
Nomen	Automat	1.458	0.293	0.780	0.215	1.866	.062		
Nomen	Container	1.053	0.269	0.120	0.202	2.775	.006		
Nomen	Pyramide	1.371	0.287	2.398	0.331	-2.346	.019	-1.311	.190
Nomen	Federball	0.979	0.265	0.417	0.206	1.673	.094		
Nomen	Kompass	1.371	0.287	0.640	0.211	2.052	.040		
Nomen	Ellenbogen	1.743	0.314	1.760	0.268	-0.042	.966	1.065	.287
Nomen	Geländer	-0.371	0.235	0.640	0.211	-3.203	.001		
Nomen	Satellit	-1.237	0.245	-1.399	0.223	0.490	.624	1.448	.147
Nomen	Orchester	-1.830	0.267	-1.978	0.254	0.402	.688	1.199	.231
Nomen	Knospe	-1.116	0.242	-0.498	0.201	-1.962	.050	-0.928	.353
Nomen	Ferse	0.445	0.246	0.331	0.204	0.358	.720	1.564	.118
Nomen	Mantel	1.053	0.269	0.079	0.201	2.900	.004		
Nomen	Ventilator	2.347	0.374	1.302	0.238	2.357	.018		
Nomen	Schleier	-1.488	0.253	-0.837	0.206	-1.995	.046	-1.045	.296
Nomen	Riegel	-2.220	0.289	-1.451	0.225	-2.099	.036	-1.326	.185

Wortart	Zielwort	Gruppe LB		Gruppe KG		Wald-Test vor Ausschluss		Wald-Test nach Ausschluss	
		<i>Beta</i>	<i>S.E.</i>	<i>Beta</i>	<i>S.E.</i>	<i>z</i> -Statistik	<i>p</i> -Wert	<i>z</i> -Statistik	<i>p</i> -Wert
Nomen	Fackel	-0.427	0.235	-0.623	0.203	0.632	.528	1.768	.077
Nomen	Wappen	-1.116	0.242	-1.503	0.228	1.166	.244		
Nomen	Klinge	-0.427	0.235	-1.015	0.210	1.865	.062		
Nomen	Absatz	-0.314	0.235	-1.107	0.213	2.499	.012		
Nomen	Armreif	-1.830	0.267	-0.086	0.200	-5.226	< .001		
Nomen	Schnalle	-1.978	0.274	-1.348	0.221	-1.788	.074	-0.961	.337
Nomen	Henkel	-1.237	0.245	-0.881	0.207	-1.108	.268	-0.112	.911
Nomen	Gipfel	1.848	0.323	2.799	0.386	-1.889	.059	-0.929	.353
Verben	fressen	-1.467	0.241	-0.249	0.202	-3.874	< .001		
Verben	zeigen	2.664	0.478	2.161	0.382	0.823	.410	1.075	.282
Verben	entgleisen	-3.368	0.377	-2.807	0.278	-1.198	.231	-1.546	.123
Verben	jonglieren	0.791	0.264	1.067	0.259	-0.748	.455	-0.537	.592
Verben	schälen	2.253	0.409	1.765	0.327	0.933	.351	1.173	.241
Verben	schieben	1.097	0.283	0.808	0.241	0.778	.437	0.991	.322
Verben	balancieren	1.683	0.335	1.765	0.327	-0.175	.861	0.052	.958
Verben	warten	1.939	0.365	0.869	0.245	2.433	.015		
Verben	demonstrieren	-1.526	0.243	-2.518	0.256	2.805	.005		
Verben	abtrocknen	-0.545	0.228	-1.011	0.200	1.535	.125	1.54	.123
Verben	abbrechen	-0.335	0.229	0.094	0.209	-1.382	.167	-1.257	.209
Verben	verblühen	-1.970	0.261	-1.559	0.211	-1.224	.221	-1.373	.170
Verben	wehen	-1.647	0.247	-0.890	0.199	-2.385	.017		

Wortart	Zielwort	Gruppe LB		Gruppe KG		Wald-Test vor Ausschluss		Wald-Test nach Ausschluss	
		<i>Beta</i>	<i>S.E.</i>	<i>Beta</i>	<i>S.E.</i>	<i>z</i> -Statistik	<i>p</i> -Wert	<i>z</i> -Statistik	<i>p</i> -Wert
Verben	wiegen	1.570	0.323	0.231	0.214	3.452	.001		
Verben	dirigieren	-2.775	0.313	-2.586	0.261	-0.465	.642	-0.772	.44
Verben	trösten	0.395	0.246	0.326	0.217	0.210	.834	0.382	.703
Verben	verbeugen	-0.755	0.229	0.094	0.209	-2.738	.006		
Verben	reiben	-1.352	0.238	-1.427	0.207	0.238	.812	0.154	.877
Verben	ziehen	1.016	0.278	1.462	0.293	-1.104	.270	-0.888	.374
Verben	schimpfen	0.457	0.248	0.139	0.211	0.977	.328	1.138	.255
Verben	abschleppen	0.720	0.260	0.581	0.228	0.403	.687	0.598	.550
Verben	stricken	-0.650	0.228	-0.081	0.205	-1.856	.063	-1.763	.078
Verben	brüllen	1.805	0.349	1.765	0.327	0.085	.932	0.315	.753
Adjektive	alt	2.425	0.411	2.482	0.355	-0.105	.916	0.069	.945
Adjektive	lieb	1.965	0.361	3.570	0.514	-2.554	.011		
Adjektive	früh	2.425	0.411	2.131	0.323	0.564	.573	0.753	.452
Adjektive	weich	2.259	0.392	2.764	0.387	-0.918	.358	-0.762	.446
Adjektive	wild	-3.641	0.431	-3.327	0.319	-0.585	.558	-0.222	.824
Adjektive	wolkenlos	-1.670	0.253	-1.656	0.229	-0.042	.967	0.525	.600
Adjektive	einfach	-0.239	0.237	-0.160	0.226	-0.242	.809	0.271	.786
Adjektive	glatt	-0.795	0.234	-2.033	0.239	3.698	< .001		
Adjektive	einfarbig	-1.606	0.251	-1.656	0.229	0.146	.884	0.715	.475
Adjektive	mutig	-1.130	0.238	-1.502	0.226	1.133	.257	1.714	.087
Adjektive	vertraut	-1.870	0.262	-1.656	0.229	-0.615	.538	-0.058	.954

Wortart	Zielwort	Gruppe LB		Gruppe KG		Wald-Test vor Ausschluss		Wald-Test nach Ausschluss	
		<i>Beta</i>	<i>S.E.</i>	<i>Beta</i>	<i>S.E.</i>	<i>z</i> -Statistik	<i>p</i> -Wert	<i>z</i> -Statistik	<i>p</i> -Wert
Adjektive	traurig	-0.239	0.237	-0.056	0.227	-0.558	.577	-0.054	.957
Adjektive	nah	0.112	0.245	0.920	0.252	-2.301	.021	-1.91	.056
Adjektive	langweilig	-0.575	0.234	-0.263	0.224	-0.961	.336	-0.437	.662
Adjektive	hoch	1.022	0.285	0.157	0.231	2.363	.018		
Adjektive	innen	0.709	0.267	1.121	0.260	-1.105	.269	-0.766	.444
Adjektive	hässlich	2.106	0.375	1.053	0.257	2.316	.021		
Adjektive	sauer	-0.630	0.234	-0.761	0.221	0.409	.682	0.969	.332
Adjektive	unzufrieden	-0.630	0.234	-0.465	0.223	-0.511	.609	0.029	.977
Adjektive	spitz	-0.352	0.236	-1.604	0.228	3.818	< .001		
Adjektive	altmodisch	-0.296	0.236	-0.314	0.224	0.054	.957	0.581	.562
Adjektive	trocken	1.835	0.348	2.358	0.343	-1.070	.285	-0.895	.371
Adjektive	ungefährlich	-1.187	0.239	-1.104	0.222	-0.253	.800	0.318	.751
Kategorien	Gemüse	1.826	0.320	1.413	0.253	1.010	.313	0.842	.400
Kategorien	Getreide	-1.180	0.239	-1.513	0.221	1.025	.306	0.818	.413
Kategorien	Pflanzen	-1.123	0.237	-0.329	0.201	-2.557	.011		
Kategorien	Kosmetika	2.764	0.413	1.163	0.238	3.362	.001		
Kategorien	Gepäck	-1.066	0.236	-0.915	0.205	-0.484	.629	-0.672	.502
Kategorien	Gewürze	0.552	0.248	-0.453	0.201	3.148	.002		
Kategorien	Instrumente	2.445	0.377	1.626	0.268	1.770	.077	1.603	.109
Kategorien	Kopfbedeckungen	-2.033	0.277	-1.722	0.230	-0.866	.387	-1.094	.274
Kategorien	Geschirr	0.748	0.256	0.832	0.223	-0.247	.805	-0.407	.684

Wortart	Zielwort	Gruppe LB		Gruppe KG		Wald-Test vor Ausschluss		Wald-Test nach Ausschluss	
		<i>Beta</i>	<i>S.E.</i>	<i>Beta</i>	<i>S.E.</i>	<i>z</i> -Statistik	<i>p</i> -Wert	<i>z</i> -Statistik	<i>p</i> -Wert
Kategorien	Sportarten	1.267	0.282	0.731	0.219	1.502	.133	1.338	.181
Kategorien	Küchengeräte	-1.010	0.235	-1.002	0.207	-0.027	.979	-0.218	.827
Kategorien	Feste	-1.180	0.239	-0.703	0.202	-1.526	.127	-1.705	.088
Kategorien	Schmuck	2.304	0.363	2.674	0.386	-0.698	.485	-0.872	.383
Kategorien	Werkzeuge	2.445	0.377	1.960	0.297	1.011	.312	0.838	.402
Kategorien	Besteck	0.682	0.253	0.682	0.218	0.001	.999	-0.16	.873
Kategorien	Baufahrzeuge	-2.374	0.302	-1.513	0.221	-2.300	.021		
Kategorien	Lebensmittel	-0.737	0.231	-1.002	0.207	0.855	.392	0.663	.507
Kategorien	Jahreszeiten	-0.200	0.231	0.992	0.230	-3.656	< .001		
Kategorien	Insekten	-0.845	0.232	-0.745	0.203	-0.327	.744	-0.509	.611
Kategorien	Möbel	0.367	0.242	0.006	0.203	1.143	.253	0.975	.330
Kategorien	Gebäude	-1.417	0.246	-0.536	0.201	-2.775	.006		
Kategorien	Zeichen	-0.630	0.230	-0.329	0.201	-0.986	.324	-1.153	.249
Kategorien	Berufe	-1.607	0.254	-1.319	0.214	-0.867	.386	-1.073	.283

*Anmerkungen.* Beta: Itemparameter des nach Gruppen getrennten Raschmodells. Die Werte für Beta und S.E. beziehen sich auf das Raschmodell mit allen Items. Für die reduzierte Version nach Ausschluss von  $k = 29$  Items ergeben sich andere Itemparameter, die hier aber nicht berichtet werden; es werden lediglich die Ergebnisse des Waldtests nach Ausschluss angegeben. Für ausgeschlossene Items werden in den letzten beiden Spalten keine Werte angeführt.



**Anhang B – Fachbeitrag (englischsprachiges Original)**



# Verbal Working Memory Processes in Students With Mild and Borderline Intellectual Disabilities: Differential Developmental Trajectories for Rehearsal and Redintegration

Gunnar Bruns\*, Birgit Ehl and Michael Grosche

Rehabilitation Sciences on Special Learning Needs, Institute of Educational Research, School of Education, University of Wuppertal, Wuppertal, Germany

## OPEN ACCESS

### Edited by:

Peter Klaver,  
University of Surrey, United Kingdom

### Reviewed by:

Yoshifumi Ikeda,  
Joetsu University of Education, Japan  
Vrinda Kalia,  
Miami University, United States

### \*Correspondence:

Gunnar Bruns  
gbruns@uni-wuppertal.de

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

Received: 21 August 2018

Accepted: 03 December 2018

Published: 09 January 2019

### Citation:

Bruns G, Ehl B and Grosche M (2019)  
Verbal Working Memory Processes in  
Students With Mild and Borderline  
Intellectual Disabilities: Differential  
Developmental Trajectories for  
Rehearsal and Redintegration.  
*Front. Psychol.* 9:2581.  
doi: 10.3389/fpsyg.2018.02581

In verbal working memory, two processes serve to retain a fading memory trace: subvocal rehearsal and lexical redintegration. While recent studies on students with mild and borderline intellectual disabilities (MBID) have yielded mixed results on rehearsal, redintegration has not been researched in MBID, yet. Furthermore, most studies have used a group-matched design which, due to methodological constraints, can only distinguish between two different development patterns. Thus, we study both rehearsal and redintegration in students with MBID using developmental trajectories that have greater potential for identifying differential developmental patterns than traditional group-matching approaches. We investigate whether three aspects in working memory develop differently in students with MBID in comparison to typically developing students: (a) the general capacity of the phonological loop, and the effectiveness of (b) rehearsal, and (c) redintegration. We use three different developmental indicators to compare trajectories: chronological age, cognitive capacity, and vocabulary size.  $N = 210$  students (87 students with MBID, 123 typically developing students) completed working memory span tasks with short and long (1- vs. 3-syllable) real words and pseudowords. The effect for word length (short vs. long) measures rehearsal, and the lexicality effect (real words vs. pseudowords) measures redintegration. Results show that developmental trajectories reveal an intercept difference but no slowed rate in rehearsal, and no impairment in redintegration. However, concerning the developmental relation between redintegration and vocabulary size, students with MBID reveal a differential pattern as redintegration appears higher for students with small vocabulary size, but unexpectedly decreases as vocabulary size increases. We conclude that students with MBID show a delayed onset in the development of capacity of the phonological loop and rehearsal and that they do not catch up in their development. Redintegration does not seem to be impaired in relation to age and cognitive capacity. However, the differential relation of redintegration with vocabulary size calls for further research. While impaired subvocal rehearsal appears to be connected to the developmental problems of students with MBID, lexical redintegration seems to be intact in relation to chronological age and cognitive capacity, making it a possible area of strength.

**Keywords:** working memory, mild intellectual disability (MID), borderline intellectual disability, rehearsal, redintegration, developmental trajectories, delayed onset, phonological loop

## INTRODUCTION

### Purpose of the Study

Mild and borderline intellectual disabilities (MBID) are defined as deficits in intellectual and adaptive functioning with an IQ between 55–70 (mild) and 70–85 (borderline). Most students with MBID have prominent deficits in academic learning. The etiology of MBID (van der Molen et al., 2014) is still largely unknown, apart from clearly diagnosable syndromes like Down syndrome or Williams syndrome.

Existing MBID etiology models are often simply lists of unconnected causal factors, describing the phenomenon only at a vague level (e.g., Schröder, 2000; Kretschmann, 2007; Shaw, 2010 as cited in Hassiotis, 2015), and lacking empirical foundation. Specific, coherent, and comprehensive theories about MBID are lacking. While most theories include cognitive processes and agree that information processing is likely to be impaired, they do not further elaborate on the nature of cognitive processes or the severity of impairment. Without such knowledge, we can neither improve our theories about MBID nor develop effective interventions.

Research on cognitive processes in MBID has focused on verbal working memory (Henry and MacLean, 2002; Hasselhorn and Mähler, 2007) and executive functions (Danielsson et al., 2010). However, different verbal working memory processes have seldom been studied together in the same sample, making it difficult to disentangle process differences from sampling differences. Furthermore, group mean matching protocols that are insensitive to differential developmental patterns have been used. Therefore, in this paper we use a novel methodology called developmental trajectories to investigate two cognitive processes in the same sample regarding verbal working memory, namely the well-researched phonological rehearsal process and the less common lexical redintegration process. Thus, the paper's contribution is two-fold: (1) the functionality of redintegration is investigated in students with MBID for the first time; and (2) the methodological extension of developmental trajectories allows to reveal differential patterns of cognitive development.

Our focus on cognitive processes does not claim to causally explain why MBID originate in the first place; nor does it neglect the relevance of other factors such as socio-cultural deprivation or inadequate education or schooling. We believe that insights in the development of cognitive processes can increase our knowledge about MBID and have the potential to effectively change theoretical models of MBID and education or intervention.

### Working Memory Theories

In our study, we focus on the overall capacity of two retention processes within verbal *working memory* (WM), namely *rehearsal* and *redintegration*. An influential framework for WM was developed by Baddeley (1986, 2000, 2012) positing four components: (a) the central executive with an (attentional) regulatory function that uses two modality-specific sub-systems; (b) the phonological loop (PL) for verbal information; (c) the visuo-spatial sketchpad for visual information; (d) the episodic

buffer for binding information from different modalities. For typically developing children (TD), various studies established positive correlations of WM with academic outcomes (e.g., Alloway and Alloway, 2010; Poloczek et al., 2012; Mähler and Schuchardt, 2016), vocabulary learning (e.g., Gathercole et al., 1997; Baddeley et al., 1998), or thinking patterns such as inferencing and classification (Craig and Lewandowsky, 2013). Thus, a well-functioning WM is seen as an important prerequisite for learning and retaining memory traces until they can be incorporated in long-term memory (LTM).

In Baddeley's conceptualization of verbal WM, the PL consists of a phonetic store and a subvocal rehearsal process for retention. The phonetic store holds only a limited amount of information, mainly restricted by time due to decay or interference. Thus, the memory trace in the phonetic store degrades already after a few seconds (1.5–2.0 s; Hasselhorn and Mähler, 2007). To maintain the trace beyond the short retention of the phonetic store, the process of rehearsal has been proposed as a repeated subvocal articulation (Baddeley et al., 1984; Hasselhorn et al., 2000), i.e., the memorandum is pronounced silently. The amount of information that can be rehearsed depends on the articulatory speed and on the automatic activation of rehearsal (Hasselhorn et al., 2000). The general capacity of the PL results from the interplay between phonetic store and subvocal rehearsal.

The capacity of the PL and the effectiveness of the retention processes in maintaining the degrading trace are assessed in span tasks (Hasselhorn and Mähler, 2007), in which participants are asked to repeat a sequence of words or digits presented in a fixed rhythm of one element every 1–1.5 s. The highest number of elements recalled in their correct position serves as indicator of span capacity. Contrasting the performance for varying conditions of the recall stimuli (e.g., word length or lexicality) provides an insight into the effectiveness of the retention processes.

The automatic activation of rehearsal is traditionally operationalized as word length effect in word span tasks (e.g., Mähler and Hasselhorn, 2003): Short words (e.g., one syllable) are more likely to be remembered (correctly) than long words (e.g., three syllables) in span tasks (Baddeley et al., 1975). The word length effect is explained in terms of the effectiveness of rehearsal: Without rehearsal, only the last few seconds could be remembered in each condition alike. Silently articulating the stimuli is more effective for short words, because more short words than long words can be subvocally articulated in the same amount of time. Findings confirming that word span performance correlates with articulatory speed (Hasselhorn et al., 2000) and that the word length effect disappears under concurrent articulation (Baddeley et al., 1984) underline the phonological basis of rehearsal. In summary, the word length effect constitutes a seminal finding that supports the concept of rehearsal as an effective means for retaining a degrading memory trace in the PL of WM (but see Campoy, 2008 and Jalbert et al., 2011b).

Although verbal WM is originally conceptualized as only phonologically based, it is not independent of LTM. For instance, lexical attributes of stimuli in word spans, which are stored in LTM, have measurable effects on span performance. These

lexical factors include word-frequency (Hulme et al., 1997), wordlikeness in pseudowords (Gathercole, 1995), lexicality (Gathercole et al., 2001), concreteness (Walker and Hulme, 1999; Allen and Hulme, 2006), and neighborhood size (Jalbert et al., 2011b; Clarkson et al., 2017).

The original PL-model does not specify a mechanism to account for these LTM effects. Lexical redintegration has therefore been proposed as a process explicitly linking WM to LTM that can account for lexical effects of word span tasks (Lewandowsky and Farrell, 2000). Redintegration aids recall in WM by using lexical information available in LTM to reconstruct a partially degraded memory trace (Schweickert, 1993; Thorn and Frankish, 2005; Roodenrys and Miller, 2008; Thorn and Page, 2009).

Redintegration is investigated empirically through word span tasks that contrast stimuli of different lexical conditions such as real words vs. pseudowords (Gathercole et al., 2001; Turner et al., 2004). Items differ in how readily they can be redintegrated “as a result of their properties in long-term memory” (Roodenrys and Miller, 2008, p. 579). The differential impact of lexical interference on words vs. pseudowords can be interpreted as a marker for redintegration (Conlin and Gathercole, 2006). As real words exist in LTM, their knowledge can be used for reconstruction. This does not apply to pseudowords, meaning that retention depends more heavily on the phonological form (Thorn et al., 2009).

Schweickert’s (1993) model of multinomial processing trees incorporates both rehearsal and redintegration processes, although the reference to rehearsal remains implicit. If a memory trace is completely intact, it can be recalled directly. If, however, the trace is partially degraded, attempting to reproduce the recall series will lead to errors (Thorn et al., 2005). Via redintegration, the sequence can be reconstructed from the remainders of the degrading trace, based on pre-existing knowledge stored in LTM (Schweickert, 1993).

To conclude, the two WM processes of rehearsal and redintegration aid children in retaining memory traces that would otherwise quickly fade. It can hence be seen as a prerequisite for establishing stable LTM traces necessary for various academic tasks, e.g., learning new words where an unknown (phonological) word form has to be remembered, reading an instruction to understand the steps to solve a problem, or maintaining a sequence of steps during problem-solving.

## Working Memory Processes in Students With Mild and Borderline Intellectual Disabilities

To sum up, rehearsal and redintegration processes both play an important role in verbal working memory and thus in academic learning. The question is, though, how students with MBID use rehearsal and redintegration processes. Empirical findings regarding the PL and rehearsal in students with MBID are still inconclusive.

On the one hand, Hasselhorn and Mähler (2007) found a delayed development in the capacity of the PL for students with MBID that was in line with their slowed general intellectual

development. Also regarding the effectiveness and automatic activation of the rehearsal process, there is evidence for a delayed development in accordance with the general intellectual development in children with mild intellectual disabilities (van der Molen et al., 2007), in adolescents with MBID (Mähler and Hasselhorn, 2003), and also when using non-verbal recall (Poloczek et al., 2016).

On the other hand, there is evidence for a more severe deficit in the PL domain that goes beyond the general poor mental development (van der Molen et al., 2007; Schuchardt et al., 2010, 2011). In adolescents with MBID, the subcomponent of the phonetic store seemed to be specifically impaired (Mähler and Hasselhorn, 2003). Regarding rehearsal, Hasselhorn and Mähler (2007) found a specific deficit in students with MBID, and Rosenquist et al. (2003) in students with MID aged 12–16 years. Other studies suggest that PL deficits seem to be connected to the severity of the intellectual disability (Henry, 2001) and the general mental development (Mähler, 2007).

Lexical redintegration has not been investigated in the population of students with MBID to date. For students with difficulties in reading and writing, Hasselhorn et al. (2010) found a trend for stronger redintegration, while children with dyscalculia showed stronger rehearsal. Grube et al. (2008) attributed differences in the phonological similarity effect to redintegration in typically developing children of 5 and 9 years. Marton and Eichorn (2014) investigated interactions between WM and LTM (without referring to the construct of redintegration) in children with specific language impairment. Henry (2010) found the interaction between LTM and WM in students with mild to moderate intellectual disabilities (IQ range 39–70) to be in line with their general development.

To conclude, we still do not know whether rehearsal and redintegration are equally available to students with MBID. Therefore, we examine these two processes in the present study to gain a better understanding of their role for students with MBID. If cognitive processes in these students turn out to be intact, they can be excluded as a causal factor in developmental models of MBID. If, on the other hand, we find these processes to be impaired, they must be considered as possible explanatory factors. It could help understand why students with MBID are facing academic challenges, and offer possible indication of training and intervention to help these students cope with limited working memory processes.

## Approaches to Assess Developmental Patterns

Research on developmental disabilities seeks to identify differential patterns of development in relation to typically developing control groups, e.g., in the “Developmental-Difference-Controversy” (Zigler and Balla, 1982). A group of students with MBID is matched with two control groups of typically developing students. While the first control group is matched for chronological age (CA), the second control group is matched for mental age (MA), i.e., raw scores in an intelligence test. This results in two scenarios: If the MBID group performs more poorly than the CA group but equally well as the MA group

(MBID = MA < CA), the deficit in the target task is in line with their general lower cognitive performance, and a developmental *delay* is inferred. If the MBID group performs even worse than the MA group (MBID < MA < CA), this reveals a form of developmental *deviance*, as the performance in the target task is more affected than expected on the grounds of the general cognitive level. Thus, developmental delay can be excluded and a deviance in development is inferred.

Although this approach has driven a number of studies with important findings on the pattern of developmental disabilities, it has several shortcomings. These are described in more detail in Thomas et al. (2009), who therefore suggest *developmental trajectories* (DTs) as an alternative and more sophisticated approach of analysis. In brief, DTs aim “to construct a function *linking performance with age* on a specific experimental task and then to assess whether this function differs between the typically developing group and the disorder group” (Thomas et al., 2009, p. 336, italics added) and thus take variability within the groups into account. This procedure allows to fit a regression model for each group and task, and then to compare intercept and slope coefficients between the groups, tasks and their interactions.

While it is possible to include non-linear functions and zero-trajectories, the resulting three linear scenarios used for this study are: (a) *delayed onset*, which can be observed when the groups differ at the intercept, i.e., the onset of development; (b) *slowed rate*, which manifests as a difference in the slopes; and (c) a *combination* of delayed onset and slowed rate. Thomas et al. (2009) demonstrate that the traditional group matching approach cannot distinguish between the first two patterns. Furthermore, DTs allow to flexibly include different indicators of mental age (which we call *developmental indicators* [DI] in our study) to assess differential developmental relations, which is not possible in the group-matching approach either. Thus, the approach of DTs provides a richer taxonomy to describe how developmental pathways can be impaired (Thomas et al., 2009), which has been used in several studies to examine various pathologies (Annaz et al., 2008; Lei et al., 2011; Carney et al., 2013), but not for students with MBID yet. It should be noted that, following Thomas et al. (2009), this paper understands the term “development” as correlations with indicators for (mental) age, as data is cross-sectional and not longitudinal.

Given that all of the studies about MBID reported above have used a group-based matching approach, the mixed results might be at least partly due to methodological reasons. To better disentangle different developmental patterns of the PL capacity, rehearsal, and redintegration, the present study uses DTs as the methodology to establish whether students with MBID show differential patterns of verbal working memory development.

## Research Questions and Hypotheses

Does verbal working memory in students with MBID develop differently from typically developing students? We examine whether three different developmental indicators (chronological age, cognitive capacity, and vocabulary size) can predict the

capacity of the PL and the effectiveness of rehearsal and redintegration processes in students with MBID compared with TD peers. Per definition, students with MBID reveal a reduced cognitive capacity, making it necessary to control for it. Vocabulary size is taken as an indicator of crystallized knowledge that is stored in LTM, allowing us to investigate the developmental relation between vocabulary size and redintegration which relies on the interaction between WM and LTM.

1. **PL capacity:** We expect to find either a delayed onset or a slowed rate in the development of the PL capacity in students with MBID because deficits that are in line with mental age have been shown in the literature. However, as noted above, these scenarios are not distinguishable in the traditional ANOVA approach.
2. **Rehearsal:** The effectiveness of rehearsal is also hypothesized to be impaired. However, as previous findings are mixed, no clear hypothesis can be built as to whether to expect a delayed onset or a slowed rate of rehearsal development, or both. Concerning the developmental relation with vocabulary size as a lexical variable, we expect no relationship with the phonologically based rehearsal process.
3. **Redintegration:** Regarding the effectiveness of redintegration, there are no studies on students with MBID from which theoretical assumptions could be derived. Since redintegration represents the use of information from LTM to infer items in the short-term store, it may be a source of cognitive impairment in students with MBID. If this is the case, a reduced redintegration effectiveness would be a differential finding that is (so far) unique to MBID. Here, we expect vocabulary size to relate to the redintegration process.

## METHODS

### Participants

The sample included  $N = 210$  German students: 87 belonged to the group of MBID (mild intellectual disability, IQ 55–70: 24 students; borderline intellectual functioning, IQ 70–85: 63 students), of whom 48 were male with a mean age of  $M = 12$  years and 11 months,  $SD = 2.39$ , ranging from 7;4–17;1 years. One hundred and twenty-three students were in the typically developing (TD) group (51 male,  $M = 8$  years and 5 months,  $SD = 1.87$ , range 6;0–13;5 years). IQ norm data were obtained from different sources: The IQ scores of students with MBID were supplied by the official assessment documents in the school records; the IQs of TD students were assessed using the Culture Fair Test family, depending on their ages (CFT 1-R, Weiß and Osterland, 2012; and CFT 20-R, Weiß, 2008, respectively). Detailed sample characteristics are described in **Table 1**. For raw score comparisons on cognitive capacity and vocabulary size, all students with MBID but only those TD students that fell into the sensitive range of the tests explained below were tested, resulting in a subsample of  $n = 102$  TD students for mental-age comparisons.

To qualify for the MBID group, students had to fulfill the following criteria: a formal diagnosis of special learning needs (in

**TABLE 1 |** Sample characteristics and descriptive statistics: means and standard deviations of all variables for students with MBID and for TD students.

	MBID (n = 87)		TD (n = 123)		Significance	
	M	SD	M	SD		
Sex (male/female)	48/39		51/72		$\chi^2_{(1)} = 3.31; p = 0.069; \varphi = 0.14$	
Age (years; months)	12;11	2;6	8;5	1;10	$t^a_{(155.74)} = 14.9; p < 0.001; d = 2.1$	
Cognitive Capacity (CFT 1-R raw scores)	63.7	13.2	65.4 <sup>b</sup>	13.5	$t_{(180)} = 0.85; p = 0.395; d = 0.13$	
Vocabulary size (WWT 6–10 raw scores)	53.1	15.9	54.7 <sup>b</sup>	14.6	$t_{(180)} = 0.70; p = 0.483; d = 0.10$	
Intelligence (IQ norm scores)	74.3	7.3	105.3	10.2	$t^a_{(207.98)} = 25.7; p < 0.001; d = 3.60$	
<b>WORD SPAN PERFORMANCE CONDITION</b>						
Word span Real words	Short	3.27	0.57	3.47	0.64	$t_{(208)} = 2.43; p = 0.016; d = 0.35$
	Long	2.71	0.44	2.79	0.55	$t_{(208)} = 1.18; p = 0.240; d = 0.17$
	Total real words	2.99	0.46	3.13	0.55	$t_{(208)} = 2.03; p = 0.043; d = 0.29$
Word span Pseudowords	Short	2.62	0.66	2.76	0.74	$t^c_{(204)} = 1.38; p = 0.169; d = 0.20$
	Long	1.71	0.43	1.79	0.54	$t^c_{(203)} = 1.09; p = 0.277; d = 0.16$
	Total pseudowords	2.16	0.51	2.28	0.59	$t^c_{(203)} = 1.44; p = 0.151; d = 0.21$

The Culture-Fair-Test 1 Revision (CFT 1-R; Weiß and Osterland, 2012) raw scores were used to assess cognitive capacity as developmental indicator (DI); the picture naming vocabulary test “Wortschatz- und Wortfindungstest für 6–10-Jährige” (WWT 6–10, Glück, 2011) raw scores were used to assess vocabulary size as DI; Intelligence norm values were obtained from different sources: for students with MBID from the official school records, for younger TD students the corresponding CFT 1-R was administered and norm values were calculated, and for older TD students (ten years and older) the CFT 20-R (Weiß, 2008) was administered and norm values were calculated.

<sup>a</sup>dfs are corrected due to unequal variances.

<sup>b</sup>Subsample of n = 102 students in the TD group to avoid ceiling effects.

<sup>c</sup>Drop-out of five students (2 with MBID, 3 TD) who did not complete the session with word span tasks with pseudowords.

German “Sonderpädagogischer Förderbedarf im Lernen”); an IQ below 85 as measured during the formal special needs assessment; no other developmental disorders, such as ADHD, ASD, and specific learning disabilities (dyslexia, dyscalculia), according to teacher report. All students with MBID were sampled from four special educational needs schools (“Förderschule Förderschwerpunkt Lernen”) in an urban environment in Western Germany. These schools constitute a special institution for students who reveal severe and continuous difficulties in academic learning, leading to the diagnosis of “special educational needs in learning.” Class size typically does not exceed 15 students. Students of the TD group attended a mainstream school (primary school or secondary school), had no diagnosis of special educational needs or developmental disorders, and their IQ and vocabulary scores had to be at least average (i.e., IQ > 85; vocabulary T-Score > 40). Class size typically ranges between 20 and 28 students in primary and 28–31 students in secondary grades. Participants were excluded if they could not follow the instructions. Data regarding monolingual status were not obtained; however, none of the participating children revealed difficulties in understanding the instructions, according to the administrators’ judgement.

The study was carried out in accordance with the recommendations of the German Psychological Society, i.e., that written informed consent be obtained from all subjects’ parents or caregivers in accordance with the Declaration of Helsinki. All participants gave their oral agreement to participate, and their parents and/or legal guardians were informed of the objectives of the study, the nature of the tasks that would be administered, and the fact that they could withdraw their agreement at any time. Permission to conduct tests in schools was obtained from the school principals.

## Procedure and Materials

Each student was tested in a group session on cognitive capacity, and in two individual sessions on the experimental tasks lasting ~30 min each. For the TD group, the first session contained the WM tasks. The second session involved a picture-naming task on vocabulary size. For the MBID group, the structure was slightly rearranged to enhance compliance: the first session involved only one WM task and the second session one WM task and the vocabulary task. The TD students at secondary schools completed the vocabulary task in a group setting.

### Word Span Tasks

We used four different word span tasks in 2 (Length: short vs. long) × 2 (Lexicality: real vs. pseudo) conditions (e.g., short real word span: “Haus–Stern–Schuh” [house, star, shoe], or long pseudo word span: “karflumen–franulich–wuralten”). For the real word condition, we used the words from the standardized German Working Memory Test Battery (AGTB 5-12, Hasselhorn et al., 2012), and for the pseudoword condition, we used the pseudowords from a study by Hasselhorn et al. (2010; see **Table A1** in the Appendix). The construction of word span sequences in all four conditions followed the principles of the AGTB 5-12. Sequences varying from two to eight words in length were presented in a 1.5 s rhythm. Each sequence was randomly assembled from a pool of nine words, spoken by a female voice, considering that the order of words within a sequence did not resemble other sequences. For each sequence length (i.e., number of words per sequence), 11 different sequences with pseudorandomized word order were provided in a list of audio files on a PC, from which they were presented aurally. The number of items per sequence was adaptively chosen

based on the child's previous performance (detailed procedure described below). The students were instructed to listen to the sequence until the final signal and to repeat the complete sequence as correctly as possible. This procedure was practiced in three trials before the actual task. While the order of lexicality was randomly assigned, the order of length conditions was fixed across all individuals: Short words were followed by long words. For the pseudowords, children were told that the words might sound unfamiliar "like in a secret language of aliens."

The word span tasks were adaptive. In two calibration items at the beginning of each condition, sequence length was adapted after every item. If it was repeated correctly, length was increased by one word; if it was not correct, length was decreased by one. The minimum length of a sequence was two words. The following eight test items were used for calculating the word span score and adaption took place after every other item: If Sequences 3 and 4 were both correct, the next two sequences would be one word longer; if both were incorrect, the next two would be one word shorter; if one was correct and one incorrect, length remained at the current level. Correct repetitions were awarded points corresponding to the length of the sequence, an incorrect repetition received points equal to the sequence length minus 1 (e.g., incorrect repetition of a four-word sequence was awarded three points). The overall word span score per condition was computed as the mean of the points of the eight test items. If all sequences were repeated incorrectly, the total score would be "1"; the theoretical maximum would be "7.5" when all test items were repeated correctly. We computed the internal consistency of the four condition scores (short real words, long real words, short pseudowords, long pseudowords) to be  $r = 0.83$  for TD children and  $r = 0.84$  for students with MBID.

### Cognitive Capacity

The CFT 1-R (Weiß and Osterland, 2012) was administered to all students with MBID and a subsample of younger students in the TD group ( $n = 102$ ). The CFT 1-R is a language-free intelligence measure, consisting of six subtests: substitutions, labyrinths, similarities, continue sequences, classification, and matrices. Norms are provided for age 5;4–9;11, with retest reliability reported to be at  $r_{tt} = 0.90$  and internal consistency at  $r = 0.97$ . As the raw score is used as a matching criterion to investigate developmental relations, norm scores are not necessary and therefore it is not problematic that the age of the MBID group exceeds the range of age for which the CFT 1-R is normed. Instead, it is more relevant that the raw scores fall within a sensitive range, i.e., that no ceiling or floor effects exist. In the primary schools, the CFT 1-R was administered as a group test of up to 15 students. The setting in the special schools (MBID group) was in single sessions or small group sessions (maximally 4 students). The group of older TD students (secondary school) was tested in a group session on the CFT 20-R (Weiß, 2008), which provides norms for their age range (8;5–60;0 years), to ensure that these students met the inclusion criterion (no intellectual disabilities). Retest-reliability is  $r_{tt} = 0.80$  and internal consistency  $r = 0.95$ .

### Vocabulary Size

Students with MBID and the younger TD students were tested at an individual session with a computer-based test on expressive vocabulary in German for primary students aged 6 to 10 years (*Wortschatz- und Wortfindungstest 6–10*, WWT 6–10, Glück, 2011). The test consists of 95 images representing (parts of) objects, activities, opposites, and categories. The child is prompted: "what is this; what is he/she doing; what is the opposite of...; what are these things altogether." Each item is shown on the screen for maximally 15 s. Reliability estimates are reported as  $r_{tt} = 0.96$ . The older TD group (secondary school) completed the vocabulary test that is part of the CFT 20-R (Weiß, 2008) during the group session. For a given word, a synonym has to be marked from a choice of 5 options (e.g., target-word: "fantasy." Options: "form," "principle," "illusion," "imagination," "apprehension") over 30 items; reliability is reported in the manual to be  $r_{tt} = 0.87$ . Note that the latter test was only used for sample selection purposes, to ensure that older control students' vocabulary was in the normal range. The group comparisons on vocabulary size were solely carried out based on the expressive WWT 6–10 raw scores that were completed by a subsample of  $n = 102$  TD students.

### Analyses

The analyses of developmental trajectories (DT) follow a series of steps of increasing complexity. The results are depicted in scatterplots with regression lines and confidence intervals. Regression analyses are carried out to test whether the effects are statistically significant (Thomas et al., 2009). In the following analyses, the word span score is always the dependent variable. In all analyses, we establish three kinds of predictors: (a) three continuous predictors for (mental) age, which we call *developmental indicators* (DIs); (b) one between-subjects predictor (*group factor*); and (c) two within-subjects predictors, called *task conditions*. In total, we construct nine developmental trajectories: one DT across groups for all three dependent variables for each of the three DIs.

The three DIs comprise chronological age (CA) and two mental age variables: cognitive capacity (COG) and vocabulary size (VOC). The DIs establish whether a measure of chronological or mental age can reliably predict performance in the word span tasks. The group factor differentiates between students with MBID and the TD group for two purposes: first, to test whether the groups differ in their overall word span scores, and second to investigate interactions with DIs and task conditions. Lastly, the task condition predictors encompass the  $2 \times 2$  different word length (short vs. long) and lexicality (real vs. pseudo) conditions. Through a task effect for word length (short words are more likely to be recalled than long words), rehearsal can be measured; the task effect for lexicality (real words are more likely to be recalled than pseudowords) estimates redintegration.

The first step of constructing a DT is to fit a regression model with the word span performance as dependent variable and a DI as predictor. This indicates whether there is any reliable relationship between the DI and word span performance. In the second step, the group factor is added to compare the groups regarding their intercepts and slopes, which are estimated

separately for each group. The main effect for group denotes whether a difference exists in the intercepts between the groups. The interaction Group × DI considers the difference in slopes to establish whether the groups differ in their correlation between DI and word span performance (i.e., whether students with MBID catch up or if the gap widens). As the third step, one or more task conditions can be included to determine whether intercepts and slopes differ for various types of tasks. Again, the main effect for task reveals whether there is a general advantage for e.g., short words over long words (word length effect) or real words over pseudowords (lexicality effect) at the intercept. The interaction Task × DI shows the difference in slopes indicating that the relationship between DI and word span performance differs for the conditions of the task (e.g., if the advantage for short words increases with age). As the fourth step, the full model with all three predictors (task, group, and DI) is analyzed to establish whether there are differential developmental patterns for different task conditions. The Task × Group interaction denotes the differences of intercepts, whether e.g., the word length effect (advantage of short words over long words) is different between the groups, and hence show whether students with MBID show a delayed onset of development. Finally, the triple interaction Task × Group × DI represents the difference in slopes indicating that the relationship between DI and task effects differs between the groups (e.g., whether the advantage for short words over long words increases with age in both groups equally). This effect denotes whether students with MBID show a slowed rate of development in comparison to the TD group.

We will compare the respective intercept and slope effects for the dependent variables capacity of the PL (Hypothesis 1), effectiveness of the rehearsal process (Hypothesis 2), and effectiveness of the redintegration process (Hypothesis 3). For each dependent variable, we report the developmental relations with all three DIs; of particular interest is the developmental relation between vocabulary size and redintegration (Hypothesis 3). To analyze the capacity of the PL, the dependent variable of word span scores is predicted by the three different DIs and the group factor, without accounting for the task conditions; thus, word span scores are averaged over all 2 × 2 task conditions. Rehearsal is operationalized by the word length effect. Therefore, the section on rehearsal takes word length as predictor into the model, and word span scores are averaged over both lexicality conditions. Analogously, lexicality is taken as predictor into the model to investigate redintegration. Here, word span scores are averaged over both length conditions.

For the computation of regression coefficients, the developmental indicators are linearly rescaled such that the minimum value of the group of children with MBID is located at 0. This only affects the size of the intercept coefficient to allow a meaningful interpretation of the intercept difference between groups (Thomas et al., 2009), because extrapolation beyond the range of the data is avoided. The slope effect is not affected by this transformation. Rescaling is carried out only for the purpose of the statistical calculation of intercept effects, while the original ages are still shown in the figures. The detailed regression models' statistics ( $F$ ,  $p$ ,  $\eta^2$ ) can be found in **Table A2** in the Appendix; as for readability, only  $p$ -values are reported in the text.

Data were prepared and figures created in R (R Studio Team, 2016; R Core Team, 2018) using the packages *ggplot2* (Wickham, 2011) and *ggpubr* (Kassambara, 2017). Regression analyses of DTs were performed in SPSS Version 25 (IBM Corp, 2017) following the procedure outlined in the electronic supplement in Thomas et al. (2009).

## RESULTS

Sample characteristics and descriptive results of mean word span scores for short and long real- and pseudoword conditions are provided in **Table 1**. Correlations between the three DIs (chronological age, cognitive capacity, vocabulary size) can be seen in **Table 2**. The order of lexicality in the WM task was randomly chosen in a subsample yielding no significant effects whether real words were administered first or second,  $t_{(118)} = 0.20$ ;  $p = 0.844$ . To test possible gender effects, two linear mixed models were estimated for each DI. A baseline model included all predictors (i.e., DI, group, length, and lexicality), and the full model additionally included the factor gender. Models were compared regarding their fit using a Chi-square statistic. None of the three models gave evidence for a significant gender effect: chronological age,  $\chi^2_{(16)} = 14.525$ ;  $p = 0.560$ ; cognitive capacity,  $\chi^2_{(16)} = 13.418$ ;  $p = 0.642$ ; and vocabulary size,  $\chi^2_{(16)} = 12.642$ ;  $p = 0.699$ . Thus, data were pooled across both gender groups.

### Capacity of the Phonological Loop

In the first analysis (**Figure 1**), we examined whether students with MBID show differential developmental patterns regarding the general capacity of the PL. This answers the question as to whether the development of PL capacity in students with MBID starts at the same onset level as in TD students; and we can determine if students with MBID tend to catch up with TD students on a possible PL capacity deficit with increasing age, cognitive capacity or vocabulary size; or if they continue to lag behind TD students. PL capacity was measured as the overall word span scores averaged over all 2 × 2 task conditions. Thus, only the group factor and the DIs were taken into the model as predictors, without distinguishing

**TABLE 2 |** Correlations for developmental indicators for students with mild and borderline intellectual disabilities (MBID below the diagonal) and typically developing students (TD above the diagonal).

	TD (n = 102)		
	Chronological age	Cognitive capacity	Vocabulary size
MBID (n = 80) Chronological age	–	0.74	0.73
Cognitive capacity	0.42 <sup>a</sup>	–	0.67
Vocabulary size	0.45 <sup>a</sup>	0.49 <sup>b</sup>	–

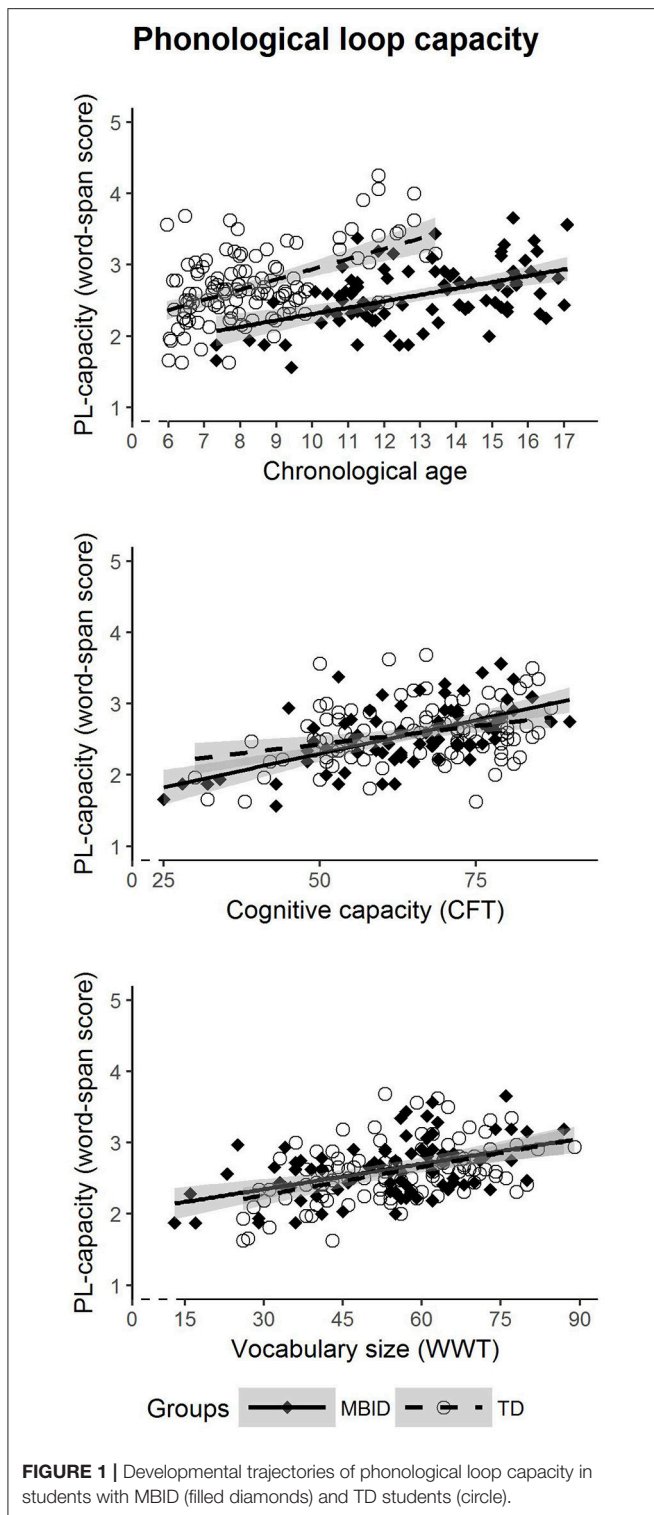
*Pearson correlations between developmental indicators (all  $p < 0.001$ ).*

*Due to unsystematic missing data in the MBID group, sample size is smaller than the full sample of n = 87 children.*

<sup>a</sup>n = 80.

<sup>b</sup>n = 73.





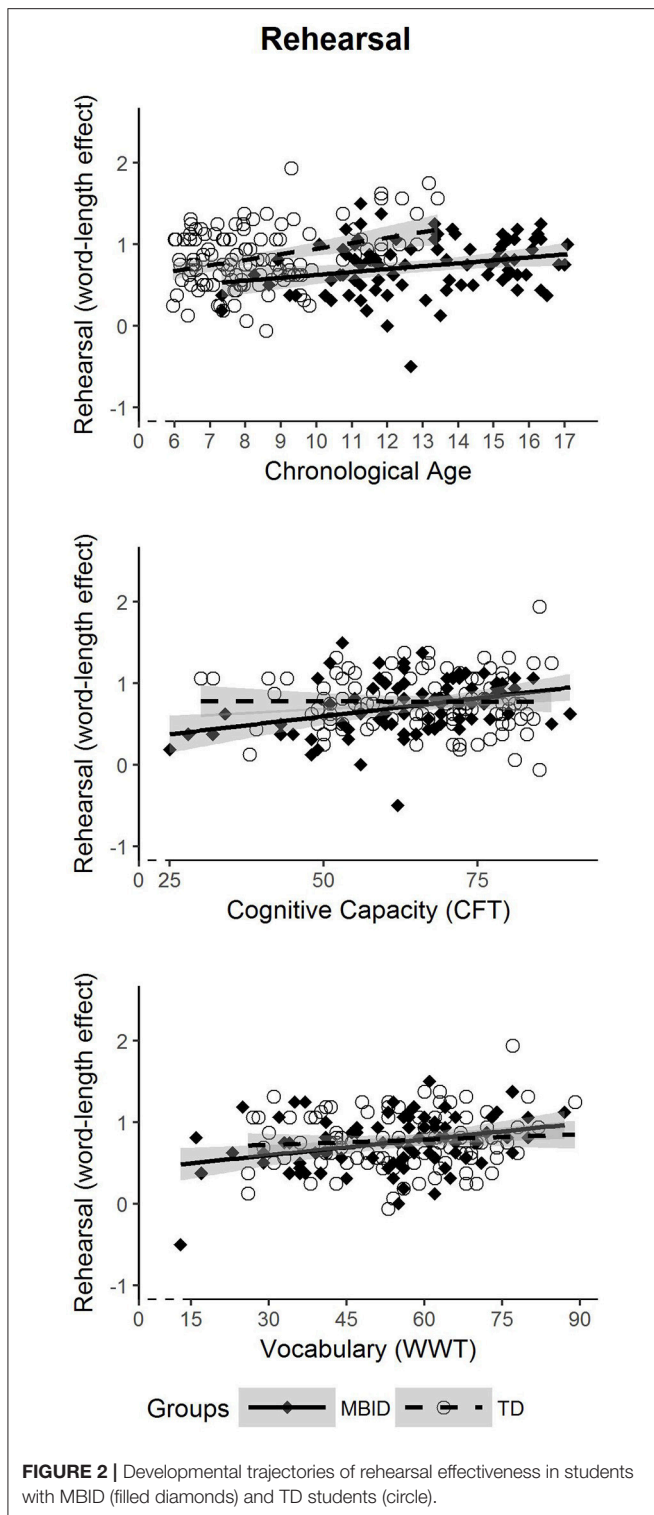
between different levels of length and lexicality. For each DI (chronological age, cognitive capacity, vocabulary size), we compared the groups concerning the intercepts (measured as group main effect) and the slopes (measured as Group × DI interaction).

When using *chronological age* as DI to predict PL capacity, we found a significant intercept effect for group ( $p < 0.001$ ), as students with MBID had a lower intercept. The interaction of Group × Age, denoting a slowed rate of development, did not reach significance ( $p = 0.068$ ). This means that students with MBID show a delayed onset in the development of PL capacity, but their rate of development in relation to chronological age was not significantly lower. With *cognitive capacity* as DI to predict PL capacity, we found a significant intercept difference for group ( $p = 0.039$ ), and a significant interaction Group × COG ( $p = 0.033$ ) showing a difference in slopes. The onset of development was significantly delayed and the slope for students with MBID was significantly steeper, indicating a stronger relationship between cognitive capacity and general PL capacity than in TD students. Thus, students with MBID reveal a delayed onset, in combination with a differential rate in PL capacity development in relation to their cognitive capacity. However, it should be noted that cognitive capacity scores may be confounded with chronological age; removing younger students with lower cognitive capacity scores (CFT 1-R raw scores < 45) from the analyses left both effects nonsignificant, i.e.,  $p = 0.081$  for the intercept difference and  $p = 0.073$  for the difference in slopes. For *vocabulary size* as DI to predict PL capacity, neither the intercept difference for group was significant ( $p = 0.331$ ) nor did the slopes differ (Group × VOC;  $p = 0.595$ ). This indicates that the developmental pattern of PL capacity does not differ in relation to vocabulary size.

In summary, regarding the general capacity of the PL as measured by the word span scores across all task conditions, students with MBID revealed only a delayed onset of their PL but showed the same rate of development over the chronological age range. Cognitive capacity turned out to be a stronger predictor for PL capacity in students with MBID than in the TD group; this did not hold, however, when children with lower raw scores of cognitive capacity were removed from the analysis. No differential effects could be observed in terms of vocabulary size on PL capacity. Vocabulary size as DI does not produce any differential patterns between the MBID and TD groups.

### Effectiveness of the Rehearsal Process

In the second analysis (Figure 2), we examined whether students with MBID show differential developmental patterns regarding the effectiveness of rehearsal. As in the first research question, this answers the question as to whether the effectiveness of rehearsal in students with MBID starts at the same onset level as in TD students; and we can determine whether developmental progress (in terms of age, cognitive capacity and vocabulary size) reflects in a similar way onto rehearsal development as in TD students. As rehearsal is operationalized as word length effect (i.e., a relative benefit for short words over long words), the task condition length is also included as predictor besides the group factor and the DIs, but without distinguishing between different levels of lexicality. For ease of interpretation, Figure 2 shows the size of the word length effect (the difference between short and long words) in the y-axis, whereas in the regression model, length is included as predictor.



Using *chronological age* as DI to predict rehearsal, a significant intercept effect for Length  $\times$  Group ( $p = 0.021$ ) was found, while the interaction of Length  $\times$  Group  $\times$  Age, denoting the rate of development, was not significant ( $p = 0.210$ ). This means that students with MBID show a delayed onset, but no slowed rate

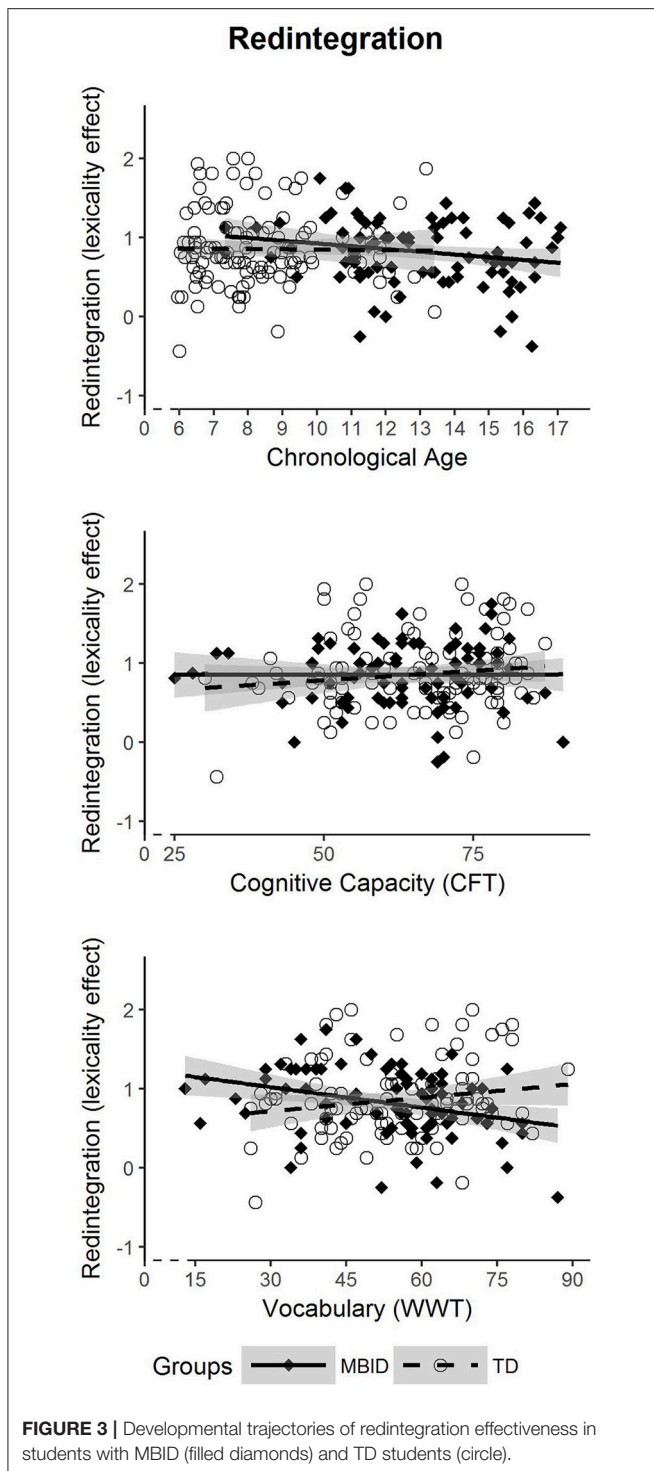
in rehearsal development in relation to chronological age. With *cognitive capacity* as DI to predict rehearsal, the Length  $\times$  Group interaction ( $p = 0.009$ ) yielded a difference in intercept, and a significant interaction Length  $\times$  Group  $\times$  COG ( $p = 0.016$ ) showed a difference in slopes. The slope for students with MBID was significantly steeper, indicating a stronger relationship between cognitive capacity and rehearsal effectiveness than in TD students. Thus, students with MBID reveal a delayed onset, in combination with a steeper rate in rehearsal development in relation to cognitive capacity. Again, removing younger students with lower cognitive capacity scores (raw scores  $< 45$ ) from the analyses left both effects nonsignificant, i.e.,  $p = 0.147$  for the intercept difference and  $p = 0.167$  for the difference in slopes. *Vocabulary size* as DI did not predict differences between the groups. Neither the intercept difference (Length  $\times$  Group;  $p = 0.212$ ) nor the slope difference (Length  $\times$  Group  $\times$  VOC;  $p = 0.192$ ) became significant. This shows that the developmental pattern of rehearsal effectiveness does not differ in relation to vocabulary size.

To summarize, rehearsal development is delayed in onset for students with MBID when using chronological age as developmental indicator. A differential rate of rehearsal development in students with MBID when accounting for cognitive capacity as DI should be interpreted with caution, as removing children with lower COG scores removed the intercept and slope differences, suggesting that the relationship between rehearsal and cognitive capacity is similar in both groups. In neither group did lexical knowledge (vocabulary size) have any significant predictive power.

### Effectiveness of the Redintegration Process

In the third analysis (Figure 3), we examined whether students with MBID show differential developmental patterns regarding the effectiveness of redintegration. As before, this answers the question as to whether the effectiveness of redintegration in students with MBID starts at the same onset level as in TD students; and we can determine whether developmental progress (in terms of age, cognitive capacity and vocabulary size) reflects in a similar way onto redintegration development as in TD students. As redintegration is operationalized as lexicality effect (i.e., a relative benefit for real words over pseudowords), the task condition lexicality is also included as predictor besides the group factor and the DIs, but without distinguishing between different levels of word length. For ease of interpretation, Figure 3 shows the size of the lexicality effect (the difference between real words and pseudowords) in the y-axis, whereas in the regression model lexicality is included as predictor.

Using *chronological age* as DI to predict redintegration, neither a significant intercept effect for Lexicality  $\times$  Group ( $p = 0.530$ ), nor a significant interaction of Lexicality  $\times$  Group  $\times$  Age ( $p = 0.617$ ) to reflect the rate of development was found. This means that students with MBID show a similar developmental pattern to TD students regarding redintegration development in relation to chronological age. With *cognitive*



capacity as DI to predict redintegration, the Lexicality × Group interaction ( $p = 0.718$ ) showed no significant difference in intercept, and a nonsignificant interaction Lexicality × Group × COG ( $p = 0.658$ ) indicated no difference in slopes. Thus, the development of redintegration also seems to be unimpaired in students with MBID in relation to cognitive capacity. In contrast, using vocabulary size as DI to predict redintegration

yielded both a significant difference in intercept (Lexicality × Group;  $p = 0.024$ ) and in slope (Lexicality × Group × VOC;  $p = 0.005$ ). The intercept difference favors students with MBID, who show greater redintegration when vocabulary size is low. In contrast, redintegration effectiveness decreases with growing vocabulary size. This reveals an unexpected differential pattern of redintegration development. Separate analyses per group indicated that in TD students, there was no significant slope effect,  $F_{(1, 100)} = 2.062$ ;  $p = 0.154$ , while for students with MBID, the negative slope effect was significant,  $F_{(1, 78)} = 8.254$ ;  $p = 0.005$ . This negative slope remained significant after one possible outlier with a high vocabulary score and low redintegration in the MBID group had been excluded from the analysis,  $F_{(1, 77)} = 5.267$ ;  $p = 0.024$ .

Regarding the effectiveness of redintegration in MBID, findings are mixed. On the one hand, there seems to be no general differential developmental pattern, as we found similar onsets and slopes between the groups when chronological age and cognitive capacity were used as DIs. On the other hand, the relationship between redintegration and vocabulary size is differentially inverse: The greater the vocabulary size in students with MBID, the weaker their redintegration effectiveness, even after one possible outlier had been excluded. This result is clearly unexpected and deserves further discussion.

## DISCUSSION

### Development of Working Memory Processes in Students With MBID

The aim of this study was to determine whether students with MBID showed differential developmental patterns in three aspects of verbal WM: (a) the capacity of the PL, (b) the effectiveness of rehearsal, and (c) the effectiveness of redintegration.

We found that in relation to *chronological age*, students with MBID had a lower developmental onset in PL capacity than their TD peers, as indicated by the significant intercept difference. Thus, students with MBID have a smaller PL capacity from early on. This finding is in line with other studies in the field (Mähler and Hasselhorn, 2003; Mähler, 2007; Schuchardt et al., 2010; Poloczec et al., 2014, 2016). Concerning the growth of PL capacity, development was not slowed in rate, as indicated by the nonsignificant slope effect for the interaction with age. This means that the PL capacity of students with MBID develops at a similar rate as in TD students, but they also did not catch up with their TD peers in the range of the current sample. However, we cannot exclude the possibility that they will at a later point in their development.

When PL capacity was set in relation to *cognitive capacity*, students with MBID revealed a steeper slope than TD students, which appears counterintuitive at first glance. Such an accelerated rate has also been reported in Henry (2001) and Schuchardt et al. (2011). One possible simple explanation could be that students with MBID might catch up in their development of PL capacity when they mentally mature (but see Colom et al., 2010), while the TD group may already have reached a plateau and does not further increase beyond the given level.

The findings on the relationship with cognitive capacity should be interpreted with caution, as removing children with low raw scores changed the results to nonsignificant differences. Thus, replicating studies are necessary before conclusions can be drawn.

Trajectories for PL capacity as predicted by *vocabulary size* as DI were equivalent for students with MBID and TD children, providing evidence that the development of the overall functioning of the PL in relation to LTM knowledge did not differ across groups. In both groups, therefore, PL capacity development is in line with the increasing vocabulary size.

The processes of rehearsal and redintegration were examined through the effects for word length and lexicality, respectively. When comparing the word length conditions (short vs. long words, across both lexicality conditions), we found evidence for an automatic activation of rehearsal in both groups. It should be noted, however, that this only holds for the conceptualization of rehearsal activation in Hasselhorn et al. (2000). Recent studies have challenged the role of rehearsal in producing the word length effect, and Poloczek et al. (2014) found evidence that the occurrence of a word length effect is also dependent on output delays.

We found a significant overall word length effect, as short words were more likely to be remembered than long words. Students with MBID revealed a delayed onset in rehearsal, as their advantage for short words was smaller than for the TD group. The effectiveness of rehearsal seemed to develop similarly with *chronological age*, meaning that older children showed stronger rehearsal in both groups equally, as there was no difference in the developmental rate of growth in rehearsal between the groups. This is in line with Mähler and Hasselhorn (2003) finding a delay but no deviance in adolescents with MBID regarding rehearsal, and other studies that found a developmental delay for students with MBID (Henry, 2001; Hasselhorn and Mähler, 2007; Mähler, 2007). It should be noted, however, that the word length effect can be subject to a proportional scaling artifact: as the word length effect is a relative measure of difference between short and long conditions, Jarrold et al. (2015) argue that it should be set in relation to the general performance, i.e., performance in the condition of short real words. Such an artifact cannot be dismissed in the present case and is out of scope of our current investigation.

Students with MBID appeared to be delayed in the onset of development when considering their *cognitive capacity*. While in the full sample, increasing cognitive capacity did not predict rehearsal effectiveness for the TD group, students with MBID increased on rehearsal effectiveness. Maybe, students with MBID catch up in their rehearsal development as they cognitively mature. As with PL capacity, this finding should be interpreted with caution, as children with lower scores in cognitive capacity seemed to affect the results. When they were removed from the analysis, there was neither a group difference in intercept nor in slope, suggesting that students with MBID and the TD group are equal in their developmental relation between rehearsal and cognitive capacity. Further studies should explore

this effect in more depth before meaningful conclusions can be drawn.

Rehearsal effectiveness depended in neither group on the *vocabulary size*, and the patterns did not differ in onset and rate of change. This was a plausible finding that vocabulary size, intended to measure lexical LTM-knowledge of words, did not predict the phonologically-based rehearsal process.

Regarding redintegration, a general benefit for real words over pseudowords could be observed, as real words are accessible in LTM making reconstruction easier. Thus, we replicated the effect that WM word span tasks depend on LTM. Regarding *chronological age*, development onset and rate of change were similar across both groups, which means that students with MBID presented neither a delayed onset nor a slowed rate of development concerning redintegration. *Cognitive capacity* did not predict redintegration effectiveness in either group. In contrast, *vocabulary size* as developmental indicator produced differential patterns in predicting redintegration. Students with MBID presented a higher effectiveness of redintegration at the lower end of vocabulary size and, as vocabulary size increased, became less effective at redintegration. Thus, students with MBID can redintegrate less efficiently than their vocabulary size would predict.

When looking at the relationship with chronological age (and when using the traditional group-matching approach), redintegration seemed to be unimpaired in students with MBID. This can be interpreted as a point of strength, indicating that they were able to make use of LTM traces for reconstruction in a similar way to their TD peers. This was qualified, however, by the negative relationship between redintegration and vocabulary size, representing a specific differential developmental pattern for MBID, because they did not seem to benefit from a higher number of words from which they could redintegrate. One should expect that a higher number of words available in the LTM would lead to a more efficient redintegration, as redintegration is conceptualized to use existing knowledge to reconstruct the trace in the short-term memory. This was not the case in students with MBID; on the contrary, a growing vocabulary size appeared to have somewhat detrimental effects on their use of LTM knowledge for reconstructing traces in the PL. Future research will have to establish whether this poses a problem of inefficient strategy use (e.g., Klauer and Lauth, 1997), poor source monitoring (e.g., Lilienthal et al., 2015), or suboptimal organization of the mental lexicon (e.g., Vitevitch et al., 2012; Kenett et al., 2016) with a greater vocabulary size causing “confusion,” or of other processes at retrieval (e.g., Dell and O’Seaghdha, 1992).

DTs served as an analysis tool to detect developmental similarities and differences concerning the WM processes of rehearsal and redintegration. They proved particularly useful in the inclusion of developmental indicators to describe and compare the development across a wider age range and investigate the relationship with other developmental indicators (in this case, cognitive capacity and vocabulary size). We would like to emphasize three innovations: First, the differentiation of a delayed onset and slowed rate for rehearsal could not be shown

by the traditional group-matching ANOVA approach. Second, DTs yielded differential developmental relations for students with MBID: Cognitive capacity did not predict rehearsal in either group when taking possible outliers into account (in the full sample of students with MBID, cognitive capacity positively predicted rehearsal), but vocabulary size negatively predicted redintegration. These detailed patterns could not be detected by the ANOVA approach either. Third, the use of a greater age range proved useful for obtaining a broader picture of development instead of comparing clusters at selected narrow age slots.

While there are several advantages to the DT approach, it should be explicitly noted that in this study, the design was cross-sectional. Therefore, “development” should not be taken literally, as we measured different individuals of a wider range of chronological age, cognitive capacity, and vocabulary size. This cross-sectional type of data is similar in the standard mental age matching approach; one important difference concerns the broader range of (mental) age, which is an explicit strength of the DT approach (Thomas et al., 2009, p. 343). A longitudinal design, which can be readily implemented in the DT framework, would be able to depict “true” development. Using correlational data on the relationship between a target task and variables taken as developmental indicators can be considered a first step requiring further validation through longitudinal designs. As Thomas et al. (2009, p. 336) point out, an “initial cross-sectional design [is ideally combined] with a longitudinal follow-up,” calling for further research.

It should be noted that the reasons why a particular developmental pattern occurs are not fully clear. DTs have been primarily used as “a descriptively more powerful empirical vocabulary” and a richer tool for the description of developmental pathways (Thomas et al., 2009, p. 355), which can even be understood as a “theory-neutral marker of atypicality” (Thomas et al., 2009, p. 340). A more general view on the underlying neuroconstructivist framework can be found in Karmiloff-Smith (1998), which provides a general theoretical context but does not make specific assumptions about which factors might be the underlying causes of different developmental patterns.

## Limitations

This study is subject to methodological and theoretical limitations. The first limitation concerns the coverage of age range. Thomas et al. (2009) recommend that the TD group should cover the whole age range from the youngest mental age to the oldest chronological age. The TD group in our sample only ranged up to the age of 13;5 years (instead of 17;1 years in the group of students with MBID). Therefore, the design may be unable to detect a possible plateau or other change in developmental rate in the TD group. While this might limit the explanatory power for the range of the older students, the general findings were not affected: The intercepts are compared at the onset of development, i.e., at the youngest age measured in the MBID group, where enough data are provided. This is similarly true for the interpretation of the comparison of slopes: The more crucial comparison is at the earlier level of development,

which is provided by the younger students in the TD group. It is useful to cover a wider age range in the MBID group, as it may require a longer time period for development to be detectable. This limitation is only relevant when chronological age is used as predictor. For the DTs cognitive capacity and vocabulary size, a subsample of the TD group in the sensitive range of the measures was used for comparison, thus spanning the whole range of the MBID group.

Second, methodological considerations must be taken into account. As students with MBID were recruited solely from special educational needs schools, findings are limited to this particular population and should not be generalized to a broader MBID population. Measurement invariance across the groups may not have been given in the different tasks, possibly affecting findings. Also, floor effects in long pseudowords, especially for (mentally) younger individual students of either group, may negatively impact reliability.

Third, besides word length and lexicality, many more influences have been shown to influence performance in verbal WM tasks. This poses a concern for the validity of the word length effect as a theoretically sound indicator of a phonologically, time-based rehearsal process. In particular, Jalbert et al. (2011a,b) showed that in all studies discovering an effect for word length, test items were confounded with neighborhood size. Neighborhood size is the number of “word[s] of the same length as the target that differs by only one letter.” (Jalbert et al., 2011b, p. 340). According to Roodenrys (2009), the neighborhood size effect can be explained in terms of a redintegration process (Clarkson et al., 2017; Derraugh et al., 2017). When experimentally controlling for neighborhood size, Jalbert et al. (2011b) succeeded in removing or even inverting the word length effect. This issue should be considered in future studies to construct stimuli to avoid confounding neighborhood size with word length.

## Implications

In this study, we investigated the capacity of the phonological loop and the effectiveness of two processes within the PL, rehearsal and redintegration, in students with MBID. In order to better guide and offer adequate instruction to these children with severe academic challenges, it is necessary to understand what cognitive processes they can rely on and in which domains they have particular difficulties. This knowledge may be used to adapt instruction, e.g., by making instructions shorter so that they do not depend as heavily on rehearsal. As theoretical models on the etiology of MBID do not specify which cognitive processes might be impaired, this study marks an attempt to shed some light on the availability of verbal working memory capacity and processes. Using developmental trajectories as analytical strategy, we showed that children with MBID seemed to have a delayed onset of PL capacity and rehearsal effectiveness. The rate of growth does not appear negatively affected, suggesting that the deficit at least does not seem to increase. Redintegration was investigated for the first time in this population, and the finding that it seems intact can be interpreted as a point of strength, promising an entry for possibly successful intervention strategies. However, the differential finding regarding the

relationship between redintegration and vocabulary size might indicate a possible difference in cognitive structure regarding the ability to use LTM information, but calls for further replicating research.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

GB and MG developed the theory and conceptualized and designed the study. GB and BE planned and conducted data

collection and preparation. GB performed the statistical analysis and wrote the first draft of the manuscript. All authors contributed to manuscript revisions, read and approved the submitted version.

## ACKNOWLEDGMENTS

We would like to thank Sebastian Poloczek for providing valuable feedback on a previous version of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02581/full#supplementary-material>

## REFERENCES

- Allen, R. J., and Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *J. Mem. Lang.* 55, 64–88. doi: 10.1016/j.jml.2006.02.002
- Alloway, T. P., and Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *J. Exp. Child Psychol.* 106, 20–29. doi: 10.1016/j.jecp.2009.11.003
- Annaz, D., Karmiloff-Smith, A., and Thomas, M. S. C. (2008). “The importance of tracing developmental trajectories for clinical child neuropsychology,” in *Child Neuropsychology. Concepts, Theory, and Practice*, eds R. Jonathan and W. R. Jody (Chichester: Wiley-Blackwell), 7–18.
- Baddeley, A., Lewis, V., and Vallar, G. (1984). Exploring the articulatory loop. *Q. J. Exp. Psychol. Sect. A* 36, 233–252. doi: 10.1080/14640748408402157
- Baddeley, A. D. (1986). *Working Memory, Vol. 11*. Oxford: Clarendon Press.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2
- Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Baddeley, A. D., Gathercole, S. E., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi: 10.1037/0033-295X.105.1.158
- Baddeley, A. D., Thomson, N., and Buchanan, M. (1975). Word length and the structure of short-term memory. *J. Verbal Learning Verb. Behav.* 14, 575–589. doi: 10.1016/S0022-5371(75)80045-4
- Campoy, G. (2008). The effect of word length in short-term memory: is rehearsal necessary? *Q. J. Exp. Psychol.* 61, 724–734. doi: 10.1080/17470210701402364
- Carney, D. P. J., Henry, L. A., Messer, D. J., Danielsson, H., Brown, J. H., and Rönnerberg, J. (2013). Using developmental trajectories to examine verbal and visuospatial short-term memory development in children and adolescents with Williams and Down syndromes. *Res. Dev. Disabil.* 34, 3421–3432. doi: 10.1016/j.ridd.2013.07.012
- Clarkson, L., Roodenrys, S., Miller, L. M., and Hulme, C. (2017). The phonological neighbourhood effect on short-term memory for order. *Memory* 25, 391–402. doi: 10.1080/09658211.2016.1179330
- Colom, R., Quiroga, M. Á., Shih, P. C., Martínez, K., Burgaleta, M., Martínez-Molina, A., et al. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence* 38, 497–505. doi: 10.1016/j.intell.2010.06.008
- Conlin, J. A., and Gathercole, S. E. (2006). Lexicality and interference in working memory in children and in adults. *J. Mem. Lang.* 55, 363–380. doi: 10.1016/j.jml.2006.05.003
- Craig, S., and Lewandowsky, S. (2013). Working memory supports inference learning just like classification learning. *Q. J. Exp. Psychol.* 66, 1493–1503. doi: 10.1080/17470218.2013.818703
- Danielsson, H., Henry, L. A., Rönnerberg, J., and Nilsson, L.-G. (2010). Executive functions in individuals with intellectual disability. *Res. Dev. Disabil.* 31, 1299–1304. doi: 10.1016/j.ridd.2010.07.012
- Dell, G. S., and O’Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition* 42, 287–314.
- Derragh, L. S., Neath, I., Surprenant, A. M., Beaudry, O., and Saint-Aubin, J. (2017). The effect of lexical factors on recall from working memory: generalizing the neighborhood size effect. *Can. J. Exp. Psychol.* 71, 23–31. doi: 10.1037/cep0000098
- Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Mem. Cogn.* 23, 83–94. doi: 10.3758/BF03210559
- Gathercole, S. E., Hitch, G. J., Service, E., and Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Dev. Psychol.* 33, 966–979. doi: 10.1037/0012-1649.33.6.966
- Gathercole, S. E., Pickering, S. J., Hall, M., and Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Q. J. Exp. Psychol. Sect. A* 54, 1–30. doi: 10.1080/02724980042000002
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige*. WWT 6–10. München: Elsevier.
- Grube, D., Lingen, M., and Hasselhorn, M. (2008). Entwicklung des phonologischen Arbeitsgedächtnisses: zur Rolle von Rehearsal und Lexikalität für den Ähnlichkeitseffekt. *Zeitschr. Entwicklungspsychol. Pädagog. Psychol.* 40, 200–207. doi: 10.1026/0049-8637.40.4.200
- Hasselhorn, M., Grube, D., and Mähler, C. (2000). “Theoretisches Rahmenmodell für ein Diagnostikum zur differentiellen Funktionsanalyse des phonologischen Arbeitsgedächtnisses,” in *Diagnostik von Lese- Rechtschreibschwierigkeiten. 1*, eds M. Hasselhorn, H. Marx, and W. Schneider (Göttingen: Hogrefe), 167–181.
- Hasselhorn, M., and Mähler, C. (2007). Phonological working memory of children in two German special schools. *Int. J. Disabil. Dev. Educ.* 54, 225–244. doi: 10.1080/10349120701330545
- Hasselhorn, M., Schuchardt, K., and Mähler, C. (2010). Phonologisches Arbeitsgedächtnis bei Kindern mit diagnostizierter Lese- und/oder Rechtschreibstörung. Zum Einfluss von Wortlänge und Lexikalität auf die Gedächtnisspanne. *Zeitschr. Entwicklungspsychol. Pädagog. Psychol.* 42, 211–216. doi: 10.1026/0049-8637/a000024
- Hasselhorn, M., Schumann-Hengstler, R., Gronauer, J., Grube, D., Mähler, C., Schmid, I., et al. (2012). *Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren*. (AGTB 5-12). Göttingen: Hogrefe.
- Hassiotis, A. (2015). Borderline intellectual functioning and neurodevelopmental disorders. Prevalence, comorbidities and treatment approaches. *Adv. Mental Health Intellect. Disabil.* 9, 275–283. doi: 10.1108/AMHID-06-2015-0028
- Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory* 9, 233–247. doi: 10.1080/09658210042000085
- Henry, L. A. (2010). The episodic buffer in children with intellectual disabilities: an exploratory study. *Res. Dev. Disabil.* 31, 1609–1614. doi: 10.1016/j.ridd.2010.04.025

- Henry, L. A., and MacLean, M. (2002). Working memory performance in children with and without intellectual disabilities. *Am. J. Mental Retard.* 107, 421–432. doi: 10.1352/0895-8017(2002)107<0421:WMPICW>2.0.CO;2
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G., Martin, S., and Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a redintegration process in immediate serial recall. *J. Exp. Psychol.* 23, 1217–1232. doi: 10.1037/0278-7393.23.5.1217
- IBM Corp (2017). *IBM SPSS Statistics for Windows*. Version 25.0. Armonk, NY: IBM Corp.
- Jalbert, A., Neath, I., Bireta, T. J., and Surprenant, A. M. (2011b). When does length cause the word length effect? *J. Exp. Psychol.* 37, 338–353. doi: 10.1037/a0021804
- Jalbert, A., Neath, I., and Surprenant, A. M. (2011a). Does length or neighborhood size cause the word length effect? *Mem. Cognit.* 39, 1198–1210. doi: 10.3758/s13421-011-0094-z
- Jarrold, C., Danielsson, H., and Wang, X. (2015). Absolute and proportional measures of potential markers of rehearsal, and their implications for accounts of its development. *Front. Psychol.* 6:e299. doi: 10.3389/fpsyg.2015.00299
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* 2, 389–398. doi: 10.1016/S1364-6613(98)01230-3
- Kassambara, A. (2017). *ggpubr. ggplot2 Based Publication Ready Plots*. Version 0.1.6. Available online at: <https://CRAN.R-project.org/package=ggpubr>
- Kenett, Y. N., Gold, R., and Faust, M. (2016). The hyper-modular associative mind: a computational analysis of associative responses of persons with Asperger syndrome. *Lang. Speech* 59, 297–317. doi: 10.1177/0023830915589397
- Klauser, K., and Lauth, G. W. (1997). “Lernbehinderungen und Leistungsschwierigkeiten bei Schülern,” in *Enzyklopädie der Psychologie (Themenbereich D, Serie I Pädagogische Psychologie, Psychologie des Unterrichts und der Schule)*, ed F. E. Weinert (Göttingen: Hogrefe), 701–738.
- Kretschmann, R. (2007). “Lernschwierigkeiten, Lernstörung und Lernbehinderung,” in *Sonderpädagogik des Lernens*, eds J. Walter and F. B. Wember (Göttingen: Hogrefe), 4–32.
- Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y., et al. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *J. Child Psychol. Psychiatry* 52, 212–220. doi: 10.1111/j.1469-7610.2010.02311.x
- Lewandowsky, S., and Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychol. Res.* 63, 163–173. doi: 10.1007/PL00008175
- Lilienthal, L., Rose, N. S., Tamez, E., Myerson, J., and Hale, S. (2015). Individuals with low working memory spans show greater interference from irrelevant information because of poor source monitoring, not greater activation. *Mem. Cogn.* 43, 357–366. doi: 10.3758/s13421-014-0465-3
- Mähler, C. (2007). Arbeitsgedächtnisfunktionen bei lernbehinderten Kindern und Jugendlichen. *Zeitschr. Entwicklungspsychol. Pädagog. Psychol.* 39, 97–106. doi: 10.1026/0049-8637.39.2.97
- Mähler, C., and Hasselhorn, M. (2003). Automatische Aktivierung des Rehearsalprozesses im phonologischen Arbeitsgedächtnis bei lernbehinderten Kindern und Erwachsenen. *Zeitschr. Pädagog. Psychol.* 17, 255–260. doi: 10.1024//1010-0652.17.34.255
- Mähler, C., and Schuchardt, K. (2016). The importance of working memory for school achievement in primary school children with intellectual or learning disabilities. *Res. Dev. Disabil.* 58, 1–8. doi: 10.1016/j.ridd.2016.08.007
- Marton, K., and Eichorn, N. (2014). Interaction between working memory and long-term memory: a study in children with and without language impairment. *Zeitschr. Psychol.* 222, 90–99. doi: 10.1027/2151-2604/a000170
- Poloczek, S., Büttner, G., and Hasselhorn, M. (2012). Relationships between working memory and academic skills: are there differences between children with intellectual disabilities and typically developing children? *J. Cogn. Educ. Psychol.* 11, 20–38. doi: 10.1891/1945-8959.11.1.20
- Poloczek, S., Büttner, G., and Hasselhorn, M. (2014). Phonological short-term memory impairment and the word length effect in children with intellectual disabilities. *Res. Dev. Disabil.* 35, 455–462. doi: 10.1016/j.ridd.2013.11.025
- Poloczek, S., Henry, L. A., Danielsson, H., Büttner, G., Mähler, C., Messer, D. J., et al., (2016). Strategic verbal rehearsal in adolescents with mild intellectual disabilities. A multi-centre European study. *Res. Dev. Disabil.* 58, 83–93. doi: 10.1016/j.ridd.2016.08.014
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org>
- Roodenrys, S. (2009). “Explaining phonological neighbourhood effects in short-term memory,” in *Interactions Between Short-Term and Long-Term Memory the Verbal Domain*, eds A. S. C. Thorn and M. Page (Hove: Psychology Press), 177–197.
- Roodenrys, S., and Miller, L. M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Mem. Cogn.* 36, 578–587. doi: 10.3758/MC.36.3.578
- Rosenquist, C., Connors, F. A., and Roskos-Ewoldsen, B. (2003). Phonological and visuo-spatial working memory in individuals with intellectual disability. *Am. J. Mental Retard.* 108, 403–413. doi: 10.1352/0895-8017(2003)108<403:PAVWMI>2.0.CO;2
- R Studio Team (2016). *RStudio. Integrated Development Environment for R*. Boston, MA: RStudio, Inc. Available online at: <http://www.rstudio.com/>
- Schröder, U. J. (2000). *Lernbehindertenpädagogik. Grundlagen und Perspektiven sonderpädagogischer Lernhilfe*. Stuttgart: W. Kohlhammer.
- Schuchardt, K., Gebhardt, M., and Mähler, C. (2010). Working memory functions in children with different degrees of intellectual disability. *J. Intellect. Disabil. Res.* 54, 346–353. doi: 10.1111/j.1365-2788.2010.01265.x
- Schuchardt, K., Mähler, C., and Hasselhorn, M. (2011). Functional deficits in phonological working memory in children with intellectual disabilities. *Res. Dev. Disabil.* 32, 1934–1940. doi: 10.1016/j.ridd.2011.03.022
- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Mem. Cogn.* 21, 168–175. doi: 10.3758/BF03202729
- Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., and Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *J. Speech Lang. Hear. Res.* 52, 336–358. doi: 10.1044/1092-4388(2009)07-0144
- Thorn, A. C., and Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *J. Exp. Psychol.* 31, 729–735. doi: 10.1037/0278-7393.31.4.729
- Thorn, A. S. C., Frankish, C. R., and Gathercole, S. E. (2009). “The influence of long-term knowledge on short-term memory: evidence for multiple mechanisms,” in *Interactions Between Short-Term and Long-Term Memory the Verbal Domain*, eds A. S. C. Thorn and M. Page (Hove: Psychology Press), 198–219.
- Thorn, A. S. C., Gathercole, S. E., and Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: an evaluation of Schweickert’s (1993) multinomial processing tree model. *Cogn. Psychol.* 50, 133–158. doi: 10.1016/j.cogpsych.2004.07.001
- Thorn, A. S. C., and Page, M. P. A. (2009). “Current issues in understanding interactions between short-term and long-term memory,” in *Interactions Between Short-Term and Long-Term Memory in the Verbal Domain*, eds A. S. C. Thorn and M. Page (Hove: Psychology Press), 1–15.
- Turner, J. E., Henry, L. A., Smith, P. T., and Brown, P. A. (2004). Redintegration and lexicality effects in children. Do they depend upon the demands of the memory task? *Mem. Cogn.* 32, 501–510. doi: 10.3758/BF03195842
- van der Molen, M. J., Henry, L. A., and van Luit, J. E. H. (2014). Working memory development in children with mild to borderline intellectual disabilities. *J. Intellect. Disabil. Res.* 58, 637–650. doi: 10.1111/jir.12061
- van der Molen, M. J., van Luit, J. H., Jongmans, M. J., and van der Molen, M. W. (2007). Verbal working memory in children with mild intellectual disabilities. *J. Intellect. Disabil. Res.* 51, 162–169. doi: 10.1111/j.1365-2788.2006.00863.x
- Vitevitch, M. S., Chan, K. Y., and Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *J. Mem. Lang.* 67, 30–44. doi: 10.1016/j.jml.2012.02.008
- Walker, I., and Hulme, C. (1999). Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term

- serial recall. *J. Exp. Psychol.* 25, 1256–1271. doi: 10.1037/0278-7393.25.5.1256
- Weiß, R. H. (2008). *CFT 20-R mit WS/ZF-R. Grundintelligenztest Skala 2 - Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R)*. Göttingen: Hogrefe.
- Weiß, R. H., and Osterland, J. (2012). *CFT 1-R. Grundintelligenztest Skala 1 - Revision*. Göttingen: Hogrefe.
- Wickham, H. (2011). *ggplot2. Elegant Graphics for Data Analysis*. New York, NY: Springer. Available online at: <http://ggplot2.org>. doi: 10.1002/wics.147
- Zigler, E., and Balla, D. A. (1982). *Mental Retardation. The Developmental-Difference Controversy*. Hillsdale, NJ: Lawrence Erlbaum.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Bruns, Ehl and Grosche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.