

dgs

Sprachheilarbeit

# Forschung Sprache

**E-Journal** für Sprachheilpädagogik, Sprachtherapie  
und Sprachförderung



DLV

## Editorial

Wilma Schönauer-Schneider & Susanne van Minnen

Kompetenzen im Fach Deutsch . . . . . 2

Andreas Mayer, Dana-Kristin Marks

Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien –  
Eine Metaanalyse zur Effektivität

Improving reading comprehension skills by mediating comprehension strategies –  
a meta-analysis of the effectiveness . . . . . 4

Anita Hönninger

Evaluation des computerbasierten Trainingsprogramms *Lautarium* bei Kindern  
mit Sprachentwicklungsstörungen

Evaluation of the computer-based training program *Lautarium* in children  
with language impairment. . . . . 38

Corinne Weyda, Wilma Schönauer-Schneider

„Halt!“ Ein Diagnoseverfahren zum Textverstehen bei siebenjährigen Kindern

„Halt!“ A diagnostic tool to evaluate text comprehension in seven year old children . . . . 59

Impressum . . . . . 77



## Editorial

### Kompetenzen im Fach Deutsch

Wilma Schönauer-Schneider & Susanne van Minnen

Kompetenzen im Fach Deutsch gründen auf einer vorschulisch erworbenen Sprachhandlungskompetenz. Sie sind in den Bildungsstandards für den Primarbereich in die vier Kompetenzbereiche *Sprechen und Zuhören*, *Schreiben*, *Lesen/mit Texten und Medien umgehen* und *Sprache/Sprache untersuchen* eingeteilt. Die Lesekompetenz wird dabei häufig als Schlüsselqualifikation verstanden, da sie von zentraler Bedeutung für den Wissenserwerb und das lebenslange Lernen ist (Böhme, Engelbert & Weirich, 2017).

In der internationalen Grundschul-Lese-Untersuchung (IGLU) wurden 2016 die Lesekompetenzen von Schülern am Ende der vierten Jahrgangsstufe im internationalen Vergleich untersucht. Dabei zeigte sich, dass die Gesamtskala in Deutschland für den Bereich Lesen sich von 2001 bis 2016 nicht signifikant veränderte, während eine Reihe anderer Staaten die Leistungen in diesen 15 Jahren verbessern konnte. „Deshalb hat sich die relative Position Deutschlands erheblich verschlechtert“ (Hußmann et al., 2018, S. 14). Besonders besorgniserregend ist, dass knapp 6 Prozent der Viertklässler in Deutschland nur ein rudimentäres Leseverständnis (Stufe I) erreichen, was in Europa nur in Frankreich, der Französischen Gemeinschaft in Belgien, der Slowakei und Malta nominell häufiger auftritt (Hußmann et al., 2018).

Der IQB Bildungstrend 2016 fokussierte mehrere Kompetenzen des Faches Deutsch in den einzelnen Bundesländern Deutschlands. In den Teilbereichen *Lesen*, *Zuhören* und *Orthografie* wurde untersucht, welches Kompetenzniveau Viertklässler erreichen. Deutschlandweit konnte gezeigt werden, dass beim Lesen 12,5 % der Schüler, beim Zuhören 10,8 % und bei der Orthografie 22,1 % die Mindeststandards nicht erreichten. Zwischen den Jahren 2011 und 2016 blieben die Ergebnisse im Bereich *Lesen* für Deutschland insgesamt weitgehend nur stabil, während in den Bereichen *Zuhören* und *Orthografie* signifikant negative Trends erkennbar waren (Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017). Die diagnostische Erfassung dieser Kompetenzen und eine anschließend effektive Förderung und Intervention sind somit unerlässlich. Diese Ausgabe von *Forschung Sprache* widmet sich mit den folgenden drei Beiträgen deshalb dem Fach Deutsch mit der Förderung des Textverstehens, einem Trainingsprogramm für Kinder mit Lese-Rechtschreibschwierigkeiten und der Erfassung des Hörverstehens von Geschichten.

In einer Metaanalyse untersuchten Andreas Mayer und Dana-Kristin Marks die Effektivität der Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien. Im zweiten Beitrag schildert Anita Hönninger Ergebnisse einer Evaluation des computerbasierten Trainingsprogramms *Lautarium* bei Kindern mit Sprachentwicklungsstörungen. Abschließend beschreiben Corrine Weyda und Wilma Schönauer-Schneider die Entwicklung eines diagnostischen Instrumentariums zur Erfassung des Hörtextverstehens.

Es bleibt zu hoffen, dass weitere Forschungen diese Kompetenzen im Fach Deutsch untersuchen und voranbringen.

## Literatur

- Böhme, K., Engelbert, M. & Weirich, S. (2017). Beschreibung der im Fach Deutsch untersuchten Kompetenzen. In: P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag. (Hrsg.) (S. 20-30). Münster/New York: Waxmann.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T.C. & Valtin, R. (Hrsg.) (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster/New York: Waxmann.
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.). (2017). *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Zusammenfassung. Münster/New York: Waxmann.

*Prof. Dr. Susanne van Minnen*, Redaktion Forschung Sprache, Professorin für Erziehungswissenschaft mit dem Schwerpunkt Erziehungswissenschaft bei Beeinträchtigung der Sprache und des Sprechens an Justus-Liebig-Universität Gießen

*Prof. Dr. Wilma Schönauer-Schneider*, Redaktion Forschung Sprache, Professorin für Sprachbehindertenpädagogik an der Pädagogischen Hochschule Heidelberg

**Bitte beachten Sie für alle Artikel in Forschung Sprache:**

Zur besseren Lesbarkeit sind personbezogene Bezeichnungen in männlicher Form dargestellt, beziehen sich aber in gleicher Weise auf beide Geschlechter.



## Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien – Eine Metaanalyse zur Effektivität\*

## Improving reading comprehension skills by mediating comprehension strategies – a meta-analysis of the effectiveness

Andreas Mayer, Dana-Kristin Marks

### Zusammenfassung

**Hintergrund:** Im deutschsprachigen Raum werden bei Kindern mit Leseschwierigkeiten aktuell v. a. phonologisch orientierte Fördermaßnahmen favorisiert, die sich primär im Hinblick auf die Worterkennung als effektiv erwiesen haben und damit eine Grundlage für sinnentnehmendes Lesen schaffen können. Insbesondere bei Kindern mit beeinträchtigtem Leseverständnis trotz angemessener Lesefertigkeit müssen diese Maßnahmen durch Konzepte ergänzt werden, die auf eine aktive Auseinandersetzung mit dem Gelesenen und das Generieren eines Situationsmodells abzielen. Eine vielversprechende Methode liegt in der Vermittlung von Strategien (Aktivieren von Vorwissen, Fragen stellen, Überwachung des Verstehensprozesses, bildhaftes Vorstellen, Zusammenfassen), die Schüler unterstützen sollen, den Inhalt des Gelesenen eigenaktiv-strategisch zu (re-)konstruieren und mit dem Vor- und Weltwissen zu verknüpfen.

**Fragestellung:** Das Ziel der vorliegenden Arbeit besteht darin zu überprüfen,

- (1) wie effektiv Interventionsmaßnahmen, die explizit die Vermittlung von Lesestrategien fokussieren, in Bezug auf das Leseverständnis sind,
- (2) inwiefern sich entsprechende Maßnahmen positiv auf das Wissen um und die Anwendung der vermittelten Verstehensstrategien auswirken und
- (3) von welchen Variablen die Effektivität beeinflusst werden könnte.

**Methodik:** Auf der Grundlage einer umfassenden Literaturrecherche in verschiedenen Datenbanken wurden alle für die Metaanalyse geeigneten Interventionsstudien hinsichtlich möglicher moderierender Variablen kodiert und in einer Übersicht zusammengeführt, die einen Überblick über die Merkmale des Studiendesigns bietet. Für die in die Analyse aufgenommenen Studien wurden für drei Outcome-Maße standardisierte Effektgrößen, die dazugehörigen Konfidenzintervalle sowie der gewichtete mittlere Gesamteffekt berechnet und in forest plots dargestellt. Darüber hinaus wurde analysiert, welche moderierenden Variablen den Effekt einer strategieorientierten Intervention beeinflussen könnten.

**Ergebnisse:** Die Überprüfungen des Textverständnisses mittels normierter Verfahren belegen auf Seiten der Interventionsgruppen einen stärkeren Leistungszuwachs im Vergleich zu den Kontrollgruppen ( $d=.21$ ). Wesentlich stärker fällt der Effekt aus, wenn informelle Überprüfungen zur Messung des Leseverständnisses eingesetzt werden ( $d=.66$ ). Auch im Bezug zum Strategiewissen und der Strategieanwendung lassen sich Effekte in einem mittleren Bereich nachweisen ( $d=.51$ ). Insbesondere Schüler der Primarstufe scheinen von diesem Interventionsansatz zu profitieren. Die Kombination mehrerer Strategien innerhalb einer Intervention ist erfolgversprechender als die Vermittlung nur einer Strategie.

**Schlussfolgerungen:** Die Vermittlung von Verstehensstrategien kann sich positiv auf das Textverständnis auswirken. Da aber insbesondere die mittels normierter Leseüberprüfungen erzielten Effekte eher in einem niedrigen Bereich angesiedelt sind, erscheint es sinnvoll, Konzepte zu entwickeln und zu evaluie-

\* Dieser Beitrag hat das Peer-Review-Verfahren durchlaufen.

ren, die auf Kinder mit Textverständnisschwierigkeiten zugeschnitten sind und einen Transfer auf alltägliche Lesesituationen ermöglichen.

#### Schlüsselwörter:

Metaanalyse, Leseverständnis, Lesestrategien, Intervention, Schulalter

#### Abstract

**Background:** Regarding reading difficulties in childhood, approaches focusing on phonology are currently favored in German-speaking countries. They have mainly proven to be effective for word recognition and, thereby, provide a good basis for meaning extracted reading. In case of *reading comprehension* difficulties in particular, this does not seem to be sufficient. There is a need to complement those approaches by support measures which aim at active examination of the text and integrate the information read into a so-called situation model. One promising option is the mediation of *strategies* (e.g. activating prior knowledge, asking questions, comprehension monitoring, mental imagery, summarizing) which enable students to gain deeper understanding of the reading content.

**Aims:** The aim of the present study is to examine

- (1) whether positive effects can be substantiated for intervention methods which primary focus on teaching reading strategies regarding reading comprehension skills and
- (2) what impact this type of support measure might have on knowing and applying the newly learned reading strategies. Furthermore,
- (3) which variables could influence the intervention's effectivity.

**Methods:** First, a comprehensive and systematic literature research within selected databases was conducted. Studies that match the search query were included and synthesized by their study features to highlight similarities and differences across the corpus of studies. Second, we conducted a meta-analysis to determine the overall effect of reading interventions on the selected outcome variables (see "aims"). Therefore, standardized effect sizes (using Cohen's  $d$  and confidential intervals) for each study as well as medium effect sizes were calculated (presented all together in forest plots). In addition to an overall point estimate of reading intervention effects, several subgroups have been created and effects relating to these potential moderating variables are reported.

**Results:** Compared with the control condition, the meta-analysis substantiates a higher performance increase of strategy-oriented approaches. When standardized measurements assessing reading comprehension skills are used, the overall effect size represents a small effect ( $d=.21$ ). The medium effect is even higher when informal diagnostic methods are used ( $d=.66$ ). This holds true for measuring the knowledge and application of the newly learned reading strategies where a medium-sized effect has been found ( $d=.51$ ). Students in primary education seem to particularly benefit from this type of intervention. Furthermore, teaching several reading strategies combined within an intervention seem to be more likely to succeed than providing students with only one single strategy.

**Conclusions:** The mediation of reading comprehension strategies is likely to improve text comprehension skills. However, the medium effect sizes remain at a rather low to modest level considering standardized assessment methods. Hence, the development and evaluation of treatment approaches, which meet the specific needs of students with reading comprehension difficulties, are still deemed to be necessary.

#### Keywords

meta-analysis, reading comprehension, reading strategies, intervention, school age

## 1 Einleitung

Die Fähigkeit, Texte lesen und verstehen zu können, stellt eine zentrale Voraussetzung für die Teilhabe am gesellschaftlichen Leben dar (Souvignier & Antoniou, 2007). In unterrichtlichen Kontexten wird das sinnentnehmende Lesen von Sach- und Erzähltexten spätestens Ende der Primarstufe von Lehrkräften üblicherweise implizit vorausgesetzt. In nahezu allen Fächern wird von Schülern erwartet, sich selbstständig neue Informationen aus gedruckten Texten anzueignen, weshalb Schwierigkeiten mit dem Textverständnis nahezu zwangsläufig zu allgemeinen Lernschwierigkeiten führen. Aber auch in der Interaktion mit Gleichaltrigen spielt das sinnentnehmende Lesen zum Zweck der Initiierung und der Aufrechterhaltung sozialer Beziehungen z. B. in sozialen Netzwerken eine wesentliche Rolle, so dass auch die soziale Integration betroffener Kinder gefährdet sein kann.

Die im vergangenen Jahr veröffentlichten Ergebnisse der letzten Internationalen Grundschul-Leseuntersuchung (IGLU) aus dem Jahr 2016 (Hußmann, Wendt, Bos, Bremerich-Vos, Kasper,

Lankes, McElvany, Stubbe & Valtin, 2017) offenbaren den umfassenden Handlungsbedarf im Bereich der Förderung des sinnentnehmenden Lesens bei Grundschulern in Deutschland. Im internationalen Vergleich zeigte sich, dass in etwa 75% der teilnehmenden EU-Staaten im Bereich des Leseverständnisses signifikant bessere Leistungen erzielt wurden als in Deutschland. Während 2001 der erreichte Mittelwert der Schüler in nur vier Staaten der EU signifikant besser ausfiel als in Deutschland, war das 2016 in 13 Ländern der Fall. Der Abstand zu den Ländern, die in der IGLU-Untersuchung am besten abschneiden, beträgt knapp ein Schuljahr.

Besonders besorgniserregend erscheint das Ergebnis der schwächsten Leser in Deutschland. 19% der teilnehmenden Kinder erreichten maximal die Kompetenzstufe 2. Nur in Israel, Frankreich, Belgien (Franz. Gemeinschaft) und auf Malta war der prozentuale Anteil an Schülern, die lediglich Kompetenzstufe 2 erreichten, höher. Diese Schüler sind bestenfalls in der Lage, „explizit angegebene Informationen zu identifizieren und eine Kohärenz auf lokaler Ebene herzustellen“ (Bremerich-Vos, Wendt & Bos, 2017, S. 92). „Für diese Gruppe ist zu erwarten, dass sie in der Sekundarstufe I mit erheblichen Schwierigkeiten beim Lernen in allen Fächern konfrontiert sein wird, wenn es nicht gelingt, sie maßgeblich zu fördern“ (Bos, Valtin, Hußmann, Wendt & Goy, 2017, S. 15).

Offensichtlich konnten insbesondere die schwächsten Leser von den in den letzten 15 Jahren eingeleiteten Unterstützungsmaßnahmen kaum profitieren. Umso notwendiger ist die Entwicklung, Evaluation und Implementierung schulischer Maßnahmen, die Schüler mit beeinträchtigtem Leseverständnis effektiv unterstützen. Der vorliegende Beitrag arbeitet die in den letzten 20 Jahren publizierten Ergebnisse von Interventionsstudien auf, die auf eine Verbesserung des Textverständnisses durch die Vermittlung von Verstehensstrategien abzielten.

## 2 Theoretischer Hintergrund

### 2.1 Ausgangslage: Fokussierung der Worterkennung

In Anlehnung an Lyon, Shaywitz und Shaywitz (2003), Tunmer und Greaney (2010) und Mayer (2016a) werden Lesestörungen als Schwierigkeiten beim Erwerb und der Anwendung der indirekten, segmentalen Lesestrategie (= phonologisches Rekodieren) und/oder der automatisierten Worterkennung charakterisiert. Diese Problematik resultiert häufig aus einer beeinträchtigten phonologischen Informationsverarbeitung oder einer gestörten Sprachentwicklung und kann sich negativ auf das Leseverständnis, die schulisch-kognitive, die sprachliche sowie die sozio-emotionale Entwicklung auswirken.

Die in dieser Definition benannte Kernsymptomatik erklärt, warum man sich bei der Erforschung der Ursachen der Lesestörung in den letzten drei Jahrzehnten primär auf die Worterkennung und die phonologische Informationsverarbeitung (v. a. die Förderung der phonologischen Bewusstheit) konzentrierte. Ebenso nachvollziehbar ist vor diesem Hintergrund, dass die meisten international publizierten Interventionsstudien auf die Worterkennung und damit auf eine höhere Lesegenauigkeit und Lesegeschwindigkeit abzielen, während Schwierigkeiten im Bereich des Textverständnisses im deutschsprachigen Raum lange Zeit vernachlässigt wurden. Im anglo-amerikanischen Raum dagegen hat die Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien seit der Publikation der Methode des reziproken Lehrens und Lernens durchaus eine längere Tradition (Palinscar & Brown, 1984).

Auch die Effektivität von Interventionsmaßnahmen wird üblicherweise daran gemessen, inwiefern es durch entsprechende Unterstützungsmaßnahmen gelingt, die Lesegenauigkeit und/oder die Lesegeschwindigkeit leistungsschwacher Schüler zu verbessern, während Auswirkungen auf das Leseverständnis kaum überprüft werden (z. B. Snellings, van der Leij, de Jong & Blok, 2009; Hinitikaa, Landerl, Aro & Lyythinen, 2008; Huemer, Landerl, Aro & Lyythinen, 2008). Auch die Studien, die in den kürzlich publizierten Metaanalysen zur Effektivität von Unterstützungsmaßnahmen bei Leseschwierigkeiten (Ise, Engel & Schulte-Körne, 2012; Galuschka & Schulte-Körne, 2015) berücksichtigt wurden, überprüften die Wirksamkeit der Interventionen üblicherweise über die Verbesserung der Worterkennung.

Dasselbe gilt für die wenigen Förderprogramme aus dem deutschsprachigen Raum, für die Galuschka und Schulte-Körne (2015) zu Folge eine empirische Evidenz ihrer Wirksamkeit vorliegt. Das Programm „Flüssig lesen lernen“ (Tacke, 2009) etwa fokussiert die Worterkennung, indem Schüler der Primarstufe beim Erwerb einer für das Lesenlernen ausreichenden phonologischen Bewusstheit, der Graphem-Phonem-Korrespondenz, dem Erlernen des synthetisierenden Lesens sowie der ganzheitlich-simultanen Verarbeitung von Silben und komplexen Konsonantenclustern



tern unterstützt werden. Die Effektivität des Förderprogramms wurde vom Autor selbst (Tacke, 2005) durch einen mittleren Effekt bei einer normierten Überprüfung der Lesegeschwindigkeit nachgewiesen, eine Erfassung des Leseverständnisses wurde nicht durchgeführt.

Ähnliches gilt für die „Lautgetreue Lese-Rechtschreibförderung“ (Reuter-Liehr, 2006), einem sprachwissenschaftlich fundierten Förderprogramm, das primär auf den Erwerb des phonologischen Rekodierens und des lautgetreuen Schreibens von Wörtern abzielt. Die Effektivität des Konzepts wurde von Klicpera, Weiss und Gasteiger-Klicpera (2013) ausschließlich über die Lesegenauigkeit und die Lesegeschwindigkeit, nicht aber über das Leseverständnis überprüft, wobei signifikante Effekte nur für die Lesegenauigkeit nachgewiesen werden konnten.

Das einzige Förderprogramm, dessen Auswirkungen auch in Bezug auf das Leseverständnis überprüft wurde, ist der Kieler Leseaufbau (Dummer-Smoch & Hackethal, 2013). Die von Groth, Hasko, Bruder, Kunze und Schulte-Körne (2013) durchgeführte Überprüfung der Effektivität des Kieler Leseaufbaus mit leseschwachen Kindern belegte deutliche Fortschritte der Interventionsgruppe im Bereich des Leseverständnisses, die sich aber nicht von den Fortschritten unterschied, die im selben Zeitraum von einer Wartekontrollgruppe gemacht wurden ( $d=.04$ ).

Die Vernachlässigung des Leseverständnisses im Kontext der Erforschung der Lesekompetenz sowie der Mangel an empirisch evaluierten Interventionsmaßnahmen zur Verbesserung des sinnentnehmenden Lesens im deutschsprachigen Raum wird unterstrichen durch die Ergebnisse einer Literaturrecherche in den Datenbanken ERIC, Medline, PsycINFO, PsycARTICLES und PSYINDEX mit dem Suchbegriff „Leseverst\*“. Sie ergab für den Zeitraum der letzten 30 Jahre insgesamt etwa 50 Publikationen deutschsprachiger Autoren in wissenschaftlichen Journalen, wobei eine differenziertere Sichtung der Beiträge zeigte, dass nur in ca. 15 Publikationen die Ergebnisse von Interventionsmaßnahmen dokumentiert wurden (z. B. Lenhard, Baier, Endlich, Schneider & Hoffmann, 2013; Mokhlesgerami, Souvignier, Rühl & Gold, 2007; Schünemann, Spörer & Brunstein, 2013; Seuring & Spörer, 2010).

Da Schwierigkeiten bei der sinnentnehmenden Verarbeitung des Gelesenen im Sinne der oben genannten Definition als Konsequenz der beeinträchtigten Worterkennung betrachtet werden, kann erwartet werden, dass Verbesserungen im Bereich der Worterkennung mit Verbesserungen im Bereich des Leseverständnisses einhergehen.

Beispielsweise gehen Berninger, Vermeulen, Abbott, McCutchen, Cotton, Cude, Dorn und Sharon (2003, S. 102) davon aus „... that if the bottleneck in word reading is eliminated through explicit code-based instruction, then reading comprehension will develop normally“.

Allerdings handelt es sich bei der Worterkennung nur um eine, wenn auch unabdingbare Komponente des Textverständnisses, das von zahlreichen weiteren sprachlich-kognitiven Faktoren beeinflusst wird. Aus diesem Grund greift die ausschließliche Fokussierung phonologisch orientierter Maßnahmen und der Worterkennung bei den Kindern zu kurz, deren Probleme mit dem sinnentnehmenden Lesen nicht primär aus einer beeinträchtigten Worterkennung resultieren, die sich vielmehr trotz durchschnittlicher Lesefertigkeit durch spezifische Schwierigkeiten mit dem Leseverständnis („poor comprehender“) charakterisieren lassen.

Ein kurzer Exkurs zum Begriff des Textverständnisses soll die Komplexität dieses Konstrukts und die damit einhergehende Vielschichtigkeit notwendiger Unterstützungsmaßnahmen verdeutlichen.

## 2.2 Exkurs: Zur Komplexität des Textverständnisses

Beim Textverständnis handelt es sich um eine komplexe Kompetenz, die weit über die Fähigkeit hinausgeht, allein die im Text explizit benannten Informationen zu identifizieren und zu verarbeiten. Vielmehr wird der Prozess des Textverstehens als eine aktive Konstruktionsleistung betrachtet, bei der mentale Repräsentationen der im Text enthaltenen inhaltlichen Aussagen generiert und aktiv mit dem Vor-, Welt- und Sprachwissen zu neuen differenzierteren Episoden verknüpft werden. „Lesen ist demnach keine passive Rezeption dessen, was im jeweiligen Text an Information enthalten ist, sondern aktive (Re-)konstruktion der Textbedeutung (Artelt, Schiefele, Schneider & Stanat, 2002, S. 7). Um diese Leistung erbringen zu können, spielen über die Worterkennung hinausgehend, spezifisch sprachliche Kompetenzen eine bedeutende Rolle. Während oder nachdem die einzelnen Wörter eines Textes in Lautsprache umgewandelt werden, müssen Wörter, Phrasen und Sätze semantisch verarbeitet und syntaktisch-morphologisch dekodiert werden.

Für die Bedeutung lexikalischer Fähigkeiten im Zusammenhang mit dem Textverständnis sprechen bspw. die von Rickets, Nation und Bishop (2007) nachgewiesenen engen korrelativen

Zusammenhänge zwischen dem Umfang des Wortschatzes und dem Leseverständnis in einer Größenordnung von  $r=.63$  ( $p<.01$ ). Die von den Autoren zusätzlich durchgeführten hierarchischen Regressionsanalysen legen nahe, dass der Umfang des Wortschatzes selbst nach Berücksichtigung nonverbaler kognitiver Fähigkeiten und der Lesefertigkeit, 17,8% der Unterschiede im Leseverständnis erklären kann. Für die Bedeutung syntaktischer und morphologischer Fähigkeiten im Kontext des Textverständnisses sprechen die Ergebnisse einer eigenen Untersuchung (Mayer, 2014). Zweitklässler mit durchschnittlichem Leseverständnis schnitten bei einer Überprüfung der rezeptiven grammatischen Fähigkeiten zwei Jahre vorher etwa 1,5 Standardabweichungen besser ab als Kinder mit beeinträchtigtem Leseverständnis.

Jedoch sichern eine angemessene Worterkennung, das Beherrschen des Wortschatzes eines gelesenen Textes sowie die grammatische Dekodierung der sprachlichen Strukturen zunächst das Verstehen auf Wort- und Satzebene. Der Verstehensprozess auf Textebene nimmt auf dieser Ebene aber erst seinen Ausgangspunkt. Die Informationen, die der Leser durch die lexikalische und grammatische Analyse einzelner Sätze gewonnen hat, müssen mit Hilfe hierarchiehöherer Prozesse in Beziehung zueinander und zu einem kohärenten Gesamtbild zusammengesetzt werden.

Dazu werden die in Form von Propositionen abstrahierten Informationen einzelner Sätze zu einem Gesamtbild zusammengefügt, indem der Textinhalt auf wesentliche Aussagen verdichtet und durch das Vor- und Weltwissen ergänzt wird. Dabei setzt sich der Leser aktiv mit den explizit genannten Informationen auseinander, ergänzt nicht explizit genannte Informationen und zieht Schlussfolgerungen („zwischen den Zeilen lesen“, Klicpera & Gasteiger-Klicpera, 1995). Er verarbeitet Kohäsionsmittel (Pronomen, Konjunktionen, Tempus, Adverbale etc.), mit denen Beziehungen zwischen Wörtern, Satzteilen und Sätzen ausgedrückt werden, wendet während und nach dem Lesen Strategien an, die es ihm ermöglichen, den Verstehensprozess zu überwachen (Comprehension Monitoring) und wandelt Gelesenes in mentale Bilder um.

Der Prozess der Bildung dieses sogenannten *Situationsmodells* („Situation Model“, Van Dijk & Kintsch, 1983) bewirkt, dass die aktuell beim Lesen verarbeiteten Informationen keine isolierten Bedeutungseinheiten darstellen, sondern mit dem bereits vorhandenen Wissen zu neuartigen Episoden im Langzeitgedächtnis verknüpft werden. Es handelt sich dabei um den Prozess, den kompetente Leser kennen, wenn sie einen Text lesen und sukzessive Bilder im Kopf entstehen, die deutlich mehr Informationen umfassen als das, was explizit beschrieben wird.

Eine angemessene Worterkennung und ein intaktes Sprachverständnis auf lexikalischer und grammatischer Ebene stellen damit eine notwendige Voraussetzung für die aktiv-strategische Rekonstruktion von Textinhalten dar. Bei spezifischen Schwierigkeiten mit dem Textverständnis müssen entsprechende Maßnahmen ergänzt werden durch solche, die auf die aktive Auseinandersetzung mit Textinhalten abzielen. Dieses Ziel fokussieren aktuell Interventionsansätze, die auf die Vermittlung von Lesestrategien abzielen. Bevor die in der Forschungsliteratur am häufigsten beschriebenen Strategien im Detail erläutert werden (Kap. 3) und die Effektivität dieses Ansatzes mittels einer systematischen Literaturanalyse bestimmt wird, werden im Folgenden auf der Basis einiger Interventionsstudien und bereits publizierter Metaanalysen (Edmonds, Vaughn, Wexler, Reutebuch, Cable, Klingler Tackett & Wick Schnakenberg, 2009; Souvignier & Antoniou, 2007) unterschiedliche Interventionsmaßnahmen bei Leseverständnisschwierigkeiten und deren Effektivität skizziert.

### 2.3 Interventionsmaßnahmen bei Leseverständnisschwierigkeiten

Den Ausführungen aus den vorangegangenen Abschnitten folgend, resultieren Schwierigkeiten mit dem Textverständnis aus einer beeinträchtigten Worterkennung (Lesegenauigkeit, Lesegeschwindigkeit), lexikalischen Defiziten, der eingeschränkten Fähigkeit, die Grammatik des Gelesenen zu dekodieren, den gelesenen Inhalt mit Vorwissen zu verknüpfen oder sich aktiv mit dem Text auseinanderzusetzen (Edmonds et al., 2009).

Souvignier und Antoniou (2007) unterscheiden hinsichtlich der Intervention bei Schwierigkeiten mit dem Textverständnis Maßnahmen, die auf hierarchieniedrige Komponenten des Leseverständnisses (z. B. phonologische Fördermaßnahmen, Unterstützung bei der Automatisierung der Worterkennung, Erweiterung lexikalischer Fähigkeiten) abzielen, von Methoden, die das Leseverständnis durch eine Optimierung der räumlichen Organisation der Textdarstellung, die Anreicherung von Texten durch Bildmaterial, lesetechnische Hilfestellungen bzw. eine sprachliche Optimierung der Texte unterstützen, sowie die Vermittlung von Verstehensstrategien, die eine aktive Auseinandersetzung mit dem Gelesenen ermöglichen sollen. Dabei erwiesen sich in einer



metaanalytischen Aufarbeitung des Forschungsstandes die Unterstützungsmaßnahmen, in denen lern- und/oder leseschwachen Schülern Strategien (z. B. Zusammenfassen, Aktivieren von Vorwissen, Überwachen des eigenen Verstehens) vermittelt werden, mit Abstand als am wirksamsten ( $d=1.38$ ), während für Ansätze zur Anreicherung von Texten nur geringe Effekte ( $d=.24$ ) nachgewiesen werden konnten.

Zu vergleichbaren Ergebnissen kamen Mastropieri, Scruggs, Bakken und Whedon (1996), die in einer Übersichtsarbeit zeigen konnten, dass strategieorientierte Interventionen deutlich höhere Effekte erzielten als Maßnahmen zur Anreicherung von Texten mit Bildern, eine gezielte Strukturierung von Texten oder Programme zur Förderung von Basisfertigkeiten.

Im deutschsprachigen Raum wurde die Effektivität eines strategieorientierten Trainings des Textverständnisses u. W. – mit Ausnahme der Interventionsstudie von Antoniou und Souvignier (2007) bislang überwiegend an Stichproben durchschnittlich lesender Kinder aus Grundschulen, Gesamtschulen, Mittelschulen und Gymnasien evaluiert (z. B. Lenhard, Baier, Endlich, Lenhard, Schneider & Hoffmann, 2012; Mokhlesgerami et al., 2007; Seuring & Spörer, 2010; Souvignier & Mokhlesgerami, 2006).

Ansätze zur Verbesserung des Textverständnisses durch die Vermittlung von Lesestrategien sollten mit zunehmender Komplexität der Textinhalte eine besondere Bedeutung erlangen, während Förderansätze zur Verbesserung der Leseflüssigkeit in Jahrgangsstufen, in denen die Worterkennung üblicherweise bereits automatisiert ist, bestenfalls marginale Auswirkungen auf das Leseverständnis haben sollten. Diese Annahme kann belegt werden durch die Metanalyse von Edmonds und Kollegen (2009), die selbst für lern- und leseschwache Schüler zwischen der sechsten und der zwölften Klasse durch Maßnahmen, die auf eine Verbesserung der Leseflüssigkeit auf Wortebene abzielten, nur geringe Effekte in Bezug auf das Leseverständnis ermitteln konnten. Dagegen konnten für Ansätze, die den Schülern Verstehensstrategien vermittelten, Effektstärken in einem durchgängig großen Bereich nachgewiesen werden.

Systematische und strukturierte Interventionsformen, die z. B. mit Hilfe phonologisch orientierter Maßnahmen auf den Erwerb und die Anwendung des alphabetischen Prinzips und die Automatisierung der Worterkennung abzielen, können aber durchaus für jüngere Kinder und Schüler mit spezifischen Schwierigkeiten im Bereich der Worterkennung hilfreich sein und zumindest eine Grundlage für die Ausbildung eines angemessenen Leseverständnisses legen, jedoch handelt es sich dabei nicht um eine zwangsläufige Entwicklung. So kommen Chard, Vaughn und Tyler (2002) durch eine Aufarbeitung des Forschungsstandes zwar zu dem Ergebnis, dass eine Verbesserung der direkten Worterkennung in vielen Fällen mit Fortschritten im Bereich des Leseverständnisses einhergeht, auch wenn das nicht das eigentliche Ziel der Intervention war. Andererseits weisen Torgesen, Alexander, Wagner, Rashotte, Voeler und Conway (2001) darauf hin, dass phonologisch orientierte Fördermaßnahmen effektiv sind, was die Vermittlung des alphabetischen Prinzips der Schriftsprache angeht, mit einer solchen Verbesserung der Lesegeschwindigkeit und Lesegenauigkeit aber nicht zwingend eine ausreichende Basis für ein altersangemessenes Leseverständnis gelegt wird.

„The data trend from the studies of fluency indicates that increased reading rate and accuracy did not always result in improved comprehension [...]. Stated more succinctly, as students improved their oral reading fluency, comprehension did not jointly improve” (Edmonds et al., 2009, S. 292).

Insbesondere die Schüler, deren primäre Probleme im Bereich lexikalischer oder grammatischer Fähigkeiten sowie der aktiv-kognitiven Auseinandersetzung mit dem Gelesenen liegen, benötigen ergänzend zur Förderung der Worterkennung spezifische Unterstützung im Bereich sprachlich-kognitiver Fähigkeiten.

Auf Grund der engen korrelativen Zusammenhänge zwischen dem Textverständnis und lexikalischen Fähigkeiten (Rickets et al., 2007) sollte eine kontinuierliche Wortschatzarbeit im Unterricht eine effektive Methode zur Förderung des Leseverständnisses darstellen. Mit dem Wortschatzsammler (Motsch, Marks & Ulrich, 2018) liegt im deutschsprachigen Raum ein evaluiertes therapeutisches Konzept vor, das durch die Vermittlung von Wortlernstrategien auf die Erweiterung und Differenzierung lexikalischer Fähigkeiten abzielt. Für die Wirksamkeit einer isolierten Wortschatzförderung in Bezug auf das Leseverständnis können der Metanalyse von Souvignier und Antoniou (2007) zu Folge Effekte in einer Größenordnung von  $d=.39$  nachgewiesen werden.

Was die Auswirkungen einer isolierten Therapie oder Förderung im Bereich syntaktischer und morphologischer Fähigkeiten in Bezug auf das Textverständnis angeht, liegen u. W. keine Ergebnisse vor.

## 2.4 Verstehensstrategien

Eines der ersten Konzepte, das für die Förderung des Textverständnisses die Vermittlung von Verstehensstrategien in den Mittelpunkt stellte, ist der Ansatz des reziproken Lehrens und Lernens nach Palinscar und Brown (1984). Während die Lehrkraft die Strategien und Aktivitäten bei der Besprechung eines Textabschnitts in einem ersten Schritt modelliert und demonstriert, übernehmen die Kinder zunehmend mehr Eigenverantwortung und wenden die Lesestrategien in Kleingruppen beim Lesen von Texten an. Dabei übernehmen die Mitglieder einer Gruppe abwechselnd die Rolle des Lehrers. Das „Lehrerkind“ entscheidet, welcher Schüler welche Lesestrategie anwendet, gibt Rückmeldung über die Qualität der Strategieverwendung und bietet gegebenenfalls Hilfe an. Im Mittelpunkt des Konzepts stehen die Strategien des Fragenstellens („questioning“), des Zusammenfassens („summarizing“), die Klärung eigener Verstehenslücken („clarifying“) und das Treffen von Vorhersagen über nachfolgende Textinhalte („predicting“). Weitere Verstehensstrategien, für die in der Forschungsliteratur Belege ihrer Effektivität vorliegen, sind darüber hinaus das bildhafte Vorstellen (Hartmann, 2006) sowie das Aktivieren von Hintergrundwissen (Dole, Valencia, Greer & Wardrobe, 1991). Die Strategien werden im Folgenden kurz skizziert (vgl. Mayer, 2016b).

Die Strategie „Aktivierung von Vorwissen“ geht von der Annahme aus, dass Texte umso besser verstanden und mit bereits vorhandenem Wissen im Langzeitgedächtnis zu neuen differenzierteren Episoden verknüpft werden können (= Konstruktion eines Situationsmodells), wenn vor der Begegnung mit einem Text, wesentliche Hintergrundinformationen zur Thematik zur Verfügung gestellt werden und auch der für das Verstehen notwendige Wortschatz erarbeitet und geklärt wird.

In engem Zusammenhang mit dieser Strategie steht das Treffen von Vorhersagen. Nach der Phase der „Aktivierung von Vorwissen“ bzw. nach dem Lesen einzelner Textabschnitte sollen die Schüler spekulieren, welche Informationen sie im Text erwarten könnten oder welche über das Vorwissen hinausgehenden Fragen der Text beantworten könnte. Die Schüler können dazu wichtige Hinweise im Text, wie die Überschrift oder illustrierendes Bildmaterial ausnutzen, um Inhalte zu antizipieren.

Die Strategie des „Comprehension Monitoring“ zielt darauf ab, die Kinder zu motivieren, die Wörter oder grammatischen Strukturen zu identifizieren, die nicht vollständig verstanden wurden und deren Bedeutung zu klären, indem sie nach Hinweisen im Text suchen, nachfragen oder in einem Wörterbuch etc. nachschlagen.

Durch das Zusammenfassen sollen Textabschnitte auf ihre Kernaussagen reduziert und damit die Herstellung globaler Kohärenz und die Konstruktion eines Situationsmodells unterstützt werden (Palinscar & Brown, 1984; Seuring & Spörer, 2010). Diese Strategie ist insbesondere in der Arbeit mit lernschwachen Schülern von besonderer Bedeutung, da sich diese Schüler beim Lesen oftmals auf vertraut wirkende Stellen unabhängig von ihrer Bedeutung im Zusammenhang konzentrieren (Miller & Keenan, 2009). Ein tieferes Verständnis und langfristiges Behalten sind aber erst dann möglich, wenn die Aussage auf das Wesentliche „verdichtet“ (Hartmann, 2006, S. 40) wird.

Beim „Fragen an den Text stellen“ können die Schüler aufgefordert werden, in die Rolle der Lehrkraft zu schlüpfen und sich zu überlegen, welche Fragen diese stellen würde, um zu überprüfen, ob ihre Schüler den Inhalt einzelner Textabschnitte oder des gesamten Textes tatsächlich verstanden haben.

Bei der Strategie des „Bildhaften Vorstellens“ handelt es sich um das Generieren mentaler Vorstellungsbilder zu gelesenen Inhalten. Kompetente Leser zeichnen sich dadurch aus, dass sie sich beim Lesen in eine aktive Interaktion mit den Inhalten begeben, während derer die expliziten Textinformationen mit dem Vorwissen zu mentalen Bildern und damit zu einer visuellen Form eines Situationsmodells kombiniert werden. Im Zusammenhang mit dem Leseverstehen sollte die Aufforderung, mentale Bilder zu Textinhalten zu erzeugen zu einer Entschleunigung des Leseprozesses führen, „wodurch die semantische Verarbeitung des Gelesenen vertieft wird“ (Hartmann, Hochstrasser & Trauffer, 2013, S. 225).

### 3 Fragestellung der Metaanalyse

In der Aufarbeitung des Forschungsstandes in Abschnitt 2.3 finden sich deutliche Hinweise darauf, dass mit strategieorientierten Fördermaßnahmen Verbesserungen im Leseverständnis von Kindern und Jugendlichen in der Primar- und Sekundarstufe erzielt werden können (Souvignier & Antoniou, 2007; Edmonds et al., 2009).

Das Ziel der vorliegenden Arbeit bestand nun darin, zu überprüfen,

- (1) welche Effekte in welcher Größenordnung sich für strategieorientierte Interventionsmaßnahmen bei leseschwachen und durchschnittlich lesenden Schulkindern in Bezug auf das Leseverständnis nachweisen lassen,
- (2) inwiefern sich derartige Unterstützungsmaßnahmen positiv auf das Wissen um und die Anwendung der vermittelten Verstehensstrategien auswirken und
- (3) inwiefern Unterschiede in der Effektivität der Maßnahmen durch Merkmale wie das Untersuchungsdesign, die Implementation der Intervention, Charakteristika der Stichprobe u.ä. erkennbar sind.

Diese Aufarbeitung des aktuellen Forschungsstandes diene mittelfristig auch als Ausgangspunkt für die Entwicklung eines unterrichtsintegrierten Förderkonzepts zur Verbesserung des Leseverständnisses durch die Vermittlung von Lesestrategien bei Kindern mit sonderpädagogischem Förderbedarf (Mayer, 2018). Die vorliegende Analyse erscheint insbesondere vor dem Hintergrund sinnvoll, da die im Rahmen der vorliegenden Publikation aufgearbeiteten Metaanalysen sich nicht ausschließlich auf die Förderung des Leseverständnisses durch die Vermittlung von Verstehensstrategien konzentrierten (Souvignier & Antoniou, 2007) bzw. die Effektivität ausschließlich für die Zielgruppe der Schüler der Sekundarstufe ermittelten (Edmonds et al., 2009). Diese Aussage kann dadurch belegt werden, dass lediglich vier der Studien, die für die vorliegende Arbeit berücksichtigt wurden, auch in der Metaanalyse von Souvignier und Antoniou (2007) Eingang gefunden haben. Die Metaanalyse von Edmonds und Kollegen (2009) und die vorliegende Arbeit stimmen nur in einer Studie überein. Während in der Arbeit von Souvignier und Antoniou (2007) Studien aufgenommen wurden, die zwischen 1967 und 2005 publiziert wurden (Edmonds et al., 2009: 1994–2004), liegt der Veröffentlichungszeitraum der für die vorliegende Analyse berücksichtigten Studien zwischen 1990 und 2016.

## 4 Methode

### 4.1 Suche und Auswahl der Studien

Um eine möglichst große Auswahl an qualitativ hochwertigen Studien für die Metaanalyse zu erhalten, wurde eine mehrgleisige systematische Literaturrecherche durchgeführt. Zwischen Juni und Juli 2017 wurde zunächst eine digitale Literaturrecherche mit deutschen und englischen Suchbegriffen über die Weboberfläche EBSCO Host in den Datenbanken ERIC, Medline, PsycINFO, PsycARTICLES und PSYNDEX durchgeführt, die im Vorfeld auf die in Tabelle 1 genannten Kriterien eingegrenzt wurde.

Tab. 1: Auswahlkriterien

Auswahlkriterium	Erläuterung/Ausschluss von ...
<b>Publikationsjahr</b>	Berücksichtigt wurden ausschließlich Studien im Publikationszeitraum von 1990–2017.
<b>Suchbegriffe/-strategien</b>	<p>Deutsch:</p> <p>Lesestrategie* OR Verstehensstrategie* OR Leseverst* (Suchfeld: Titel; keine weiteren Einschränkungen)</p> <p>Englisch:</p> <ol style="list-style-type: none"> <li>1. read* AND *strateg* (Suchfeld: Titel),</li> <li>2. kombiniert mit: intervention OR treatment OR training OR teach* OR instruction OR therapy (Suchfeld: Abstract),</li> <li>3. kombiniert mit folgenden Suchbegriffen in fünf separaten Suchdurchläufen: (1) "summary*"; (2) „background knowledge“ OR „prior knowledge“; (3) questioning OR „asking question*“; (4) „comprehension monitoring“ OR monitoring; (5) „mental imagery“ OR „visuali*“ (Suchfeld: Abstract)</li> </ol>
<b>Zielgruppe</b>	Ausschluss von Studien, die Erwachsene als Zielgruppe definierten (über die Funktion „not“: adult* (Suchfeld: Titel))

Parallel zur digitalen Recherche erfolgte unter Berücksichtigung dieser Kriterien eine manuelle Suche in gängigen internationalen Fachzeitschriften (*Annals of Dyslexia*, *Journal of Learning Disabilities*, *Scientific Studies of Reading*, *Reading and Writing*, *Reading Research Quarterly*). Schließlich wurde auch einschlägigen Querverweisen aus gesichteten Arbeiten nachgegangen.

Das Suchergebnis der digitalen Recherche umfasste zunächst 302 Publikationen. Um diese im ersten Schritt identifizierten Publikationen auf relevante Studien zu reduzieren, wurden in einem ersten Schritt alle Arbeiten ausgeschlossen, bei denen es sich dem Titel und dem Abstract zu Folge nicht um eine Interventionsstudie handelte. Schließlich wurden für die Metaanalyse im folgenden Schritt ausschließlich Arbeiten aufgenommen,

- 1) bei denen ein Volltextzugriff möglich war,
- 2) die neben einer Experimentalgruppe auch eine Kontrollgruppe beinhalten (Evidenzlevel I-III nach CEBM 2009),
- 3) deren Interventions- und Kontrollgruppe(n) sich ausschließlich aus Probanden ohne Primärbeeinträchtigungen (genetische Syndrome, Autismus, sensorische Beeinträchtigungen wie Gehörlosigkeit) zusammensetzte,
- 4) die als Outcomevariable das Leseverständnis mittels informeller oder normierter Verfahren überprüften,
- 5) deren Ergebnisdarstellung die Berechnung von Effektstärken ermöglichte.

Nach Anwendung dieser Kriterien konnten für die vorliegende Analyse 23 Studien berücksichtigt werden. Publikationen, in denen die Ergebnisse von zwei oder mehr strategieorientierten Interventionen mit einer Kontrollgruppe verglichen (in Tabelle 2 mit \* markiert) wurden, wurden als einzelne Studien interpretiert, sodass es insgesamt möglich war, 38 Einzelvergleiche vorzunehmen.

#### 4.2 Kodierung der Studien

Jede in der Metaanalyse berücksichtigte Studie wurde in Bezug auf relevante deskriptive Kriterien (Merkmale des Untersuchungsdesigns, Stichprobenmerkmale, Merkmale der Implementation und der Intervention) sowie die interessierenden Ergebnismaße (Leseverständnis, Strategiewissen) kodiert und in eine Datenmaske (IBM SPSS Statistics, 2014) übertragen. Die Autoren kodierten die Studien weitgehend in Anlehnung an die Arbeit von Souvignier und Antoniou (2007) zunächst unabhängig voneinander und verglichen die Ergebnisse im Anschluss. Uneinigkeiten in der Kodierung wurde diskutiert, bis zu 100 % Konsens gefunden wurde. Auf die Bestimmung einer Interrater-Reliabilität mit weiteren Kodierern wurde verzichtet. Folgende Kriterien liegen der Kodierung zugrunde.

*Merkmale des Untersuchungsdesigns:* In Bezug auf das Untersuchungsdesign wurde die Anzahl der Messzeitpunkte (Prä-, Post-, Follow-up-Testung), der Untersuchungsgruppen sowie der Probanden je Untersuchungsgruppe festgehalten.

*Stichprobenmerkmale:* Neben dem Alter der Probanden wurden die Studien hinsichtlich der besuchten Klassenstufe (Primarstufe: Klasse 1/2, 3/4, Sekundarstufe) sowie dem Lern- bzw. Leseneiveau der Schüler zum Beginn der Intervention kategorisiert. Den von den Autoren vorgenommenen Einteilungen folgend wurden Stichproben mit durchschnittlichen Lesern, leseschwachen Probanden (Beeinträchtigungen in der Worterkennung und im Leseverständnis), Probanden mit spezifischen Schwierigkeiten im Leseverständnis bei gleichzeitig durchschnittlicher Lesefertigkeit, Probanden mit allgemeinen Lernbeeinträchtigungen, die sich auch im Leseverständnis niederschlagen, und unselektierte Stichproben differenziert.

*Merkmale der Implementation:* Als wesentliches Merkmal der umgesetzten Maßnahmen wurde der zeitliche Umfang und damit die Intensität der Förderung festgehalten. Neben der Angabe des totalen Umfangs der Förderung (in Stunden) und der Frequenz (täglich, mehrmals oder einmal pro Woche) wurden die Maßnahmen als kurze (<12 Fördereinheiten) oder lange Interventionen (>12 Fördereinheiten) klassifiziert. Darüber hinaus wurde das Setting, in dem die Förderung durchgeführt wurde (Einzel-, Kleingruppen-, Klassensetting) sowie die Durchführenden der Förderung (Forschungsteam, angeleitete Lehrkraft) kodiert.

*Merkmale der Intervention:* Was die Inhalte der Förderung angeht, wurden die Art der Vermittlungsmethode (computergestützt, direkte Instruktion durch die Lehrkraft, reziprokes Lehren und Lernen, Palinscar & Brown, 1984) oder eine Kombination aus direkter Instruktion und reziprokem Lehren) sowie die Anzahl der vermittelten Strategien kodiert. Zusätzlich wurde festgehalten, welche Strategie(n) isoliert oder in Kombination mit anderen Strategien vermittelt wurden. Das

Kodierschema für diese Variable ergab sich nach Sichtung der Studien und wurde nicht im Vorfeld festgelegt.

*Ergebnismaße:* Als zentraler Parameter für die Effektivität der Förderung wurden die Ergebnisse der Überprüfungen des Leseverständnisses berücksichtigt, die entweder mittels normierter oder informeller Verfahren ermittelt wurden. Darüber hinaus wurden die Resultate von Messinstrumenten berücksichtigt, die das Wissen um und/oder die Anwendung von Verstehensstrategien während des Leseprozesses erfassen.

Berücksichtigt wurden ausschließlich Outcome-Variablen, die eine Berechnung von Effektstärken ermöglichten. Diese wurden auf der Grundlage der in den Studien angegebenen Werte (Mittelwerte, Standardabweichungen, Stichprobengröße) berechnet und für die nachfolgenden, vergleichenden Analysen herangezogen (statistische Berechnungen: vgl. 4.3). Wurden mehrere normierte oder informelle Überprüfungen des Leseverständnisses durchgeführt, wurden die Daten der Verfahren gewählt, die das Konstrukt Leseverständnis bzw. Lesestrategie am besten abbilden.

### 4.3 Statistische Berechnungen

Um die Effektivität der Förderung zu bestimmen, wurde für jeden Vergleich zwischen Experimental- und Kontrollgruppe das Effektstärkenmaß Cohen's  $d$  berechnet, das den Unterschied zwischen zwei Gruppen in Form der Standardabweichung ausdrückt. Die Verwendung von Cohen's  $d$  entspricht der üblichen Vorgehensweise bei Metaanalysen (Edmonds et al., 2009; Souvignier & Antoniou, 2007; Swanson, Zheng & Jerman, 2009). Alternativ wäre es möglich gewesen für die Bestimmung der Effektstärke auf das Maß Hedge's  $g$  zurückzugreifen, das die ermittelten Effekte insbesondere bei kleinen Stichproben ( $n < 20$ ) genauer schätzt. Dies betrifft in der vorliegenden Analyse allerdings nur eine Studie (Aarnoutse, Brand-Gruwel & Oduber, 2007). Die unterschiedlichen Stichprobengrößen in den einzelnen Studien wurden berücksichtigt, indem für jede Studie eine gewichtete Effektstärke ( $d_w$ ) berechnet wurde (s. u.). Den Konventionen von Cohen (1988) folgend werden Effektstärken von  $d \geq .20$  als kleiner Effekt,  $d \geq .50$  als mittlerer und  $d \geq .80$  als großer Effekt interpretiert.

Die primär interessierende Effektstärke war für alle Outcome-Variablen der *Leistungszuwachs über die Zeit*, also die Veränderung der Experimentalgruppe (strategieorientierte Förderung des Leseverständnisses) vom Präzum Posttest bzw. zum Follow-up-Test im Vergleich zur Kontrollgruppe. Die dazu relevanten Daten in Bezug auf die unterschiedlichen Outcome-Variablen (normierte oder experimentelle Überprüfung des Leseverständnisses, Strategiewissen-/anwendung) wurden für die einzelnen Messzeitpunkte (Prätest vs. Posttest, Prätest vs. Follow-up-Test) zunächst aus den Originalstudien extrahiert und anschließend online in eine Datenmaske eingegeben, um die Effektstärken zu generieren (Lenhard & Lenhard, 2016). Dazu wurde auf die von Morris (2008) empfohlene Formel zur Berechnung der Effektstärke cohen's  $d$  (sog.  $d_{ppc2}$ ) zurückgegriffen, bei der die Differenz zwischen der Prä- und Posttest-Messung der Experimentalgruppe von der Differenz zwischen der Prä- und Posttest-Messung der Kontrollgruppe subtrahiert und durch die gepoolte Standardabweichung dividiert wird (vgl. auch Lenhard & Lenhard, 2016), also der Leistungszuwachs der Experimentalgruppe mit derjenigen der Kontrollgruppe verglichen wurde:

$$d_{ppc2} = C_p \left[ \frac{(M_{post,T} - M_{pre,T}) - (M_{post,C} - M_{pre,C})}{SD_{pre}} \right], \text{ wobei}$$

$$SD_{pre} = \sqrt{\frac{(n_T - 1) SD_{pre,T}^2 + (n_C - 1) SD_{pre,C}^2}{n_T + n_C - 2}} \quad \text{und} \quad C_p = 1 - \frac{3}{4(n_T + n_C - 2) - 1_{ist}}$$

(Morris, 2008, S. 369).

Mit dieser Berechnung werden zum einen mögliche Prätestunterschiede zwischen Experimental- und Kontrollgruppe sowie die Stichprobengröße kontrolliert. Damit sollte es sich hier um das aussagekräftigste Maß zur Beurteilung der Effektivität der erfolgten Maßnahmen handeln.



Darüber hinaus wurden die zu den mittleren standardisierten Differenzen (Cohen's  $d$ ) gehörigen Konfidenzintervalle berechnet (CI, unter Annahme des 95 %-Niveaus):

$$CI = d \pm Z_{\alpha/2} \times SE_d, \text{ wobei}$$

$$SE_d = \sqrt{V_d} \text{ und } V_d = \frac{n_T + n_C}{n_T \times n_C} + \frac{d^2}{2 \times (n_T + n_C)} \text{ ist}$$

(Borenstein, Hedges, Higgins & Rothenstein, 2009, S. 27).

Von einem signifikanten Unterschied zwischen den untersuchten Versuchsbedingungen kann ausgegangen werden, wenn das Konfidenzintervall ausschließlich im positiven oder im negativen Bereich liegt. Beinhaltet das Intervall dagegen auch den Wert Null kann keine gesicherte Aussage über die Richtung des Effekts getroffen werden. Die Konfidenzintervalle werden zusammen mit den Effektstärken in den Abbildungen 1–6 in *forest plots* dargestellt.

Zur Aggregation der Effektstärken zu einem Gesamteffekt wurde auf das Modell fester Effekte („fixed-effectmodel“) als statistisches Auswertungsverfahren zurückgegriffen (Döhring & Bortz, 2016). Dem „Fixed-Effect-Modell“ für Metanalysen folgend ist davon auszugehen, dass eine Studie mit geringem Stichprobenumfang einen höheren Stichprobenfehler beinhaltet als eine Studie mit großer Stichprobe und bei der Berechnung des Effekts deshalb mit geringerem Gewicht eingehen sollte. Dieser Annahme wurde Rechnung getragen, indem für die einzelnen Studien eine gewichtete, den Stichprobenumfang berücksichtigende Effektstärke ( $d_w$ ) berechnet wurde. Dazu wurden die Effektstärken jeder einzelnen Studie mit der Größe der jeweiligen Stichprobe multipliziert. Diese Einzelwerte wurden addiert sowie anschließend durch die Gesamtzahl der Probanden dividiert, für die diese Outcome-Variable erhoben wurde.

$$d_w = \frac{\sum N \times d}{\sum N}$$

Der prozentuale Anteil, mit dem auf dieser Basis jeder Einzelvergleich den Gesamteffekt der verschiedenen Outcome-Variablen beeinflusst, ist in den *forest plots* unter der Angabe des relativen Gewichts (*relative weight*) abzulesen.

Zur Ermittlung möglicher moderierender Variablen wurden weitere aggregierte Effektstärken für die zu diesem Zweck gebildeten Subgruppen unter Verwendung desselben statistischen Verfahrens berechnet (Vergleich der Effekte von Studien mit Probanden aus der Primar- und Sekundarstufe; Vergleich der Effekte von Studien mit unterschiedlichen Förderzeiträumen; Vergleich der Effekte von Studien, in denen eine oder mehrere Strategien vermittelt wurden etc.; vgl. 5.2.4).

## 5 Ergebnisse

### 5.1 Deskriptive Darstellung

Auf der Grundlage der beschriebenen Recherchestrategien verblieben 23 Studien, die in die Metaanalyse aufgenommen und nach dem skizzierten Schema kodiert wurden. Eine Zusammenfassung der Studien findet sich in Tabelle 2. Da in Studien mit mehreren Experimentalgruppen die Ergebnisse jeder EG mit der Kontrollgruppe in Form von Einzelvergleichen berücksichtigt wurden (siehe Kennzeichnung \* in Tabelle 2), konnten für die Metaanalyse insgesamt  $N=38$  Einzelvergleiche berücksichtigt werden.



Tab. 2: Kriterien geleitete Übersicht über die in der Metaanalyse aufgenommenen Studien (alphabetisch sortiert, Evidenzlevel orientiert an CEBM 2009)

	Autoren (Jahr) / Evidenzlevel	Studien-design	N	Alter	Charakteristika der Stichprobe	Intervention				Outcome-Variablen
						Intensität	Setting	Strategien	Vermittlungsmethode:	
1	Aarnoutse et al., 1997* <i>IIb (matched samples)</i>	T1 – I – T2	47	Primarstufe Kl. 3/4	1.) lernschwach: n=16, EG: n=8, KG: n=8 2.) Beeintr. LV: n=31 EG: n=16, KG: n=15	10 Std. 2–3x p.W.	Kleingruppe	Kombination von 4 Strategien	Kombination aus direkter Instruktion & reziprokem Lehren	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
2	Antoniou & Souvignier, 2007 <i>Ib (cluster-randomisiert)</i>	T1 – I – T2 T3	73	Sekundarstufe	lernschwach: EG: n=45, KG: n=28	22 Std. Frqz.: k.A.	Klassen-setting	Kombination von 5 Strategien	Kombination aus direkter Instruktion & reziprokem Lehren	<ul style="list-style-type: none"> <li>LV (informell)</li> <li>Strategieanwendung</li> </ul>
3	Berkeley et al., 2011* <i>Ib (cluster-randomisiert)</i>	T1 – I – T2 – T3	79	Sekundarstufe	1.) lernschwach: n=40, EG: n=20, KG: n=20 2.) lernschwach: n=39 EG: n=19, KG: n=20	6 Std. 1x p.W.	Klassen-setting	Kombination von 5 Strategien	Kombination aus direkter Instruktion & reziprokem Lehren	<ul style="list-style-type: none"> <li>LV (informell)</li> <li>Strategieanwendung</li> </ul>
4	Berkeley & Riccimini, 2011* <i>Ib (stratifizierte Randomisierung)</i>	T1-I-T2*	319	Sekundarstufe	1.) durchschnittl. Leser: n=288, EG: n=156, KG: n=132 2.) lernschwach: n=31 EG: n=21, KG: n=10	1 Std. 3x20 min an 3 Tagen in 1 W.	Klassen-setting	Kombination von 4 Strategien	Kombination aus direkter Instruktion & reziprokem Lehren („QRAC-the-Code“)	<ul style="list-style-type: none"> <li>LV (informell)</li> </ul>
5	Brand Gruwel et al., 1998 <i>IIb</i>	T1 – I – T2 – T3	157	Primarstufe Kl. 3/4	Leseschwache SuS: EG: n=83, KG: n=74	15 Std. 2–3x p.W.	Kleingruppe	Kombination von 4 Strategien	Kombination aus direkter Instruktion & reziprokem Lehren	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
6	Center et al., 1999 <i>Ib (paarweise Randomisierung)</i>	T1 – I – T2	66	Primarstufe Kl. 1/2	Beeintr. LV: EG: n=33, KG: n=33	12 Std. 2–3x p.W.	Kleingruppe	1 Strategie	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (normiert)</li> </ul>
7	Chan et al., 1990* <i>Ib (randomisiert)</i>	T1 – I T2	104	Primarstufe Kl. 1/2  Primarstufe Kl. 3/4	1.) Durchschnittl. Leser: A: n=26 (EG: n=13, KG: n=13) B: n=26 (EG: n=13, KG: n=13) 2.) leseschwach: A: n=26 (EG: n=13, KG: n=13) B: n=26 (EG: n=13, KG: n=13)	2 Std. täglich	Kleingruppe	1 Strategie	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (informell)</li> <li>Strategieanwendung</li> </ul>
8	Jitendra et al., 2000 <i>Ib (stratif. Randomisierung)</i>	T1 – I – T2 – T3	33	Sekundarstufe	beeintr. LV: EG: n=18, KG: n=15	8,75 Std. tägl.	Kleingruppe	1 Strategie	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (informell)</li> </ul>
9	Johnson-Glenberg 2000* <i>IIb</i>	T1 – I – T2	73	Primarstufe Kl. 3/4	1.) beeintr. LV: n=37 EG: n=23, KG: n=14 2.) beeintr. LV: n=36 EG: n=22, KG: n=14	13 Std. 2–3x p.W.	Kleingruppe	Kombination von 2 Strategien	Reziprokes Lehren	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>LV (informell)</li> <li>Strategieanwendung</li> </ul>
10	Lenhard et al., 2012* <i>IIb</i>	T1 – I – T2	282	Sekundarstufe	1.) leseschwach: n=152 EG: n=80, KG: n=72 2.) leseschwach: n=130 EG: n=58, KG: n=72	15 Std. 1x p.W. 13,5 Std. 1x p.W.	Kleingruppe	1. Kombination von 3 Strategien 2. 1 Strategie	1. direkte Instruktion durch Lehrkraft 2. computergestützt	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
11	Lucariello et al., 2012 <i>Ib (stratifizierte Randomisierung)</i>	T1 – I – T2	45	Primarstufe Kl.3/4	unselektiert: n=45 EG: n=23, KG: n=22	12 Std. 2–3x p.W.	Klassen-setting	1 Strategie	vorrangig Lernen am Modell	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
12	Lysynchuk et al., 1990* <i>Ib (paarweise Randomisierung)</i>	T1 – I – T2	72	Primarstufe Kl.4; Sekundarstufe Kl.7	1.) Beeintr. im LV: n= 36, EG: n=18; KG: n=18 2.) Beeintr. im LV: n= 36, EG: n=18; KG: n=18	13 Std. täglich	Kleingruppe	Kombination von 4 Strategien	Reziprokes Lehren	<ul style="list-style-type: none"> <li>LV (normiert)</li> </ul>
13	Mistar et al., 2016 <i>IIb</i>	T1 – I – T2	71	Sekundarstufe	Unselektiert: n=71 EG: n=36, KG: n=35	18 Std. 2x p.W.	Klassen-setting	Kombination von 2 Strategien	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (informell)</li> </ul>

	Autoren (Jahr) / Evidenzlevel	Studien-design	N	Alter	Charakteristika der Stichprobe	Intervention				Outcome-Variablen
						Intensität	Setting	Strategien	Vermittlungsmethode:	
14	Mokhesgerami et al., 2007 <i>IIb</i>	T1 – I – T2	617	Sekundarstufe	Unselektiert: n=617 EG: n=387, KG: n=230	21 Std. 1x p.W.	Klassensetting	Kombination von 5 Strategien	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
15	Mokhesgerami et al., 2006	T1 – I – T2 – T3	452	Sekundarstufe	Unselektiert: n=452 EG: n=367, KG: n=85	13 Std. 1x p.W.	Klassensetting	Kombination von 3 Strategien	vorrangig Lernen am Modell	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
16	Ponce et al., 2012 <i>IIb</i>	T1 – I – T2	825	Primarstufe Kl. 3/4	Unselektiert: n=825 EG: n=742, KG: n=83	45 Std. 1x p.W.	Klassensetting	Kombination von 4 Strategien	Computergestützt	<ul style="list-style-type: none"> <li>LV (normiert)</li> </ul>
17	Schünemann et al., 2013* <i>IIb (teilweise klassenweise Randomisierung der beiden EG)</i>	T1 – I – T2 – T3	368	Sekundarstufe	1.) unselektiert: n=189 EG: n=127, KG*: n=62 2.) unselektiert: n=179 EG: n=117, KG*: n=62	10,5 Std. 2–3xp.W.	Klassensetting	Kombination von 4 Strategien	Reziprokes Lehren Reziprokes Lehren + selbstreguliertes Lernen	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
18	Seuring & Spörer, 2010* <i>IIb</i>	T1 – I – T2 – T3	506	Sekundarstufe	1.) unselektiert: n=253 EG: n=127, KG*: n=126 2.) unselektiert: n=253 EG: n=127, KG*: n=126	9 Std. 2–3x p.W.	Klassensetting	Kombination von 4 Strategien bzw. 3 Strategien	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
19	Souvignier & Mokhesgerami, 2006* <i>IIb</i>	T1 – I – T2 – T3	582	Sekundarstufe	1.) unselektiert: n=179 EG: n=95, KG*: n=84 2.) unselektiert: n=230 EG: n=146, KG*: n=84 3.) unselektiert: n=173 EG: n=89, KG*: n=84	15 Std. 2–3x p.W.	Klassensetting	Kombination von 6 Strategien	Direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>Strategieanwendung</li> </ul>
20	Spörer et al., 2008 <i>IIb</i>	T1 – I – T2 – T3	45	Sekundarstufe	unselektiert: n=45 EG: n=21, KG: n=24	14,25 Std. 2–3x p.W.	Klassensetting	Kombination von 2 Strategien	Reziprokes Lehren	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>LV (informell)</li> </ul>
21	Spörer et al., 2009* <i>Ib (cluster-randomisiert)</i>	T1 – I – T2 – T3	342	Primarstufe Kl. 3/4	1.) unselektiert: n=108 EG: n=42, KG*: n=66 2.) unselektiert: n=126 EG: n=60, KG*: n=66 3.) unselektiert: n=108 EG: n=42, KG*: n=66	10,5 Std. 11 Std. 11 Std. 2–3x p.W.	Kleingruppe	Kombination von 4 Strategien	1. Tradition. Reciprocal Teaching (RT) 2. RT in Paaren 3. Direkte Instruktion durch Lehrkraft	<ul style="list-style-type: none"> <li>LV (informell)</li> <li>Strategieanwendung</li> </ul>
22	Spörer & Brunstein, 2009 <i>Ib (cluster-randomisiert)</i>	T1 – I – T2	186	Sekundarstufe	unselektiert: EG: n=95, KG: n=91	10 Std. 2–3x p.W.	Klassensetting	Kombination von 2 Strategien	Reziprokes Lehren	<ul style="list-style-type: none"> <li>LV (normiert)</li> <li>LV (informell)</li> <li>Strategieanwendung</li> </ul>
23	Wassenburg et al., 2015 <i>IIb</i>	T1 – I – T2	115	Primarstufe Kl. 3/4	durchschnittl. Leser: EG: n=64, KG: n=51	4 Std. 2–3x p.W.	Kleingruppe	1 Strategie	direkte Instruktion durch Lehrer	<ul style="list-style-type: none"> <li>LV (normiert)</li> </ul>

### 5.1.1 Studiendesigns und Evidenzlevel

Bei den meisten aufgenommenen Studien handelt es sich um Interventionsstudien mit einem klassischen Prä-Posttest-Design mit mindestens einer Kontrollgruppe.

Bei circa einem Drittel der berücksichtigten Studien (n=10) wurde zusätzlich ein Follow-up-Test (T3) durchgeführt, um mögliche Langzeiteffekte zu bestimmen (n=2637, s. Tabelle 2: Studiendesign). Der Zeitpunkt der Follow-up-Testung lag zwischen sechs Wochen und fünf Monaten nach Durchführung des Posttests (T2).

Als Kontrollbedingung fungierten in der Regel reguläre Lesestunden ohne spezifische Instruktionen (im Sinne eines *treatment as usual*). Die Gruppenzuweisung erfolgte in zehn Fällen randomisiert, wobei überwiegend eine Cluster-Randomisierung auf Schul- oder Klassenebene stattfand. Teilweise wurden auch Forschungszwillinge („matched pairs“) gebildet, die dann randomisiert auf die Versuchsbedingungen aufgeteilt wurden (Center et al., 1999; Lysynchuk et al., 1990). In den verbleibenden Studien, in denen die Zuteilung zur Versuchsbedingung nicht randomisiert erfolgte, wurde versucht, unter Berücksichtigung der Prätestergebnisse und der Erhebung weiterer Stichprobenmerkmale vergleichbare Ausgangsbedingungen in den Gruppen zu schaffen.

Unter Berücksichtigung der formalen Angaben zum Studiendesign sind die aufgenommenen Studien dem Evidenzlevel Ib-IIb zuzuordnen (randomisierte, kontrollierte Studien/RCTs und gut geplante Kontrollstudien ohne Randomisierung, vgl. CEBM 2009, Tabelle 2: Evidenzlevel). Damit handelt es sich bei dem für die Metaanalyse genutzten Studienkorpus um eine aussagekräftige Datenbasis, deren Ergebnisse und gewonnenen Erkenntnisse für die Planung zukünftiger Interventionen grundsätzlich berücksichtigt werden sollten.

### 5.1.2 Stichprobencharakteristika

*Alter und Klassenstufe:* Für die vorliegende Metaanalyse wurden die Daten von insgesamt  $N=5.459$  Probanden im Alter von 7;7-14;7 Jahren ( $MW=11;2$ ,  $SD=1.77$ ) verarbeitet ( $n_1=3.397$ ,  $n_2=2.062^1$ ). Die Stichprobengröße lag zwischen 16 und 825 Probanden mit durchschnittlich 143,66 Teilnehmern ( $SD: 169,24$ ). 1.810 Probanden mit einem durchschnittlichen Alter von 9;6 Jahren (7;7-11;1 Jahre,  $SD=1.1$ ;) aus zehn Studien besuchten die Primarstufe (Tabelle 2: Alter). Dabei handelt es sich überwiegend um ältere Grundschüler der dritten und vierten Jahrgangsstufe ( $n=1.705$ ), lediglich  $n=105$  Probanden aus zwei Studien besuchten die erste oder zweite Klasse (Center et al., 1999; Chan et al., 1990). Einen wesentlich größeren Anteil der Stichprobe dieser Metaanalyse stellen mit  $n=3.649$  aus insgesamt 14 Studien Schüler der Sekundarstufe mit einem durchschnittlichen Alter von 12;1 Jahren dar (10;0-14;9,  $SD=1.3$ ) (Tabelle 2: Alter).

*Lern- und Lesekompetenzen:* Die Stichproben der inkludierten Studien weisen nicht nur eine große Heterogenität in Bezug auf das Alter auf, sondern lassen sich auch hinsichtlich der Lesefähigkeiten und des kognitiven Entwicklungsstandes der Probanden differenzieren (Tabelle 2: Charakteristika der Stichprobe). So bestanden die Zielgruppen der Interventionsstudien aus

- durchschnittlichen Lesern ( $n=455$ ),
- leseschwachen Probanden (Beeinträchtigungen in der Worterkennung und im Leseverständnis,  $n=491$ ),
- Probanden mit spezifischen Schwierigkeiten im Leseverständnis bei gleichzeitig durchschnittlicher Lesefertigkeit ( $n=275$ ),
- Probanden mit allgemeinen Lernbeeinträchtigungen, die sich auch im Leseverständnis niederschlagen ( $n=199$ ) und
- unselektierten Stichproben ( $n=4.039$ ).

### 5.1.3 Charakterisierung der durchgeführten Interventionen

*Setting der Intervention:* In 14 Studien wurde die Förderung im Klassensetting durchgeführt (Tabelle 2: Setting; EG:  $n=2.898$ ). Dabei handelte es sich überwiegend um Klassen der Sekundarstufe. Nur in zwei Studien, in denen Schüler der Primarstufe gefördert wurden, wurde die Maßnahme ebenfalls im Klassensetting umgesetzt (EG:  $n=848$ ; Lucariello et al., 2012; Ponce et al., 2012). In den übrigen neun Studien mit Schülern der Primarstufe wurde die Förderung im Kleingruppensetting mit zwei bis maximal acht Schülern durchgeführt ( $n=463$ ; Aarnoutse et al., 1997; Brand-Gruwel et al., 1998; Center et al., 1999; Chan et al., 1990; Jitendra et al., 2000; Johnson-Glenberg, 2000; Lenhard et al., 2012; Lysynchuk et al., 1990; Spörer et al., 2009; Wassenburg et al., 2015). In zwei Studien waren die Probanden im Kleingruppensetting Schüler der Sekundarstufe ( $n=69$ , Jitendra et al., 2000; Lysynchuk et al., 1990).

*Vermittelte Strategien (Anzahl, Art, Kombination):* In sechs der aufgenommenen Studien wurde nur eine Strategie vermittelt ( $n=493$ , Center et al., 1999; Chan et al., 1990; Jitendra et al., 2000; Lenhard et al., 2012: Teilstudie 2; Lucariello et al., 2012; Wassenburg et al., 2015). Dabei handelte es sich bei drei Studien um die Strategie des „Mental Imagery“ (Center et al., 1999; Chan et al., 1990; Lucariello et al., 2012), in zwei Studien um die Strategie des Zusammenfassens (Jitendra et al., 2000; Lenhard et al., 2012/2) und in einer Studie um die Strategie des Comprehension Monitorings (Wassenburg et al., 2015). Mit Ausnahme der Studie von Lucariello und Kollegen (2012), bei der die Strategie des Mental Imagery im Klassensetting und vorrangig durch ein Lernen am Modell vermittelt wurde, und der computergestützten Vermittlung bei Lenhard und Kollegen (2012), wurde die ausgewählte Strategie über die direkte Instruktion durch die Lehrkraft unterrichtet.

1 Durch die Aufspaltung einer Studie in mehrere Einzelvergleiche wurden bei mehr als einer Experimentalgruppe die Daten von insgesamt ca. 650 Probanden aus den jeweiligen Kontrollgruppen doppelt gezählt.

In 17 Studien (n=4966) wurden hingegen bis zu sechs Strategien in die Förderung integriert. In den meisten Studien wurden vier Strategien unterrichtet (Aarnoutse et al., 1997; Berkeley & Riccomini, 2011; Brand Gruwel et al., 1998; Lysynchuk et al., 1990; Ponce et al., 2012; Schünemann et al., 2013; Spörer et al., 2009), wobei in den meisten Fällen das Zusammenfassen, das Stellen von Fragen an den Text, das Comprehension Monitoring sowie das Treffen von Vorhersagen fokussiert wurden. In jeweils drei Studien wurden zwei (Spörer & Brunstein, 2009; Johnson-Glenberg, 2000; Mistar et al., 2016), drei (Lenhard et al., 2012; Mokhlesgerami et al., 2006; Seuring & Spörer, 2010) resp. fünf Strategien (Antoniou & Souvignier, 2006; Berkeley et al., 2011; Mokhlesgerami et al., 2007) vermittelt.

*Methode:* Bei knapp der Hälfte der Studien (n=10) erfolgte die Vermittlung der Strategien durch die direkte Instruktion der Lehrkraft (n=2354, Center et al., 1999; Chan et al., 1990; Jitendra et al., 2000; Lenhard et al., 2012; Mistar et al., 2016; Mokhlesgerami et al., 2007; Seuring & Spörer, 2010; Souvignier & Mokhlesgerami, 2006; Spörer et al., 2009; Wassenburg et al., 2015). In sechs Studien kam das reziproke Lehren und Lernen in Anlehnung an Palinscar und Brown (1984) als Vermittlungsmethode zum Einsatz (n=978, Johnson-Glenberg, 2000; Lysynchuk et al., 1990; Schünemann et al., 2013; Spörer et al., 2008; Spörer et al., 2009; Spörer & Brunstein, 2009). Eine Kombination aus den beiden genannten Methoden fand sich in fünf Studien wieder (n=356, Aarnoutse et al., 1997; Antoniou & Souvignier, 2006; Berkeley et al., 2011; Berkeley & Riccomini, 2011; Brand Gruwel et al., 1998). Spörer und Kollegen (2009) verglichen in ihrer Studie bewusst drei verschiedene didaktische Methoden und tauchen daher in dieser Analyse mehrfach auf. Darüber hinaus fand das Lernen am Modell als Methode Erwähnung (n=497, Lucariello et al., 2012; Mokhlesgerami et al., 2006) sowie in zwei Fällen eine computergestützte Vermittlung der Strategien (n=955, Lenhard et al., 2012; 2. Experimentalgruppe; Ponce et al., 2012).

*Implementation der Intervention:* In 14 Studien erfolgte die Implementation der Intervention in der Experimentalgruppe durch das Forscherteam (n=1.218; Aarnoutse et al., 1997; Berkeley & Riccomini, 2011; Brand Gruwel et al., 1998; Center et al., 1999; Chan et al., 1990; Jitendra et al., 2000; Johnson-Glenberg, 2000; Lucariello et al., 2012; Lysynchuk et al., 1990; Schünemann et al., 2013; Seuring & Spörer, 2010; Spörer et al., 2008; 2009; Wassenburg et al., 2015). In den neun übrigen Studien mit n=2.179 Probanden wurde die strategieorientierte Förderung durch Lehrkräfte im Kleingruppen- oder Klassensetting durchgeführt (Antoniou & Souvignier, 2007; Berkeley et al., 2011; Lenhard et al., 2012; Mistar et al., 2016; Mokhlesgerami et al., 2006, 2007; Ponce et al., 2012; Souvignier & Mokhlesgerami, 2006; Spörer & Brunstein, 2009).

*Intensität der Förderung:* Im Durchschnitt lag die Interventionszeit bei 11.42 Zeitstunden (SD=7.67). Dabei variiert die Interventionsdauer zwischen gerade einmal einer Zeitstunde (Berkeley & Riccomini, 2011: 3x 20 Minuten) bis zu einem Förderzeitraum von 45 Zeitstunden (Ponce et al., 2012).

In circa der Hälfte der aufgenommenen Studien (n=12) ist der Förderzeitraum als Kurzzeitin-tervention von maximal zwölf Zeitstunden zu charakterisieren (n=2210; Aarnoutse et al., 1997; Berkeley et al., 2011; Berkeley & Riccomini, 2011; Center et al., 1999; Chan et al., 1990; Jitendra et al., 2000; Lucariello et al., 2012; Schünemann et al., 2013; Seuring & Spörer, 2010; Spörer et al., 2009; Spörer & Brunstein, 2009; Wassenburg et al., 2015). In elf Studien lag der zeitliche Umfang der Förderung bei mehr als zwölf Zeitstunden (n=3249 Probanden; Brand Gruwel et al., 1998; Johnson-Glenberg, 2000; Lenhard et al., 2012; Lysynchuk et al., 1990; Mistar et al., 2016; Mokhlesgerami et al., 2006, 2007; Souvignier & Mokhlesgerami, 2006; Spörer et al., 2008, Antoniou & Souvignier, 2007; Ponce et al., 2012).

#### 5.1.4 Erhobene Outcome-Variablen

In insgesamt 17 Studien wurde das *Leseverständnis über normierte Verfahren* erhoben (Tabelle 2: Outcome Variablen, n=4780, davon n=8 Studien mit n=1707 Probanden der Primarstufe, sowie n=8 Studien mit n=3073 Probanden aus der Sekundarstufe, zzgl. der Studie von Lysynchuk et al., 1990 mit beiden Altersstufen als Probandengruppen). Als Aufgabenformate dienten in der Regel Fragen zu Kurzgeschichten/Textpassagen, die im Multiple-Choice-Format beantwortet werden sollen.

In zehn Studien wurde das *Leseverständnis über informelle Verfahren* erhoben (Tabelle 2: Outcome Variablen,  $n=1325$  davon  $n=3$  Studien mit 519 Probanden der Primarstufe und  $n=7$  Studien mit 806 Probanden der Sekundarstufe). In den üblicherweise speziell für die jeweilige Studie konstruierten Verfahren wurde das Leseverständnis vergleichbar zu den standardisierten Verfahren in der Regel ebenfalls über Fragen im Multiple-Choice-Format erhoben, die auf explizit genannte oder implizite Informationen zu einem gelesenen Text abzielen. Teilweise sollten stattdessen oder zusätzlich offene Fragen beantwortet oder die Hauptideen einzelner Textabschnitte schriftlich zusammengefasst werden.

In 15 Studien wurden das *Wissen um bzw. die Anwendung von Verstehensstrategien* während des Leseprozesses überprüft (Tabelle 2: Outcome Variablen,  $n=4232$  davon  $n=6$  Studien mit 768 Probanden der Primarstufe und  $n=9$  Studien mit 3464 Probanden der Sekundarstufe). Die Messinstrumente wurden überwiegend für die Studien selbst konzipiert und decken daher eine große Spannbreite an Aufgabenformaten ab. Zum einen kamen Aufgabenformate zum Einsatz, die auf die *Überprüfung einer direkten Anwendung der Strategien an unbekannten Lesetexten* abzielten. Die Qualität der Strategieanwendung wurde dann in der Regel anhand von Ratingskalen und unter Angabe der Interraterreliabilität bewertet. Neben der Überprüfung der unmittelbaren Anwendung der Strategien während der Auseinandersetzung mit einem neuen Text wurde in anderen Studien eher das *Wissen über die Anwendung von Lesestrategien* überprüft (im Sinne einer Überprüfung des *metakognitiven* Lesestrategiewissens, vgl. Mokhlesgerami et al., 2007).

## 5.2 Metaanalytische Auswertung der Effektgrößen

### 5.2.1 Effekte auf das Leseverständnis in normierten Testverfahren

Zur Beurteilung der Effektivität strategieorientierter Förderkonzepte in Bezug auf Verbesserungen in normierten Leseverständnistests werden nachfolgend die standardisierten Unterschiede zwischen den Experimentalgruppen (strategieorientierte Interventionen) und den jeweiligen Kontrollgruppen im Bezug zur Entwicklung des Leseverständnisses berichtet (Cohen's  $d$ ). Berücksichtigt wurden alle Studien, in denen das Leseverständnis nach Beendigung der Intervention mittels normierter Verfahren erfasst wurde. Wie in Kapitel 4.3 dargestellt, handelt es sich dabei um einen Vergleich des Leistungszuwachses zwischen der Experimentalgruppe und der Kontrollgruppe unter Berücksichtigung der Stichprobengröße von T1 zu T2 bzw. T3, sodass mögliche Prätestunterschiede kontrolliert wurden.

**Unmittelbare Effekte (T1-T2):** Die Berechnung der mittleren, gewichteten Effektstärke für kurzfristig nachweisbare Effekte (PräPosttest) erfolgte auf der Grundlage von 24 Datensätzen mit insgesamt  $n=4.438$  Probanden, in denen der Fortschritt der Stichprobe mittels normierter Verfahren erfasst wurde. Abbildung 1 zeigt den so ermittelten Gesamteffekt sowie die in diesem Zusammenhang berücksichtigten Studien mit Angabe der jeweiligen Effektstärke und deren relativem Anteil (relative weight) am Gesamteffekt. Visualisiert werden die mittlere, gewichtete Effektstärke und die einzelnen Effektstärken unter Angabe des 95 %-Konfidenzintervalls in dem rechtsseitigen *forest plot*.



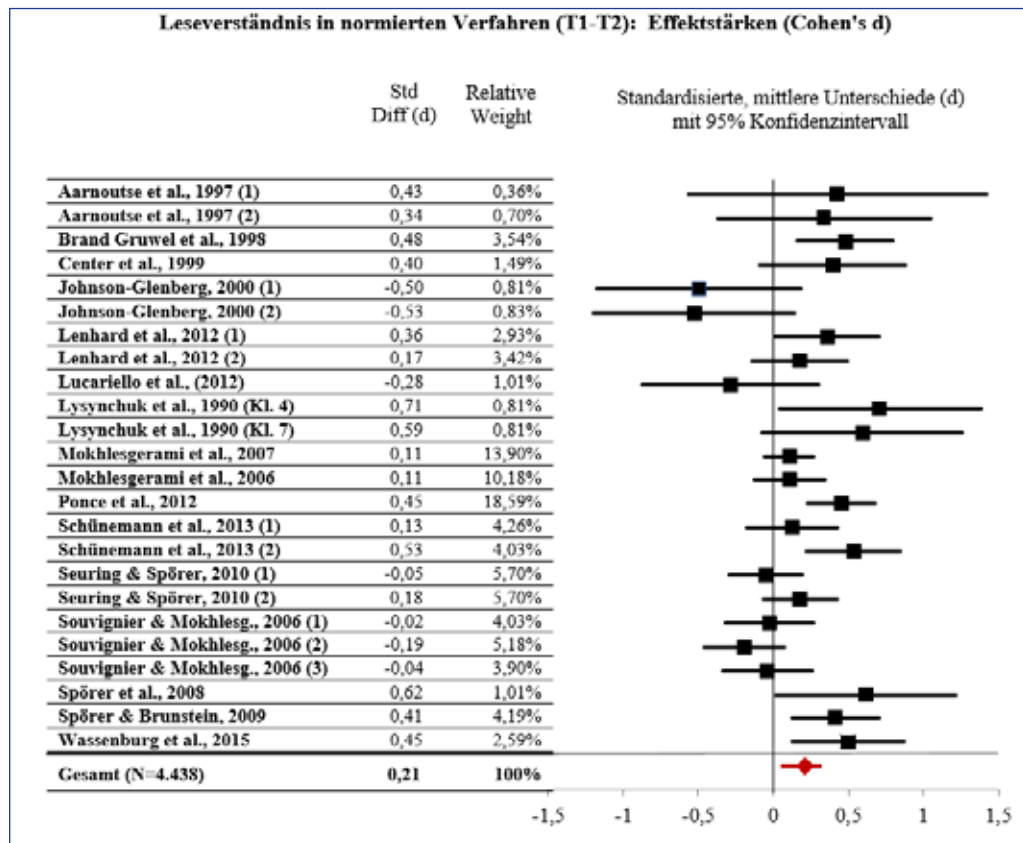


Abb. 1: Verbesserungen im Leseverständnis unmittelbar nach Interventionsende (T1-T2), gemessen mit normierten Verfahren – forest plot zur Darstellung der Effektstärken unter Angabe des Konfidenzintervalls sowie der gemittelten gewichteten Gesamteffektstärke.

Der Gesamteffekt, für dessen Berechnung die Effektstärken gemittelt und entsprechend der Stichprobengrößen gewichtet wurden, liegt bei  $d=.21$  [.15, .27]. Das Konfidenzintervall schließt den Wert Null aus, sodass über alle Studien hinweg, ein signifikanter Vorteil strategieorientierter Maßnahmen im Vergleich zu alternativen Angeboten angenommen werden kann. Diese Meta-Effektgröße entspricht den Konventionen von Cohen (1988) folgend einem *kleinen* Effekt.

Für neun Studien kann durch die metaanalytische Aufbereitung der Daten kein statistisch bedeutsamer Vorteil einer Strategievermittlung nachgewiesen werden (vgl. Abb. 1;  $-.19 \leq d \leq .18$ ; Anteil an Gesamteffekt: 56,27%). In drei (Teil-) Studien deuten sich auf den ersten Blick negative Effekte kleiner bis mittlerer Effektstärke an. Dabei handelt es sich um die beiden Teilstudien von Johnson-Glenberg (2000; Teilstudie 1:  $d=-0,50$  [-1.18, .18]; Teilstudie 2:  $d=-0,53$  [-1.21, .15]) und um die Studie von Lucariello und Kollegen (2012,  $d=-.28$  [-.87, .30]). Bei beiden Studien handelt es sich um eine mittellange, mehrmals wöchentlich stattfindende Intervention mit Probanden der Primarstufe (Kl. 3/4), wobei zum einen Kinder mit spezifischen Beeinträchtigungen im Leseverständnis (Johnson-Glenberg, 2000) zum anderen eine unselektierte Stichprobe (Lucariello et al., 2012) die Zielgruppe der Intervention darstellte.

Gemeinsamkeiten zwischen den Studien konnten bis auf die genannte Dauer und Frequenz der Intervention nicht identifiziert werden (vgl. Tabelle 2). Das Konfidenzintervall der ermittelten Effektstärken umfasst jedoch jeweils den Wert Null, sodass für beide Studien keine gesicherte Aussage über die Richtung des Effekts gemacht werden kann. Ein statistisch relevanter Nachteil einer strategieorientierten Förderung im Vergleich zur Kontrollbedingung zeigt sich unter Berücksichtigung des Konfidenzintervalls demnach nicht. Aufgrund der Stichprobengrößen fallen diese drei Studien insgesamt zudem nur wenig ins Gewicht (Anteil am Gesamteffekt: 2,65 %).

In insgesamt zwölf Studien wurden positive Effekte kleiner bis mittlerer Effektstärke zugunsten der Strategievermittlung gefunden. Diese Studien tragen zu 41,05 % zum Gesamteffekt bei. Davon offenbaren acht (Teil-)Studien ausschließlich im positiven Bereich liegende Konfidenzintervalle, die auf einen signifikanten Vorteil einer strategieorientierten Förderung im Vergleich zu einer Alternativbehandlung oder Lesestunden ohne spezifische Instruktionen hindeuten. Unter diesen Studien liegt der größte Unterschied zwischen Experimental- und Kontrollgruppe mit



$d=.71$  [.04, 1.39] und wurde in der Teilstudie von Lysynchuk und Kollegen (1990) für die  $n=36$  Probanden der vierten Klasse gefunden, die alle spezifischen Beeinträchtigungen im Leseverständnis aufwiesen. In täglichen Förderstunden wurden hier sukzessive vier Strategien eingeführt und fortwährend über einen Zeitraum von zwei Wochen eingeübt (vgl. Tab. 2).

**Langfristige Effekte (T1-T3):** Zur Berechnung der langfristigen Effekte auf das Leseverständnis standen nur noch 13 (Teil-)Studien zur Verfügung, die auf den Daten von insgesamt  $n=2.452$  Probanden beruhen. Insbesondere in den Studien, in denen die größten Effekte unmittelbar nach Beendigung der Intervention nachgewiesen werden konnten, wurde keine follow-up Untersuchung durchgeführt (z. B. Ponce et al., 2012; Lysynchuk et al., 1990; Wassenburg et al., 2015). Dafür kann für die folgende Analyse erstmalig die Studie von Spörer und Kollegen (2009) berücksichtigt werden, in der normierte Verfahren ausschließlich zur Messung von Langzeiteffekten eingesetzt wurden.

Analog zur ersten Abbildung zeigt Abbildung 2 die metaanalytische Aufbereitung der Daten in einem *forest plot* mit den statistisch relevanten Kennwerten.

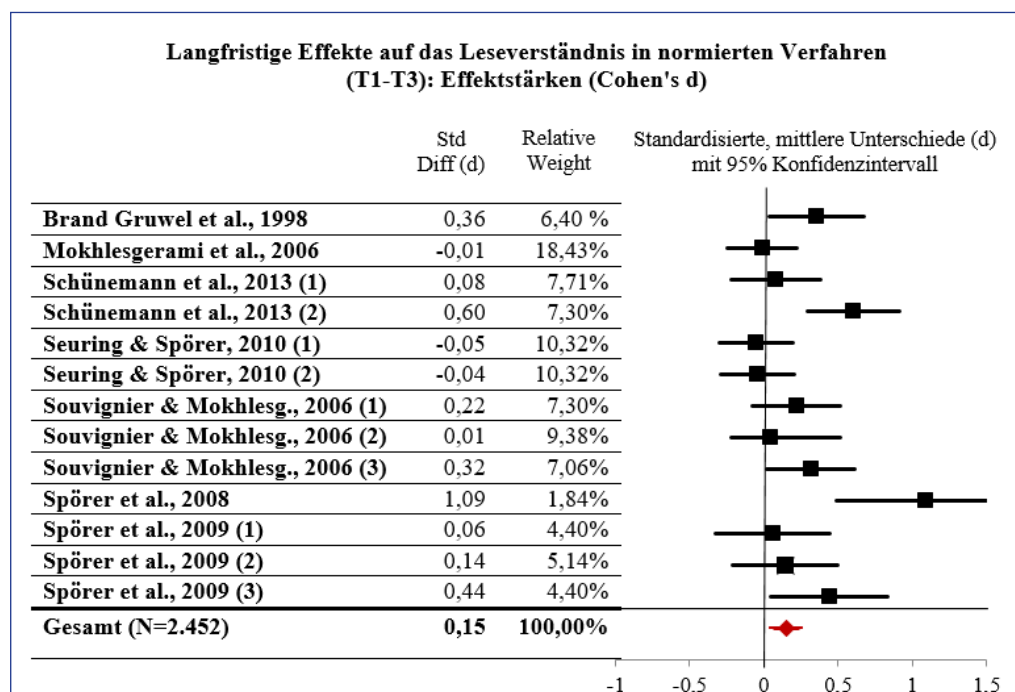


Abb. 2: Langfristige Verbesserungen im Leseverständnis (T1-T3), gemessen mit normierten Verfahren – forest plot zur Darstellung der Effektstärken unter Angabe des Konfidenzintervalls sowie der gemittelten gewichteten Gesamteffektstärke.

Übereinstimmend mit dem Ergebnis für unmittelbare Interventionseffekte lassen sich unter Berücksichtigung der Konfidenzintervalle auch langfristig keine *negativen* Effekte eines strategienorientierten Förderkonzepts im Vergleich zu alternativen Maßnahmen nachweisen.

Aus dem Vergleich mit den unmittelbaren Effekten (T1-T2) lässt sich aber auch die Annahme einer langfristig stabilen oder sogar zunehmenden Überlegenheit von Strategieansätzen nicht unmittelbar ableiten. Der Gesamteffekt von  $d=.15$  [.07, .23] deutet lediglich positive Tendenzen an.

In sieben der für diese Analyse berücksichtigten (Teil-)Studien lässt sich *kein* statistisch bedeutsamer Effekt (vgl. Abb. 2;  $-.05 \leq d \leq .22$ ; Anteil an Gesamteffekt: 73%) belegen. Die Teilstudie 1 von Souvignier und Mokhlesgerami (2006) weist zwar tendenziell mit  $d=.22$  auf einen kleinen Effekt zugunsten des Strategiekonzepts hin. Wird jedoch zusätzlich das Konfidenzintervall berücksichtigt [-.07, .51], relativiert sich dieser Effekt. Im Vergleich zu den unmittelbaren Effekten (T1-T2) zeigt sich in den verschiedenen Teilstudien aber eine tendenzielle Verbesserung der Probanden im Falle einer langfristigen Leistungsbeobachtung (T1-T3). Evaluiert wurde in dieser Studie das für den deutschsprachigen Raum entwickelte Förderkonzept der „Text-Detektive“ (Gold, Mokhlesgerami, Rühl, Schreblowski & Souvignier, 2004), in dem die Strategien (Aktivie-

ren von Hintergrundwissen, Zusammenfassen, Unklarheiten beseitigen, mental imagery) sukzessive eingeführt und von metakognitiven Übungselementen begleitet werden.

In den verbleibenden fünf Studien zeigen sich positive Langzeiteffekte von kleiner bis großer Effektstärke mit im positiven Bereich liegenden Konfidenzintervallen. Den stärksten Effekt weist die Studie von Spörer und Kollegen (2008) mit  $d=1.09$  [.49, 1.69] auf, deren Ergebnisse aufgrund der geringen Anzahl an Probanden ( $n=45$ ) den Gesamteffekt aber nur minimal beeinflusst. Im Vergleich zum unmittelbaren Effekt mittlerer Stärke ( $d=.62$  [.02, 1.22]) konnten die Probanden dieser Studie ihre Leistung sechs Wochen nach Interventionsabschluss im Vergleich zur Kontrollgruppe nochmals deutlich steigern. Bei den Probanden handelte es sich um eine unselektierte Stichprobe deutschsprachiger Siebtklässler, denen hochfrequent über fünf Wochen die Strategien des Zusammenfassens und des Vorhersagens vermittelt wurden.

### 5.2.2 Effekte auf das Leseverständnis in informellen Verfahren

Analog zum vorangegangenen Abschnitt werden im Folgenden die standardisierten Unterschiede zwischen den Experimentalgruppen (strategieorientierte Interventionen) und den Kontrollgruppen in Bezug zum Leseverständnis berichtet, die über informelle Verfahren erhoben wurden.

**Unmittelbare Effekte (T1-T2):** Die Grundlage für die Berechnung der mittleren, gewichteten Effektstärke für unmittelbare Trainingseffekte (T1-T2) bildeten 18 Datensätze mit insgesamt  $N=1.325$  Probanden. Abbildung 3 auf der nachfolgenden Seite zeigt den Gesamteffekt, die einzelnen berücksichtigten Studien (mit standardisierter Differenz  $d$  und relativem Gewicht) sowie das dazugehörige *forest plot* (unter Angabe des 95 %-Konfidenzintervalls).

Erkennbar ist ein grundsätzlich zur ersten Abbildung ähnliches Bild, mit überwiegend noch deutlicher im positiven Bereich liegenden Effektstärken und Konfidenzintervallen. Unter entsprechender Berücksichtigung aller Studien ergibt sich hier im Durchschnitt ein positiver Effekt zugunsten strategieorientierter Fördermaßnahmen von *mittlerer* Effektstärke ( $d=.66$  [.54, .77]).

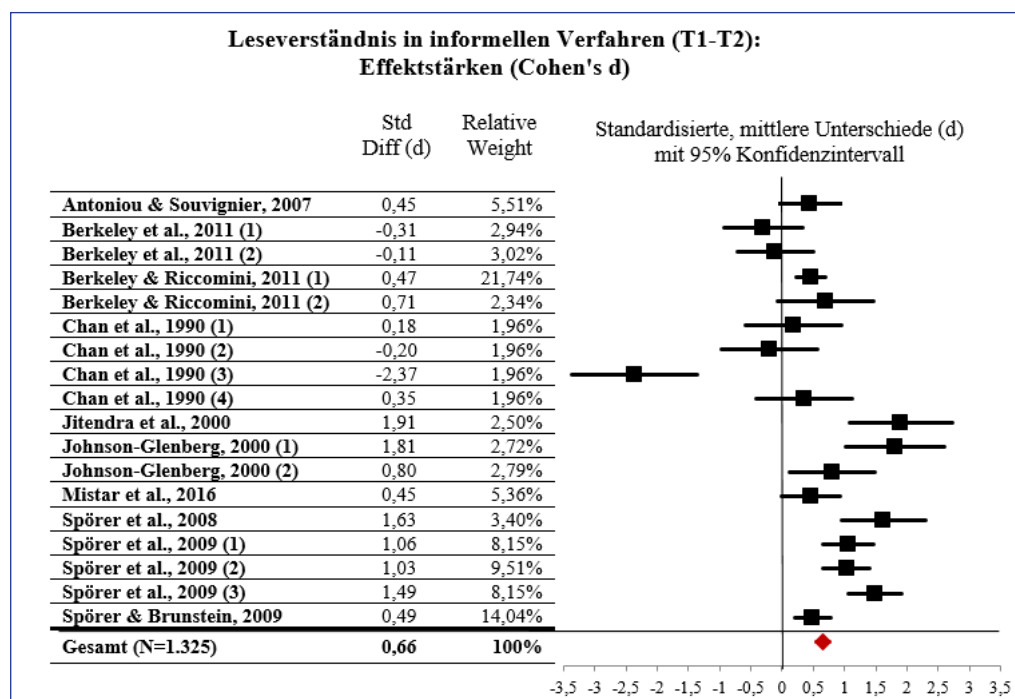


Abb. 3: Verbesserungen im Leseverständnis unmittelbar nach Interventionsende (T1-T2), gemessen mit informellen Verfahren – forest plot zur Darstellung der Effektstärken unter Angabe des Konfidenzintervalls sowie der gemittelten gewichteten Gesamteffektstärke.

Auffallend sind alle (Teil-)Studien von Chan und Kollegen (1990): Hier konnte entweder *kein* oder ein *negativer* Effekt eines Strategietrainings nachgewiesen werden (Anteil am Gesamteffekt: 7,84%). Zwar liegt nur bei der Teilstudie 3 von Chan und Kollegen (1990) mit  $n=26$  Probanden der Effekt mit einer 95%-Wahrscheinlichkeit tatsächlich im negativen Bereich,  $d=-2.37$  [-3.37,

-1.37], jedoch ist unter Betrachtung der Konfidenzintervalle auch der kleine, positive Effekt der Teilstudie 4,  $d=.35$  [-.43, 1.24] nur als statistische Tendenz zu werten. Bei den Studien von Chan und Kollegen (1990) wurde allein die Strategie des mental imagery in zwei Fördereinheiten vermittelt. Die Zielgruppe bestand aus durchschnittlichen Lesern (Teilstudie 1 und 2) und allgemein leseschwachen Primarstufenschülern.

Ähnlich verhält es sich mit den Studien von Antoniou und Souvignier (2007), Berkley und Kollegen (2011), Berkeley und Riccomini (2011: Teilstudie 2) und Mistar und Kollegen (2016), bei denen das Konfidenzintervall den Wert Null einschließt und damit die Irrtumswahrscheinlichkeit, dass ein positiver Effekt vorliegt, höher als 5 % liegt. Bei diesen Studien handelt es sich überwiegend um Kurzzeitinterventionen mit allgemein lernschwachen Schülern aus der Sekundarstufe. Hinsichtlich der Strategievermittlung ist insbesondere die Arbeit von Berkeley und Riccomini (2011) zu erwähnen, in der ausschließlich die Strategie des Comprehension Monitorings vermittelt wurde.

Die übrigen neun Studien zeigen bei der Messung des Leseverständnisses über informelle Verfahren im positiven Bereich liegende Konfidenzintervalle und mittlere bis große Effektstärken von  $.49 \leq d \leq 1.81$ . Im Vergleich zu den Studien, in denen normierte Verfahren zum Einsatz kamen, können hier demnach stärkere, positive Effekte nachgewiesen werden. Dies bestätigt sich bei den vier Studien, die das Leseverständnis sowohl über normierte als auch über informelle Verfahren erhoben haben (Johnson-Glenberg, 2000: Teilstudie 1 und 2; Spörer et al., 2008; Spörer & Brunstein, 2009). Bei Spörer und Kollegen (2008) lag der Effekt unter Verwendung von normierten Verfahren mit  $d=.62$  [.02, 1.22] in einem mittleren Bereich während unter Anwendung informeller Verfahren ein Vorteil strategieorientierter Maßnahmen von *großer* Effektstärke belegt werden konnte ( $d=1.63$  [.95, 2.30]). Ein noch größerer Vorteil der Strategievermittlung in Form einer Effektstärke von  $d=1.81$  [1.02, 2.60] wurde in der Teilstudie 2 von Johnson-Glenberg (2000) mit Hilfe informeller Verfahren nachgewiesen, während die mit einer normierten Überprüfung ermittelten Effektstärke tendenziell im negativen Bereich lag (Teilstudie 2:  $d=-0.53$  [-1.21, .15], vgl. 5.2.1).

**Langfristige Effekte (T1-T3):** In acht der zuvor aufgeführten (Teil-)Studien wurde auch die langfristige Leistungsentwicklung über informellen Verfahren evaluiert. Die Berechnung des Gesamteffekts basiert demnach auf den Daten von insgesamt  $N=572$  Probanden (vgl. Abb. 4).

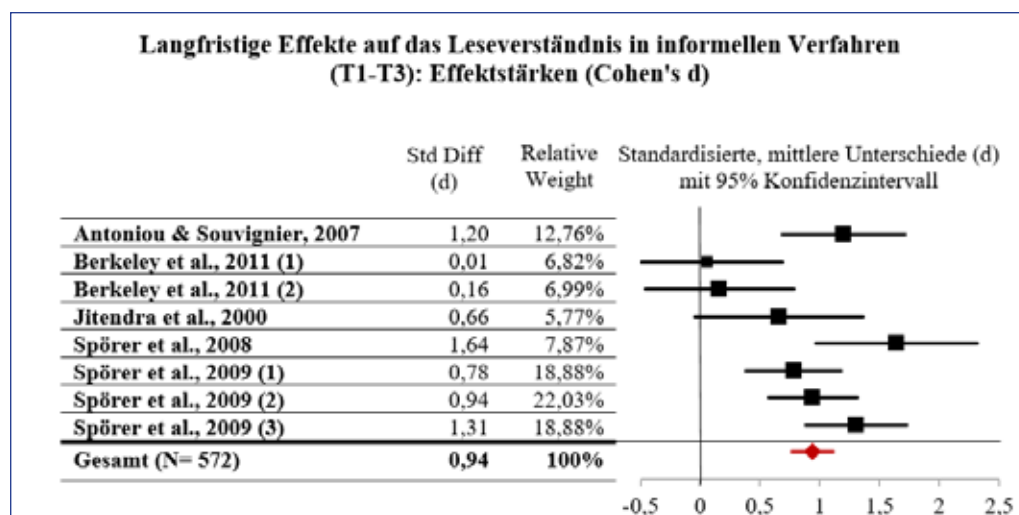


Abb. 4: Langfristige Verbesserungen im Leseverständnis (T1-T3), gemessen mit informellen Verfahren – forest plot zur Darstellung der Effektstärken unter Angabe des Konfidenzintervalls sowie der gemittelten gewichteten Gesamteffektstärke.

Anders als bei den normierten Verfahren belegt die Meta-Effektgröße hier einen deutlichen langfristigen Vorteil eines Strategietrainings in Form eines großen Effekts,  $d=.94$  [.76, 1.11].

In drei Studien (Berkley et al., 2011, Teilstudien 1 und 2, Jitendra et al., 2000) konnte kein langfristiger Vorteil eines Strategietrainings gefunden werden (vgl. Abb. 4,  $.01 \leq d \leq .66$ ; Anteil am Gesamteffekt: 19,58 %). Die Ergebnisse von Jitendra und Kollegen (2000) weisen zwar mit  $d=.66$

auf einen mittelgroßen, positiven Effekt hin. Bei gleichzeitiger Berücksichtigung des Konfidenzintervalls  $[-.05, 1.36]$  wird jedoch deutlich, dass die Wahrscheinlichkeit eines positiven Effekts über der tolerierten 5 %-Grenze liegt, sodass hier nur von einer positiven Tendenz zugunsten der Strategievermittlung auszugehen ist.

Vergleicht man die berechneten langfristigen Ergebnisse zu T3 mit den Effekten, die unmittelbar nach Beendigung der Intervention erfasst wurden, so zeigt sich in drei Vergleichen (Antoniou & Souvignier, 2007; Berkeley et al., 2011, Teilstudie 1 und 2) in der follow-up Untersuchung ein deutlicherer Effekt als zu T2, wobei lediglich bei Antoniou und Souvignier (2007) die Verbesserung von statistischer Relevanz scheint ( $d=.45$  vs.  $d=1.20$ ). Die von Spörer und Kollegen (2008) für die follow-up Untersuchung ermittelten Effekte sind identisch für die Werte nach der Intervention, sodass von einer hohen Stabilität des Gelernten ausgegangen werden kann. In den anderen Fällen weisen die kleineren Effektgrößen zu T3 eine leichte Annäherung der Kontrollgruppen an, wobei die mittlere Differenz immer noch einem mittleren bis großen Effekt entspricht ( $d=.31$ -.78).

### 5.2.3 Effekte auf das Wissen um und die Anwendung von Verstehensstrategien während des Leseprozesses

Als letzte relevante Outcome-Variable wurden die Effekte strategieorientierter Fördermaßnahmen in Bezug auf das Wissen um und die Anwendung von Verstehensstrategien während des Leseprozesses ermittelt.

**Unmittelbare Effekte (T1-T2):** Die Berechnung der mittleren, gewichteten Effektstärke für kurzfristig feststellbare Effekte (T1-T2) erfolgte auf der Grundlage von 24 Datensätzen mit insgesamt  $N=3.809$  Probanden (vgl. Abb. 5).

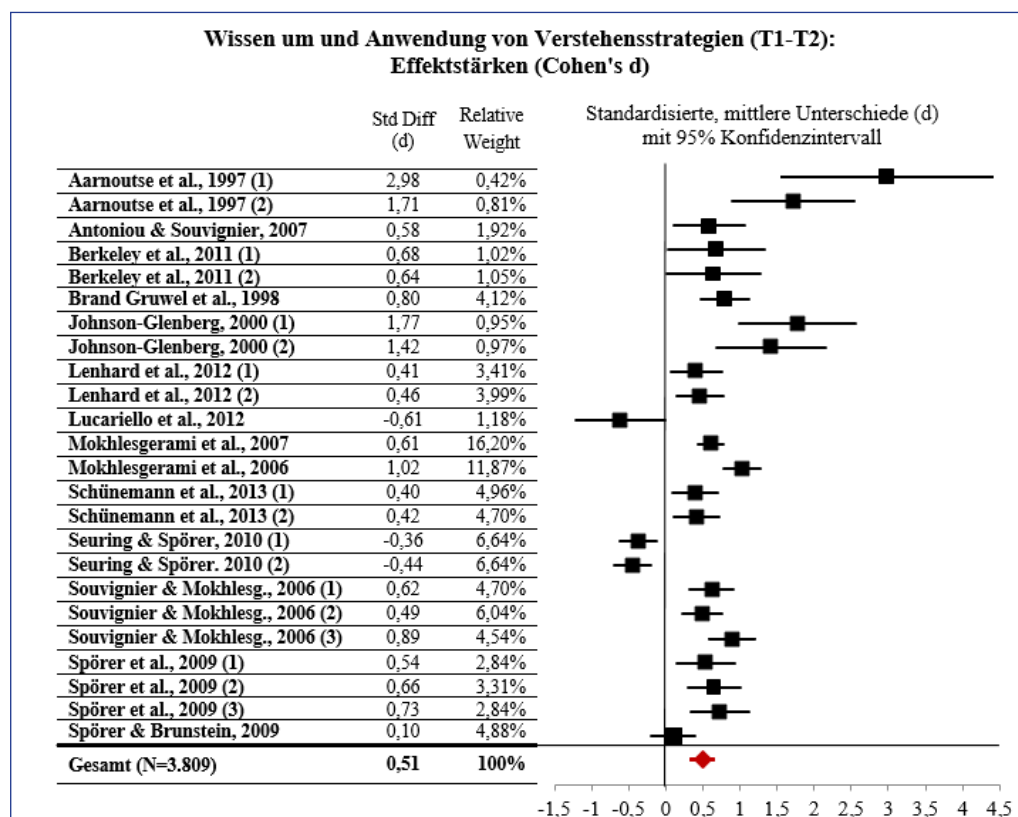


Abb. 5: Wissen um und Anwendung von Verstehensstrategien während des Leseprozesses unmittelbar nach Interventionsende (T1-T2), gemessen mit experimentellen Verfahren – forest plot zur Darstellung der Effektstärken unter Angabe des Konfidenzintervalls sowie der gemittelten gewichteten Gesamteffektstärke.

Den Erwartungen entsprechend zeigen sich überwiegend positive Effekte in Bezug auf diese relativ nahe Transferleistung der in der strategieorientierten Förderung gelernten Inhalte: Die Probanden der Experimentalgruppen sind im Wissen um und in der Anwendung von Verste-

hensstrategien den Probanden der Kontrollgruppen überlegen. Etwas deutlicher als in Bezug auf die zwei zuvor berichteten Outcome-Variablen zeigen sich hier positive Effekte kleiner bis großer Effektstärke von  $.41 \leq d \leq 2.98$ . Die Meta-Effektgröße über alle Studien hinweg liegt bei  $d=.51$  [.44, .57], was einem Gesamteffekt mittlerer Größe entspricht.

Den „Ausreißer nach oben“ stellt die Teilstudie 1 von Aarnoutse und Kollegen (1997) mit  $n=16$  Probanden dar, deren Stichprobe aus Kindern mit Defiziten in der Worterkennung bei gleichzeitig intaktem Sprachverständnis darstellten ( $d=2.98$  [1.56, 4.40]). Aber auch die andere Teilgruppe dieser Studie mit Schwierigkeiten in der Worterkennung und beeinträchtigtem Sprachverständnis war der Kontrollgruppe mit einer Effektstärke von  $d=1.71$  [.89, 2.54] deutlich überlegen. Die Autoren führten die Strategien sukzessive ein und ergänzten die Förderung durch Übungsstunden, in denen die bis zu diesem Zeitpunkt vermittelten Strategien gemeinsam angewendet wurden.

Erwartungswidrige Ergebnisse liefern die Studien von Lucariello und Kollegen (2012) ( $d=-.61$  [-1.21, -.01]) und Seuring und Spörer (2010) ( $d=-.36$  [-.61, -.12], bzw.  $d=-.44$  [-.69, -.20]), in denen die Kontrollgruppe der Experimentalgruppe in Bereich dieser Outcome Variable tendenziell überlegen war und damit *negative* Effekte geringer Größe gefunden wurden. Diese Ergebnisse spiegeln die in beiden Studien nicht nachweisbaren Effekte auf das Leseverständnis (erhoben über normierte Verfahren) wider. Ebenso konnte in der Studie von Spörer und Brunstein (2009) kein Effekt ( $d=.10$  [-.19, .39]) nachgewiesen werden. Der Anteil der drei genannten Studien am Gesamteffekt liegt bei 19,34%. In diesen Studien bestand die Stichprobe aus überwiegend durchschnittlichen Lesern der dritten, fünften und siebten Jahrgangsstufe. In den übrigen 20 (Teil-) Studien liegen sowohl die Effektstärken als auch die Konfidenzintervalle im positiven Bereich.

**Langfristige Effekte (T1-T3):** In Bezug auf die langfristige Entwicklung zeigt sich ein ähnlich positives Gesamtbild zugunsten der Vermittlung von Verstehensstrategien. Auf der Basis von 15 Datensätzen ( $n=2.559$ ) kann insgesamt weiterhin eine Überlegenheit in Bezug auf Strategiewissen und -anwendung belegt werden (vgl. Abb. 6). Die Überlegenheit drückt sich zum einen durch einen positiven, mittleren gewichteten Gesamteffekt von  $d=.57$  [-.48, .67] aus und wird zum anderen durch insgesamt 12 Datensätze mit ausschließlich im positiven Bereich liegenden Effektstärken und Konfidenzintervallen untermauert (Anteil an Gesamteffekt: 84,37%).

In drei (Teil-)Studien konnte mit  $d<.20$  *kein* statistisch bedeutsamer Langzeiteffekt nachgewiesen werden, wobei das jeweilige Konfidenzintervall zusätzlich den Wert Null einschließt (Berkeley et al., 2011: Teilstudie 1; Seuring & Spörer, 2010: Teilstudie 2; Spörer et al., 2009: Teilstudie 1; Anteil am Gesamteffekt: 15,63%), sodass die Überlegenheit der Trainingsgruppen nicht sicher gegen den Zufall abgesichert werden kann. Während bei Berkeley und Kollegen (2011) und Seuring und Spörer (2010) bereits in Bezug auf die anderen Outcome-Variablen keine Vorteile gegenüber der Kontrollgruppe belegt werden konnten, sind bei Spörer und Kollegen (2009) trotz der fehlenden Effekte auf die Strategieanwendung in der Teilstudie 3 durchaus nennenswerte langfristige Verbesserungen im Leseverständnis von mittlerer Effektstärke zu verzeichnen (vgl. Abb. 4, 5.2.2). Im Zusammenhang mit der Arbeit von Spörer und Kollegen (2009) ist von besonderem Interesse, dass sich die drei Teilstudien v.a. in der Methodik der Strategievermittlung unterschieden. Während in den beiden Teilstudien 2 und 3, in denen durchaus positive Effekte in Bezug auf die Kenntnis und Anwendung der vermittelten Strategien nachweisbar sind ( $d=.42$  bzw.  $d=.72$ ), die Strategien mittels der Methode des reziproken Lehrens und Lernens vermittelt wurden, wurden kam in der Teilstudie 1 ausschließlich die direkte Instruktion der Lehrkraft zum Einsatz.

Vergleicht man bei diesen drei Studien nun die langfristige Entwicklung mit den Effekten unmittelbar nach der Intervention zeigt sich langfristig eher eine abnehmende Differenz zwischen den Leistungen der Kontroll- und der Treatmentgruppe in Form von kleineren Effektgrößen als zu T2. Bei insgesamt sieben Datensätzen aus drei Studien ist jedoch Gegenteiliges, also größere Mittelwertsdifferenzen zwischen EG und KG zu T3 als zu T2, zu beobachten (Schünemann et al., 2013; Seuring & Spörer, 2010; Souvignier & Mokhesgerami, 2006; Anteil am Gesamteffekt: 56,9%). Zusammenfassend ist also von einer hohen Stabilität für diese Outcome Variable auszugehen.

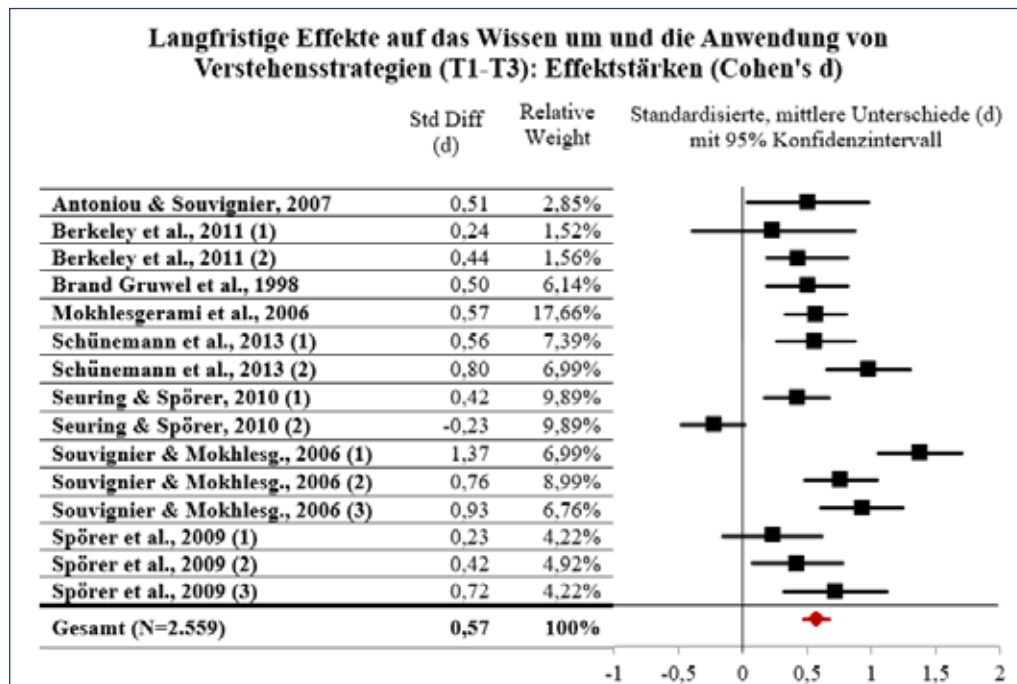


Abb. 6: Langfristige Effekte auf das Wissen um und die Anwendung von Verstehensstrategien während des Leseprozesses (T1-T3), gemessen mit experimentellen Verfahren – forest plot zur Darstellung der Effektstärken unter Angabe des Konfidenzintervalls sowie des gemittelten, gewichteten Gesamteffekts.

#### 5.2.4 Subgruppenanalysen

Um besonders wertvolle Merkmale für die Entwicklung von Förderprogrammen zu ermitteln, sollte in einem letzten Schritt die Frage beantwortet werden, ob sich Studien mit unterschiedlichen Stichproben, Interventions- und Implementationsmerkmalen in ihren Auswirkungen hinsichtlich des Leseverständnisses differenzieren lassen. Einen Überblick über die standardisierten Mittelwertdifferenzen in der Leistungsentwicklung unmittelbar nach Interventionsende (T1 bis T2), separiert nach unterschiedlichen Stichprobenmerkmalen, Merkmalen der Intervention und deren Implementation gibt Tabelle 3. Zu berücksichtigen ist dabei, dass es sich hierbei lediglich um eine Analyse mit deskriptivem Charakter handelt und die Vergleiche zwischen den dokumentierten Effektstärken lediglich Tendenzen wiedergeben. Darüber hinaus unterscheiden sich die Studien in der Regel in mehr als einem Kriterium voneinander, die bei der vorliegenden Zusammenfassung der Studien hinsichtlich eines einzelnen Parameters außer Acht gelassen wurden. Insgesamt sind die Ergebnisse auch vor dem Hintergrund der jeweils geringen Anzahl berücksichtigter Studien mit Vorsicht zu interpretieren (vgl. 6.2).



Tab. 3: Übersicht über die mittleren, gewichteten Effekte unmittelbar nach Interventionsende (T1-T2) in Bezug auf die drei relevanten Outcome-Variablen (Leseverständnis, gemessen mit normierten und informellen Verfahren; Wissen um/Anwendung von Verstehensstrategien) – separiert nach interessierenden Studiencharakteristika und unter Angabe zentraler statistischer Kennwerte (siehe Legende zu den Abkürzungen unterhalb der Tabelle).

	Normierte Verfahren			Informelle Verfahren					
Studiencharakteristika	$n_i / n \ (n_i + n_c)$	$d_w$	CI	$n_f / n \ (n_i + n_c)$	$d_w$	CI	$n_i / n \ (n_i + n_c)$	$d_w$	CI
Stichprobenmerkmale									
Alter/Klassenstufe									
Primarstufe	10 / 1364 (1032/332)	.39	[.26, .51]	9 / 519 (241/278)	.86	[.68, 1.04]	9 / 664 (319/345)	.81	[.65, .96]
Sekundarstufe	14 / 3074 (1854/1220)	.13	[.06, .21]	9 / 806 (431/375)	.54	[.40, .68]	15 / 3145 (1899/1246)	.44	[.37, .52]
Lern- und Lesekompetenzen									
Durchschnittliche Leser	1 / 115 (64/51)	.50	[.13, .87]	3 / 340 (182/158)	.39	[.18, .61]	---	---	---
Leseschwache Probanden mit Beeinträchtigungen in Worterkennung und LV	3 / 439 (221/218)	.34	[.15, .53]	2 / 52 (26/26)	-1.01	[-1.59, -.43]	3 / 439 (221/218)	.57	[.38, .76]
Probanden mit Beeinträchtigungen im LV bei durchschnittlicher Lesefertigkeit	6 / 242 (130/112)	.19	[-.06, .44]	3 / 106 (63/43)	1.49	[1.05, 1.92]	3 / 104 (61/43)	1.63	[1.18, 2.08]
Probanden mit allg. Lernbeeinträchtigungen, die sich auch im LV niederschlagen	1 / 16 (8/8)	.43	[-.56, 1.42]	4 / 183 (105/78)	.21	[-.08, .50]	4 / 168 (92/76)	.85	[.53, 1.16]
Unselektierte Stichprobe	13 / 3626 (2463/1163)	.19	[.12, .26]	6 / 644 (296/348)	.94	[.75, 1.12]	14 / 3098 (1844/1254)	.44	[.37, .52]
Merkmale der Intervention des Treatments									
Setting der Intervention									
Klassensetting	15 / 3908 (2601/1307)	.19	[.13, .26]	8 / 773 (413/360)	.48	[.34, .62]	16 / 3190 (1922/1268)	.43	[.36, .50]
Kleingruppensetting	9 / 530 (285/245)	.35	[.18, .52]	10 / 552 (259/293)	.92	[.75, 1.10]	8 / 619 (296/323)	.90	[.74, 1.08]
Vermittelte Strategien									
Vermittlung einer Strategie	4 / 356 (178/178)	.33	[.12, .54]	5 / 137 (70/67)	.07	[-.26, .41]	2 / 175 (81/94)	.15	[-.15, .45]
Kombination von zwei Strategien	4 / 1084 (882/202)	.38	[.23, .54]	3 / 259 (140/119)	.72	[.47, .97]	3 / 259 (140/119)	.52	[.27, .77]
Kombination von mehr als zwei Strategien	16 / 2998 (1826/1172)	.14	[.06, .21]	10 / 929 (462/467)	.74	[.60, .87]	19 / 3375 (1997/1378)	.53	[.46, .60]
Methode									
Direkte Instruktion durch die Lehrkraft	9 / 2038 (1148/890)	.08	[-.01, .16]	7 / 316 (148/168)	.50	[.27, .72]	8 / 1965 (1093/872)	.34	[.25, .43]
Computergestützte Strategievermittlung	2 / 955 (800/155)	.44	[.27, .61]	---	---	---	1 / 130 (58/72)	.41	[.06, .76]
Lernen am Modell	2 / 497 (390/107)	.07	[-.14, .29]	---	---	---	2 / 497 (390/107)	.88	[.66, 1.10]
Reziprokes Lehren und Lernen	8 / 744 (441/303)	.31	[.17, .46]	6 / 538 (263/275)	1.02	[.84, 1.20]	7 / 861 (486/375)	.52	[.38, .66]
Kombination aus direkter Instruktion und reziprokem Lehren und Lernen	3 / 204 (107/97)	.46	[.18, .73]	3 / 152 (84/68)	.11	[-.21, .43]	6 / 356 (191/165)	.90	[.68, 1.12]
Merkmale der Implementation des Treatments									
Durchführender									
Forscherteam	15 / 1494 (827/667)	.24	[.13, .34]	13 / 916 (457/459)	.81	[.67, .94]	13 / 1538 (817/721)	.31	[.21, .41]
Lehrkraft	9 / 2944 (2059/885)	.20	[.12, .28]	5 / 409 (215/194)	.34	[.15, .54]	11 / 2271 (1401/870)	.64	[.55, .73]
Förderzeitraum									
Kurzzeitintervention (bis max. 12 Zeitstd.)	10/1333 (737/596)	.24	[.12, .35]	13 / 1053 (525/538)	.61	[.49, .74]	13 / 1573 (823/750)	.20	[.10, .30]
Intervention zwischen 12–24 Zeitstd.	14 / 3105 (2149/956)	.25	[.17, .33]	5 / 262 (147/115)	.89	[.63, 1.15]	11 / 2236 (1395/841)	.73	[.64, .82]

Legende: LV= Leseverständnis;  $n_f$ = Anzahl der berücksichtigten Fälle/(Teil-)Studien;  $n$  ( $n_T + n_C$ )= Stichprobengröße gesamt;  $n_T$  = Stichprobengröße Trainingsgruppe,  $n_C$  = Stichprobengröße Kontrollgruppe  $d_w$ = gewichtete, mittlere Effektstärke; CI= 95%-Konfidenzintervall

**Stichprobenmerkmale:** In Bezug auf das Alter der Stichproben macht Tabelle 3 deutlich, dass in den für diese Analyse aufgearbeiteten Studien die Interventionsmaßnahmen, die in der Primarstufe durchgeführt wurden deutlich effektiver sind als diejenigen, die in der Sekundarstufe umgesetzt wurden. Dies gilt für alle drei erhobenen Outcome-Variablen. Diese Aussage kann dahingehend spezifiziert werden, dass es offensichtlich v. a. Schüler der dritten und vierten Klasse sind, die von einer entsprechenden Förderung profitieren, da in 14 von insgesamt 17 Studien, die in der Primarstufe durchgeführt wurden, ältere Grundschulkinder der dritten und vierten Jahrgangsstufe die Zielgruppe waren.

In Bezug auf die unterschiedlichen Lernvoraussetzungen basieren die berechneten, mittleren Effektstärken aufgrund der insgesamt fünf Subgruppen auf relativ wenigen Datensätzen und sind nur bedingt aussagekräftig. Verbesserungen des Leseverständnisses in normierten Verfahren finden sich bei allen Probandengruppen, auch wenn die Effektstärken lediglich klein ausfallen. Konkretere Schlussfolgerungen können gezogen werden, wenn die beiden anderen Outcome-Variablen betrachtet werden. Es wird deutlich, dass sich Probanden mit spezifischen Schwierigkeiten im Bereich des Leseverständnisses bei normal ausgeprägter Worterkennung im Vergleich zur Kontrollgruppe am positivsten entwickeln (Leseverständnis über informelle Verfahren:  $d=1.49$  [ $1.05, 1.92$ ] bzw. Strategiewissen/-anwendung:  $d=1.63$  [ $1.18, 2.08$ ]). Hier scheint es sich um die Zielgruppe zu handeln, die die vorrangige Zielgruppe entsprechender Unterstützungsmaßnahmen darstellen sollte. Im Gegensatz dazu profitieren Kinder mit Beeinträchtigungen in der Worterkennung und dem Leseverständnis von einem strategieorientierten Ansatz in deutlich geringerem Ausmaß. Der auf zwei Studien basierende Effekt auf das über informelle Verfahren gemessene Leseverständnis liegt deutlich im negativen Bereich,  $d=-1.01$  [ $-1.59, -.43$ ].

Wie sich schon in der bisherigen Ergebnisdarstellung andeutete (vgl. 5.2.2), findet sich kein eindeutiger Effekt von Strategiekonzepten im Sinne einer Verbesserung des Leseverständnisses für Probanden, die im Zuge kognitiver Beeinträchtigungen Schwierigkeiten im Bereich des Leseverständnisses entwickeln (Antoniou & Souvignier, 2007; Berkley et al., 2011; Berkeley & Riccomini, 2011: Teilstudie 2; Mistar et al., 2016). Eine deutliche Überlegenheit von großer Effektstärke in der dritten Outcome-Variable (Wissen um und in der Anwendung von Verstehensstrategien während des Leseprozesses) konnte aber wie für alle anderen Subgruppen auch für diese Probandengruppe nachgewiesen werden ( $d=.85$  [ $.53, 1.16$ ]).

**Merkmale der Intervention:** Werden die Studien nach dem Setting der Intervention zusammengefasst und miteinander verglichen, so zeigen sich in Bezug auf alle drei Outcome-Variablen stärkere Effekte zugunsten einer Durchführung im Kleingruppensetting im Vergleich zum Klassenunterricht ( $.35 \leq d \leq .90$  vs.  $.19 \leq d \leq .48$ ).

In Studien, in denen nur eine Strategie vermittelt wurde, konnte ausschließlich in normierten Verfahren eine Verbesserung des Leseverständnisses in Form eines kleinen Effektes ( $d=.33$  [ $.12, .54$ ]) nachgewiesen werden. Eine Kombination von Strategien erbrachte in Bezug auf das informell erfasste Leseverständnis und die Anwendung von Strategien Effekte mittlerer Größe ( $.52 \leq d \leq .74$ ) im Vergleich zur Kontrollintervention. Diese fallen deutlich höher aus als in den Studien, in denen lediglich eine Strategie vermittelt wurde. Tendenziell zeigt sich das auch für Verbesserungen in normierten Verfahren.

Was die unterschiedlichen Methoden angeht, lassen sich lediglich hypothetische Aussagen formulieren, da die Anzahl der Studien für die gebildeten Subgruppen zu gering ist. Am auffälligsten ist, dass in Studien, in denen die Strategien alleine über die direkte Instruktion durch die Lehrkraft vermittelt wurden ( $n=2038$  Probanden), kein Effekt auf die Leseverständnisleistung in normierten Verfahren nachgewiesen werden konnte ( $d=.08$  [ $-.01, .16$ ]). Auch die Effektstärken für die anderen Outcome Variablen fallen tendenziell niedriger aus, als bei den anderen Vermittlungsmethoden. Ein ähnliches Bild zeigt sich nur noch bei Studien, die bei der Vermittlung das Modelllernen fokussierten ( $d=.07$  [ $-.14, .29$ ]). Besonders wirksam scheinen Interventionen zu sein, die Verstehensstrategien mit Hilfe des reziproken Lehrens und Lernens oder softwaregestützt vermitteln.

**Implementation des Treatments:** Die letzte Subgruppenbildung bezieht sich auf Merkmale der Implementation des Treatments (*Person, die die Intervention durchführte*). Kein eindeutiger Trend lässt sich aus diesem Vergleich hinsichtlich der Personengruppen, die die Interventionsmaßnahmen umsetzten (*Forscherteam vs. Lehrkraft*) sowie der *Intensität der Förderung* ableiten. Während die Ergebnisse in Bezug auf das Leseverständnis in normierten Verfahren noch vergleich-

bar ausfallen, zeigt sich in Bezug auf die Leseverständnisentwicklung in informellen Verfahren ein deutlich größerer Effekt, wenn Mitglieder des Forscherteams die Intervention durchführten ( $d=.81$  [.67, .94] vs.  $d=.34$  [.15, .54]). Gegenteiliges lässt sich für das Wissen um und die Anwendung der gelernten Strategien feststellen (Durchführung durch Forscherteam:  $d=.31$  [.21, .41] vs. Durchführung durch Lehrkraft:  $d=.64$  [.55, .73], vgl. Tab. 3). In Bezug auf den letzten Parameter (Förderzeitraum) lässt sich tendenziell ein Vorteil einer längeren Begleitung in der Vermittlung und Übung der Strategien ableiten, der insbesondere für die Überprüfung der Kenntnis und der Anwendung der vermittelten Strategien offensichtlich wird.

## 6 Diskussion

### 6.1 Zusammenfassung und inhaltliche Reflexion der Ergebnisse

Die Ergebnisse der vorgelegten Metaanalyse machen deutlich, dass die Vermittlung von Verstehensstrategien eine Maßnahme darstellt, die sich zusammenfassend betrachtet positiv auf das Leseverständnis von Schülern auswirken kann. Mit einer mittleren gewichteten Effektstärke von  $d=.21$  [.15, .27] bei der Überprüfung des Leseverständnisses mittels normierter Verfahren handelt es sich zwar nur um einen kleinen Effekt, dessen Bedeutung aber aufgrund der Komplexität des Textverständnisses nicht unterschätzt werden sollte. Neben der aktiv-strategischen Auseinandersetzung mit einem Text wird dieses zumindest auch von der Worterkennung sowie lexikalischen und syntaktisch-morphologischen Fähigkeiten in der rezeptiven Modalität beeinflusst. Die Ergebnisse sollten aber auch dahingehend interpretiert werden, dass sich die Förderung des Textverständnisses nicht auf die Vermittlung von Verstehensstrategien reduzieren darf. Das aktiv-strategische Lesen stellt innerhalb einer mehrdimensionalen Förderung des Textverstehens nur ein, wenn auch ein sehr wichtiges Element, dar, welches durch Maßnahmen, die auf die Worterkennung abzielen und insbesondere bei spracherwerbsgestörten Kindern durch eine Förderung grammatischer und lexikalischer Fähigkeiten in der rezeptiven Modalität ergänzt werden muss.

Aus der zuletzt durchgeführten Subgruppenanalyse kann mit einer gewissen Vorsicht abgeleitet werden, dass die vorrangige Zielgruppe für die Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien ältere Schüler der Primarstufe mit durchschnittlicher Worterkennung und spezifischen Schwierigkeiten im Bereich des Leseverstehens darstellen. Die am besten geeignete Methode scheint das reziproke Lehren und Lernen in der Kleingruppe zu sein.

Dass im Gegensatz dazu durchschnittlich lesende Schüler aus der Sekundarstufe bzw. unselektierte Stichproben in deutlich geringerem Ausmaß profitieren, ist wenig überraschend, da bei unselektierten Stichproben, deren Schüler sich zu einem großen Teil durch ein unauffälliges Leseverständnis charakterisieren lassen, Lernfortschritte nur bedingt möglich sind. Evtl. ist dies auch ein Ausdruck dessen, dass normalbegabte Schüler von den seit der Veröffentlichung der ersten PISA Studie (2003) implementierten Unterstützungsmaßnahmen in ausreichendem Maße profitieren konnten, während diese für die schwächsten Schüler noch nicht ausreichend waren. Dafür sprechen auch die in IGLU 2016 ermittelten Ergebnisse, dass der prozentuale Anteil guter und sehr guter Leser in Deutschland den Vergleichswerten der EU und der OECD entsprechen, während aber 19 % der teilnehmenden Kinder aus Deutschland maximal die Kompetenzstufe 2 erreichten, ein Wert, der – wie einleitend bereits erwähnt – nur in vier Länder übertroffen wurde (Bos et al., 2017).

Auch spezifisch leseschwache Schüler, deren Schwierigkeiten primär im Bereich der Worterkennung offensichtlich werden, scheinen von einer Strategievermittlung deutlich weniger zu profitieren als Kinder mit spezifischen Schwierigkeiten im Bereich des Leseverständnisses bei durchschnittlicher Worterkennung („poor comprehender“), auch wenn dieses Bild nicht für alle erhobenen Outcome-Variablen, sondern v.a. für die Erfassung der Effektivität mittels informeller Verfahren Gültigkeit besitzt. Diese Ergebnisse legen nahe, dass für eine effektive Vermittlung und Anwendung von Strategien eine angemessene Worterkennung eine notwendige Voraussetzung darstellt. Erst wenn die Worterkennung weitgehend automatisiert ist, dürften Kinder in der Lage sein, ihre Ressourcen auf die aktiv-strategische Auseinandersetzung mit den Inhalten des Gelesenen zu lenken. Je weniger bewusste Aufmerksamkeit auf die Lesetechnik gelenkt werden muss, desto mehr kognitive Ressourcen stehen für die hierarchiehöheren Fähigkeiten der aktiv-strategischen Auseinandersetzung auf Textebene zur Verfügung. Muss ein Leser ein zu hohes Ausmaß der vorhandenen Ressourcen auf die Lesetechnik lenken, stehen dagegen insbesondere für die Sinnentnahme auf Textebene und die aktive Auseinandersetzung mit gelesenen Inhalten kaum

mehr Kapazitäten zur Verfügung. Mit anderen Worten: Bei beeinträchtigter Worterkennung erreichen die Schüler gar nicht erst die Stufe im Verstehensprozess, auf der ihnen das strategieorientierte Lesen eine Hilfe darstellen könnte.

Gegen diese Annahme könnten auf den ersten Blick die stärkeren Effekte sprechen, die für Studien mit Schülern aus der Primarstufe ermittelt werden konnten. Jedoch wurden auch diese Studien zum großen Teil in dritten und vierten Klassen durchgeführt, also zu einem Zeitpunkt, zu dem zumindest in unselektierten Stichproben bei einem Großteil der Schüler eine ausreichende Worterkennung angenommen werden kann. Bei Schülern mit massiven Schwierigkeiten in der Worterkennung sollten dagegen zunächst phonologisch orientierte Fördermaßnahmen umgesetzt werden, um die für das Textverständnis notwendige Lesefertigkeit und Lesegeschwindigkeit zu gewährleisten. Selbst Edmonds und Kollegen (2009, S. 293) kommen trotz der von ihnen ermittelten Ergebnisse, dass für Maßnahmen, die auf die Leseflüssigkeit abzielen, bei Schülern der Sekundarstufe bestenfalls marginale Auswirkungen auf das Leseverständnis nachgewiesen werden können, zu der Schlussfolgerung, „[that] the evidence from this synthesis would not suggest forgoing instruction in reading skills such as fluency or advanced decoding strategies with secondary struggling readers – particularly for students whose word reading skills are exceedingly low“.

Im Vergleich mit der von Souvignier und Antoniou (2007) vorgelegten Metaanalyse ( $d=.50$ ) fallen die für die vorliegende Arbeit berechneten Effekte für Strategieprogramme insgesamt deutlich niedriger aus. Zwei Gründe könnten für diese Unterschiede verantwortlich sein: So wurden in der Metaanalyse von Souvignier und Antoniou (2007) für die Berechnung der Effektstärken nach eigenen Aussagen mögliche Prätestunterschiede nicht durchgängig kontrolliert. Ein weiterer Grund könnte in den für die Analysen berücksichtigten Stichproben liegen: Während sich die Studien in der vorliegenden Metaanalyse zu einem großen Teil aus unselektierten Stichproben zusammensetzte ( $n=4.039$ ) und nur in vier Studien ( $n=199$ ) der Fokus auf Kindern mit allgemeinen Lernschwierigkeiten lag, wurden für die Analyse von Souvignier und Antoniou (2007) ausschließlich Studien mit lernschwachen Schülern berücksichtigt. Interessanterweise konnte in den Studien, die in der vorliegenden Analyse allgemein lernschwache Schüler berücksichtigten (Aarnoutse et al., 1997; Antoniou & Souvignier, 2007; Berkeley & Riccomini, 2011), annähernd dieselbe Effektstärke ( $d=.43-.47$ ) ermittelt werden wie bei Souvignier und Antoniou (2007).

Sowohl die über alle Studien hinweg ermittelten gewichteten Effektstärken als auch die Ergebnisse der Subgruppenanalyse machen deutlich, dass Auswirkungen auf das Leseverständnis, die über informelle Verfahren ermittelt wurden, deutlicher ausfallen als diejenigen, die über normierte Verfahren bestimmt wurden ( $d=.66$  [.54, .77] vs.  $d=.21$  [.15, .27]). Zu einem vergleichbaren Ergebnis kommen Souvignier und Antoniou (2007, S. 53), die in ihrer Metaanalyse deutlich machen konnten, dass „höhere Effekte dann zu erwarten sind, wenn keine standardisierten Testverfahren, sondern von den Untersuchungsleitern selbst konstruierte Inventare eingesetzt werden.“ Es ist davon auszugehen, dass informelle Verfahren eine größere Nähe zu den Inhalten der Intervention aufweisen als normierte Lesetests und somit eher das während der Intervention unmittelbar Gelernte überprüfen („teaching-to-the-test“), während normierte Verfahren den Generalisierungseffekt auf alltägliche Lesesituationen besser abbilden sollten. Zumindest dürften normierte Verfahren eine geringere inhaltliche Nähe zu den Inhalten von Fördermaßnahmen aufweisen, als die von Untersuchern entwickelten Tests (Souvignier & Antoniou, 2007). Das in der vorliegenden Analyse gefundene Ergebnis, dass die über normierte Verfahren ermittelten Effekte in Bezug auf das Textverständnis deutlich geringer ausfallen, als die über informelle Instrumente ermittelten, legen damit nahe, dass in Bezug auf den Transfer einer Interventionsmaßnahme auf das Leseverständnis im Allgemeinen durchaus noch Handlungsbedarf besteht.

Um die Anwendung der Strategien auf neue Lesekontexte übertragen zu können und damit einen maximalen Transfer der Trainingsinhalte erreichen zu können, sollte sich die Unterstützung insbesondere der Schüler mit spezifischen Schwierigkeiten im Bereich des Textverständnisses nicht auf die Interventionsmaßnahme selber reduzieren. Vielmehr sollten Hilfen für die aktive-strategische Auseinandersetzung mit Textinhalten ein Unterrichtsprinzip darstellen, das nicht nur im Deutschunterricht, sondern auch in Sachfächern umgesetzt wird.

Vergleichbar äußern sich Hartmann et al. (2013, S. 225f.): „Die Kinder müssen Verstehensstrategien gründlich erlernen, verinnerlichen und automatisieren, damit sie diese in verschiedenen Lesesituationen selbstständig und mühelos auf diverse Texte anwenden können.“ Sie betonen die Notwendigkeit, die Schüler zum selbstständigen, flexiblen und kombinierten Strategiegebrauch in authentischen Lesesituationen vorzubereiten, indem die Anwendung der Strategien fächer-

übergreifend und auch in außerschulischen Situationen motiviert wird. Durch die wiederholte Übung und Anwendung der Strategien über einen längeren Zeitraum sollte es möglich sein, dass „deklaratives Wissen über Strategien in prozedurales Wissen überführt wird, das schließlich mühelos und sicher angewendet werden kann“ (Hartmann, 2006, S. 61).

In engem Zusammenhang mit der Forderung nach einem gründlichen Erlernen der Strategien, das auf die Automatisierung der Anwendung abzielt, steht das Ergebnis der vorliegenden Metaanalyse, dass die Vermittlung mehrerer Strategien über einen längeren Förderzeitraum effektiver ist als die Vermittlung lediglich einer Strategie in einer Kurzintervention. Ein Grund dafür liegt auf der Hand und ist weitgehend trivial: In einer längeren Intervention sollte es besser möglich sein, mehrere Strategien und deren flexible Anwendung zu vermitteln und systematisch einzuüben. Dass die Vermittlung nur einer Strategie nicht gewinnbringend ist, kann exemplarisch an der Studie von Lucariello und Kollegen (2012) konkretisiert werden. In einer Kurzzeitintervention von 12 Stunden wurde ausschließlich die Strategie des „mental imagery“ vermittelt. Sowohl bei einer Überprüfung des Leseverständnisses als auch der Erfassung der Strategiekennntnis schnitten die Kinder der Interventionsgruppe nach der Intervention schlechter ab als die Kinder der Kontrollgruppe ( $d = -.31$  bzw.  $-.61$ ). Die nicht nachweisbaren Effekte in dieser Studie dürften aber auch mit dem Anspruch der Strategie des „mental imagery“ zusammenhängen. Trotz der unmittelbar einsichtigen Bedeutung der Umwandlung gedruckter Texte in mentale Bilder für das langfristige Behalten gelesener Inhalte, muss für diese Strategie berücksichtigt werden, dass der effektive Einsatz das richtige Verstehen der Textinhalte voraussetzt. Das Generieren korrekter mentaler Bilder ist nur möglich, wenn die Inhalte des Textes in lexikalischer und syntaktisch-morphologischer Hinsicht richtig verarbeitet wurden. Aus diesem Grund sollte die Strategie des „mental imagery“ in der Praxis mit einer Förderung lexikalischer und grammatischer Kompetenzen verknüpft bzw. insbesondere in Kombination mit der Strategie des „Comprehension Monitorings“ vermittelt werden, da auch die bedeutsame Strategie des Zusammenfassens das korrekte Verstehen der Textinhalte in sprachlicher Hinsicht voraussetzt.

## 6.2 Grenzen der Studie

Um die Relevanz und Qualität dieser Metaanalyse adäquat einschätzen zu können, sollen abschließend Grenzen insbesondere methodischer Art angesprochen werden. Diese beziehen sich auf die Ermittlung und Kodierung der Datensätze sowie die gewählten statistischen Verfahren der Datenauswertung.

**Publication-Bias:** Die Literaturrecherche zielte ausschließlich auf deutsch- oder englischsprachig publizierte Studien in etablierten Fachzeitschriften, bei denen ein Volltextzugriff möglich war. Dies ist kein grundsätzlich unübliches Vorgehen; jedoch kann sich daraus eine systematische Verzerrung der Ergebnisse der Metaanalyse ergeben. Da Studien mit nicht-signifikanten oder theoretisch erwartungswidrigen Ergebnissen häufiger nicht publiziert werden, ist eine Verzerrung in Richtung einer Überschätzung des interessierenden Effekts nicht auszuschließen (Döring & Bortz, 2016). Weiter sind Verzerrungen durch Language-Bias, Availability-Bias und Cost-Bias zu bedenken. Da die Beschaffung unpublizierter Studien jedoch auch oft einer gewissen Willkür unterliegen, sollte bei einer Replikation dieser Studie die Durchführung statistischer Verfahren zur Abschätzung der Verzerrung (funnel plot) in Erwägung gezogen werden (Döring & Bortz, 2016).

**Umgang mit abhängigen Untersuchungsergebnissen in der Kodierung und Auswertung der Daten:** Einige der in der vorliegenden Metaanalyse aufgearbeiteten Studien umfassten mehrere Experimentalgruppen, deren Fortschritte mit einer Kontrollgruppe verglichen wurden. Bei diesen Studien wurden, wie im Methodenteil erläutert, die Kontrollgruppen doppelt berücksichtigt, um Einzelvergleiche mit jeder Experimentalgruppe vornehmen zu können. Dieses Vorgehen ist nicht unproblematisch und muss kritisch betrachtet werden, da die so konstruierten Datensätze nur scheinbar unabhängig voneinander sind und die angegebene Gesamtstichprobengröße höher ist als die reale Versuchspersonenzahl. Zudem wird die Arbeit einzelner Arbeitsgruppen mit der jeweils eingesetzten Methodik durch dieses Vorgehen möglicherweise stärker gewichtet. Dies könnte zu einer Verzerrung der Ergebnisse geführt haben. Die Problematik abhängiger Untersuchungsergebnisse ist in der Praxis nicht unbekannt. Döring und Bortz (2016) nennen verschiedene Lösungsstrategien, wie beispielsweise den Einbezug nur des wichtigsten Teilergebnisses der Primärstudie (hier: der Vergleich der Kontrollgruppe mit nur einer der Experimentalgrup-

pen), die Aggregation der relevanten Teilergebnisse der Primärstudien-Stichprobe zu einem Gesamtwert (hier: die Zusammenfassung der Ergebnisse aller Experimentalgruppen zu einem Wert) oder auch die Durchführung multivariater Metaanalysen, welche die Interkorrelationen der Effektgrößen einbezieht. Die erstgenannten Optionen erschienen aus theoretisch inhaltlicher Perspektive für die vorliegende Studie weniger geeignet, da der mit diesem Vorgehen einhergehender Informationsverlust in Bezug auf vorhandene Unterschiede in der Auswahl der Strategien und/oder deren Vermittlungsmethode von den Autoren nicht in Kauf genommen werden wollte. Letztgenannte Option – die multivariaten Metaanalysen – sind von diesem Problem weniger betroffen, werden aber gleichzeitig auch als „sehr komplexe und voraussetzungsvolle statistische Analysetechniken“ (Döhring & Bortz, 2016, 913) beschrieben. Trotz dieser Herausforderungen sollte diese Variante für zukünftige Forschungsarbeiten (zunehmend) bedacht werden (zu den Herausforderungen statistischer Auswertungsverfahren in Abhängigkeit von komplexen Fragestellungen wie den vorliegenden: siehe nachfolgender Abschnitt).

**Wahl der statistischen Modelle in der Berechnung des Gesamteffektes und möglicher Moderatorvariablen:** In der vorliegenden Untersuchung wurde unter Anwendung des Fixed-Effect-Modells ein Gesamteffekt für die jeweils interessierende Outcome-Variable berechnet. Neben der Interpretation der Konfidenzintervalle hätte eine Signifikanzprüfung des Gesamteffekts zusätzlich Aufschluss darüber geben können, ob es sich um einen überzufälligen Populationseffekt handelt. Dieser Signifikanztest sowie die Anwendung des Fixed-Effect-Modells werden bei vorliegenden *homogenen* Primärstudien bzw. bei vorliegender Homogenität der untersuchungsspezifischen Effektgrößen für sinnvoll erachtet (Döhring & Bortz, 2016). Die Ähnlichkeit der Effekte ist neben dem Einsatz statistischer Homogenitätstests auch durch eine „sachgerechte Auswahl geeigneter Studien“ und die „Konstruktion einer einheitlichen Strukturierungs- und Auswertungsschablone für den gesamten Pool der Primärstudien“ zu prüfen (Döhring & Bortz, 2016, 903). In der vorliegenden Studie wurde diesem sogenannten Uniformitäts- oder Inkommensurabilitäts-Problem durch die Berücksichtigung der beiden letztgenannten Aspekte begegnet. Angestrebt wurde demnach eine Auswahl von Studien, die als einander inhaltlich und methodisch in Bezug auf die unabhängigen, abhängigen Variablen sowie die Kontrollvariablen ähnlich genug eingestuft wurden, um sie im Sinne der formulierten Forschungsfrage miteinander vergleichen zu können. Darüber hinaus wurde viel Zeit in die Kodierung der Studien und Konsensfindung in der Kodierung investiert. In Folgestudien oder Replikationen könnten zur Optimierung des methodischen Vorgehens jedoch mehr als zwei Personen in den Kodierungsprozess einbezogen werden, sodass auch die Interrater-Reliabilität bestimmt sowie der genannte Homogenitätstest durchgeführt (und ggf. ein anderes statistisches Modell ausgewählt) werden kann. In der vorliegenden Arbeit wurde das Fixed-Effect-Modell auch deshalb beibehalten, da der potenziell vorhandenen Heterogenität im Gesamtpool der ausgewählten Studien insofern Rechnung getragen wurde, als dass eine Gesamteffektschätzung (nach dem Fixed-Effect-Modell) für inhaltlich homogenere Teilstichproben durchgeführt wurde (vgl. Abb. 1-6, Effektschätzungen für die Studien, die die als relevant erachteten Outcome-Variablen und Messzeitpunkte beinhalten).

Zusätzlich wurde einer möglichen Heterogenität der Studien und deren Effektstärken zu begegnen versucht, indem mögliche moderierende Variablen durch Subgruppenanalysen ermittelt werden sollten. Einschränkend anzumerken ist, dass es sich dabei um rein deskriptive Analysen handelt, bei denen nur *ein* inhaltlich plausibler Einflussfaktor isoliert betrachtet und eine Schätzung des Effekts bei unterschiedlicher Ausprägung dieses Faktors vorgenommen wurde. Es wurde keine Moderatorvariablenanalyse im eigentlichen Sinne durchgeführt, die z. B. über Meta-Regressionen oder Varianzanalysen versucht, die Unterschiedlichkeit der Studien, also die Varianz der Effektgrößen aufzuklären. Zu bedenken gilt zudem, dass die gebildeten Subgruppen sich nicht nur hinsichtlich eines Faktors voneinander unterscheiden. Eine Analyse der Wechselwirkung mit anderen Faktoren erfolgte jedoch nicht. Demnach müssen die gewonnenen Erkenntnisse aus diesen Berechnungen wie bereits mehrfach erwähnt zurückhaltend interpretiert werden und mithilfe komplexerer statistischer Verfahren für hierarchisch strukturierte Daten (Multilevel Modeling, genannte Meta-Regressionen) überprüft werden. Dies setzt jedoch eine ausreichend große Probandenzahl in den berücksichtigten Ebenen voraus und scheint für eine Vielzahl der in diesem Themenbereich durchgeführten Studien nur bedingt anwendbar.

Die Vergleiche zwischen den dokumentierten Effektstärken der Subgruppen geben daher lediglich Tendenzen wieder und sollen einen Beitrag dazu leisten, mehr über die Versuchsbedingungen zu erfahren, unter denen ein Effekt größer oder kleiner ausfällt.



### 6.3 Ausblick

Auch wenn die Ergebnisse der vorliegenden Metaanalyse in Bezug zur optimalen *Vermittlungsmethode* aufgrund der geringen Anzahl an Studien in den einzelnen Subgruppen mit großer Vorsicht zu interpretieren sind, legen die Resultate tendenziell nahe, dass durch eine Kombination aus expliziter Strategievermittlung und selbstständigen Übungsphasen (z. B. auf der Basis des reziproken Lehrens und Lernens) die besten Ergebnisse hinsichtlich des Textverständnisses erzielt werden könnten. Während zu Beginn der Vermittlung einer Strategie die Lehrperson eine zentrale Rolle übernimmt, die die Bedeutung einer Strategie erklärt und deren Anwendung modellhaft demonstriert, übernehmen die Schüler sukzessive mehr Verantwortung, indem sie zunächst von der Lehrkraft bei der Anwendung angeleitet werden und sie anschließend selbstständig umsetzen. Dafür sprechen z. B. die Ergebnisse der drei Teilstudien von Spörer und Kollegen (2009). So erzielten die beiden Experimentalgruppen, in denen die Strategien mittels reziproken Lehrens und Lernens eingeübt wurden, in allen drei Outcome Variablen tendenziell bessere Leistungen als die Kinder der Gruppe, in denen die Strategien ausschließlich über das Modell der Lehrkraft vermittelt wurden.

Auf der Grundlage der Ergebnisse der vorliegenden Metaanalyse entstand am Lehrstuhl für Sprachheilpädagogik der LMU München ein Förderkonzept, für dessen Entwicklung die im Rahmen dieser Studie ermittelten zentralen Ergebnisse handlungsleitend waren. Bei den „Lesetricks von Professor Neugier“ (Mayer, 2018), handelt es sich um ein für den Klassenunterricht konzipiertes Förderkonzept für Kinder mit spezifischen Schwierigkeiten im Bereich des Leseverständnisses ab der vierten Klasse. In einem zeitlichen Umfang von 25 Unterrichtsstunden werden insgesamt sechs Verstehensstrategien (Aktivierung von Hintergrundwissen, Vorhersagen treffen, Comprehension Monitoring, mental imagery, Fragen an den Text stellen und Zusammenfassen) vermittelt. Bei jeder Strategie steht zu Beginn die explizite, direkte Instruktion durch die Lehrkraft im Vordergrund. Die Lehrkraft erläutert die Bedeutung der Strategie und modelliert deren Anwendung an einem Satz oder einem kurzen Text. Es folgen Übungen in Kleingruppen oder Partnerarbeit, während derer die Unterstützungsmaßnahmen sukzessive abgebaut und der Schwierigkeitsgrad gesteigert wird, sodass eine zunehmend selbstständige, flexible Strategieverwendung möglich werden sollte. Dabei stellt die Anwendung der Strategien beim selbstständigen Lesen von Texten das Ziel, aber nicht zwingend den Ausgangspunkt der Förderung dar. Die meisten Strategien werden zu Beginn an kleineren schriftsprachlichen Einheiten (Wörter, Sätze, Absätze) oder auch in Hörsituationen eingeübt. Zudem erhalten die Schüler zu Beginn der Strategievermittlung zahlreiche Hilfestellungen bei der ersten selbstständigen Anwendung der Strategie, die schrittweise ausgeblendet werden und das Ziel verfolgen, die Kinder beim Erreichen des Ziels ein Stück weit „an die Hand zu nehmen“ und zu begleiten. Die Effektivität des Förderprogramms wird im Schuljahr 2018/19 in vierten Klassen von Sonderpädagogischen Förderzentren in Bayern evaluiert.

### Literatur

Die mit \* gekennzeichneten Publikationen wurden für die Metanalyse berücksichtigt.

- \* Aarnoutse, C., Brand-Gruwel, S., & Oduber, R. (1997). Improving reading comprehension strategies through listening. *Educational Studies*, 23, 209-227.
- \* Antoniou, F., & Souvignier, E. (2007). Strategy instruction in reading comprehension: An intervention for students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 5, 41-57.
- Artelt, C., Schiefele, U., Schneider, W., & Stanat, P. (2002). Leseleistungen deutscher Schülerinnen und Schüler im internationalen Vergleich (PISA). Ergebnisse und Erklärungsansätze. *Zeitschrift für Erziehungswissenschaft*, 5, 6-27.
- \* Berkeley, S., Mastropieri, A. M., & Scruggs, E.T. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *Journal of Learning Disabilities*, 44, 18-32.
- \* Berkeley, S., & Riccomini, J.P. (2011). QRAC-the-Code: A comprehension monitoring strategy for middle school social studies textbooks. *Journal of Learning Disabilities*, 46, 154-165.
- Berninger, V.W., Vermeulen, K., Abbott, R.D., McCutchen, D., Cotton, S., Cude, J., Dorn, S., & Sharon, T. (2003). Comparison of three approaches to supplementary reading instruction for low-achieving second grade readers. *Language, Speech and Hearing Services in Schools*, 34, 101-116.
- Borenstein, M., Hedges, L., Higgins & J., Rothstein, H. (2009). *Introduction into Meta-Analysis*. Chichester: Wiley.

- Bos, W., Valtin, R., Hußmann, A., Wendt, H. & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kaspar, E. Lankes, N. McElvany, T. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 13-28). Münster, New-York: Waxmann.
- \* Brand-Gruwel, S., Aarnoutse, C.A.J., & Van den Bos, K.P. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and Instruction*, 8, 63-81.
- Bremerich-Vos, A., Wendt, H., & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kaspar, E. Lankes, N. McElvany, T. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79-142). Münster, New-York: Waxmann.
- CEBM, Oxford Centre for Evidence-based Medicine (2009). *Levels of evidence*. Abgerufen von <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/> [07.08.2018]
- \* Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in Year 2. *Journal of Research in Reading*, 22, 241-256.
- \* Chan, L.K.S., Cole, P.G., & Morris, J.N. (1990). Effects of instruction in the use of a visual-imagery strategy on the reading – comprehension competence of disabled and average readers. *Learning Disability Quarterly*, 13, 2-11.
- Chard, D.J., Vaughn, Sh., & Tyler, B.J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, 35, 386-406.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Döhring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. Auflage. Heidelberg: Springer.
- Dole, J., Valencia, S., Greer, E., & Wardrope, J. (1991). Effects of two types of prereading instruction on the comprehension of narrative and expository text. *Reading Research Quarterly*, 26, 142-159.
- Dummer-Smoch, L., & Hackethal, R. (2013). *Kieler Leseaufbau*. Kiel: Veris Verlag.
- Edmonds, M.S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Klingler Tackett, K., & Wick Schnakenberg, J. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, 79, 262-300.
- Galuschka, K., & Schulte-Körne, G. (2015). Evidenzbasierte Interventionsansätze und forschungsbasierte Programme zur Förderung der Leseleistung bei Kindern und Jugendlichen mit Lesestörung – Ein systematischer Review. *Zeitschrift für Erziehungswissenschaft*, 18, 473-487.
- Gold, A., Mokhesgerami, J., Rühl, K., Schreblowski, S., & Souvignier, E. (2004). *Wir werden Textdetektive*. Göttingen: Vandenhoeck & Ruprecht.
- Groth, K., Hasko, S., Bruder, J., Kunze, S., & Schulte-Körne, G. (2013). Interventionseffekte bei Lese-Rechtschreibstörung: Evaluation von zwei Förderkonzepten unter besonderer Betrachtung methodischer Aspekte. *Lernen und Lernstörungen*, 2, 161-175.
- Hartmann, E. (2006). In *Bildern denken – Texte besser verstehen. Lesekompetenz strategisch stärken*. München: Reinhardt Verlag.
- Hartmann, E., Hochstrasser, M., & Trauffer, U. (2013). Förderung des aktiv-strategischen Textverstehens am Beispiel des bildhaften Vorstellens. *Sprachförderung und Sprachtherapie in Schule und Praxis*, 2, 224-233.
- Hinitikaa, S., Landerl, K., Aro, M., & Lyythinen, H. (2008). Training reading fluency: is it important to practice reading aloud and is generalisation possible? *Annals of Dyslexia*, 58, 59-79.
- Huemer, S., Landerl, K., Aro, M., & Lyythinen, H. (2008). Training reading fluency among poor readers of German: many ways to the goal. *Annals of Dyslexia*, 58, 115-137.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E., McElvany, N., Stubbe, T., & Valtin, R. (Hrsg.). (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster, New-York: Waxmann.
- Ise, E., Engel, R.R., & Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung? Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger Förderansätze. *Kindheit und Entwicklung*, 21, 122-136.
- \* Jitendra, A.K., Hoppes, M.K., & Xin, Y.P. (2000). Enhancing main idea comprehension for students with learning problems: The role of a summarization strategy and self-monitoring instruction. *The Journal of Special Education*, 34, 127-139.
- \* Johnson-Glenberg, M.C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visual strategies. *Journal of Educational Psychology*, 92, 772-782.
- Klicpera, C., & Gasteiger-Klicpera, B. (1995). *Psychologie der Lese- und Schreibschwierigkeiten Entwicklung, Ursachen, Förderung*. Weinheim: Beltz Psychologie Verlags Union.
- Klicpera, C., Weiss, J., & Gasteiger-Klicpera, B. (2013). Erfolg einer schulischen Legasthenikerförderung für Kinder der 3. und 4. Schulstufe nach dem Programm von Reuter-Liehr. *Heilpädagogische Forschung*, 39, 87-95.
- \* Lenhard, W., Baier, H., Endlich, D., Lenhard, A., Schneider, W., & Hoffmann, J. (2012). Computerunterstützte Leseverständnisförderung: Die Effekte automatisch generierter Rückmeldungen. *Zeitschrift für pädagogische Psychologie*, 26, 135-148.

- Lenhard, W., Baier, H., Endlich, D., Schneider, W., & Hoffmann, J. (2013). Rethinking strategy instruction: direct reading strategy instruction versus computer-based guided practice. *Journal of Research in Reading*, 36, 223-24.
- Lenhard, W., & Lenhard, A. (2016). *Berechnung von Effektstärken*. <https://www.psychometrica.de/effektstaerke.html>. Dettelbach: Psychometrica (02.1.2018)
- \* Lucariello, J.M., Butler, A.G., & Tine, M.T. (2012). Meet the "Reading Rangers": Curriculum for teaching comprehension strategies to urban third graders. The University of Pennsylvania Graduate School of Education's Online Urban Education Journal <https://www.urbanedjournal.org/archive/volume-9-issue-2-fall-2012/meet-%e2%80%9creading-rangers%e2%80%9d-curriculum-teaching-comprehension-strategi>, Zugriff am 9.1.2018
- Lyon, G.R., Shaywitz, S.E., & Shaywitz, B.A. (2003). Defining Dyslexia, Comorbidity, Teachers' Knowledge of Language and Reading. A Definition of Dyslexia. *Annals of Dyslexia*, 53, 1-14.
- \* Lysynchuk, L.M., Pressley, M., & Vye, N.T. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal*, 90, 469-484.
- Mastropieri, M., Scruggs, T. E., Bakken, J. P., & Whedon, C. (1996). Reading Comprehension. A synthesis of research in learning disabilities. *Advances in Learning and Behavioural Disabilities*, 10b, 201-227.
- Mayer, A. (2018). Die Lesetricks von Professor Neugier – ein strategieorientiertes Konzept zur Förderung des Textverständnisses. *Sprachförderung und Sprachtherapie in Schule und Praxis*, 7, 149-155.
- Mayer, A. (2016a). *Lese-Rechtschreibstörungen*. München: Reinhardt Verlag.
- Mayer, A. (2016b). Vermittlung strategischer Lesekompetenz im inklusiven Unterricht. In D. Gebele & A. Zepter (Hrsg.), *Inklusion: Sprachdidaktische Perspektiven. Theorie, Empirie, Praxis*. Reihe KöBeS (Kölner Beiträge zur Sprachdidaktik) (Band 11, S. 228-255). Duisburg: Gilles & Francke Verlag.
- Mayer, A. (2014). *Früherkennung und Prävention von Schriftspracherwerbsstörungen im inklusiven Unterricht*. Universität zu Köln: Unveröffentlichte Habilitationsschrift.
- Miller, A., Keenan, J. (2009). How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia*, 59, 99-113.
- \* Mistar, J., Zuhairi, A., & Yanti, N. (2016). Strategies training in the teaching of reading comprehension for EFL learners in indonesia. *English Language Teaching*, 9, 49-56.
- \* Mokhlesgerami, J., Souvignier, E., & Gentsch, S. (2006). Förderung von Lesestrategien Erprobung eines Unterrichtsprogramms in Haupt-, Real- und Gesamtschulen. *Empirische Pädagogik*, 20, 70-9.
- \* Mokhlesgerami, J., Souvignier, E., Rühl, K., & Gold A. (2007). Naher und weiter Transfer eines Unterrichtsprogramms zur Förderung der Lesekompetenz in der Sekundarstufe I. *Zeitschrift für pädagogische Psychologie*, 21, 169-18.
- Morris, S. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11, 364-386.
- Motsch, H.-J., Marks, D.-K., & Ulrich, T. (2018). *Der Wortschatzsammler. Evidenzbasierte Strategietherapie lexikalischer Störungen im Kindesalter* (3. Aufl.). München: Ernst Reinhardt Verlag.
- Palinscar, A., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- \* Ponce, H.R., Lopez, M.J., & Mayer, R.E. (2012). Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies. *Computers & Education*, 59, 1170-1183.
- Reuter-Liehr, C. (2006). *Lautgetreue Lese-Rechtschreibförderung. Band 2/1. Elementartraining Phonemstufe I*. Bochum: Winkler Verlag.
- Ricketts, J., Nation, K., & Bishop, D. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11, 235-257.
- \* Schünemann, N., Spörer, N., & Brunstein, J. (2013). Integrating self-regulation in whole-class reciprocal teaching: A moderator-mediator analysis of incremental effects on fifth graders' reading comprehension. *Contemporary Educational Psychology*, 38, 289-305.
- \* Seuring, V.A., & Spörer, N. (2010). Reziprokes Lehren in der Schule: Förderung von Leseverständnis, Leseflüssigkeit und Strategieanwendung. *Zeitschrift für pädagogische Psychologie*, 24, 191-205.
- Snellings, P., van der Leij, A., de Jong, P., & Blok, H. (2009). Enhancing the reading fluency of Children with Reading Disabilities in an orthographically transparent Language. *Journal of Learning Disabilities*, 42, 291-304.
- Souvignier, E., & Antoniou, F. (2007). Förderung des Leseverständnisses bei Schülerinnen und Schülern mit Lernschwierigkeiten – eine Metaanalyse. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 76, 46-62.
- \* Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction*, 16, 57-71.
- \* Spörer, N., & Brunstein, J. (2009). Fostering the reading comprehension of secondary school students through peer-assisted learning: Effects on strategy knowledge, strategy use, and task performance. *Contemporary Educational Psychology*, 34, 289-297.
- \* Spörer, N., Brunstein, J.C., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and instruction*, 19, 272-286.

- \* Spörer, N., Seuring, V., Schünemann, N., & Brunstein, J. (2008). Förderung des Leseverständnisses von Schülern der 7. Klasse. Effekte peer-gestützten Lernens in Deutsch und Englisch. *Zeitschrift für Pädagogische Psychologie*, 22, 247–259.
- Swanson, H.L., Zheng, X. & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities*, 42, 260–289.
- Tacke, G. (2005). Evaluation eines Lesetrainings zur Förderung lese-rechtschreibschwacher Grundschüler der zweiten Klasse. *Psychologie in Erziehung und Unterricht*, 52, 198–209.
- Tacke, G. (2009). *Flüssig lesen lernen. Übungen, Spiele und spannende Geschichten. 1. und 2. Kl. Lehrerband mit Arbeitsblättern*. Stuttgart: Klett Verlag.
- Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeler, K.K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities. Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities* 34, 33–58. 78
- Tunmer, W.E., & Greaney, K. (2010). Defining Dyslexia. *Journal of Learning Disabilities*, 43, 229–243.
- Van Dijk, T., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- \* Wassenburg, S.I., Bos, L.T., de Koning, B.B., & Van der Schoot, M. (2015). Effects of an inconsistency-detection training aimed at improving comprehension monitoring in primary school children. *Discourse Processes*, 52, 463–488.

## Zu den Autoren

*Prof. Dr. Andreas Mayer* ist Inhaber des Lehrstuhls für Sprachheilpädagogik (Sprachtherapie und Förderschwerpunkt Sprache) an der LMU München. Arbeits- und Forschungsschwerpunkte sind die Diagnostik von Sprach- und Sprechstörungen, die spezifische Akzentuierung des Unterrichts im Förderschwerpunkt Sprache sowie die Theorie und Praxis gestörter Schriftspracherwerbsprozesse.

*Dr. Dana-Kristin Marks* ist wissenschaftliche Mitarbeiterin am Lehrstuhl von Prof. Mayer. Aktuelle Forschungsschwerpunkte liegen im Bereich der Diagnostik und Behandlung von Spracherwerbsstörungen im Kontext von Ein- und Mehrsprachigkeit.

## Korrespondenzadresse

Prof. Dr. Andreas Mayer, Dr. Dana-Kristin Marks  
LMU München  
Lehrstuhl für Sprachheilpädagogik (Sprachtherapie und Förderschwerpunkt Sprache)  
Leopoldstraße 13  
80802 München  
E-Mail:  
Andreas.Mayer@edu.lmu.de  
D.Marks@lmu.de



## **Von Angsthase bis Zirkuszebra ...**

### **Lautgeschichten zu jedem Buchstaben des Alphabets**

Seien Sie dabei, wenn Finkenjunge Friedrich das Fliegen lernt, gehen Sie mit Dackel Detlef auf Spurensuche oder erleben Sie zusammen mit Papageienfisch Peter ein spannendes Abenteuer.

Schul- und Vorschulkinder benötigen vielfältige Übungsanlässe, um alle Buchstaben und Lautverbindungen sicher erkennen, benennen, schreiben und erlesen zu können.

Das Buch „Von Angsthase bis Zirkuszebra“ bietet Lehrern, Therapeuten und Erziehern 33 fantasievolle, lustige und spannende Laut- und Buchstabengeschichten, um Kinder in diesem Lernprozess zu unterstützen. Jeder Buchstabe/Laut hat seine eigene Geschichte. Im Vordergrund stehen das Kernalphabet sowie die Lautverbindungen (sch, sp, st) und die Umlaute (ä, ö, ü).

Die Geschichten sind vielfältig einsetzbar, sie können im Unterricht, in der Sprachtherapie oder einfach nur zum Vorlesen genutzt werden.

Karen Schneider, 1. Auflage 2019,  
kartoniert: ISBN 978-3-8248-1250-9,  
80 Seiten,  
E-Book: ISBN 978-3-8248-9951-7,  
PDF, 2.354 KB,  
EUR 10,00 [D]



Tel.: +49 (0) 6126 9320-13  
Fax: +49 (0) 6126 9320-50



bestellung@schulz-kirchner.de  
www.skvshop.de





## Evaluation des computerbasierten Trainingsprogramms *Lautarium* bei Kindern mit Sprachentwicklungsstörungen\*

## Evaluation of the computer-based training program *Lautarium* in children with language impairment

Anita Hönninger

### Zusammenfassung

**Hintergrund:** Kinder mit Sprachentwicklungsstörungen sind häufig von Schriftsprachproblemen betroffen. Die Wirksamkeit von Programmen zur Förderung der phonologischen Bewusstheit und Kombinationstrainings mit schriftsprachlichen Anteilen zur Unterstützung des Schriftspracherwerbs im Vorschul- und frühen Grundschulalter zeigen unterschiedliche Ergebnisse. Befunde von Trainingsstudien im höheren Grundschulalter liegen kaum vor.

**Ziele:** In der vorliegenden Studie wurde untersucht, ob ein computerbasiertes Trainingsprogramm (*Lautarium*) zur Förderung bei Lese-Rechtschreibschwierigkeiten auch bei Kindern mit Sprachentwicklungsstörungen im höheren Grundschulalter wirksam ist.

**Methode:** Dritt- und Viertklässler (N=33) mit Sprachentwicklungsstörungen eines sonderpädagogischen Förderzentrums wurden mit standardisierten Testverfahren zur Überprüfung phonologischer Leistungen (P-ITPA), zum Lesen (ELFE 1-6) und zum Rechtschreiben (WRT 2+, WRT 3+) überprüft. Die Evaluationsstudie wurde als Wartekontrollgruppendesign mit drei Messzeitpunkten (Vortest, Nachtest 1, Nachtest 2) durchgeführt. Die Kinder der Trainingsgruppe (N=16) übten täglich für ca. 15–20 Minuten individuell mit Kopfhörern an Laptops über einen Zeitraum von ca. neun Wochen parallel zum regulären Unterricht. Zur Untersuchung, ob es kurz- oder längerfristige Trainingseffekte gab, erfolgten Kovarianzanalysen mit den Vortestergebnissen als Kovariate (ANCOVA) sowie t-Tests auf Unterschiede der Differenzwerte der Nachtests zum Vortest (Gain-Scores) und die Berechnung der Effektstärke nach Cohen.

**Ergebnisse:** Die Trainingsgruppe zeigte in einem der Subtests zur phonologischen Bewusstheit signifikante und anhaltende Trainingseffekte mit mittlerer bis hoher Effektstärke (Cohen's  $d$  bzw.  $d_{\text{korr}}$  zwischen 0,56 und 1,29). Es zeigten sich Tendenzen zu Leistungssteigerungen in einem zweiten Subtest zur phonologischen Bewusstheit sowie zu positiven Trainingseffekten auf das Lesen und auf das lautgetreue Schreiben.

Schlussfolgerungen: Die Ergebnisse der Studie deuten darauf hin, dass Kinder mit Sprachentwicklungsstörungen von der Förderung mit *Lautarium* profitieren könnten. Weitere Studien sind nötig.

### Schlüsselwörter

Sprachentwicklungsstörungen, Lese-Rechtschreibschwierigkeiten, *Lautarium*, phonologische Bewusstheit

\* Dieser Beitrag hat das Peer-Review-Verfahren durchlaufen.



## Abstract

**Background:** Children with language impairment are often affected by reading and writing disorders. Studies investigating the effectiveness of phonological awareness intervention and combination trainings with written language components in preschool and during the first two years of literacy acquisition show different results. So far, there are hardly any findings from training studies in higher primary school age.

**Aims:** The present study examined whether a computer-based training program for children with reading and spelling disorders is also effective in children with language impairment.

**Methods:** Third- and fourth-graders (N=33) with language impairment of a school for special education were tested with standardized phonological examinations (P-ITPA), reading (ELFE 1-6) and spelling (WRT 2+, WRT 3+). The evaluation study was carried out as a waiting control group design with three measurement times (pre-test, post-test 1, post-test 2). The children of the training group (N=16) practiced daily for about 15-20 minutes individually using headphones on laptops over a period of about nine weeks parallel to attending their regular classes. To investigate whether there were short-term or longer-term training effects, analyzes of covariance (ANCOVA) were performed with the pre-test scores as a covariates. As a further test, the differences between post-tests and pre-test scores (gain scores) between the training and control groups were tested for significance using independent sample t-tests. Cohen's d was calculated to evaluate the effect size.

**Results:** The training group showed significant and sustained training effects in one of the subtests on phonological awareness with medium to large effect sizes (Cohen's *d* and  $d_{\text{kor}}$  between 0,56 and 1,29). There was a trend for the children in the training group to show more improvements in a second subtest on phonological awareness as well as in reading and spelling compared to the children in the control group.

**Conclusion:** The results of the study suggest that children with language impairment could benefit from *Lautarium*. Further studies are necessary.

## Keywords

language impairment, developmental dyslexia, *Lautarium*, phonological awareness

## 1 Einleitung

Seit ungefähr drei Jahrzehnten ist bekannt, dass Kinder mit Sprachentwicklungsstörungen häufig Schwierigkeiten beim Schriftspracherwerb zeigen (Mayer, 2013). Da die Schriftsprache ein zentrales Element im Unterricht darstellt, besteht die Gefahr, dass Beeinträchtigungen beim Lesen und Schreiben zu Problemen beim schulischen Lernen insgesamt führen, die dann auch den weiteren Bildungs- und Berufsweg maßgeblich beeinflussen (vgl. Dannenbauer, 2002). Neben den schulischen Auffälligkeiten können Lese-Rechtschreibschwierigkeiten auch langfristig negative Auswirkungen auf die sozial-emotionale Entwicklung dieser Kinder und Jugendlichen haben (McGee, Williams, Share, Anderson & Silva, 1986; Silva, Williams & McGee, 1987). Torgesen, Wagner und Rashotte (1997) betonten bereits 1997, dass Schulen geeignete präventive Maßnahmen einsetzen sollten, um die großen Folgedefizite beim Lesen und Schreiben möglichst gar nicht erst aufkommen zu lassen.

## 2 Hintergrund

### 2.1 Sprachentwicklungsstörungen und Schriftsprachprobleme

Laut der interdisziplinären S2k-Leitlinie (De Langen-Müller, Kauschke, Kiese-Himmel, Neumann & Noterdaeme, 2012) entwickeln 40–75 % der Kinder mit spezifischer Sprachentwicklungsstörung (SSES; Synonym: umschriebene Sprachentwicklungsstörung, USES) Probleme im Schriftspracherwerb, die sich bis ins Jugendlichen- und Erwachsenenalter auswirken können.

Allerdings zeigen Kinder mit Sprachentwicklungsstörungen verschiedenartige Leistungen beim Lesen und Schreiben. Aus diesem Grund wird auch von einem unterschiedlichen Bedingungsgefüge ausgegangen. Schöler spricht von einem Anteil von ca. 60 % von Kindern mit SSES, die in der Schule Probleme beim Schriftspracherwerb haben (Schöler & Scheib, 2004). Die übrigen 40 % ohne Probleme beim Schriftspracherwerb profitieren besonders im Anfangsunterricht sogar vom Lesen- und Schreibenlernen. Es wird vermutet, dass die Schrift diesen Kindern bei Segmentierungsproblemen in der Lautsprache kompensatorisch hilft. Schöler plädiert deshalb für eine Unterscheidung dieser beiden Gruppen, da diese auch unterschiedliche Kompensationsstrategien benötigen.

Unterschiede in den Prävalenzangaben können darauf beruhen, dass unter dem Begriff Sprachentwicklungsstörungen teils unterschiedliche Störungsbilder subsumiert werden (Schöler & Scheib, 2004). Nicht nur Kinder mit SSES sind anfällig für Probleme beim Schriftspracherwerb, sondern auch Kinder mit anderen Sprachentwicklungsstörungen (Bird, Bishop & Freeman, 1995; Nathan, Stackhouse, Goulondris & Snowling, 2004; Peterson, Pennington, Shriberg & Boada, 2009; Puranik, Petscher, Al Otaiba, Catts & Lonigan, 2008; Raitano, Pennington, Tunick, Boada & Shriberg, 2004).

Für Störungen im Schriftspracherwerb gelten Beeinträchtigungen im Bereich der phonologischen Informationsverarbeitung als entscheidende Ursache (Snowling, 2000; Stanovich, 1988). Das Konstrukt der phonologischen Informationsverarbeitung (Wagner & Torgesen, 1987) bezieht sich auf die Repräsentation und Verarbeitung gesprochener oder geschriebener Sprache. In Anlehnung an Wagner und Torgesen (1987) werden drei Komponenten unterschieden: die phonologische Bewusstheit, das phonologische Arbeitsgedächtnis und die Abrufgeschwindigkeit phonologischer Repräsentationen aus dem Langzeitgedächtnis.

Hinsichtlich ihrer Fähigkeiten im Bereich der phonologischen Bewusstheit stellen Kinder mit Sprachentwicklungsstörungen eine heterogene Gruppe dar und haben nicht generell ein Entwicklungsdefizit (vgl. Magnusson & Nauclér, 1993).

Bei einer SSES wird insbesondere ein Funktionsdefizit in der phonologischen Schleife des Arbeitsgedächtnisses als mitverursachend für die Ausbildung einer spezifischen Sprachentwicklungsstörung angenommen (Hasselhorn & Werner, 2000; Janczyk, Schöler & Grabowski, 2004). Ausgehend von diesen Erkenntnissen zum phonologischen Arbeitsgedächtnisdefizit wird zunehmend die Frage untersucht, ob die gefundenen Abweichungen zur normalen Entwicklung lediglich quantitativer Natur oder aber qualitativ struktureller Art sind (Schuchardt, Worgt & Hasselhorn, 2012, S. 80). Neben diesen rein phonologischen Speicherschwierigkeiten gibt es in jüngeren Studien auch vermehrt Erkenntnisse zu zentral-exekutiven Verarbeitungsdefiziten (Schuchardt et al., 2012). Wie sich die jeweiligen Arbeitsgedächtnisbeeinträchtigungen bei Kindern mit SSES sowohl auf die Sprach- als auch auf den später folgenden Schriftspracherwerb und auf die schulische Entwicklung insgesamt auswirken, bleibt zu klären.

Bezüglich der Abrufgeschwindigkeit phonologischer Repräsentationen zeigen leseschwache Kinder in der Benennungsgeschwindigkeit deutliche Unterschiede zu durchschnittlich lesenden Kindern (Mayer, 2018).

Im angloamerikanischen Raum finden Untersuchungen zu Zusammenhängen zwischen Sprachentwicklungsstörungen und Lese-Rechtschreibstörungen (LRS) bereits seit mehreren Jahrzehnten statt (beispielsweise Bishop & Adams, 1990; Catts, 1993; McArthur, Hogben, Edwards, Heath & Mengler, 2000). Durch die Zusammenführung von Forschungen aus dem Bereich des Sprach- und des Schriftspracherwerbs mit den jeweiligen Auffälligkeiten und Störungen zeigte sich, dass bestimmten Störungen des Sprach- und des Schriftspracherwerbs gemeinsame Bedingungsgefüge zugrunde liegen könnten (Schöler, 2011).

Ramus, Marshall, Rosen und van der Lely (2013) untersuchten mehrere in der Literatur beschriebene Arbeitsmodelle zur Beziehung zwischen phonologischen Defiziten, SSES und LRS: Das sog. „severity model“ und das „additional deficit model“ gehen davon aus, dass Kinder mit SSES und LRS ein gemeinsames phonologisches Defizit aufweisen. Diese beiden Modelle unterscheiden sich darin, dass das „severity model“ lediglich von einer unterschiedlichen Ausprägtheit des Defizits ausgeht (Kamhi & Catts, 1986; Tallal, 2003), während das „additional deficit model“ ein zusätzliches Defizit bei Kindern mit SSES vermutet (vgl. Bishop & Snowling, 2004). Aus beiden Modellen folgt, dass Kinder mit SSES immer auch eine LRS haben müssten, was nach neueren Erkenntnissen nicht der Fall ist. In einem dritten Modell, dem sog. „component model“, wird davon ausgegangen, dass LRS und SSES unterschiedliche Störungsbilder mit eigenen Ursachen sind. Die Tatsache, dass sie überzufällig häufig gemeinsam auftreten, wird einer Komorbidität auf ätiologischer Ebene zugeschrieben (Catts, Adlof, Hogan & Weismer, 2005). Bei diesem Modell entwickeln Kinder mit SSES nicht unbedingt eine LRS und haben auch nicht notwendigerweise phonologische Defizite. Falls sie dennoch phonologische Defizite haben, müssen diese nicht unbedingt derselben Natur sein wie bei Kindern mit LRS. Als Ergebnis ihrer eigenen Untersuchung postulieren Ramus und Kollegen (2013) eine Erweiterung des „component model“, das sie als „multiple component model“ bezeichnen. Dabei wird von zwei verschiedenen Kategorien phonologischer *Fähigkeiten* („phonological skills“, „phonological representations“) ausgegangen. Als Fazit hinsichtlich der Beziehung zwischen phonologischer Defizite, SSES und LRS stellen die Autoren fest, dass Kinder mit SSES nicht zwangsläufig immer ein phonologisches De-

fizit und auch nicht immer zwangsläufig eine Lesestörung haben. Auch wenn SSES und LRS häufig komorbid auftreten, gibt es reine SSES und reine LRS.

## 2.2 Wirksamkeit von Trainingsprogrammen bei Sprachentwicklungsstörungen

Phonologische Defizite bei Kindern mit Sprachentwicklungsstörungen können also sowohl bei der Verursachung einer Sprachentwicklungsstörung als auch für Probleme beim Schriftspracherwerb eine Rolle spielen. In der Praxis sollten phonologische Probleme möglichst frühzeitig identifiziert, die phonologische Entwicklung beobachtet und bei Bedarf gezielt gefördert werden (Al Otaiba, Puranik, Ziolkowski & Montgomery, 2009). Es scheint, dass sich bei Vorschulkindern mit Sprachentwicklungsstörungen die phonologische Bewusstheit und die Phonembewusstheit langsamer entwickeln als bei unbeeinträchtigten Gleichaltrigen (Bird et al., 1995). Aus diesem Grund werden bei Kindern mit Sprachentwicklungsstörungen im Vorschul- oder frühem Grundschulalter häufig Trainingsprogramme mit dem Förderschwerpunkt der phonologischen Bewusstheit eingesetzt. Allerdings existiert nur eine geringe Anzahl an evaluierten Trainingsstudien bei diesen Kindern (Al Otaiba et al., 2009). Dabei kann zwischen Förderprogrammen unterschieden werden, die sich ausschließlich auf die phonologische Bewusstheit konzentrieren (Al Otaiba et al., 2009; Hartmann & Studer, 2013; Segers & Verhoeven, 2004; Van Kleeck, Gillam & McFadden, 1998; Warrick, Rubin & Rowe-Walsh, 1993) und Trainingsstudien, die die Wirksamkeit von Kombinationstrainings bei Kindern mit Sprachentwicklungsstörungen, d.h. Trainings mit phonologischen und schriftsprachlichen Komponenten, untersuchen (Gillon, 2000, 2002; Hartmann, 2002, 2003; Hartmann & Studer, 2013; Hund-Reid & Schneider, 2013; Marx, Weber & Schneider, 2005; Ritter, Park, Saxon & Colson, 2013).

Hartmann (2002, 2003) überprüfte die Wirksamkeit eines metaphonologischen Vorschultrainings unter Einbezug von Laut-Buchstabenbeziehungen bei Kindern mit Sprachentwicklungsstörung im Halbjahr vor der Einschulung. Es zeigte sich zwar ein unmittelbarer Trainingseffekt auf die phonologische Bewusstheit, dieser war jedoch nicht stabil. In einem zweiten Nachtest vier Monate nach Beginn der ersten Klasse hatten die trainierten Kinder keinen Vorteil mehr hinsichtlich ihrer Leistungen im Bereich der phonologischen Bewusstheit, ähnlich war es beim Lesen und Schreiben.

In der Trainingsstudie von Marx und Kollegen (2005) wurden die phonologischen Fähigkeiten von Kindern mit einer Sprachentwicklungsstörung vor und nach einem kombinierten Training zur phonologischen Bewusstheit mit Phonem-Graphem-Zuordnung („Hören, Lauschen, Lernen 1 und 2“) untersucht. Dabei konnten die Kinder mit Sprachentwicklungsstörungen unmittelbar in vergleichbarem Ausmaß wie die sprachlich unauffälligen Kinder vom Training profitieren, dies galt jedoch nicht im Hinblick auf die schwierigeren Aufgaben zur phonologischen Bewusstheit.

Einen Überblick über die Forschungslage hinsichtlich der Wirksamkeit von frühen Interventionen zur phonologischen Bewusstheit bei Kindern mit Sprech- und Sprachentwicklungsstörungen in Studien der Jahre 1990 bis 2006 findet sich in der Metaanalyse von Al Otaiba und Kollegen (2009). Hartmann und Studer (2013) berücksichtigen in ihrer Übersichtsarbeit weitere veröffentlichte Studien im Zeitraum bis zum Jahr 2010. Die Ergebnisse der Metaanalyse von Al Otaiba und Kollegen (2009) werden dabei weitestgehend bestätigt und teils ergänzt: Am meisten scheinen Kinder mit Sprachentwicklungsstörungen, zumindest kurzzeitig, von Trainings zu profitieren, die Übungen zur phonologischen Bewusstheit mit Übungen zur systematischen Verknüpfung der Phonem-Graphem-Korrespondenz verbinden. Dies entspricht aktuellen Befunden aus Metaanalysen zur Wirksamkeit von Förderprogrammen bei LRS (Galuschka, Ise, Krick & Schulte-Körne, 2014; Ise, Engel & Schulte-Körne, 2012). Isolierte Trainings der phonologischen Bewusstheit reichen bei umfassenden Beeinträchtigungen der Schriftsprache bzw. bei LRS nicht aus, um Verbesserungen im Bereich der Schriftsprache zu bewirken. Positivere Effekte gibt es bei kombinierten Trainings, bei denen eine systematische Vermittlung der Phonem-Graphem-Beziehungen mit Übungen zur phonologischen Bewusstheit erfolgt. Diese frühzeitige Verbindung zwischen phonologischen und schriftsprachlichen Inhalten entspricht der phonologischen Verknüpfungshypothese (Hatcher, Hulme & Ellis, 1994).

Kinder mit persistierenden Lautsprachproblemen und großen Rückständen in der phonologischen Bewusstheit sollten als Subgruppe besondere Beachtung finden (Hartmann & Studer, 2013). Sie scheinen ein noch weitaus höheres Risiko für das Ausbilden von Schriftsprachproblemen bzw. einer LRS zu haben.

In einer Trainingsstudie bei Drittklässlern mit SSES untersuchten Cholewa, Nierhaus, Tiro & Unger (2013), ob und wie persistierende Defizite des alphabetischen bzw. sublexikalischen Schreibens durch ein gezieltes Training der phonologischen Bewusstheit (Onset-Reim- und Phonem-Graphem-Training) vermindert werden könnte. Es ergaben sich jedoch nur wenige Hinweise für eine differenzielle Wirksamkeit der beiden erprobten Trainingsverfahren. Die Ergebnisse der Studie deuten allerdings darauf hin, dass ein Training der phonologischen Bewusstheit und sublexikaler Schreibleistungen auch im höheren Grundschulalter für Kinder mit SSES und LRS indiziert sein kann (Cholewa et al., 2013, S. 336f).

Diese zusammengefassten Erkenntnisse und der spürbare Mangel an einem motivierenden und wirksamen Förderprogramm für die schulische Arbeit bei Kindern mit Sprachentwicklungsstörungen waren die Motivation, das bereits für Grundschulkinder erfolgreich evaluierte PC-Förderprogramm *Lautarium* auch bei Kindern mit Sprachentwicklungsstörungen einzusetzen und zu evaluieren. Die Untersuchung wurde mit einer Entwicklungsversion von *Lautarium* im Rahmen einer wissenschaftlichen Abschlussarbeit durchgeführt (Hönninger, 2017).

### 3 Das computerbasierte Trainingsprogramm *Lautarium*

*Lautarium* ist ein computerbasiertes Trainingsprogramm zur Förderung von Grundschulkindern mit LRS oder einem erhöhtem Risiko dafür (Klatte, Steinbrink, Bergström & Lachmann, 2017). *Lautarium* erfüllt für diese Zielgruppe allgemeine Gütekriterien für Interventionsprogramme, d. h. es ist bereichsspezifisch wirksam, die Ergebnisse der bisherigen Untersuchungen weisen auf einen Transfer auf nicht trainierte Leistungen und auch auf eine Nachhaltigkeit von Trainingseffekten hin (Steinbrink & Lachmann, 2014).

#### 3.1 Theoretischer Hintergrund des Programms

Das Programm basiert auf der Erkenntnis, dass die phonologische Informationsverarbeitung eine wesentliche kognitiv-sprachliche Voraussetzung für einen ungestörten Schriftspracherwerb ist (Bradley & Bryant, 1983; Lundberg, Olofsson & Wall, 1980) und setzt soweit wie möglich inhaltliche und formale Aspekte von Interventionsmaßnahmen um, die sich in der internationalen und nationalen Forschung als effektiv erwiesen haben. So haben Metaanalysen zur Wirksamkeit deutschsprachiger Förderkonzepte bei LRS gezeigt, dass symptomspezifische Ansätze deutlich wirksamer sind als Funktions- und Wahrnehmungstrainings, dass bestimmte Merkmale (z. B. Dauer und Umfang der Förderung) die Wirksamkeit von LRS-Förderprogrammen beeinflussen, oder dass der wichtigste Förderansatz bei Leseschwierigkeiten eine strukturierte Einführung der Laut-Buchstabenbeziehung („phonics instruction“) ist (Galuschka et al., 2014; Ise et al., 2012).

Über die drei Funktionen der phonologischen Informationsverarbeitung (phonologische Bewusstheit, phonologische Arbeitsgedächtnis, Abrufgeschwindigkeit phonologischer Repräsentationen aus dem Langzeitgedächtnis) hinaus wird von den Autoren des Förderprogramms auch auf die Bedeutung der Phonemwahrnehmung für den Schriftspracherwerb hingewiesen (Klatte et al., 2017). Es wird davon ausgegangen, dass eine intakte Phonemwahrnehmung eine Voraussetzung für die Ausbildung qualitativ hochwertiger Repräsentationen im Langzeitgedächtnis darstellt, die sich wiederum positiv auf die Entwicklung der phonologischen Bewusstheit auswirken können (vgl. Steinbrink & Lachmann, 2014).

Der theoretische Hintergrund des Konzepts von *Lautarium* wird im Manual (Klatte et al., 2017) ausführlich erläutert und im Folgenden mit Verweis auf die genannten Studien kurz zusammengefasst. Das phonologische Arbeitsgedächtnis (Melby-Lervåg & Hulme, 2013) und der schnelle Abruf von Wortrepräsentationen aus dem Langzeitgedächtnis (Kirby, Georgiou, Martinussen & Parrila, 2010) lassen sich kaum trainieren. Der Bereich der phonologischen Bewusstheit lässt sich gut trainieren und in Metastudien konnten vielfach Transfereffekte auf den späteren Schriftspracherwerb nachgewiesen werden (Bus & van IJzendoorn, 1999; Ehri et al., 2001; Fischer & Pfost, 2015). Ein Training der Phonemwahrnehmung führt zwar zu Verbesserungen (Bischof et al., 2002; McArthur, Ellis, Atkinson & Coltheart, 2008; Strehlow et al., 2006), bei isoliertem Training der Phonemwahrnehmung scheint es jedoch keine Transfereffekte auf schriftsprachliche Leistungen zu geben (McArthur et al., 2008; Strehlow et al., 2006). Positive Transfereffekte von Phonemtrainings auf schriftsprachliche Leistungen zeigten sich dann, wenn diese mit einem Training zur phonologischen Bewusstheit (Gonzalez, Garcia Espinel & Guzman Rosquete, 2002) oder einer Laut-Buchstabe-Zuordnung kombiniert waren (Ecalte, Magnan, Bouchafa & Gombert, 2009; Magnan, Ecalte, Veillet & Collet, 2004).

### 3.2 Konzeption und Aufbau

*Lautarium* wurde so konzipiert, dass Aufgaben, die ursprünglich für andere Sprachen entwickelt wurden, der deutschen Phonologie und einem computerbasierten Format angepasst wurden (Klatte et al., 2017). Das Programm basiert auf einem umfangreichen Bild- (Rossion & Pourtois, 2004; Snodgrass & Vanderwart, 1980) und Sprachmaterial sowie Bausteinen, die Laute und die zugehörigen Basisgrapheme (Thomé, 2000) repräsentieren. Die Übungen bauen aufeinander auf und es erfolgt eine möglichst frühzeitige Verbindung zwischen phonologischen und schriftsprachlichen Inhalten gemäß der phonologischen Verknüpfungshypothese (Hatcher et al., 1994). Die Zielgruppen des Programms sind Kinder der ersten und zweiten Klasse (Version Klasse 1 und 2) sowie der dritten und vierten Klasse (Version Klasse 3 und 4). Diese Unterscheidung nach Klassenstufen gab es für die in der Studie verwendete Entwicklungsversion noch nicht.

Das Programm umfasst aufeinander aufbauende Übungen zur Phonemwahrnehmung, zur phonologischen Bewusstheit, zur Graphem-Phonem-Zuordnung, zum Lesen und Schreiben lautgetreuer Wörter sowie zur schnellen Worterkennung. Der Bereich *Phonemwahrnehmung* umfasst ein Training der Konsonanten- und Vokalwahrnehmung. Im Bereich der Konsonanten (Diskrimination und Rekognition) finden dabei besonders schwer differenzierbare Plosivlaute (/b/, /p/, /d/, /t/, /g/, /k/) Beachtung, die mit ansteigendem Schwierigkeitsgrad (von zunächst einfachen Silbenstrukturen bis hin zur Einbettung in Konsonantencluster) präsentiert werden. Ergänzt wird dieser Bereich um ein Training der Vokallängenwahrnehmung (Diskrimination und Identifikation). Vokale spielen für den Schriftspracherwerb im Deutschen eine besondere Rolle (Klicpera & Gasteiger-Klicpera, 1998) und Befunde aus Studien weisen darauf hin, dass ein Training der Vokallängenwahrnehmung sinnvoll ist (Landerl, 2003; Steinbrink, Klatte & Lachmann, 2014). Der Bereich *phonologische Bewusstheit* umfasst Übungen zur Lautanalyse (z. B. Laute erkennen, Laute zählen), zur Lautsynthese (z. B. Wörter aus Lautbausteinen bauen) und zur Lautklassifikation („Odd-One-Out“-Aufgaben). Der Bereich *Lesen und Schreiben* lautgetreuer Wörter wird mit Hilfe von Graphembausteinen (z. B. Wörter aus einer Folge von Graphembausteinen erlesen oder mit Graphembausteinen schreiben), Leseübungen (z. B. das zu einem vorgesprochenen Wort passende Schriftwort unter visuell oder phonologisch ähnlichen Distraktoren finden) sowie durch schnelle Worterkennung (z. B. Blitzlesen) trainiert. Eine differenzierte Beschreibung der Übungen des Programms findet sich im Manual.

Das Programm ist selbsterklärend und adaptiv so aufgebaut, dass Kinder die Aufgaben in der Regel ohne weitere Hilfe einzeln am Computer bearbeiten können sollten. Das Trainingsprogramm ist mit einem virtuellen Belohnungssystem in Form von Tokens ausgestattet, die die Kinder für das Ausgestalten eines persönlichen Aquariums verwenden können. Diese Form der Motivation über Token-Systeme hat sich in einer Metaanalyse (vgl. Ise et al., 2012) als besonders wirksam bei Trainingsprogrammen herausgestellt. Es zeigte sich, dass die mittlere Effektstärke der Studien, in denen Token eingesetzt wurden deutlich höher lag als die Effektstärke der Studien, in denen keine Token zum Einsatz kamen. *Lautarium* soll als Intensivtraining fünfmal pro Woche für ca. 20–30 Minuten über einen Zeitraum von ungefähr acht Wochen durchgeführt werden.

## 4 Fragestellung

Die Wirksamkeit des Trainingsprogramms *Lautarium* wurde bereits bei Drittklässlern mit LRS sowie bei Erstklässlern in der Schuleingangsphase nachgewiesen (Klatte et al., 2014; Klatte, Steinbrink, Bergström & Lachmann, 2016). In der vorliegenden Studie wird die Wirksamkeit bei Kindern mit Sprachentwicklungsstörungen untersucht.

Es wird der Fragestellung nachgegangen, ob die Interventionsgruppe nach Abschluss des Trainings bereichsspezifisch einen höheren Lernzuwachs bei Aufgaben zur phonologischen Bewusstheit zeigt als die Kontrollgruppe. Weiterhin wurde überprüft, ob Transfereffekte auf nicht trainierte Bereiche beim Lesen und Rechtschreiben beobachtbar sind und ob es nachhaltig wirksam ist (Dignath, Buettner & Langfeldt, 2008; Klauer, 2001). Darüber hinaus wurden Erfahrungen gesammelt, ob sich das Trainingsprogramm *Lautarium* für den schulischen Einsatz unter den Rahmenbedingungen eines Förderzentrums eignet.



## 5 Methoden

### 5.1 Stichprobenumfangsplanung

Die Ergebnisse einer bereits durchgeführten Evaluationsstudie zu *Lautarium* (Klatte et al., 2014) zeigten bei schwachen Lesern der ersten Klasse signifikante Trainingseffekte mittlerer bis hoher Effektstärken auf die schriftsprachlichen und phonologischen Fähigkeiten. Für die vorliegende Untersuchung wurde aufgrund dieser Ergebnisse zumindest in Teilbereichen ebenfalls mit einem mittleren bis großen Effekt gerechnet.

Die Stichprobenumfangsplanung wurde für eine Kovarianzanalyse (ANCOVA) mit zwei Gruppen (Trainings- und Kontrollgruppe) und einer Kovariate (Vortestleistung) durchgeführt. Das Signifikanzniveau  $\alpha$  wurde auf 0,05 und die a priori gewünschte Teststärke ( $1-\beta$ ) auf 0,90 gesetzt entsprechend der Empfehlungen von Rasch, Frieze, Hofmann und Naumann (2010, S. 80). Die Berechnung erfolgte mit G\*Power (Version 3.1.9.2). Für einen großen Effekt ( $f=0,40$ ) ergibt sich eine Stichprobengröße von 68, für einen mittleren Effekt ( $f=0,25$ ) eine Stichprobengröße von 171. Diese Stichprobengröße war im Rahmen dieser Studie nicht möglich. Eine mögliche im Vorfeld maximal zu erwartende Größe der Stichprobe bei Einbeziehung aller Kinder aus allen dritten und vierten Klassen lag bei 52. Für einen starken Effekt ergibt sich bei dieser Anzahl eine Teststärke von 0,81.

### 5.2 Stichprobe und Untersuchungsdesign

Die Studie fand im Laufe des Schuljahres 2015/16 an einem Sonderpädagogischen Bildungs- und Beratungszentrum mit dem Förderschwerpunkt Sprache in Baden-Württemberg statt. Zu Beginn des Schuljahres 2015/16 wurden für die Erfassung der Stichprobe alle 52 Kinder der dritten und vierten Klassen des Sonderpädagogischen Bildungs- und Beratungszentrums angefragt (Rücklaufquote ca. 83 %). Die Ausgangsstichprobe umfasste zunächst insgesamt 43 Kinder, davon besuchten 25 Kinder die Klassenstufe drei aus drei Parallelklassen und 18 Kinder die Klassenstufe vier aus zwei Parallelklassen. Alle Kinder der Ausgangsstichprobe hatten bei Aufnahme in das Sonderpädagogische Bildungs- und Beratungszentrum einen diagnostizierten sonderpädagogischen Förderbedarf im Bereich Sprache mit der Diagnose *Sprachentwicklungsstörung* und einen IQ im durchschnittlichen Bereich. Die kognitiven Fähigkeiten der Kinder wurden bei Aufnahme in die Schule mit einem standardisierten Testverfahren (K-ABC; Melchers & Preuss, 2009) im Rahmen der Gutachtenerstellung erfasst (nonverbale Skala:  $M=99$ ;  $SD=9,7$ ). Die Ausgangsstichprobe reduzierte sich im Laufe der Untersuchung wegen zwei Studienausfällen und acht Studienausschlüssen (die Diagnose Sprachentwicklungsstörung war aufgrund anderer begleitend diagnostizierter Entwicklungsstörungen nicht mehr gültig) auf 33 Kinder (27,3 % weiblich) im Alter von 8;1 bis 10;6 Jahren ( $M=9;4$ ,  $SD=0;6$ ). Die Geschlechterverteilung entspricht in etwa dem statistischen Verhältnis zwischen Jungen und Mädchen, die eine Sprachentwicklungsstörung haben (de Langen-Müller et al., 2011). 14 Kinder der untersuchten Stichprobe sind einsprachig deutsch, 19 der Kinder haben einen Migrationshintergrund und sind mehrsprachig. In der Studie wurden bewusst auch mehrsprachige Kinder aufgenommen, um eine repräsentative Gruppe an Schulen für Sprachbehinderte bzw. Sonderpädagogischen Bildungs- und Beratungszentren mit Förderschwerpunkt Sprache in einem großstädtischen Einzugsgebiet abzubilden.

Bis auf drei Kinder zeigten alle Kinder entweder Auffälligkeiten im Weingartener Grundwortschatz Rechtschreib-Test (WRT) in der jeweiligen Variante der entsprechenden Klassenstufe (Prozentrang  $\leq 16$ ) und/oder auffällige Ergebnisse (Prozentrang  $\leq 16$ ) beim Lesetest ELFE 1-6 (für die Testverfahren s.u.). Drei Kinder, die Ergebnisse mit einem Prozentrang über 16 zeigten, wurden dennoch auf Wunsch der Lehrkräfte mit in die Untersuchung aufgenommen, da die gezeigten Leistungen tagesformabhängig seien.

Das Merkmal Aufmerksamkeit wurde erfasst, da Auffälligkeiten in der Aufmerksamkeitsentwicklung oder eine Aufmerksamkeitsstörung häufig als komorbide Begleitstörung bei Sprachentwicklungsstörungen auftritt (Mueller & Tomblin, 2012). Die Erfassung hinsichtlich des Merkmals Aufmerksamkeit erfolgte über eine Rückmeldung durch die Lehrkräfte. Bei acht Kindern liegt eine von einem Kinder- und Jugendpsychiater diagnostizierte Aufmerksamkeitsstörung vor. Bei neun weiteren Kindern hatten die Lehrkräfte den Eindruck, dass es Auffälligkeiten in der Aufmerksamkeitsentwicklung gibt bzw. möglicherweise eine Aufmerksamkeitsstörung vorliegt. Das Urteil von Lehrkräften hinsichtlich einer Aufmerksamkeitsstörung ist in der Regel aussagekräftig (Frölich, Döpfner & Banaschewski, 2014).



Bei der vorliegenden Untersuchung handelt es sich um eine quasi-experimentelle Untersuchung (Eid, Gollwitzer & Schmitt, 2013, S. 63f) in einer natürlichen Umgebung (Schule). Das Forschungsdesign ist eine Evaluationsstudie mit Vortest, Nachtest 1 und Nachtest 2 mit einer Wartekontrollgruppe. Aus organisatorischen Gründen war im vorgefundenen schulischen Alltag nur dieses Wartekontrollgruppendesign und keine aktive Kontrollgruppe möglich. Ein Nachteil des Wartekontrollgruppen-Designs ist eine zusätzliche Konfundierung, z. B. mit dem Faktor erhöhte Aufmerksamkeit (Bortz & Döring, 2006).

### 5.3 Angewandte Testverfahren und Erhebungsinstrumente

Da in der vorliegenden Studie Auswirkungen eines Lernprogramms auf phonologische und schriftsprachliche Leistungen untersucht werden sollen, wurden standardisierte Testverfahren zur Überprüfung phonologischer Leistungen (P-ITPA), zum Lesen (ELFE 1-6) und zum Rechtschreiben (WRT 2+, WRT 3+) verwendet. Alle genannten Verfahren wurden zu allen drei Testzeitpunkten (Vortest, Nachtest 1, Nachtest 2) eingesetzt. Zwei Verfahren (P-ITPA und ELFE 1-6) wurden bereits in vorherigen Evaluationsstudien zu *Lautarium* eingesetzt, so dass für diesen Bereich ein direkter Vergleich der Ergebnisse möglich ist.

Der P-ITPA (Esser et al., 2010) ist ein Verfahren für Kinder ab einem Alter von vier Jahren bis zum Ende der fünften Klasse mit umschriebenen Entwicklungsstörungen der Laut- und Schriftsprache. Bei der hier durchgeführten Untersuchung wurden die beiden Subtests „Vokale ersetzen“ und „Konsonanten auslassen“ aus dem Bereich *Phonologische Bewusstheit* in Einzeluntersuchungen durchgeführt. Bei beiden Subtestaufgaben steht die Manipulation von Lauten im Vordergrund, die der phonologischen Bewusstheit im engeren Sinne zugeordnet wird. Hinsichtlich der psychometrischen Gütekriterien zeigt der P-ITPA für die interne Konsistenz (Cronbachs Alpha) sowohl insgesamt als auch differenziert für die beiden durchgeführten Subtests hohe Werte (Vokale ersetzen: 0,94 und Konsonanten auslassen: 0,95). Die Objektivität ist bei Einhaltung der vorgegebenen Durchführungs- und Auswertungsanweisungen gewährleistet. Die Validität ist durch Vergleiche mit anderen als valide anerkannten Tests ( $r = 0,37$  bis  $r = 0,79$ ) und durch weitere externe kriterienbezogene Einschätzungen (z. B. Lehrerurteile  $r = 0,75$ ) belegt.

Der ELFE 1-6 (Lenhard, Schneider & Schneider, 2006) ist ein Leseverständnistest für Erst- bis Sechstklässler. Das Leseverständnis wird auf den Ebenen Wortverständnis (Dekodieren, Synthese), Satzverständnis (sinnentnehmendes Lesen, syntaktische Fähigkeiten) und Textverständnis (Auffinden von Informationen, satzübergreifendes Lesen, schlussfolgerndes Denken) erfasst. Für jeden Untertest gibt es ein Zeitlimit. Die interne Konsistenz (Cronbachs Alpha) ist mit 0,92 bis 0,97 für die drei Untertests als hoch anzusehen. Die Validität ist durch hohe Übereinstimmung mit einem anderen standardisierten Verfahren (Würzburger Leise Leseprobe/WLLP) belegt. Bei der vorliegenden Untersuchung wurde die Papier-Version als Gruppentest verwendet.

Im Bereich des Rechtschreibens wurde der Weingartener Grundwortschatz Rechtschreib-Test (WRT) als Gruppentest durchgeführt, um eine Aussage zu den Rechtschreibleistungen in Bezug zu der jeweiligen Klassennorm zu erhalten. Für die Klassenstufe drei kam der WRT 2+ (Birkel, 2007b) und für die Klassenstufe vier die Kurzform des WRT 3+ (Birkel, 2007a) zum Einsatz. Die Objektivität des WRT ist aufgrund vorgegebener Instruktionen und differenzierter Hinweise für die Auswertung und die Interpretation der Ergebnisse gewährleistet. Die interne Konsistenz (Cronbachs Alpha) ist beim WRT 2+ mit 0,94 als hoch anzusehen. Beim WRT 3+ kann die interne Konsistenz für die verwendete Kurzform mit einem geringeren Wert von 0,85 als hinreichend betrachtet werden. Die Validität ist sowohl beim WRT 2+ ( $r = 0,84$ ) als auch beim WRT 3+ ( $r = 0,80$ ) durch eine hohe Übereinstimmung mit anderen standardisierten Verfahren gewährleistet.

Zusätzlich wurden im Rahmen der Evaluationsstudie Erfahrungswerte im Umgang der Kinder mit dem Programm gewonnen. Um neben Beobachtungen während des Trainings weitere Rückmeldungen zu bekommen, wurde nach dem Ende des Trainingszeitraums von allen Kindern der Trainingsgruppe eine individuelle Rückmeldung zum Lernprogramm eingeholt. Dabei sollten die Kinder sowohl eine Bewertung des Programms auf einer fünfstufigen Likert-Skala (in Form von Smileys) ankreuzen, als auch inhaltliche Rückmeldungen zum Programm geben. Um authentische Antworten der Kinder zu erhalten, die nicht von individuellen Anstrengungen bei Lese- oder Schreibantworten überdeckt werden sollten, wurde für alle übrigen Fragen kein Fragebogen zum Ankreuzen bzw. Schreiben eingesetzt, sondern die Fragen vorgelesen und die Antworten der Kinder dokumentiert.

#### 5.4 Bildung der Trainings- und der Kontrollgruppe

Eine komplett zufällige Zuweisung der Kinder in die Kontroll- und Trainingsgruppe war nicht möglich. Ein Teil der Kinder der Trainingsgruppe wurde in enger Absprache mit den jeweiligen Lehrkräften ausgewählt. Merkmale für die Auswahl dieser Kinder durch die Lehrkräfte in die Trainingsgruppe waren (a) ein sehr großer Förderbedarf im Bereich des Lesens und Schreibens (Testwertergebnisse beim Vortest und Einschätzung nach Leistungen im Fach Deutsch), und (b) keine zusätzliche außerschulische Förderung im Bereich des Lesens und Schreibens.

Um eine hohe interne Validität zu erreichen, wurde versucht, mögliche Störvariablen im Vorfeld zu identifizieren und statistisch zu kontrollieren. Als Wartekontrollgruppe fungierten die Kinder der jeweiligen Klassenstufe, die nicht am Trainingsprogramm teilnahmen.



Abb. 1: Versuchsplan

#### 5.5 Ablauf der Studie

Die Fördereinheiten mit *Lautarium* fanden parallel zum regulären Unterricht statt (vgl. Abb. 1). Für die Förderung standen in einem Fachraum mit Einzelplatztischen insgesamt sechs Laptops und jeweils Kopfhörer zur Verfügung. Für einen Arbeitsplatz in einem sich räumlich direkt anschließenden kleineren Raum gab es zusätzlich auch Lautsprecher. Dieser Raum wurde genutzt, wenn einzelne Kinder individuelle Unterstützung benötigten. Die Trainingsgruppe wurde aufgrund der schulischen Rahmenbedingungen (z. B. begrenzte Anzahl an der in der Schule zur Verfügung stehenden Laptops, Bedürfnisse der unterschiedlichen Klassen aufgrund des Stundenplans etc.) in drei Gruppen aufgeteilt. Eine Fördergruppe mit sechs Kindern aus der Klassenstufe drei, eine Fördergruppe mit sechs Kindern aus der Klassenstufe vier und eine gemischte Fördergruppe mit vier Kindern aus der Klassenstufe drei und einem Kind aus der Klassenstufe vier.

Nach der Einführung zu Beginn des Trainings kamen die Kinder immer selbstständig und meist sehr motiviert zur verabredeten Zeit in den vorbereiteten Raum. Die Kinder hatten feste Plätze, sie waren schnell im Umgang mit dem Computerprogramm vertraut und konnten so bereits nach kurzer Zeit zügig mit ihren jeweiligen Übungen beginnen. Die Dauer der Förderung pro Einheit war auf max. 20 Minuten begrenzt. Die Fördereinheiten wurden von der Untersuchungsleiterin betreut.

In einer bereits durchgeführten Evaluationsstudie bei Drittklässlern mit LRS (Klatte et al., 2016) arbeiteten die Kinder acht Wochen mit dem Programm, entsprechend ca. 40 Fördereinheiten pro Kind. In der hier untersuchten Stichprobe gab es aufgrund von schulischen Veranstaltungen mehrere Tage, an denen keine Förderung stattfinden konnte. Aus diesem Grund wurde der Trainingszeitraum auf ca. neun Schulwochen ausgedehnt, so dass die meisten der Kinder der Trainingsgruppe am Ende des Untersuchungszeitraums ebenfalls ca. 40x am Programm teilgenommen hatten.

Im Anschluss an den Nachtest 2 erhielten die Kinder der Wartekontrollgruppe in Absprache mit den jeweiligen Klassenlehrerinnen die Möglichkeit am Förderprogramm *Lautarium* teilzunehmen. Die Kinder der Trainingsgruppe, die das Programm im Laufe der Trainingsphase noch nicht beenden hatten, konnten nach dem Nachtest 2 weiter am Förderprogramm teilnehmen und es beenden.

## 5.6 Statistische Analysen

Die Datenanalyse erfolgte mit Hilfe der Software IBM SPSS Statistics 24. Da es bei dieser Untersuchung nicht um eine Ergebnisinterpretation hinsichtlich Gesamtpopulationen geht, sondern um die Ermittlung von Lernzuwächsen, wurden die Rohwertdaten der durchgeführten Testverfahren verwendet. Zur Untersuchung, ob es kurz- oder längerfristige Trainingseffekte gab, erfolgten Kovarianzanalysen mit den Vortestergebnissen als Kovariate (ANCOVA). Dieses statistische Vorgehen eignet sich für experimentelle Untersuchungsdesigns mit Vortest, Nachtest 1 und 2, die Hypothesen über Interventionseffekte überprüfen (Rausch, Maxwell & Kelley, 2003; Read, Kendall, Carper & Rausch, 2013). Es wurde geprüft, ob sich (a) die Gruppenleistungen vom Vortest zum Nachtest verändert haben, und ob sich (b) die Leistungen der Gruppen im Vergleich von Vortest zu Nachtest 2 unterschieden. Per Konvention wird ein Signifikanzniveau von  $p < 0,05$  verwendet. Als weitere statistische Analyse wurden zur Vergleichbarkeit mit Vorläuferstudien zu *Lautarium* die Differenzwerte der Nachtests zum Vortest (Gain-Scores) zwischen der Trainings- und Kontrollgruppe mittels t-Tests für unabhängige Stichproben auf signifikante Unterschiede geprüft. t-Tests sind robust gegen Verletzungen der Voraussetzungen (Rasch et al., 2010) und gegen Messfehler, während bei der Kovarianzanalyse Messfehler der Kovariate (Vortestleistung) unberücksichtigt bleiben (Eid et al., 2013, S. 670). Die Voraussetzungen für die Anwendbarkeit der Kovarianzanalyse (Rausch et al., 2003; Read et al., 2013) und des t-Tests (Rasch et al., 2010) wurden statistisch geprüft (Levene-Test zur Prüfung der Varianzhomogenität, Shapiro-Wilk-Test zur Prüfung der Normalverteilungsannahme).

Bei signifikanten Gruppenunterschieden ( $p < 0,05$ ) wurden für die Ergebnisse der Kovarianzanalyse die um die Vortestleistungen korrigierten Effektstärken  $d_{\text{kor}}$  (Klauer, 2001) berechnet bzw. für die t-Tests Cohen's  $d$  als Effektstärkemaß (Cohen, 1988). Werte um 0,2 können als kleine, Werte ab 0,5 als mittlere und Werte ab 0,8 als starke Effekte interpretiert werden (vgl. Cohen, 1988). Die Berechnungen von  $d_{\text{kor}}$  bzw. Cohen's  $d$  erfolgte mit den Online-Programmen von Psychometrica – Institut für psychologische Diagnostik (Lenhard & Lenhard, 2014).

Im Bereich Rechtschreiben wurden die Ergebnisse der Kinder der Klassen drei und vier getrennt voneinander betrachtet, da klassenstufenspezifische Aufgaben (WRT2+ bzw. WRT3+) angewandt wurden.

## 6 Ergebnisse

### 6.1 Vortestleistungen

Die Vortestleistungen der gesamten untersuchten Stichprobe (ohne Aufteilung in Trainings- und Kontrollgruppe) wurden bezüglich der möglichen Einflussfaktoren Klassenstufe, Mehrsprachigkeit und Aufmerksamkeit betrachtet (siehe Tabelle 1). Zudem wurde ein Vergleich der Trainings- und Kontrollgruppe durchgeführt (siehe Tabelle 2, Spalte Vortest).

Bezüglich der Klassenstufe gab es signifikante Unterschiede bei Leistungen in dem Subtest „Vokale ersetzen“ des P-ITPA (t-Test für unabhängige Stichproben,  $p < 0,05$ ). Daher wird die Klassenstufe als eine potenziell relevante Kovariate behandelt. Die Vortestleistungen der mehrsprachigen Kinder der Gesamtgruppe unterschieden sich in den beiden Subtests des P-ITPA und den Subtests des ELFE 1-6 nicht signifikant ( $p > 0,05$ ) von den Vortestleistungen der Kinder mit Deutsch als Muttersprache. Dies gilt auch bei Aufteilung nach Klassenstufen bis auf den Subtest „Konsonanten auslassen“ des P-ITPA in der Klassenstufe vier, der aufgrund der kleinen Anzahl von nur vier einsprachig deutschen Kindern in der Klassenstufe vier jedoch nicht sicher interpretiert werden kann. Auch bezüglich des Merkmals Aufmerksamkeit unterschieden sich die Vortestleistungen nicht signifikant ( $p > 0,05$ ). Zwischen der Trainingsgruppe und der Wartekontrollgruppe gab es zum Zeitpunkt des Vortests keine signifikanten Unterschiede ( $p < 0,05$ ) hinsichtlich des mittleren Alters (innerhalb der jeweiligen Klassenstufe) sowie der mittleren Leistungen, außer beim WRT 2+ in der Klassenstufe drei. Die Ergebnisse sind teilweise nur eingeschränkt interpretierbar, da in einzelnen Fällen die Varianzhomogenität oder die Normalverteilungsannahme verletzt waren.

Tab. 1: Vortestleistungen in Bezug auf mögliche Einflussfaktoren. Signifikante Unterschiede sind mit \* markiert

	Klassenstufe		Mehrsprachigkeit Klasse 3		Mehrsprachigkeit Klasse 4		Aufmerksamkeit Klasse 3		Aufmerksamkeit Klasse 4	
	Mittelwert (SD)		Mittelwert (SD)		Mittelwert (SD)		Mittelwert (SD)		Mittelwert (SD)	
	3 (n=21)	4 (n=12)	nein (n=10)	ja (n=11)	nein (n=4)	ja (n=8)	nein (n=9)	ja (n=12)	nein (n=7)	ja (n=5)
P-ITPA „Vokale ersetzen“	6,00* (3,89)	10,08* (3,78)	5,90 (4,20)	6,09 (3,78)	9,75 (4,11)	10,25 (3,88)	7,89 (4,65)	4,58 (2,58)	10,43 (4,39)	9,60 (3,13)
P-ITPA „Konsonanten auslassen“	8,90 (4,89)	12,08 (4,70)	9,60 (4,27)	8,27 (5,52)	7,75* (5,78)	14,25* (2,12)	10,89 (4,37)	7,42 (4,89)	13,00 (5,57)	10,80 (3,27)
ELFE WV „Wortverständnis“	24,29 (10,63)	30,08 (6,32)	24,80 (7,51)	23,82 (13,22)	32,50 (8,10)	28,88 (5,41)	22,56 (7,59)	25,58 (12,62)	31,14 (7,73)	28,25 (3,85)
ELFE SV „Satzverständnis“	8,19 (4,27)	11,08 (3,37)	9,00 (3,94)	7,45 (4,61)	12,50 (4,12)	10,38 (2,97)	7,89 (4,08)	8,42 (4,58)	11,14 (4,18)	11,00 (2,24)
WRT 2+	10,57 (7,67)	–	12,90 (9,30)	8,45 (5,43)	–	–	10,89 (5,93)	10,33 (9,02)	–	–
WRT 3+	–	1,83 (2,08)	–	–	2,50 (3,00)	1,50 (1,60)	–	–	2,14 (1,77)	1,40 (2,61)

## 6.2 Analyse des Trainingseffekts

Tabelle 2 enthält die Ergebnisse (Rohwerte) für die Teilstichproben im Zeitverlauf (Vortest, Nachtest 1, Nachtest 2). Erwartungsgemäß nehmen die Leistungen der Kinder vom Vortest bis zum Nachtest 2 zu. Es zeigt sich eine Tendenz zu höheren Leistungszuwächsen in der Trainingsgruppe im Vergleich zur Kontrollgruppe bezogen auf den Vortest in allen untersuchten Subtests bis auf den WRT 2+ im Nachtest 1 und den WRT 3+. Besonders ausgeprägt ist dies bei dem Subtest „Konsonanten auslassen“ des P-ITPA zwischen Vortest und Nachtest 1.

Aufgrund der zum Zeitpunkt des Vortests teilweise signifikanten Unterschiede der Leistungen zwischen der Klassenstufe drei und vier wurden die Testergebnisse auch getrennt nach Klassenstufen ausgewertet. Abbildung 2 zeigt die Ergebnisse für den Subtest „Konsonanten auslassen“ des P-ITPA. Abbildung 3 zeigt das im Trainingszeitraum erreichte Level (Nummer des letzten gespielten Spiels) und die mittlere Anzahl der Durchgänge je Level. Die Kinder der Klassenstufe vier haben erwartungsgemäß höhere Levels erreicht und tendenziell weniger Wiederholungen der Übungen benötigt.

Tab. 2: Ergebnisse für die Teilstichproben im Zeitverlauf (Trainingsgruppe N=16, Kontrollgruppe N=17), RW: Rohwerte, LZ: Leistungszuwachs im Vergleich zum Vortest, d: Cohen's d; alle Angaben auf zwei Nachkommastellen gerundet.

		Vortest	Nachtest 1			Nachtest 2		
		RW M (SD)	RW M (SD)	LZ M (SD)	d	RW M (SD)	LZ M (SD)	d
P-ITPA „Vokale ersetzen“	Training	7,44 (4,24)	11,00 (3,62)	3,65 (2,76)	0,43 (n.s.)	12,25 (3,66)	4,81 (2,32)	0,25 (n.s.)
	Kontrolle	7,53 (4,45)	9,88 (4,03)	2,35 (3,26)		11,59 (3,85)	4,06 (3,56)	
P-ITPA „Konsonanten auslassen“	Training	8,69 (5,43)	12,19 (4,31)	3,50 (2,13)	1,29	13,25 (4,93)	4,56 (3,08)	1,20
	Kontrolle	11,35 (4,32)	12,12 (4,77)	0,76 (2,11)		12,41 (5,23)	1,06 (2,77)	
ELFE „Wortverständnis“ (WV)	Training	23,44 (6,36)	31,06 (6,15)	7,63 (3,34)	0,49 (n.s.)	35,50 (7,96)	12,06 (5,20)	0,41 (n.s.)
	Kontrolle	29,18 (11,39)	33,41 (10,59)	4,24 (9,11)		39,12 (9,92)	9,94 (5,07)	
ELFE „Satzverständnis“ (SV)	Training	8,06 (4,19)	11,31 (4,01)	3,25 (2,60)	0,45 (n.s.)	13,63 (2,68)	5,56 (2,66)	0,41 (n.s.)
	Kontrolle	10,35 (3,94)	12,06 (4,91)	1,71 (4,00)		14,65 (4,97)	4,29 (3,46)	

		Vortest	Nachtest 1			Nachtest 2		
		RW M (SD)	RW M (SD)	LZ M (SD)	d	RW M (SD)	LZ M (SD)	d
WRT 2+	Training (N=9)	6,67 (4,90)	8,00 (5,52)	1,33 (1,94)	-0,21 (n.s.)	12,00 (7,04)	5,33 (3,08)	0,74 (n.s.)
	Kontrolle (N=12)	13,50 (8,23)	15,67 (8,85)	2,17 (5,25)		15,17 (9,23)	1,67 (6,20)	
WRT 3+	Training (N=7)	1,43 (2,30)	2,29 (2,69)	0,86 (1,57)	-0,98 (n.s.)	3,14 (2,73)	1,71 (2,50)	-0,88 (n.s.)
	Kontrolle (N=5)	2,40 (1,82)	5,00 (3,39)	2,60 (1,95)		5,80 (2,17)	3,40 (1,14)	

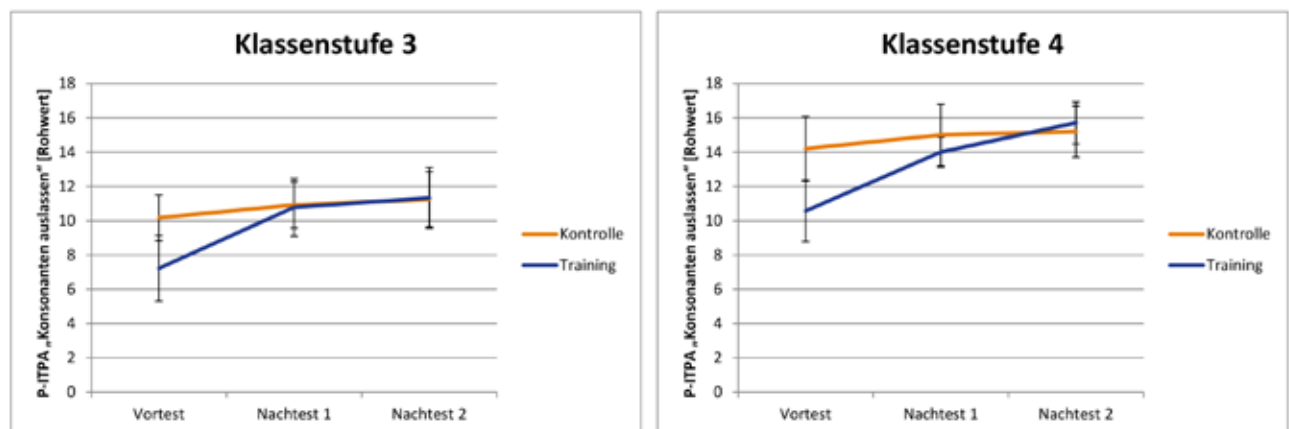


Abb. 2: Leistungen im P-ITPA „Konsonanten auslassen“ der Trainings- und Kontrollgruppe im Zeitverlauf getrennt nach Klassenstufen (Mittelwerte und Standardfehler).

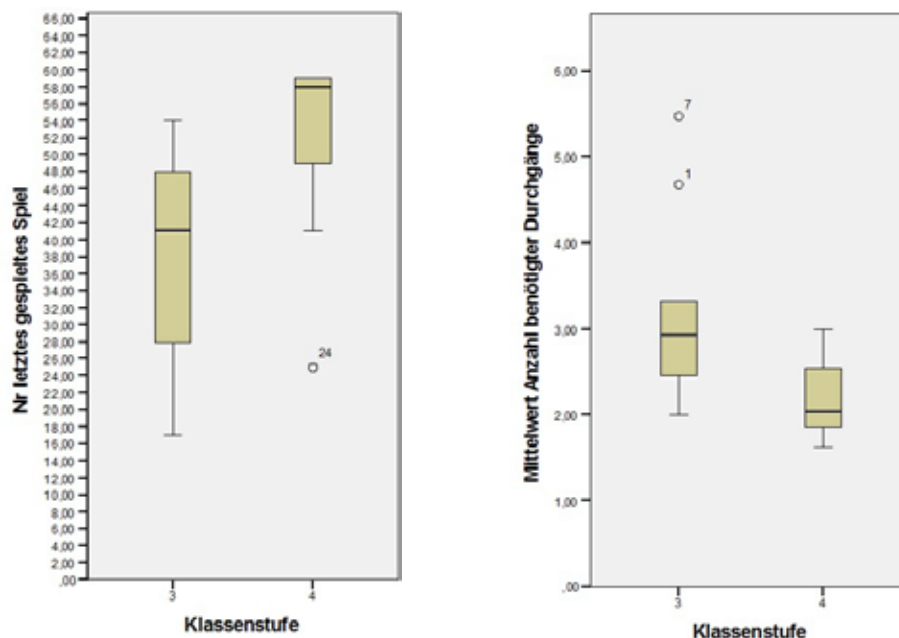


Abb. 3: Erreichtes Level des Trainingsprogramms und mittlere Anzahl der Durchgänge getrennt nach Klassenstufen

Die Kovarianzanalyse wurde ohne eine Aufteilung nach Klassenstufe durchgeführt, da davon ausgegangen wird, dass der Einfluss der Klassenstufe durch die Kovariate Vortestleistung bereits berücksichtigt ist. Für den Subtest „Konsonanten auslassen“ des P-ITPA lassen sich hochsignifikante Gruppenunterschiede ( $p < 0,01$ ) zu beiden Nachtest-Zeitpunkten mit mittlerer bis hoher



Effektstärke ( $d_{\text{kor}} = 0,56$  bzw.  $0,71$ ) nachweisen. Bei allen anderen Testverfahren sind die Unterschiede nicht signifikant.

Die Lernzuwächse (Gain-Scores) in Bezug auf den Vortest wurden, analog zur Kovarianzanalyse, ohne eine Aufteilung nach Klassenstufe durchgeführt, da wieder davon ausgegangen wird, dass der Einfluss der Klassenstufe durch die Subtraktion der Vortestleistung bereits berücksichtigt ist. Die Ergebnisse zeigen, wie bei der Kovarianzanalyse, für den Subtest „Konsonanten auslassen“ des P-ITPA hochsignifikante Gruppenunterschiede ( $p < 0,01$ ) zu beiden Nachtest-Zeitpunkten. Die Effektstärke nach Cohen ist mit  $d = 1,29$  bzw.  $1,20$  hoch. Im Mittel liegt der Lernzuwachs der Trainingsgruppe zum Zeitpunkt des Nachtest 1 um 2,73 Rohwert-Punkte höher als der Lernzuwachs der Kontrollgruppe. Zum Zeitpunkt des Nachtests 2 beträgt die Differenz 3,5 Rohwert-Punkte.

Bei allen anderen Leistungstests sind keine signifikanten Gruppenunterschiede nachweisbar. Tendenzen zu höheren Leistungszuwächsen der Trainingsgruppe im Vergleich zur Kontrollgruppe zeigten sich beim Subtest „Vokale ersetzen“ des P-ITPA, bei den Subtests „Wortverständnis“ und „Satzverständnis“ des ELFE 1-6 und beim Nachtest 2 des WRT2+.

Da sich die Vortestleistungen der Klassenstufen drei und vier teilweise signifikant unterschieden, wurde geprüft, ob das Merkmal Klassenstufe einen Einfluss auf die Lernzuwächse hat. t-Tests für unabhängige Stichproben zeigten in allen Fällen keine signifikanten Unterschiede in den Lernzuwächsen zwischen den Klassenstufen sowohl in der Trainings- als auch in der Kontrollgruppe ( $p > 0,05$ ).

Auch hinsichtlich des Merkmals Mehrsprachigkeit zeigten sich keine signifikanten Unterschiede in den Lernzuwächsen sowohl in der Trainings- als auch in der Kontrollgruppe ( $p > 0,05$ ).

Obwohl sich die Vortestleistungen bezüglich der Variable Aufmerksamkeit nicht signifikant unterschieden, wurde geprüft, ob das Merkmal Aufmerksamkeit einen Einfluss auf die Lernzuwächse hat, weil in der Praxis häufig ein solcher Zusammenhang vermutet wird. Hierzu wurden t-Tests für unabhängige Stichproben zum Vergleich der Mittelwerte sowohl von Kindern mit diagnostizierter Aufmerksamkeitsstörung zu den Kindern ohne diagnostizierte Aufmerksamkeitsstörung als auch von Kindern mit Auffälligkeiten in der Aufmerksamkeit (diagnostizierte oder vermutete Aufmerksamkeitsprobleme) zu den Kindern ohne Auffälligkeiten in der Aufmerksamkeit für alle Leistungstests durchgeführt. Es zeigten sich keine signifikanten Unterschiede in den Lernzuwächsen bis auf den Vergleich mit/ohne diagnostizierte Aufmerksamkeitsstörung bei der Trainingsgruppe beim Nachtest 1 des Subtests „Wortverständnis“ des ELFE 1-6.

Die Bewertungen des Programms durch die Kinder anhand der Rückmeldefragen waren durchweg sehr positiv. Bei der Likert-Skala kreuzten 15 Kinder das erste Smiley (beste Bewertung) an, ein Kind kreuzte das mittlere Smiley an. Bei den Rückmeldungen zu den Übungsaufgaben war zu beobachten, dass die Kinder ihre eigenen Schwierigkeiten (z. B. bei Odd-One-Out-Aufgaben) auch im Rückblick noch gut einschätzen konnten. Es scheint, dass das Arbeiten mit dem Programm bei einem Teil der Kinder dazu führte, dass die Selbstwahrnehmung und Selbsteinschätzung gestärkt wurde. Die Bewertung von *Lautarium* durch die Kinder deckte sich mit der Beobachtung der Untersuchungsleiterin. Die Kinder kamen fast durchgängig sehr motiviert zu den Fördereinheiten und zeigten während der Übungsphasen eine engagierte Lernhaltung.

## 7 Diskussion

Das Trainingsprogramm *Lautarium* wurde erstmals bei Kindern mit Sprachentwicklungsstörungen evaluiert. Zur Fragestellung, ob das Trainingsprogramm bereichsspezifisch auf Leistungen der phonologischen Bewusstheit bei Kindern mit Sprachentwicklungsstörungen wirksam ist, zeigten sich bei der Interventionsgruppe der dritten und vierten Klassen sowohl kurz- als auch längerfristig signifikante Leistungssteigerungen in einem Teilbereich (P-ITPA: Konsonanten auslassen). In einem zweiten überprüften Leistungsbereich (P-ITPA: Vokale ersetzen) konnten keine signifikanten Leistungssteigerungen beobachtet werden, eine Tendenz zu Leistungssteigerungen war jedoch erkennbar.

Ein Teil der Kinder der Interventionsgruppe hatte die Übungen zur Analyse der Vokallänge während des Trainingszeitraums noch nicht oder nicht vollständig durchgeführt. Damit lässt sich teilweise erklären, warum in diesem Bereich keine signifikanten Leistungssteigerungen beobachtbar waren. Eine weitere mögliche Erklärung könnte sein, dass Kinder mit Sprachentwicklungsstörungen auch im höheren Grundschulalter noch langanhaltende Probleme in ihren pho-



nologischen Fähigkeiten haben und daher mehr Übung benötigen (Klicpera, Graeven, Schabmann & Gasteiger-Klicpera, 1993). Bei einer Untersuchung von Kindern mit Sprachentwicklungsstörungen der zweiten bis vierten Klassenstufe hinsichtlich ihrer Lese- und Rechtschreibleistungen zeigte sich, dass Kinder mit Sprachentwicklungsstörungen auch zum Ende der Grundschulzeit noch große Probleme in ihren phonologischen Fähigkeiten hatten (Gasteiger-Klicpera & Klicpera, 2005). Hübner (2015) bestätigt diese Beobachtungen und kommt in ihren Untersuchungen von möglichen Wirkzusammenhänge zwischen laut- und schriftsprachlichen Leistungen bei Kindern mit multimodalen Sprachentwicklungsstörungen im höheren Grundschulalter (vierte Klasse) zu folgendem Schluss: „Auch im höheren Grundschulalter besteht offensichtlich bei einem nicht unerheblichen Teil von Kindern mit Sprachentwicklungsstörungen dringender Bedarf an spezifischer und vertiefter Förderung der phonologischen Fähigkeiten. Dies bezieht sich sowohl auf Intensität der Förderung als auch auf die zeitliche Ausdehnung der Förderangebote, denn Kompetenzen in der phonologischen Bewusstheit und Phonemdifferenzierung können nur erworben und dem Arbeitsspeicher zugeführt werden, wenn sie häufig aktiviert und geübt werden“ (Hübner, 2015, S. 436).

Zur Ermittlung, ob Transfereffekte auf nicht trainierte Bereiche beim Lesen und Schreiben beobachtbar waren, lassen sich folgende Ergebnisse zusammenfassen: Eine Tendenz zu positiven Trainingseffekten findet sich im Bereich Lesen zu beiden Nachtestzeitpunkten sowohl für den Subtest „Wortverständnis“ als auch für den Subtest „Satzverständnis“ des ELFE 1-6, im Bereich Schreiben für die Klassenstufe drei im Nachtest 2. Dies bestätigt in der Tendenz Ergebnisse der bisher durchgeführten Evaluationsstudien zu *Lautarium* (Klatte et al., 2016). Bei der Bewertung der Ergebnisse ist zu berücksichtigen, dass nur drei Kinder das Programm komplett durchlaufen hatten, die meisten Kinder hatten das Programm im untersuchten Zeitraum also noch nicht beendet. Ein Teil der Kinder hatte sogar weniger als die Hälfte der Übungslevels geschafft. Es ist daher zu vermuten, dass hierdurch die tatsächlichen Effekte unterschätzt werden. Womöglich wären bei vollständiger Durchführung des Trainings weitere Entwicklungsprozesse angestoßen worden, wie z. B. die Kenntnis der Lautstruktur und der Laut-Buchstabe-Zuordnung, die erst das Erlesen von unbekannten Wörtern ermöglichen und wie eine Art „Selbstlernmechanismus“ wirken (vgl. dazu Klatte et al., 2016). Die vorgefundenen Ergebnisse zum Lesen bestätigen einen Befund der Studie von Ritter und Kollegen (2013) bei Kindern mit Sprachentwicklungsstörungen der Klassenstufen eins bis drei, bei der die Auswirkungen eines Trainings zur phonologischen Bewusstheit verknüpft mit Phonem-Graphem-Zuordnungen auf die phonologische Bewusstheit, auf das Wortlesen und auf das Textverständnis untersucht wurden. Der Effekt hinsichtlich des Lesens war dort auf allen Klassenstufen schwächer als im Bereich der phonologischen Bewusstheit.

Neben phonologischen Fähigkeiten wirken sich allgemein auch linguistische Kompetenzen (z. B. Morphologie, Syntax, Semantik) auf das Leseverstehen aus (Ennemoser, Marx, Weber & Schneider, 2012). Die hier untersuchten Kinder mit Sprachentwicklungsstörungen hatten auf unterschiedlichen sprachlichen Ebenen Beeinträchtigungen. Hatz verweist in seiner Untersuchung darauf, dass sich Beeinträchtigungen beim Wortschatz und bei grammatischen Strukturen „auf die weiterführenden schriftsprachlichen Kompetenzen (Leseverständnis, schriftsprachlicher Ausdruck) nachteilig auswirken“ (Hatz, 2015, S. 74).

Möglicherweise wurden auch durch die gewählten Testverfahren zum Lesen und Schreiben Leistungssteigerungen nicht entsprechend erfasst. Beim ELFE 1-6 liegt das Hauptaugenmerk beim Leseverständnis. Der durchgeführte WRT 2+ beinhaltet Leistungsmessungen bzgl. des Schreibens lautgetreuer Wörter und des Anwendens erster orthografischer Regeln. Möglicherweise spiegeln sich Leistungsverbesserungen der Trainingsgruppe beim Schreiben lautgetreuer Wörter in den Testergebnissen nicht ausreichend wider. Beim WRT 3+ wird ausschließlich orthografisches Wissen geprüft, das durch das Förderprogramm nicht trainiert wird. Geeignete Verfahren zum Lesen und Schreiben könnten für diesen Zweck Subtests des P-ITPA (Esser et al., 2010) oder des Salzburger Lese- und Rechtschreibtests II (SLRT II) (Moll & Landerl, 2014) sein. Teils sind diese Verfahren aber nur als Individualtests möglich. Aus zeitökonomischen Gründen war dies im Rahmen der vorliegenden Untersuchung allerdings nicht leistbar.

Hinsichtlich der Frage, ob sich das Programm für den schulischen Einsatz unter den Rahmenbedingungen eines Förderzentrums eignet, lassen die praktischen Erfahrungen, die während der Studie gemachten Beobachtungen und die Rückmeldungen der Kinder anhand des Fragebogens den Schluss zu, dass es eine motivierende, ressourcenadäquate und individuelle Übungs- und Fördermöglichkeit für bestimmte Kinder mit Sprachentwicklungsstörungen im Rahmen eines schulischen Gesamtkonzepts der phonologischen und schriftsprachlichen Förderung sein kann.

Bei der Ergebnisinterpretation sollten mehrere methodische Einschränkungen berücksichtigt werden. Aufgrund der schulischen Rahmenbedingungen gab es für die Datenerhebung nur eine natürlich vorgefundene Gruppe mit einer kleinen Stichprobe ( $n = 33$ ). Der Nachteil eines quasi-experimentellen Untersuchungsdesigns ist die Beeinträchtigung der internen Validität durch systematische Störvariablen. Da eine Studie in der schulischen Praxis allerdings auch die Bedürfnisse vor Ort berücksichtigen muss, um Unterstützung und Akzeptanz der Beteiligten zu erhalten, war eine komplett zufällige Zuordnung von Kindern zur Trainings- und Kontrollgruppe nicht möglich, Einschränkungen in der Erfüllung der Voraussetzungen der statistischen Tests waren zu erwarten und in Kauf zu nehmen.

Hinsichtlich der eingesetzten Testverfahren könnten bei künftigen Studien Verfahren gewählt werden, die das Konstrukt der phonologischen Bewusstheit differenzierter erfassen (z. B. BAKO 1-4) als beim eingesetzten Verfahren der beiden Subtests des P-ITPA. Bei der durchgeführten Studie wurde nur das Leseverständnis überprüft. Eine Überprüfung der Lesegeschwindigkeit (z. B. mit der WLLP-R) könnte differenziertere Aussagen treffen.

## 8 Schlussfolgerungen

Die Ergebnisse der Studie deuten darauf hin, dass Kinder mit Sprachentwicklungsstörungen von der Förderung mit *Lautarium* in Teilbereichen der phonologischen Bewusstheit profitieren können. Da sich Tendenzen zu positiven Trainingseffekten auch beim lautgetreuen Lesen und Schreiben (im Nachtest 2 des WRT2+) beobachten lassen, sollten weitere Studien unter Berücksichtigung der oben genannten Aspekte durchgeführt werden.

Die Kinder mit Sprachentwicklungsstörungen der hier beschriebenen Trainingsstudie scheinen mehr Übungszeit benötigt zu haben als die Kinder mit LRS ohne Sprachentwicklungsstörungen (vgl. Ergebnisse zur Trainingszeit Klatte et al., 2017, S. 61f). Dies entspricht Befunden von anderen Studien bei Kindern mit Sprachentwicklungsstörungen (Al Otaiba et al., 2009). Bei künftigen Evaluationsstudien sollte das experimentelle Design so angelegt werden, dass die Trainingsgruppe das komplette Programm durchführt. So könnte überprüft werden, ob sich die gefundenen Tendenzen zu Trainingseffekten in den einzelnen Bereichen bestätigen lassen. Möglicherweise wirkt das Training in einem ersten Schritt auf phonologische Leistungen bei Kindern mit Sprachentwicklungsstörungen und zeigt dann erst in einem zweiten Schritt Auswirkungen auf schriftsprachliche Leistungen im Bereich des lautgetreuen Lesens und Schreibens.

Für künftige Studien könnte es interessant sein zu untersuchen, ob es möglicherweise Subgruppen von Kindern mit Sprachentwicklungsstörungen gibt, bei denen sich besonders stark ausgeprägte Trainingseffekte beobachten lassen. Dafür könnte eine differenziertere Analyse der jeweils vorliegenden Sprachentwicklungsstörung hilfreich sein. Schnitzler (2014) weist in ihrem Überblick beispielsweise auf einen Subtyp von Kindern hin, bei dem primär die phonologische Ebene betroffen ist. Cholewa (2010) fordert für die Evidenz der Wirksamkeit von Sprachtherapie eine Betrachtung von spezifischen Zielgruppen mit möglichst homogenen Gruppen, räumt aber auch ein, dass dies sehr aufwändig sei. Keine näheren Aussagen gibt es in der vorliegenden Studie z. B. zu individuellen Beeinträchtigungen in der Phonemwahrnehmung, im phonologischen Arbeitsgedächtnis und beim Abruf phonologischer Informationen aus dem Langzeitgedächtnis.

Eine genauere Analyse individueller Lernvoraussetzungen hinsichtlich der phonologischen Bewusstheit bzw. deren Einschränkungen könnte möglicherweise differenziertere Rückschlüsse auf festgestellte Fördereffekte zulassen.

Eine Förderung der phonologischen Bewusstheit ist nur dann sinnvoll, wenn tatsächlich Defizite in diesem Bereich vorliegen. Hinsichtlich der untersuchten Klassenstufen drei und vier könnte eine phonologisch orientierte Förderung auch um Ansätze ergänzt werden, die automatisierte Leseprozesse unterstützen, z. B. Maßnahmen, die auf die Automatisierung der Worterkennung abzielen. Mayer (2018) verweist auf den Einfluss der Benennungsgeschwindigkeit auf unterschiedliche Teilkompetenzen des Schriftspracherwerbs und betont als Konsequenz für die Prävention und Förderung, dass über die phonologisch orientierten Maßnahmen hinaus, auch automatisierte Leseprozesse in den Mittelpunkt gestellt werden sollten.

Ein zeitlich früherer Einsatz des Programms zur Prävention bei Risikokindern mit Sprachentwicklungsstörungen könnte dazu führen, dass Folgedefizite bei diesen Kindern abgemildert werden. Dazu werden derzeit erste Praxiserfahrungen mit Zweitklässlern gesammelt. Möglicherweise eignet sich das Förderprogramm auch als individuelle Differenzierungsmaßnahme für inklusive Settings.

*Lautarium* wurde als Trainingsprogramm für Grundschulkinder mit Lese-Rechtschreibschwierigkeiten konzipiert. Beim Einsatz des Programms bei Kindern mit Sprachentwicklungsstörungen ist sicherzustellen, dass die Kinder, die in Teilbereichen besonders gravierende Probleme haben, durch manche Aufgaben des Programms nicht überfordert werden. Diese Kinder benötigen qualifizierte Lehrkräfte, die gezielte und individuelle Unterstützung bieten können, um „mögliche ungünstige Nebenwirkungen“ positiver intendierter Maßnahmen (vgl. Brezing, 2000, S. 15) zu verhindern. In einer möglichen Weiterentwicklung des Programms mit technischen Modifizierungen einzelner Aufgabenformate für individualisierte Förderprofile könnten besondere Bedarfe für Kinder mit Sprachenentwicklungsstörungen Berücksichtigung finden.

Hartmann (2009) weist darauf hin, dass für eine fundierte Auswahl, Implementation und Evaluation von computergestützten Maßnahmen diagnostisches und therapeutisches Fachwissen, Erfahrungen sowie Analyse- und Kritikfähigkeit nötig sind. Eine engere Zusammenarbeit zwischen Wissenschaft und Praxis könnte dabei zu an den schulischen Rahmenbedingungen orientierten Lösungen führen.

## Literatur

- Al Otaiba, S., Puranik, C. S., Ziolkowski, R. A. & Montgomery, T. M. (2009). Effectiveness of early phonological awareness interventions for students with speech or language impairments. *The Journal of Special Education*, 43(2), 107-128. <https://doi.org/10.1177/0022466908314869>
- Bird, J., Bishop, D. V. & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech, Language, and Hearing Research*, 38(2), 446-462.
- Birkel, P. (2007a). *WRT 3+ Weingartener Grundwortschatz Rechtschreib-Test für dritte und vierte Klassen*. Göttingen: Hogrefe, 2., neu normierte und vollständig überarbeitete Auflage.
- Birkel, P. (2007b). *WRT2+ Weingartener Grundwortschatz Rechtschreib-Test für zweite und dritte Klassen* (2., neu normierte und vollständig überarbeitete Auflage). Göttingen: Hogrefe.
- Bischof, J., Gratzka, V., Strehlow, U., Haffner, J., Parzer, P. & Resch, F. (2002). Reliabilität, Trainierbarkeit und Stabilität auditiv diskriminativer Leistungen bei zwei computergestützten Mess- und Trainingsverfahren. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 30(4), 261-270.
- Bishop, D. V. & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, 31(7), 1027-1050.
- Bishop, D. V. & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, 130(6), 858-886.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (4. überarb. Aufl.). Heidelberg: Springer.
- Bradley, L. & Bryant, P. E. (1983). Categorizing sounds and learning to read – a causal connection. *Nature*, 301(5899), 419-421. <https://doi.org/10.1038/301419a0>
- Brezing, H. (2000). Welche Bedürfnisse haben Anwender (innen), und wie werden sie in der Forschung abgedeckt? Die Bedeutung von Evaluationsstandards und von Effektivitätskriterien für die Praxis. *Evaluation psychologischer Interventionsmassnahmen. Standards und Kriterien: Ein Handbuch*. Huber, Bern, 8-18.
- Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91(3), 403-414.
- Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech, Language, and Hearing Research*, 36(5), 948-958.
- Catts, H. W., Adlof, S. M., Hogan, T. P. & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, 48(6), 1378-1396.
- Cholewa, J. (2010). Empirische Sprachheilpädagogik: Strategien der Sprachtherapieforschung bei Störungen der Sprachentwicklung. *Empirische Sonderpädagogik*, 2(3), 48-68.
- Cholewa, J., Nierhaus, I., Tiro, M. & Unger, J. (2013). Entwicklungsstörungen des sublexikalischen (alphabetischen) Schreibens bei deutschsprachigen Drittklässlern mit Sprachentwicklungsstörungen. Welchen Effekt hat ein Training der phonologischen Bewusstheit? *Empirische Sonderpädagogik*, 5(4), 315-342.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 edition). Hillsdale, N.J.: Routledge.
- Dannenbauer, F. M. (2002). Spezifische Sprachentwicklungsstörung im Jugendalter. *Die Sprachheilarbeit*, 47(1), 10-17.
- de Langen-Müller, U., Kauschke, C., Kiesel-Himmel, C., Neumann K. & Noterdaeme, M. (2011). *Diagnostik von Sprachentwicklungsstörungen (SES) unter Berücksichtigung umschriebener Sprachentwicklungsstörungen (USES)*. Abgerufen von <http://www.awmf.org/leitlinien/detail/ll/049-006.html> [18.04.2019]

- Dignath, C., Buettner, G. & Langfeldt, H.-P. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review*, 3(2), 101-129.
- Ecalte, J., Magnan, A., Bouchafa, H. & Gombert, J. E. (2009). Computer-based training with ortho-phonological units in dyslexic children: new investigations. *Dyslexia*, 15(3), 218-238. <https://doi.org/10.1002/dys.373>
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z. & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250-287.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden* (3., korrigierte Auflage). Weinheim, Basel: Beltz.
- Ennemoser, M., Marx, P., Weber, J. & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(2), 53-67. <https://doi.org/10.1026/0049-8637/a000057>
- Esser, G., Ballaschk, K., Hammill, D. D., Mather, N., Roberts, R. & Wyschkon, A. (2010). *Potsdam-Illinois Test für psycholinguistische Fähigkeiten: P-IPTA; deutsche Fassung des Illinois test of psycholinguistic abilities, (ITPA-3) von DD Hammill, N. Mather & R. Roberts; Manual*. Hogrefe.
- Fischer, M. Y. & Pfost, M. (2015). Wie effektiv sind Maßnahmen zur Förderung der phonologischen Bewusstheit? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(1), 35-51. <https://doi.org/10.1026/0049-8637/a000121>
- Frölich, J., Döpfner, M. & Banaschewski, T. (2014). *ADHS in Schule und Unterricht: Pädagogisch-didaktische Ansätze im Rahmen des multimodalen Behandlungskonzepts*. Stuttgart: Kohlhammer.
- Galuschka, K., Ise, E., Krick, K. & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PloS one*, 9(2), e89900. <https://doi.org/10.1371/journal.pone.0089900>
- Gasteiger-Klicpera, B. & Klicpera, C. (2005). Lese- und Rechtschreibschwierigkeiten bei sprachgestörten Kindern der 2. bis 4. Klassenstufe. In P. Arnoldy & B. Traub (Hrsg.). *Sprachentwicklungsstörungen früh erkennen und behandeln* (S. 77-95). Karlsruhe: von Loepfer Literaturverlag.
- Gillon, G. T. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech, and Hearing Services in Schools*, 31(2), 126-141.
- Gillon, G. T. (2002). Follow-up study investigating the benefits of phonological awareness intervention for children with spoken language impairment. *International Journal of Language & Communication Disorders*, 37(4), 381-400.
- Gonzalez, M. del R. O., Garcia Espinel, A. I. & Guzman Rosquete, R. (2002). Remedial interventions for children with reading disabilities: speech perception – An effective component in phonological training? *Journal of Learning Disabilities*, 35(4), 334-342.
- Hartmann, E. (2002). Phonologische Bewusstheit und vorschulische LRS-Prävention. *Leseforum* 11 (2002), 3-6.
- Hartmann, E. (2003). LRS-Prävention bei sprachentwicklungsgestörten Kindern durch vorschulische Förderung phonologischer Bewusstheit: Grundlagen, Forschungsbilanz und Perspektiven. *Mitsprache*, 35, 7-38.
- Hartmann, E. (2009). Erfolg versprechende Computerprogramme für Kinder und Jugendliche mit Lese-Rechtschreibschwierigkeiten: Ein Überblick. *Sprache · Stimme · Gehör*, 33(04), 203-211. <https://doi.org/10.1055/s-0029-1242739>
- Hartmann, E. & Studer, F. (2013). Wie effektiv sind metaphonologische Vorschultrainings zur LRS-Prävention bei Kindern mit lautsprachlichen Beeinträchtigungen? *Empirische Sonderpädagogik*, 1, 42-68.
- Hasselhorn, M. & Werner, I. (2000). Zur Bedeutung des phonologischen Arbeitsgedächtnisses für die Sprachentwicklung. In: H. Grimm (Hrsg.): *Sprachentwicklung. Enzyklopädie der Psychologie. Bd. 3*, (S. 363-378). Göttingen/Bern/Toronto/Seattle: Hogrefe.
- Hatcher, P. J., Hulme, C. & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65(1), 41-57.
- Hatz, H. (2015). *Phonologische Bewusstheit und Schriftspracherwerb: Auswirkungen eines Trainings phonologischer Bewusstheit und eines um Rechtschreibinhalte erweiterten Trainings im ersten Schuljahr auf den Erwerb des Lesens und Rechtschreibens bei Schülerinnen und Schülern mit gering ausgebildeten schriftsprachspezifischen Vorläuferfertigkeiten*. Pädagogische Hochschule, Heidelberg.
- Hönninger, A. (2017). *Eine schulische Evaluation des computerbasierten Trainingsprogramms „Lautarium“ bei Schülerinnen und Schülern mit Sprachentwicklungsstörung*. Unveröffentlichte Masterarbeit. Technische Universität Kaiserslautern.
- Hübner, K. (2015). *Schriftspracherwerb unter sprachlich erschwerten Bedingungen. Am Übergang von phonologischen zu orthographischen Rechtschreibleistungen. Eine Untersuchung zur sprachlichen Verarbeitung und zur Rechtschreibung bei Kindern mit Sprachentwicklungsstörungen im höheren Grundschulalter*. Frankfurt am Main: Peter Lang Verlag.

- Hund-Reid, C. & Schneider, P. (2013). Effectiveness of phonological awareness intervention for kindergarten children with language impairment. *Canadian Journal of Speech-Language Pathology & Audiology*, 37(1), 6-25.
- Ise, E., Engel, R. R. & Schulte-Körne, G. (2012). Was hilft bei der Lese-Rechtschreibstörung? Ergebnisse einer Metaanalyse zur Wirksamkeit deutschsprachiger Förderansätze. *Kindheit und Entwicklung*, 21(2), 122-136. <https://doi.org/10.1026/0942-5403/a000077>
- Janczyk, M., Schöler, H. & Grabowski, J. (2004). Arbeitsgedächtnis und Aufmerksamkeit bei Vorschulkindern mit gestörter und unauffälliger Sprachentwicklung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36(4), 200-206.
- Kamhi, A. G. & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders*, 51(4), 337-347.
- Kirby, J. R., Georgiou, G. K., Martinussen, R. & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, 45(3), 341-362.
- Klatte, M., Steinbrink, C., Bergström, K. & Lachmann, T. (2016). „Lautarium“ – Ein computerbasiertes Trainingsprogramm für Grundschul Kinder mit Lese-Rechtschreibschwierigkeiten. In W. Schneider & M. Hasselhorn (Hrsg.), *Förderprogramme für Vor- und Grundschule*. Bd. 14 (S. 115-142). Hogrefe Verlag.
- Klatte, M., Steinbrink, C., Bergström, K. & Lachmann, T. (2017). *Lautarium – Ein computerbasiertes Trainingsprogramm für Grundschul Kinder mit Lese-Rechtschreibschwierigkeiten*. Göttingen: Hogrefe.
- Klatte, M., Steinbrink, C., Pröhl, A., Estner, B., Christmann, C. & Lachmann, T. (2014). Effekte des computerbasierten Trainingsprogramms „Lautarium“ auf die phonologische Verarbeitung und die Lese-Rechtschreibleistungen bei Grundschulkindern. In G. Schulte-Körne (Hrsg.), *Legasthenie und Dyskalkulie – Neue Methoden zur Diagnostik und Förderung* (S. 127-144). Bochum: Winkler.
- Klauer, K. J. (2001). *Handbuch Kognitives Training* (2., überarb. u. erw. A.). Göttingen: Hogrefe-Verlag.
- Klicpera, C. & Gasteiger-Klicpera, B. (1998). *Psychologie der Lese- und Schreibschwierigkeiten: Entwicklung, Ursachen, Förderung*. Weinheim: Beltz.
- Klicpera, C., Graeven, M., Schabmann, A. & Gasteiger-Klicpera, B. (1993). Wieweit haben sprachentwicklungsgestörte Kinder spezielle Probleme beim Lesen und Schreiben? Ein Vergleich mit guten und schwachen Lesern in der Grundschule sowie lernbehinderten Kindern. *Die Sprachheilarbeit*, 38, 231-244.
- Landerl, K. (2003). Categorization of vowel length in German poor spellers: An orthographically relevant phonological distinction. *Applied Psycholinguistics*, 24(4), 523-538. <https://doi.org/10.1017/S0142716403000262>
- Lenhard, W. & Lenhard, A. (2014). Computation of Effect Sizes. *ResearchGate*. <https://doi.org/10.13140/RG.2.1.3478.4245>
- Lenhard, W., Schneider, W. & Schneider, W. (2006). *ELFE 1-6: ein Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe Göttingen.
- Lundberg, L., Olofsson, A. & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology*, 21(1), 159-173.
- Magnan, A., Ecalte, J., Veuillet, E. & Collet, L. (2004). The effects of an audio-visual training program in dyslexic children. *Dyslexia*, 10(2), 131-140. <https://doi.org/10.1002/dys.270>
- Magnusson, E. & Nauclér, K. (1993). The development of linguistic awareness in language-disordered children. *First language*, 13(37), 93-111.
- Marx, P., Weber, J. & Schneider, W. (2005). Phonologische Bewusstheit und ihre Förderung bei Kindern mit Störungen der Sprachentwicklung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(2), 80-90. <https://doi.org/10.1026/0049-8637.37.2.80>
- Mayer, A. (2013). *Gezielte Förderung bei Lese- und Rechtschreibstörungen*. München: Reinhardt.
- Mayer, A. (2018). Benennungsgeschwindigkeit und Lesen\* Rapid automatized Naming (RAN) and Reading. *Forschung Sprache*, 6(1), 20-41.
- McArthur, G. M., Ellis, D., Atkinson, C. M. & Coltheart, M. (2008). Auditory processing deficits in children with reading and language impairments: Can they (and should they) be treated? *Cognition*, 107(3), 946-977. <https://doi.org/10.1016/j.cognition.2007.12.005>
- McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M. & Mengler, E. D. (2000). On the “specifics” of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry*, 41(7), 869-874.
- McGee, R., Williams, S., Share, D. L., Anderson, J. & Silva, P. A. (1986). The relationship between specific reading retardation, general reading backwardness and behavioural problems in a large sample of Dunedin boys: A longitudinal study from five to eleven years. *Journal of Child Psychology and Psychiatry*, 27(5), 597-610.
- Melby-Lervåg, M. & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49(2), 270-291.
- Melchers, P. & Preuss, U. (2009). *Kaufman Assessment Battery for Children*. Frankfurt/Main. Germany: Pearson Assessment.
- Moll, K. & Landerl, K. (2014). *Lese- und Rechtschreibtest (SLRT-II). Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)*. (2. korrigierte Auflage). Bern: Huber.
- Mueller, K. L. & Tomblin, J. B. (2012). Examining the comorbidity of language disorders and ADHD. *Topics in Language Disorders*, 32(3), 228-246. <https://doi.org/10.1097/TLD.0b013e318262010d>

- Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004). The development of early literacy skills among children with speech difficulties: a test of the „critical age hypothesis“. *Journal of Speech, Language, and Hearing Research*, 47(2), 377–391. [https://doi.org/10.1044/1092-4388\(2004/031\)](https://doi.org/10.1044/1092-4388(2004/031))
- Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research*, 52(5), 1175–1188. [https://doi.org/10.1044/1092-4388\(2009/08-0024\)](https://doi.org/10.1044/1092-4388(2009/08-0024))
- Puranik, C. S., Petscher, Y., Al Otaiba, S., Catts, H. W., & Lonigan, C. J. (2008). Development of oral reading fluency in children with speech or language impairments: a growth curve analysis. *Journal of Learning Disabilities*, 41(6), 545–560. <https://doi.org/10.1177/0022219408317858>
- Raitano, N. A., Pennington, B. F., Tunick, R. A., Boada, R., & Shriberg, L. D. (2004). Pre-literacy skills of subgroups of children with speech sound disorders. *Journal of Child Psychology and Psychiatry*, 45(4), 821–835. <https://doi.org/10.1111/j.1469-7610.2004.00275.x>
- Ramus, F., Marshall, C. R., Rosen, S. & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: towards a multidimensional model. *Brain*, 136(2), 630–645. <http://dx.doi.org/10.1093/brain/aws356>
- Rasch, B., Fries, M., Hofmann, W. & Naumann, E. (2010). *Quantitative Methoden Band 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Berlin Heidelberg.
- Rausch, J. R., Maxwell, S. E. & Kelley, K. (2003). Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design. *Journal of Clinical Child and Adolescent Psychology*, 32(3), 467–486.
- Read, K. L., Kendall, P. C., Carper, M. M. & Rausch, J. R. (2013). Statistical methods for use in the analysis of randomized clinical trials utilizing a pretreatment, posttreatment, Follow-up (PPF) Paradigm. *The Oxford Handbook of Research Strategies for Clinical Psychology*, 253–261.
- Ritter, M. J., Park, J., Saxon, T. F. & Colson, K. A. (2013). A phonologically based intervention for school-age children with language impairment: Implications for reading achievement. *Journal of Literacy Research*, 45(4), 356–385. doi: 10.1177/1086296X13504156.
- Rossion, B. & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33(2), 217–236.
- Schnitzler, C. D. (2014). LRS bei Kindern mit überwundenen phonologischen Aussprachestörungen. *Logos*, 22(4), 254–264.
- Schöler, H. (2011). Prognose schriftsprachlicher Leistungen und Risiken im Vorschulalter am Beispiel des Heidelberger Auditiven Screening in der Einschulungsuntersuchung (HASE). In M. Hasselhorn & W. Schneider (Hrsg.). *Tests & Trends, NF Bd. 9 Frühprognose schulischer Kompetenzen* (S. 13–31). Göttingen: Hogrefe.
- Schöler, H. & Scheib, K. (2004). Desiderate und Thesen zur Diagnostik bei Sprachentwicklungsstörungen. *Sprache · Stimme · Gehör*, 28(01), 37–41.
- Schuchardt, K., Worgt, M. & Hasselhorn, M. (2012). Besonderheiten im Arbeitsgedächtnis bei Kindern mit Sprachauffälligkeiten. In M. Hasselhorn & C. Zoelch (Hrsg.). *Funktionsdiagnostik des Arbeitsgedächtnisses* (S. 77–93). Göttingen: Hogrefe.
- Segers, E. & Verhoeven, L. (2004). Computer-supported phonological awareness intervention for kindergarten children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, 35(3), 229–239.
- Silva, P. A., Williams, S. & McGee, R. (1987). A longitudinal study of children with developmental language delay at age three: Later intelligence, reading and behaviour problems. *Developmental Medicine & Child Neurology*, 29(5), 630–640. <https://doi.org/10.1111/j.1469-8749.1987.tb08505.x>
- Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215.
- Snowling, M. J. (2000). *Dyslexia*. Oxford: Blackwell publishing.
- Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, 21(10), 590–604. <https://doi.org/10.1177/002221948802101003>
- Steinbrink, C., Klatte, M. & Lachmann, T. (2014). Phonological, temporal and spectral processing in vowel length discrimination is impaired in German primary school children with developmental dyslexia. *Research in Developmental Disabilities*, 35(11), 3034–3045. <https://doi.org/10.1016/j.ridd.2014.07.049>
- Steinbrink, C. & Lachmann, T. (2014). *Lese-Rechtschreibstörung*. Berlin Heidelberg: Springer.
- Strehlow, U., Haffner, J., Bischof, J., Gratzka, V., Parzer, P. & Resch, F. (2006). Does successful training of temporal processing of sound and phoneme stimuli improve reading and spelling? *European Child & Adolescent Psychiatry*, 15(1), 19–29.
- Tallal, P. (2003). Language learning disabilities integrating research approaches. *Current Directions in Psychological Science*, 12(6), 206–211.
- Thomé, G. (2000). Linguistische und psycholinguistische Grundlagen der Orthografie: die Schrift und das Schreibenlernen. In R. Valtin (Hrsg.). *Rechtschreiben lernen in den Klassen 1–6. Grundlagen und didaktische Hilfen*. (S. 12–16). Frankfurt am Main: Grundschulverband.
- Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific studies of reading*, 1(3), 217–234.



- Van Kleeck, A., Gillam, R. B. & McFadden, T. U. (1998). A study of classroom-based phonological awareness training for preschoolers with speech and/or language disorders. *American Journal of Speech-Language Pathology*, 7(3), 65-76.
- Wagner, R. K. & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192-212.
- Warrick, N., Rubin, H. & Rowe-Walsh, S. (1993). Phoneme awareness in language-delayed children: Comparative studies and intervention. *Annals of Dyslexia*, 43(1), 153-173.

## Danksagung

Ich möchte mich an dieser Stelle ganz herzlich bei der Schulleitung, den beteiligten Lehrkräften und den Eltern für ihr Vertrauen und ihre Unterstützung bedanken. Den Kindern gebührt der größte Dank für ihre Mitwirkung an der Studie und ihr tägliches Üben mit *Lautarium*. Nur so wurde diese Arbeit möglich.

## Zur Autorin

Anita Hönninger ist stellvertretende Leiterin eines Sonderpädagogischen Bildungs- und Beratungszentrums mit den Förderschwerpunkten Hören und Sprache.

## Korrespondenzadresse

Anita Hönninger  
Hermann-Gutzmann-Schule  
Sonderpädagogisches Bildungs- und Beratungszentrum  
mit den Förderschwerpunkten Hören und Sprache  
Anemonenweg 4  
68305 Mannheim  
anita.hoenninger@hgs.ma.schule-bw.de



Materialien zur Therapie,  
Annette V. Fox-Boyer (Hrsg.),  
Version 1.0.2 (2019),  
Windows-Anwendung mit Manual  
(PDF, 44 Seiten), Grammatik- und  
Wortschatztest, Ergebnis-Protokoll und  
Auswertung: Art.-Nr. 5000, 33.784 KB  
Lizenz je Gerät EUR 98,50 [D]  
(unverbindliche Preisempfehlung)

**Manual (Auszug) und Demo zur  
Anwendung finden Sie auf der Artikel-  
Detailseite in unserem Online-Shop.**

## Digitalversion für Windows

### TROG-D – Test zur Überprüfung des Grammatikverständnisses

Der TROG-D liegt erstmalig auch in einer digitalen Anwendung für Windows vor (konzipiert für Windows 7, 8, 8.1). Die digitale Version vereint die inhaltliche Qualität der gedruckten Ausgabe des Grammatik- und Wortschatztests mit den Vorteilen digitaler Medien. Die Touchfunktion erlaubt es, den seit vielen Jahren in der Praxis erprobten Test noch einfacher und kindgerechter durchzuführen. Nach der Testdurchführung automatisiert erzeugte Protokollbögen, die alle relevanten Auswertungsdaten des Klienten enthalten, sparen dem Therapeuten wertvolle Zeit und ermöglichen ihm eine einfache und standardisierte Form der Auswertung und Dokumentation. Die Daten sind als PDF- oder CSV-Datei abrufbar. Das dazugehörige Manual beschreibt alle Funktionen zur Anwendung.

- Programm führt durch den Test
- Erfassung und Auswertung erfolgen automatisiert
- digitale Dokumentationen integrierbar in z.B. bestehende Praxis-Software
- einfache und kindgerechte Durchführung
- erprobt und bereits in der Anwendung



**Bestellung:** [www.skvshop.de](http://www.skvshop.de)  
**Fragen & Support:** Tel.: +49 (0) 6126 9320-26/17  
[support@schulz-kirchner.de](mailto:support@schulz-kirchner.de)





## „Halt!“ Ein Diagnoseverfahren zum Textverstehen bei siebenjährigen Kindern\*

## „Halt!“ A diagnostic tool to evaluate text comprehension in seven year old children

Corinne Weyda, Wilma Schönauer-Schneider

### Zusammenfassung

Hörverstehen von Texten ist eine wichtige Kompetenz für Lern- und Bildungserfolg, die Kindern mit Sprachverständnisstörungen oftmals Probleme bereitet und deshalb überprüft werden sollte. Skarakis-Doyle (2002) entwickelte für Kinder im Kindergartenalter eine mehrdimensionale Diagnostik, die neben Fragen zu einer Geschichte auch das Entdecken von Erwartungsverletzungen umfasst. Für das Schulalter fehlen hier jedoch insbesondere im deutschsprachigen Raum noch derartige diagnostische Instrumentarien. In der vorliegenden Studie wurde deshalb ein Diagnoseverfahren zum Hörtextverstehen im Grundschulalter entwickelt. Als Grundlage diente die Adaption der Geschichte *„Halt!“, rief der Bär* von Manfred Mai (1995). Nach dem Hören der Geschichte müssen Kinder Inhalts- und Inferenzfragen im „Multiple Choice-Format“ beantworten sowie in einer Version mit Inkonsistenzen diese Fehler entdecken und korrigieren.

Das Verfahren sowie weitere Vergleichstests, u. a. zum Satz- und Textverstehen wurden 2012 an 58 monolingual deutschsprachigen Kindern im Alter von 7;0–7;11 Jahren überprüft. 13 Kinder davon wiesen rezeptive Auffälligkeiten auf. Im Vordergrund stand die Frage, ob sich Kinder mit rezeptiven Auffälligkeiten von sprachlich unauffälligen Kindern bei dem vorliegenden Verfahren unterscheiden. Zusätzlich wurden die Items hinsichtlich Schwierigkeit, Trennschärfe und Reliabilität untersucht.

Die Ergebnisse zeigen, dass Kinder mit rezeptiven Auffälligkeiten sowohl beim Beantworten von Fragen als auch beim Erkennen und Korrigieren von Erwartungsverletzungen signifikant schlechter abschneiden als sprachlich unauffällige Kinder. Die Itemanalyse zeigte für die meisten Items einen ausreichenden Schwierigkeitsgrad bzw. Trennschärfe und eine zufriedenstellende Reliabilität. Insgesamt scheint somit das Verfahren geeignet, Kinder mit nicht ausreichendem Verstehen von Hörtexten zu erfassen. Da jedoch nur 13 Kinder mit rezeptiven Auffälligkeiten überprüft wurden, sind diese Ergebnisse in weiteren, größer angelegten Studien zu evaluieren.

### Schlüsselwörter

Sprachverstehen, Monitoring des Sprachverstehens, Textverstehen, Diagnostik, Sprachentwicklungsstörungen

### Abstract

Text comprehension is a crucial skill for successful learning and education. As children with receptive language disorders often struggle with understanding texts, it is important to assess listening comprehension skills. Skarakis-Doyle (2002) developed a multidimensional diagnostic instrument for children at the age of kindergarten, which consists of comprehension questions and the detection of expectancy violations. There are no corresponding diagnostic procedures for school age children in the German-speaking area. Therefore, in this study, a diagnostic instrument was developed in order to evaluate text comprehension for elementary school. It is based on an adaption of the story *„Halt!“, rief der Bär* by

\* Dieser Beitrag hat das Peer-Review-Verfahren durchlaufen.

Manfred Mai (1995). After listening to the story, children had to answer several multiple choice questions. Some of the questions asked about literal information, others required the listener to draw an inference. Additionally the children had to detect and correct expectancy violations. This procedure and other comparative tests were carried out with 58 monolingual German speaking children at the age of 7;0–7;11 years. 13 of these children showed receptive language disorders. The diagnostic procedure was evaluated regarding item analyses and the suitability of the two subtests.

Results demonstrated that children with receptive language disorders revealed significant poorer skills in both subtests. The item analyses for most of the items indicated sufficient difficulty and selectivity indices and good reliabilities. Overall, „Halt!“ seems to be appropriate to detect children with difficulties in text comprehension, but this result should be verified with a larger sample of children.

#### Keywords

language comprehension, comprehension monitoring, assessment, language impairment

## 1 Einleitung

Der Wissenserwerb im Kindesalter hängt stark vom Hörverständnis für Texte ab, da bereits im Kindergarten und dann besonders in der Grundschule viel Wissen in Erzählform vermittelt und daher über das Sprachverstehen angeeignet wird (Zollinger, 2010). In Längsschnittstudien konnte zudem gezeigt werden, dass die Kompetenzen im Hörverständnis für Geschichten und Schlussfolgern mit späteren Leistungen im Leseverstehen zusammenhängen (Kim, 2015; van den Broek, White, Kendeou & Carlson, 2009; Oakhill & Cain, 2007). Daher haben Textverständnisstörungen gravierende Auswirkungen auf die gesamte Entwicklung des Kindes. Defizite im Text- und Leseverstehen können bis ins Erwachsenenalter persistieren und stehen mit Problemen im sozialen Bereich im Zusammenhang (Rißling, Ronniger, Petermann & Melzer, 2016; Clegg, Hollis, Mawhood & Rutter, 2005). Die Diagnostik sprachlicher Leistungen konzentriert sich aber oft auf produktive Fähigkeiten und vernachlässigt rezeptive Leistungen, insbesondere die Erfassung des Textverstehens. Da das Sprachverständnis als Fähigkeit nicht direkt beobachtbar und nur durch sichtbar werdende Reaktionen auf Sprache zu beurteilen ist, sind Interpretationsfehler bei dessen Erfassung häufig (Hachul & Schönauer-Schneider, 2016). Die rezeptiven Leistungen von Kindern werden dadurch oftmals überschätzt und Störungen werden übersehen.

Im deutschen Sprachraum bestehen nur wenige Verfahren zur Erfassung des Hörtextverstehens im Grundschulalter, z. B. die Mäusegeschichte von Gebhard (2008) oder der Untertest Fragen zum Text des SET 5-10 (Petermann, Metz & Fröhlich, 2010), die wichtige Aspekte des Textverstehens wie das Monitoring des Sprachverstehens, Ziel-Plan-Strukturen oder Perspektivenübernahme vernachlässigen. Ziel ist es aus diesen Gründen, ein geeignetes Verfahren zur Erfassung des Textverstehens im Grundschulalter zu entwickeln, das zwischen auffälligem und nicht-auffälligem Textverstehen unterscheidet, und dieses an siebenjährigen Kindern zu überprüfen.

## 2 Theoretischer Hintergrund

### 2.1 Textverstehen

Das Verstehen von Texten stellt die komplexeste Form des Sprachverstehens dar und beinhaltet sowohl Aspekte des Wort-, als auch des Satzverstehens, geht aber noch weit darüber hinaus. Aus kognitionspsychologischer Perspektive werden beim Textverstehen im Sinne eines Konstruktions-Integrations-Modells (van Dijk & Kintsch, 1983) mentale Repräsentationen auf drei hierarchischen Ebenen angenommen, wobei die ersten beiden Ebenen textbasiert sind (Richter & Schnotz, 2018; Kintsch & Rawson, 2007).

Auf der ersten Ebene wird eine **Oberflächenstruktur** erstellt. Es werden Wörter und Phrasen, die im Text enthalten sind, v.a. linguistisch verarbeitet. Das Verstehen von Wörtern, wörtlichen Formulierungen und syntaktischen Strukturen liefert alle sprachlichen Details für die Textoberfläche. Die zweite Ebene bildet textbasierte Repräsentationen als **Textbasis** ab. Sie stellt den semantischen Gehalt bzw. die semantische Struktur eines Textes in Form von Propositionen dar. Darin sind Wortbedeutungen enthalten, die kombiniert und zu komplexen Netzwerken zusammengefügt werden, z. B. Prädikat-Argument-Strukturen (GRABEN [agent: HASE; objekt: LOCH]). Lediglich eine kleine Anzahl an Inferenzen ist enthalten, die für die Herstellung von lokaler Textkohärenz notwendig ist. Zum Beispiel werden pronominale Referenten mit der passen-

den Person, auf das sich das Pronomen bezieht, innerhalb und zwischen Sätzen verbunden (Snyder & Caccamise, 2010). Diese Textbasis auf Grundlage expliziter Informationen reicht jedoch für ein tieferes und umfassenderes Verstehen nicht aus. Auf einer dritten, nicht textbasierten Repräsentationsebene muss der Rezipient das im Text enthaltene Wissen mit sachbezogenem Weltwissen, eigenen Erfahrungen bzw. Erwartungen verknüpfen und ein **mentales Modell** (Johnson-Laird, 1983) oder **Situationsmodell** (van Dijk & Kintsch, 1983) konstruieren. Erst nach der Erstellung dieses Situationsmodells hat der Rezipient eine wirkliche Vorstellung des Textes konstruiert.

Dazu werden verschiedene zugrundeliegende kognitive Fähigkeiten bzw. Prozesse angenommen. Der Hörer muss u. a. implizite Bedeutungen ableiten oder fehlende Informationen ergänzen, denn in Texten sind Informationen oftmals nicht explizit formuliert oder fehlen. Beispielsweise müssen zwischen verschiedenen Propositionen neue Verknüpfungen erstellt werden oder müssen durch Weltwissen zusätzliche Informationen integriert werden. Diese Verknüpfungen bzw. Schlussfolgerungen werden auch als **Inferenzen** bezeichnet, die vorhandene Textinformationen anreichern, strukturieren oder verdichten (Richter & Schnotz, 2018). Man unterscheidet zwischen lokalen und globalen Inferenzen. Lokale Inferenzen sind textverbindend bzw. referentiell, z. B. Pronomen mit dem jeweiligen Nomen oder Ableiten von Wortbedeutungen aus dem Kontext, während globale Inferenzen wissensbasiert die Verknüpfung von Informationen im Text mit Erfahrungen oder Weltwissen des Hörers erfordern (Hall & Barnes, 2017). In der hier untersuchten Geschichte muss der Hörer z. B. eine kausale Inferenz ziehen, warum ein Fuchs, der einen Bären weglocken will, dem Bären von einem Honig in der Nähe erzählt.

Neben Inferenzen spielen Prozesse der Validierung von Textinformationen eine große Rolle. Hörer bzw. Leser nutzen ihr Vorwissen auch zur Überprüfung der verstandenen Inhalte im Hinblick auf Erwartungsverletzungen und mangelnde Plausibilität. Bei inkonsistenten Textinformationen oder nicht zutreffenden Erwartungen reagieren sie beispielsweise mit Stirnrunzeln bzw. „Hä?“ oder beim Lesen mit verlängerten Blickbewegungen und Lesezeiten (Richter & Schnotz, 2018; Skarakis-Doyle & Dempsey, 2008). Diese metakognitive Strategie wird auch Comprehension Monitoring oder **Monitoring des Sprachverstehens** genannt (Marks, Mayer & Schönauer-Schneider, 2018; Wassenburg, Beker, van den Broek & van der Schoot, 2015).

In vielen Texten ist es erforderlich, andere Perspektiven einzunehmen und sich in die Gedanken und Verhaltensweisen anderer hineinzuversetzen, die möglicherweise von der eigenen Sichtweise abweichen. Rezipienten müssen somit die Fähigkeit der **Theory of Mind (ToM)** nutzen. Insbesondere das Erkennen mentaler und emotionaler Zustände bestimmter Personen ist für das vertiefte Verstehen von Texten unerlässlich. Studien konnten belegen, dass Messungen der ToM für das Verstehen von Texten prädiktiv sind (Dore, Amendum, Golinkoff & Hirsh-Pasek, 2018; Kim, 2015).

Das Bilden von Inferenzen, das Monitoring des Sprachverstehens und auch ToM sind kognitiv anspruchsvolle Fähigkeiten, die an hierarchieniedrigere kognitive Kompetenzen wie dem phonologischen Arbeitsgedächtnis ansetzen. Es wird vermutet, dass das phonologische Arbeitsgedächtnis für die Konstruktion des Situationsmodells und die dafür erforderlichen Fähigkeiten grundlegend bedeutsam ist, indem Informationen aufrechterhalten und vernetzt werden, die für eine lokale und globale Kohärenzbildung erforderlich sind (u. a. Kim, 2015).

Somit erfordert das Textverstehen multiple Repräsentationen, die auf jeder Ebene unterschiedliche sprachliche und kognitive Prozesse erfordern. Dazu gehören u. a. Wortverstehen, syntaktisches Verstehen, phonologisches Arbeitsgedächtnis, Comprehension Monitoring, Inferenzen ziehen oder auch Theory of Mind (Kim, 2015). Für eine Überprüfung des Textverstehens wäre es deshalb hilfreich, nicht nur das erstellte Situationsmodell durch Fragen zum Text abzu prüfen, sondern auch Prozesse der Erstellung wie die Evaluation des Textes (Monitoring des Sprachverstehens) oder Fähigkeiten der Perspektivenübernahme/ToM in der Diagnostik mit zu berücksichtigen.

## 2.2 Methoden zur Erfassung des Textverstehens

Es gibt verschiedene Methoden, um das Textverstehen eines Kindes zu überprüfen. Im deutschsprachigen Raum dominiert vor allem die Methode der Fragenbeantwortung. Selten lassen sich die Methoden des (gemeinsamen) Nacherzählens und des Entdeckens von Erwartungsverletzungen finden (Hachul & Schönauer-Schneider, 2016). Im Folgenden wird im Hinblick auf das erarbeitete Verfahren auf die Methoden der Fragenbeantwortung und des Entdeckens von Erwartungsverletzungen eingegangen.



Bei den Verständnisfragen können offene Fragen (W-Fragen), „Fragen“ im Multiple-Choice-Format bzw. Mehrfachwahlaufgaben oder Entscheidungsfragen gestellt werden. Vorteile der Multiple Choice-Aufgaben sind ihre hohe Auswertungsobjektivität, die einfache ökonomische Auswertung und die geringen Anforderungen an die Sprachproduktion im Gegensatz zu den offenen Fragen. Nachteilig wirkt sich jedoch die höhere Ratewahrscheinlichkeit aus (Hachul & Schönauer-Schneider, 2016).

Bei der Fragestellung kann man zwischen faktischen Informationsfragen und schlussfolgernden bzw. Inferenzfragen unterscheiden. Informationsfragen können in der Regel mit einem Wort beantwortet werden und betreffen meist explizit im Text genannte Informationen (Gebhard, 2008). Bei den Inferenzfragen müssen Schlussfolgerungen gezogen werden, um verschiedene im Text genannte Informationen sinnvoll zu verknüpfen. Diese Frageart bezieht sich eher auf den Verstehensprozess und ist somit eine geeignete Methode, um kohärentes Verstehen bzw. Inferenzen und somit das Situationsmodell zu erfassen (Kim, 2016; Mathieu, 1995). In zahlreichen Studien mit sprachunauffälligen und sprachauffälligen Kindern konnte gezeigt werden, dass die sprachauffälligen Kinder meist größere Probleme mit Inferenzfragen aufweisen (u. a. Dodwell & Bavin, 2008, Karasinski & Ellis Weismer, 2010) und daher besondere Schwierigkeiten beim Verstehen von Texten haben.

Um die Fähigkeiten des Monitorings des Spracherstehens beim Verstehen von Texten zu erfassen, entwickelte Skarakis-Doyle (2002) vorrangig für Kinder im Kindergartenalter die Methode der *Expectancy Violation Detection Task*, d. h. das *Entdecken von Erwartungsverletzungen*. Hierbei wird die Leistung der Kinder überprüft, Unstimmigkeiten zwischen Elementen in der Geschichte und ihren Erwartungen zu entdecken und zu korrigieren. In bekannten und thematisch vertrauten Geschichten werden verschiedene inhaltliche Verletzungen, wie Zielunterbrechungen, prosodische Veränderungen oder Ersetzungen innerhalb der Geschichte vorgenommen (Skarakis-Doyle & Dempsey, 2008). Bei der Dokumentation der Reaktionen auf die Erwartungsverletzungen werden sowohl verbale als auch nonverbale Reaktionen, die innerhalb von fünf Sekunden oder am Ende der Phrase auftreten, erfasst und in die Auswertung einbezogen. Bei den verbalen Antworten wird zwischen unspezifischen Protesten („Nein“) und Protesten mit Korrekturen („Nein, das war ...“) unterschieden. Korrekturen beziehen sich auf die verletzten Textstellen und den richtigen Inhalt (Skarakis-Doyle, 2002, Skarakis-Doyle & Dempsey, 2008). Bei diesem Test handelt es sich um eine online-Methode zur Überprüfung der frühen Monitoring-Leistungen, da die Kinder während des Verstehensprozesses auf Unstimmigkeiten reagieren müssen. Dadurch sind sowohl die Anforderungen an die expressiven Leistungen als auch an den Arbeitsspeicher gering. Skarakis-Doyle, Dempsey und Lee (2008) berichten moderate Korrelationen der Leistungen im primären Verstehen und dem Entdecken von Erwartungsverletzungen. Jüngere Kinder im Alter von 30 bis 48 Monaten können Ersetzungen von Objekten und Personen besser entdecken als Ersetzungen von Handlungen (Skarakis-Doyle, 2002). Auch bei Heel und Janda (2011, 2014), die Kinder im Alter von vier bis sechs Jahren untersuchten, stellte die Ersetzung der Handlungen eines Protagonisten die schwierigste Erwartungsverletzung dar.

### 2.3 Konzeption des Verfahrens

Als Grundlage der Überprüfung dient die kurze Geschichte *„Halt!“, rief der Bär* von Manfred Mai (1995). Sie handelt von einem Bären, der den Eingang eines sonst rundherum eingezäunten Waldes bewacht. Er ist der Meinung, dass ihm der Wald gehöre und er daher kein Tier in den Wald lassen dürfe. Als Erstes kommt ein Hase und möchte in den Wald, was ihm der Bär verbietet. Der Hase hoppelt allerdings ein Stück weiter, gräbt eine Mulde unter den Zaun und erhält so unbemerkt Zutritt. Als Nächstes kommt ein Stinktier, das ebenfalls um Einlass bittet. Der Bär verweigert dies abermals und das Stinktier lässt daraufhin eine Stinkbombe los, sodass der Bär ohnmächtig wird und das Stinktier in den Wald gehen kann. Als der Bär aus seiner Ohnmacht erwacht, kann er sich an das Geschehene nicht erinnern und fühlt sich noch etwas schwach. Diesen Umstand nutzt ein Fuchs aus, der dem Bären rät, etwas Honig essen zu gehen. Dieser übergibt die Verantwortung den Wald zu bewachen dem Fuchs, der allerdings alle vorbeikommenden Tiere in den Wald lässt. Als der Bär gestärkt zurückkehrt, gewährt er dem Fuchs zur Belohnung Eintritt in den Wald. Der Bär bewacht den Wald weiter und wundert sich darüber, dass kein Tier mehr vorbeikommt.

Die Geschichte ist linear aufgebaut, folgt dem gängigen Geschichtenschema und enthält viele schlussfolgernde Elemente sowie komplexe Ziel-Plan-Strukturen mit drei Episoden und vier Protagonisten. Die drei Episoden bestehen jeweils aus einem verursachenden Geschehen, bei



dem die Tiere nacheinander in den Wald gehen wollen und der Bär ihnen das untersagt. Darauf folgt eine Aktion bzw. ein Lösungsversuch der Tiere, um in den Wald zu kommen. Diese Aktionen beinhalten jedes Mal eine Täuschung des Bären und gelingen, ohne dass der Bär dies bemerkt. Am Ende der Geschichte wird daher noch die Verwunderung als Reaktion des Bären beschrieben. Nach Mathieu (1998) kann ein Kind ab sechs Jahren zwei Ereignisse mit etwa fünf Protagonisten innerhalb einer Erzählung verstehen und Verknüpfungsfragen dazu beantworten. Für Grundschul Kinder sollte die Komplexität der Geschichte daher adäquat sein. Die Hauptfiguren weisen miteinander konkurrierende Zielstrukturen auf: Der Bär bewacht seinen Wald, damit niemand hineingehen kann. Alle anderen Tiere hingegen verfolgen das Ziel in den Wald zu kommen. Daher denkt sich jedes Tier einen anderen Plan aus, um Zutritt in den Wald zu erhalten. Der Bär erreicht sein Ziel nicht, da er nicht bemerkt, wie ihn die anderen Tiere täuschen. Der Hase, das Stinktier und der Fuchs erreichen jedoch ihr Ziel ohne dass der Bär die Täuschung bemerkt. Die Wahl der Geschichte hängt mit dieser recht komplexen Ziel-Plan-Struktur der Protagonisten zusammen, da in vielen Studien (u. a. Graesser, Olde & Klettke, 2002) die Bedeutsamkeit von Zielstrukturen in Geschichten für das Verständnis betont wurde. Nur wenn die Ziele eines Charakters entdeckt und die dazugehörigen Motive und Pläne durch Perspektivenübernahme bzw. ToM geschlussfolgert werden können, ist es möglich, dessen Handlungen zu interpretieren (Kim, 2015, 2016; Schank & Abelson, 1977).

Sprachlich ist die Geschichte im Präteritum erzählt, beinhaltet jedoch keine unbekannten oder niedrigfrequenten Vergangenheitsformen. Bezüglich des verwendeten Wortschatzes und der Grammatik wurden kleine Veränderungen an der Geschichte vorgenommen. Zudem wurde nach Behrens (2010) darauf geachtet, dass die Geschichte nicht größeren Schulgruppen bekannt ist und die angesprochenen Themen nicht bestimmte Gruppen, wie beispielsweise Jungen oder Mädchen, systematisch bevorzugen oder benachteiligen. Außerdem sind das Setting *Wald* und die vorkommenden Tiere wohl den meisten Vorschul- und Grundschulkindern bekannt.

### 2.3.1 Untertest Verständnisfragen

Dieser Untertest wurde in Anlehnung an die Mäusegeschichte von Gebhard (2008) konzipiert. Bei den Verständnisfragen wurde das Multiple Choice-Format mit drei Antwortmöglichkeiten gewählt, da die produktiven Sprachfähigkeiten weitgehend ausgeschlossen werden sollen (Weyda, 2012). Bei den Antwortmöglichkeiten wurde darauf geachtet, dass bei Unkenntnis der Geschichte alle drei Antworten gleich wahrscheinlich sind. Außerdem wurde ein ausbalanciertes Antwortformat verwendet, indem die Positionen der richtigen Antwortmöglichkeiten variiert wurden. Die 25 Verständnisfragen bestehen in relativ ausgewogenem Verhältnis aus elf Informationsfragen (z. B. Der Fuchs ließ ... a) kein Tier in den Wald; b) alle Tiere in den Wald; c) ein Tier in den Wald) und vierzehn schlussfolgernden Fragen (z. B. Das Stinktier wollte ... a) den Wald bewachen; b) in den Wald gehen; c) mit dem Bären spielen). Letztere setzen sich aus verschiedenen Arten der Inferenzen zusammen (referentielle Inferenzen, Verstehen der Ziel-Plan-Struktur bzw. der Täuschung, Perspektivenübernahme, kausale Inferenzen, Schlussfolgerungen zum übergeordneten Ziel). Die schlussfolgernden Fragen können in lokale und globale Kohärenzbildung eingeteilt werden. Dabei handelt es sich um neun schlussfolgernde Fragen, die die Herstellung von lokaler Kohärenz verlangen und fünf Inferenzfragen, die die globale Kohärenzbildung erfordern. Somit wurden zum einen hierarchieniedrige bzw. einfache Schlussfolgerungen, bei denen Informationen aus benachbarten Sätzen verknüpft werden müssen, miteinbezogen. Zum anderen wurden hierarchiehohe bzw. komplexe Inferenzen, bei denen Informationen aus mehreren Sätzen und mehreren Abschnitten mit Weltwissen zusammengeführt werden müssen, integriert.

### 2.3.2 Untertest Erwartungsverletzungen

Der Untertest zur Überprüfung der Fähigkeiten im Monitoring des Sprachverstehens wurde in Anlehnung an den Untertest *Expectancy Violation Detection Task* von Skarakis-Doyle (2002) konzipiert. Zu Beginn wurden zwei Übungsaufgaben eingefügt. Die in die Geschichte integrierten Erwartungsverletzungen bestehen aus drei Zielunterbrechungen, einer Änderung des Refrains mit Zielunterbrechung, vier Ersetzungen innerhalb der Geschichte und einer widersprüchlichen Ersetzung. Die Ersetzungen innerhalb der Geschichte im Verfahren *Halt* wurden in Anlehnung an Skarakis-Doyle (2002) aufgeteilt in Ersetzungen von Protagonisten, Objekten, Handlungen eines anderen Protagonisten und Handlungen desselben Protagonisten. Die Auftretenshäufigkeit der einzelnen Kategorien wurde nicht übernommen, da die vorgenommenen Ersetzungen der Geschichte angepasst wurden. Um den Schwierigkeitsgrad für das Alter der in diesem Verfahren

getesteten Kinder anzupassen, wurde nur eine Verletzung bezüglich des Protagonisten (z. B. „Das kann ich doch für dich tun, solange du weg bist“, bot **das Stinktier** an (anstatt: *der Fuchs*) und des Objekts (z. B. „Dagegen hilft am besten **eine Stinkbombe** (anstatt *Portion Honig*) in den Untertest einbezogen. Da dem Verstehen von Zielstrukturen bei der Rezeption von Geschichten ein besonderer Stellenwert zukommt, wurden vier Zielunterbrechungen integriert. Das Ausmaß, in dem ein Rezipient eine Geschichte versteht, hängt mit dem Verfolgen der Ziele und Pläne der Charaktere zusammen (Kim, 2015; Magliano, Taylor & Kim, 2005). Daher ist es besonders wichtig, dass Kinder überwachen, ob die Handlungen der Protagonisten deren Plänen folgen und in einer erfolgreichen Problemlösung resultieren. In den eingefügten Zielunterbrechungen verhalten sich die Protagonisten jedoch nicht ihrem Ziel entsprechend, sondern handeln entgegen ihres Plans (z. B. Der Fuchs legte sich vor den Eingang und ließ **kein Tier in den Wald** (anstatt *ließ jedoch alle Tiere in den Wald, die hineinwollten*).

Für eine möglichst hohe Durchführungs- und Auswertungsobjektivität wurden vor der Erprobung der Tests Instruktionen für die genaue Vorgehensweise sowie Auswertungsregeln ausgearbeitet. Bei den Verständnisfragen erhält das Kind für jede korrekt gelöste Frage einen Punkt, bei einer falschen Antwort sind null Punkte zu vergeben. Der Untertest Entdecken von Erwartungsverletzungen ist in der Auswertung zweigeteilt. Zuerst wird ein Punkt für die Entdeckung der Erwartungsverletzung (durch einen eindeutigen verbalen Kommentar des Kindes, z. B. „Nein!“ oder „Stimmt nicht!“) gegeben. Kann das Kind dem Inhalt weiter durch eine darauffolgende Spezifizierung des entdeckten Fehlers oder durch eine Korrektur widersprechen, werden zwei Punkte vergeben (z. B. „Nein, keine Stinkbombe, Honig!“). Eine nonverbale Reaktion wird zwar protokolliert, da sie für die qualitative Einschätzung von Bedeutung ist, wird aber nicht bewertet.

### 3 Fragestellungen

Im Rahmen der Studie wurde das Verfahren Halt zur Überprüfung des Textverstehens erprobt (Weyda, 2012). Dazu wurden die einzelnen Untertests und Items des zu überprüfenden Verfahrens auf ihre Eignung hin untersucht und die Ergebnisse in den Untertests von rezeptiv auffälligen Kinder mit denen unauffälliger Kinder verglichen, um mögliche Unterschiede feststellen zu können. Es wurden explizit rezeptiv auffällige Kinder als Subgruppe ausgewählt, da Kinder mit überwiegend expressiven Sprachentwicklungsstörungen nicht zwingend auch Auffälligkeiten im Hörtextverstehen aufweisen. Vorstudien mit ähnlich konzipierten Verfahren bei Vorschulkindern belegen bessere Leistungen beim Beantworten von Fragen und beim Entdecken von Erwartungsverletzungen bei Kindern mit expressiven Sprachentwicklungsstörungen (Schulze, 2010; Hagemester, 2011). Im Folgenden werden vorrangig Fragestellungen und Ergebnisse zum neu konzipierten Verfahren im Vergleich der beiden Gruppen berichtet, während genauere Analysen zu Zusammenhängen mit dem phonologischen Arbeitsgedächtnis und der Perspektivenübernahme hier nicht näher dargestellt werden.

#### ■ Fragestellung 1: Finden sich Zusammenhänge zwischen den Untertests von Halt?

Die beiden Untertests untersuchen verschiedene Aspekte des Textverstehens, die sich allerdings hinsichtlich ihrer Anforderungen auch überschneiden. Skarakis-Doyle und Dempsey (2008), Skarakis-Doyle und Kollegen (2008) sowie Heel und Janda (2011, 2014) fanden Zusammenhänge zwischen den Leistungen beim Beantworten von Verständnisfragen und Entdecken von Erwartungsverletzungen für jüngere Kinder. Daher wird untersucht, inwieweit die Leistungen in den beiden Untertests bei siebenjährigen Kindern zusammenhängen.

#### ■ Fragestellung 2: Finden sich Unterschiede zwischen den rezeptiv auffälligen und unauffälligen Kindern in den beiden Untertests von Halt?

Andere Geschichtentests fanden unterdurchschnittliche Leistungen von sprachgestörten Kindern beim Beantworten von Verständnisfragen (z. B. Gebhard, 2008). Weitere Studien (u. a. Skarakis-Doyle & Dempsey, 2008; Heel & Janda, 2011, 2014; Hagemester, 2011) konnten darüber hinaus zeigen, dass Vorschulkinder mit rezeptiven Sprachstörungen beim Entdecken von Erwartungsverletzungen schlechter abschnitten als Gleichaltrige mit normaler Sprachentwicklung bzw. auch Kinder mit expressiven Sprachentwicklungsstörungen.

■ **Fragestellung 3: Finden sich Unterschiede zwischen den rezeptiv auffälligen und unauffälligen Kindern beim Beantworten der verschiedenen Fragearten?**

Es ist von Interesse, ob sich Gruppenunterschiede hinsichtlich der Beantwortung sowohl von beiden Arten der Inferenz- als auch Informationsfragen zeigen. Nach dem Konstruktions-Integrations-Modell nach van Dijk und Kintsch (1983) wird angenommen, dass die für die Beantwortung globaler Inferenzfragen benötigten Verarbeitungsprozesse höher und komplexer einzustufen sind. Karasinski und Ellis Weismer (2010) fanden signifikant geringere Leistungen sprachauffälliger Jugendlicher bei der Bildung von lokaler und globaler Kohärenz im Vergleich zu sprachunauffälligen. Dies wird auch für die jüngeren Kinder dieser Studie untersucht.

■ **Fragestellung 4: Finden sich Unterschiede zwischen den rezeptiv auffälligen und unauffälligen Kindern in den Reaktionsmustern im Untertest *Entdecken von Erwartungsverletzungen*?**

Bei Skarakis-Doyle und Dempsey (2008) reagierten Kinder mit einer Sprachentwicklungsstörung im Gegensatz zu den elaborierten Antworten der unauffälligen Kinder hauptsächlich nonverbal. Daher ist von Interesse, ob auch rezeptiv auffällige Kinder im Alter von sieben Jahren ein anderes Muster hinsichtlich der Reaktionen auf die Erwartungsverletzungen zeigen als gleichaltrige Kinder mit altersgemäßen Fähigkeiten im Sprachverständnis.

■ **Fragestellung 5: Sind die Items der beiden Untertests hinsichtlich ihrer Schwierigkeit, Trennschärfe und Reliabilität geeignet?**

Im Rahmen der Itemanalyse werden Itemschwierigkeiten, Trennschärfen und Reliabilitätskoeffizienten analysiert.

## 4 Methode

### 4.1 Untersuchungsablauf und Beschreibung der Untersuchungsinstrumente

In der Untersuchung (Weyda, 2012) wurden fünf verschiedene Tests verwendet, die mit jedem Kind einzeln durchgeführt wurden. Das Verfahren *Halt* wurde zu Beginn jeder Testung durchgeführt. Um die teilnehmenden Kinder hinsichtlich ihrer rezeptiven Leistungen einschätzen und differenzieren zu können, wurden zwei Vergleichstests herangezogen. Ziel war es, in Kinder mit unauffälligem bzw. auffälligem Sprachverstehen einzuteilen. Hier wurden Verfahren zum grammatischen Verstehen und Textverstehen verwendet, um auffälliges Sprachverstehen auf mehreren Ebenen zu erfassen. Zum Wortverstehen waren zum Testzeitpunkt für die Altersgruppe keine ausreichend normierten Verfahren im deutschsprachigen Raum vorhanden. Der TROG-D (Test zur Überprüfung des Grammatikverständnisses) von Fox (2006) wurde zur Erfassung des Sprachverstehens auf Satzebene verwendet, da dieser etablierte und standardisierte Test Normwerte für den gewünschten Altersbereich zwischen 7;0 und 7;11 Jahre bietet und grammatisches Verstehen in enger Verbindung mit Textverstehen steht (u. a. Kim, 2015). Als zweites Verfahren wurde die Mäusegeschichte von Gebhard (2008) eingesetzt, um die Textverstehensleistung zu überprüfen. Dabei handelt es sich um einen informellen Test zum Hörverständnis für Geschichten bei Kindern zwischen 7 und 9 Jahren, bei dem das Textverstehen mit 12 Fragen im Multiple Choice-Format mit jeweils drei Antwortmöglichkeiten und einer zusätzlich offenen Frage erfasst wird. Das Verfahren beinhaltet drei schlussfolgernde Fragen (z. B. Gefährlich für das Mäuslein war in Wirklichkeit ... a) das erste Tier; b) das zweite Tier; c) gar kein Tier) und neun Informationsfragen (z. B. Das erste Tier war auf dem Kopf ... a) schneeweiß; b) mausgrau; c) feuerrot). Alle korrekt beantworteten Fragen werden jeweils mit einem Punkt bewertet, während die offene Frage der qualitativen Einschätzung dient. Der Test liefert ein quantitatives Maß, ob ein Kind zu den schwächsten 15 % seiner Altersgruppe gehört (Gebhard, 2008). Die Mäusegeschichte wurde als Vergleichstest herangezogen, da der erste Untertest im Verfahren *Halt* in Anlehnung an die Mäusegeschichte entwickelt wurde und beide Verfahren für das Grundschulalter konzipiert wurden.

Ferner wurden zwei weitere Fähigkeiten im Zusammenhang mit Textverstehen untersucht: das phonologische Arbeitsgedächtnis und ToM. Für die Erfassung des phonologischen Arbeitsgedächtnisses wurde der Mottier-Test, ein Bestandteil des Zürcher Lesetests (Linder & Grisse mann, 2000), zur Einschätzung der auditiven Merkspanne durch das Nachsprechen von Sinnlossilben eingesetzt. Ein Wert unter 12 Rohpunkten galt als auffällig. Da die Studie 2012 durchgeführt wurde, stand die 2. Auflage des Tests noch nicht zur Verfügung.

ToM wurde durch eine Aufgabe zur falschen Überzeugung überprüft, da die Geschichte *Halt* komplexe Ziel-Plan-Strukturen und Täuschungen beinhaltet. Dazu wurde die Testung zum falschen Inhalt nach Gopnik und Astington (1988) verwendet, um die Anforderungen an expressive und rezeptive Sprachfähigkeiten relativ gering zu halten. In der Untersuchung wurde den teilnehmenden Kindern eine Smarties-Schachtel präsentiert, in der sich entgegen der Erwartungen Stifte befanden. Es sollte nun bestimmt werden, welchen Inhalt ein Stofftier in der verschlossenen Schachtel vermutete und was sich tatsächlich darin verbarg („Was denkt die Ratte ist in der Schachtel?“ und „Was ist wirklich in der Schachtel?“). Konnte das Kind die zweiteilige Testfrage richtig beantworten, erhielt es einen Punkt. Bei inkorrekt beantworteter Frage wurde kein Punkt vergeben, so dass lediglich die Unterscheidung ein oder kein Punkt im Sinne einer vorhandenen bzw. nicht vorhandenen ToM möglich war.

Die Tests wurden in 35 bis 45 Minuten in der jeweiligen Schule bzw. der sprachtherapeutischen Praxis in den Regionen München bzw. Stuttgart in einem separaten Raum durchgeführt und anschließend anhand einer statistischen Analyse mit SPSS (SPSS Statistics 19) ausgewertet.

#### 4.2 Stichprobenbeschreibung

Die Stichprobe der Untersuchung setzte sich aus  $N = 58$  Probanden zusammen. Davon waren 22 (37,9 %) Kinder weiblich und 36 (62,1 %) männlich. Das Durchschnittsalter betrug 7;57 Jahre. Voraussetzungen für die Teilnahme war ein Alter von 7;0 bis 7;11 Jahre, eine altersgemäße allgemeine Entwicklung, unauffällige Hörfähigkeit, durchschnittliche kognitive Fähigkeiten und ein monolingual deutscher Spracherwerb. 31 Kinder besuchten eine Regelschule (53,4 %) und 27 Kinder eine Sprachheilschule (46,6 %). Der hohe Anteil von Kindern aus Sprachheilschulen sollte möglichst viele rezeptiv auffällige Kinder sicherstellen.

Für die Studie sollten die Kinder in eine Gruppe mit rezeptiv unauffälligen und rezeptiv auffälligen Kindern eingeteilt werden. Kinder wurden als rezeptiv auffällig eingestuft, wenn sie im TROG-D und/oder der Mäusegeschichte (MG) unterdurchschnittlich waren, d. h. bei einem T-Wert unter 40 beim TROG-D und bei einem Rohwert unter 7 bei der Mäusegeschichte. Da das grammatische Verstehen eine bedeutsame Rolle für das Textverstehen spielt, wurden auch Kinder, die nur im TROG-D auffielen, als rezeptiv auffällig eingestuft. Zudem wurde die Mäusegeschichte zwar an 117 sieben-, acht- und neunjährigen Kindern überprüft, die Untergruppe der Siebenjährigen ist jedoch nicht ausreichend untersucht, um lediglich von diesem Test auf rezeptiv auffällig zu schließen.

Die Testung mit den Vergleichstests ergab, dass 45 Kinder keine rezeptiven Defizite aufwiesen, während 13 Kinder durch unterdurchschnittliche Leistungen im TROG-D und/oder der Mäusegeschichte (MG) als rezeptiv auffällig eingestuft wurden (vgl. Tab. 1). In der Mäusegeschichte fielen insgesamt nur fünf Kinder rezeptiv auf, wobei davon 3 auch im TROG-D unterdurchschnittlich waren. Acht Kinder zeigten nur im TROG-D auffällige Ergebnisse. Somit stuft der TROG-D elf Kinder als rezeptiv auffällig ein, während die Mäusegeschichte nur fünf Kinder als rezeptiv auffällig erkannte. Für die weiteren Analysen wurden alle rezeptiv auffälligen Kinder als Untersuchungsgruppe ( $N = 13$ ) gewertet.

Tab. 1: Einteilung der Probanden in rezeptiv auffällig bzw. unauffällig

	unauffällig	nur Mäusegeschichte auffällig	nur TROG-D auffällig	Mäusegeschichte u. TROG-D auffällig
Anzahl der Kinder	45	2	8	3

## 5 Ergebnisse

Der Schwerpunkt der Ergebnisdarstellung liegt auf der Beantwortung der Fragestellungen im Hinblick auf die Halt-Geschichte im Vergleich der beiden Untersuchungsgruppen (rezeptiv auffällig und rezeptiv unauffällig). Zur Vollständigkeit sind in der folgenden Tabelle (vgl. Tab. 2) jedoch auch die Ergebnisse in den Vergleichstests TROG-D bzw. Mäusegeschichte und den Zusatztests Mottier-Test bzw. Aufgabe zur Perspektivenübernahme/ToM (falscher Inhalt) dargestellt. Dabei wurden beim TROG-D die Rohwerte verwendet, da einige Kinder so gut abschnitten, dass ihnen kein T-Wert zugewiesen werden konnte.

Tab. 2: Deskriptive Statistik der Untersuchungsgruppen in den Vergleichs- und Zusatztestverfahren

Test	TROG-D		Mäusegeschichte		Mottier		Falsch. Inhalt (ToM)	
rezeptive Auffälligkeit	unauff.	auffällig	unauff.	auffällig	unauff.	auffällig	unauff.	auffällig
<b>M</b>	16,31	8,92	10,69	7,62	16,33	12,08	0,89	0,38
<b>SD</b>	2,37	2,75	1,59	2,06	4,49	2,78	0,32	0,51
<b>Max</b>	21	14	12	11	26	18	1	1
<b>Min</b>	12	5	7	4	9	8	0	0
<b>Anzahl (%) auffälliger Kinder</b>	s.o.	s.o.	s.o.	s.o.	6 (13,3 %)	6 (46,2 %)	8 (11,1 %)	5 (61,5 %)

Die rezeptiv auffälligen Kinder sind erwartungsgemäß im TROG-D und in der Mäusegeschichte deutlich schwächer als die Kontrollgruppe, wobei in der Mäusegeschichte auch einige rezeptiv auffällige Kinder höhere Werte erreichten, was der Subgruppeneinteilung (s. 4.2) geschuldet ist.

Im phonologischen Arbeitsgedächtnis liegen der Mittelwert und der Maximalwert der rezeptiv auffälligen Kinder deutlich unter den Werten der Vergleichsgruppe. Vergleicht man den prozentualen Anteil auffälliger Kinder in den Untersuchungsgruppen, so weisen 46,2 % der rezeptiv auffälligen Kinder und nur 13,3 % der rezeptiv unauffälligen Kinder unterdurchschnittliche Leistungen im auditiven Kurzzeitgedächtnis auf.

Bei der Aufgabe zur ToM (Erkennen des falschen Inhalts) schneiden die rezeptiv auffälligen Kinder im Mittelwert deutlich schlechter ab als die unauffälligen Kinder, wobei die Einteilung in 0 oder 1 Punkt sehr wenig differenziert. Betrachtet man den prozentualen Anteil auffälliger Kinder im Bereich ToM, so haben 61,5 % der rezeptiv auffälligen Kinder Schwierigkeiten bei der Perspektivenübernahme, jedoch nur 11,1 % der unauffälligen. Diese Ergebnisse sprechen für größere Schwierigkeiten rezeptiv auffälliger Kinder im phonologischen Arbeitsgedächtnis und vor allem in der ToM, die das Textverstehen negativ beeinflussen könnten. Diese wichtigen Aspekte werden hier jedoch nicht weiter vertieft.

Im Folgenden sind die Ergebnisse in Bezug auf die Fragestellungen zur Halt-Geschichte beschrieben (vgl. 3). Die rezeptiv auffälligen Kinder schneiden in beiden Untertests im Mittelwert deutlich schlechter ab als die rezeptiv unauffälligen Kinder (vgl. Tab. 3). Dabei zeigt die Standardabweichung jeweils eine größere Streuung bei den rezeptiv auffälligen Kindern.

Die rezeptiv unauffälligen Kinder konnten bei den Verständnisfragen mindestens 14 Fragen korrekt beantworten, die rezeptiv auffälligen zeigten einen Minimalwert von 5, so dass insbesondere die schwächeren der jeweiligen Gruppe sich deutlich unterscheiden. Allerdings liegt der Maximalwert bei beiden Gruppen bei der Gesamtzahl von 25, so dass auch in der rezeptiv auffälligen Gruppe die Beantwortung altersgemäß gelang. Die Mittelwerte belegen jedoch deutliche Unterschiede in der Anzahl der korrekt beantworteten Fragen.

Beim Erkennen der Erwartungsverletzungen liegt der Mittelwert der rezeptiv auffälligen Kinder wiederum erkennbar unter dem der Vergleichsgruppe. Während die rezeptiv unauffälligen Kinder den höchstmöglichen Wert von 18 erreichen, gelingt dies in der rezeptiv auffälligen Gruppe nicht. Der Minimalwert von 0 in der rezeptiv auffälligen Gruppe belegt zudem, dass hier teilweise keine einzige Erwartungsverletzung erkannt wurde. Der hohe Anteil an expressiven Auffälligkeiten in der Gruppe der rezeptiv unauffälligen Kinder kann möglicherweise die extrem niedrigen Werte der rezeptiv unauffälligen Kinder beim Entdecken von Erwartungsverletzungen erklären.

Tab. 3: Ergebnisse der Untersuchungsgruppen in den Untertests der Halt-Geschichte

Untertest Halt	Verständnisfragen		Erwartungsverletzungen	
rezeptive Auffälligkeit	unauffällig	auffällig	unauffällig	auffällig
<b>M</b>	22	15,54	15,64	9,15
<b>SD</b>	2,71	5,74	3,59	5,79
<b>Max</b>	25	25	18	16
<b>Min</b>	14	5	3	0



Genauere Analysen zu den Subgruppen finden sich nun in der Beantwortung der Fragestellungen. Da für die rezeptiv unauffälligen Kinder in beiden Untertests keine Normalverteilung der Messwerte angenommen werden kann und die Gruppengröße der rezeptiv auffälligen Kinder mit  $N=13$  sehr gering ist, werden zur inferenzstatistischen Überprüfung der Fragestellungen nicht-parametrische Verfahren verwendet.

#### ■ Fragestellung 1: Zusammenhänge zwischen den Untertests von *Halt*

Um die Zusammenhänge der Leistungen aller getesteten Kinder in den beiden Untertests von *Halt* überprüfen zu können, wurden die prozentualen Anteile der richtig gelösten Aufgaben in den beiden Untertests miteinander korreliert. Dazu wurde der Rangkorrelationskoeffizient nach Spearman verwendet. Im folgenden Streudiagramm (vgl. Abb. 1) werden die Ergebnisse aller Kinder zwischen den beiden Untertests von *Halt* verglichen.

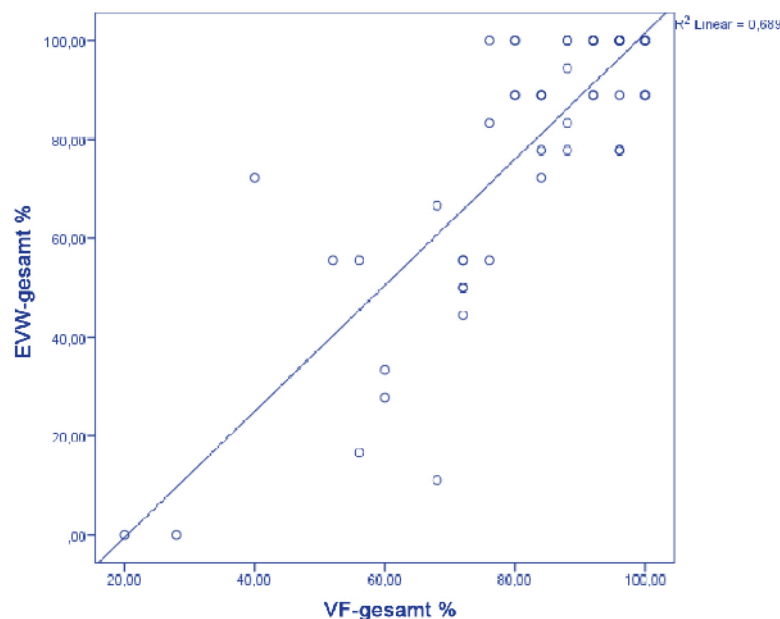


Abb. 1: Zusammenhang der Untertests von *Halt*

Durch die Ergebnisse in den Erwartungsverletzungen werden 68,9% der Varianz bei den Verständnisfragen aufgeklärt. Es zeigt sich eine hohe positive Korrelation von  $r(56)=.724$  ( $p<.001$ ) zwischen den Leistungen in den beiden Untertests von *Halt*.

#### ■ Fragestellung 2: Unterschiede zwischen den Untersuchungsgruppen in den Untertests von *Halt*

Im Untertest *Verständnisfragen* zeigen die rezeptiv unauffälligen Kinder einen Mittelwert von 22,0 ( $SD=2.71$ ). Die rezeptiv auffälligen hingegen weisen einen Mittelwert von 15,5 ( $SD=5.74$ ) auf. Ein Mittelwertvergleich mittels des Mann-Whitney-U-Tests ergab einen signifikanten Unterschied ( $U=85$ ,  $p<.001$ ). Im Untertest *Entdecken von Erwartungsverletzungen* zeigen die rezeptiv unauffälligen Kinder einen Mittelwert von 15,6 ( $SD=3.59$ ), während die rezeptiv auffälligen einen Mittelwert von 9,2 ( $SD=5.79$ ) aufweisen. Ein Mittelwertvergleich mittels der Mann-Whitney-U-Tests für zwei unabhängige Stichproben ergab ebenfalls einen signifikanten Unterschied ( $U=83$ ,  $p<.001$ ). Das folgende Boxplotdiagramm verdeutlicht diese Gruppenunterschiede für beide Untertests (vgl. Abb. 2).



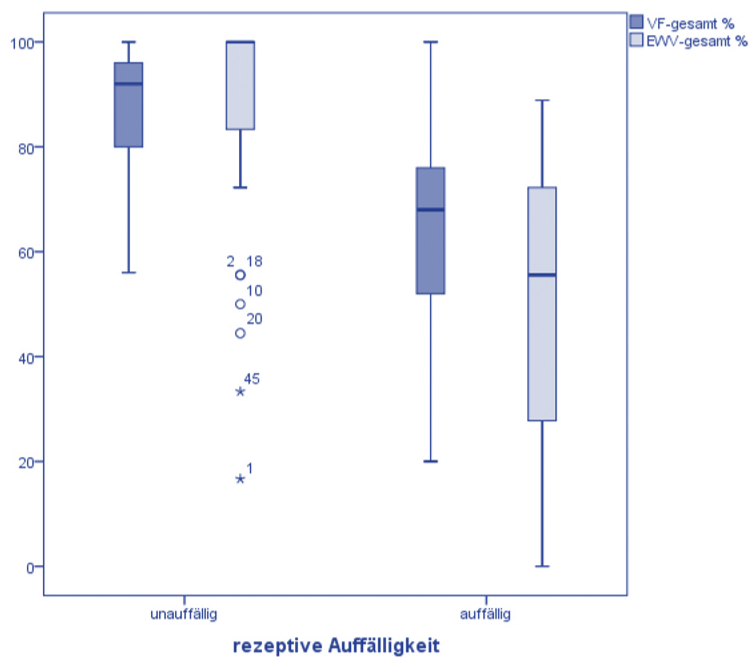


Abb. 2: Gruppenunterschiede für die Untertests von Halt

Die rezeptiv auffälligen Kinder erzielen in beiden Untertests niedrigere Werte als die unauffälligen. Die Mediane der sprachverstehensauffälligen Kinder liegen in einem geringeren Prozentbereich. Die Streuung der Rohwerte ist bei den rezeptiv auffälligen Probanden besonders im Untertest *Entdecken von Erwartungsverletzungen* deutlich größer. Allerdings sind in der Gruppe der unauffälligen Kinder bei den Erwartungsverletzungen Ausreißer und Extremwerte zu finden.

### ■ Fragestellung 3: Unterschiede zwischen den Untersuchungsgruppen bei den unterschiedlichen Fragearten im Untertest *Verständnisfragen*

Des Weiteren wurden die Leistungen bei den verschiedenen Fragearten zwischen den Untersuchungsgruppen der rezeptiv unauffälligen und auffälligen Kinder verglichen. Beide Gruppen weisen bessere Leistungen bei den Informationsfragen als bei den Inferenzfragen auf. Die rezeptiv unauffälligen Kinder schneiden sowohl bei den Informations- als auch bei den Inferenzfragen besser ab als die Kinder mit rezeptiver Auffälligkeit. Ein Mittelwertvergleich zwischen den Gruppen mittels des Mann-Whitney-U-Tests für zwei unabhängige Stichproben ergab sowohl für Informationsfragen einen signifikanten Unterschied ( $U=104$ ,  $p<.001$ ), als auch für Inferenzfragen ( $U=86$ ,  $p<.001$ ). Beim Vergleich der lokalen und globalen Inferenzfragen fällt auf, dass beide Gruppen bessere Leistungen bei den lokalen Inferenzfragen zeigen. Die rezeptiv unauffälligen Kinder weisen bei beiden Fragearten bessere Leistungen auf. Die niedrigsten Prozentwerte weisen die rezeptiv auffälligen Kinder bei den globalen Inferenzfragen auf. Ein Mittelwertvergleich zwischen den Gruppen mittels der Mann-Whitney-U-Tests für zwei unabhängige Stichproben ergab bei lokalen Inferenzfragen einen signifikanten Unterschied ( $U=95$ ,  $p<.001$ ), ebenso bei globalen Inferenzfragen ( $U=121$ ,  $p<.001$ ).

### ■ Fragestellung 4: Unterschiede zwischen den Untersuchungsgruppen in den Reaktionsmustern im Untertest *Entdecken von Erwartungsverletzungen*

Es wurde überprüft, ob sich die Reaktionen der rezeptiv auffälligen von den unauffälligen Kindern im Hinblick auf die verschiedenen Erwartungsverletzungen unterscheiden (vgl. Tab. 4). Der prozentuale Vergleich der Reaktionen der rezeptiv unauffälligen und auffälligen Kinder auf die Erwartungsverletzungen ergibt, dass beide Gruppen am häufigsten mit einer Spezifizierung bzw. Korrektur der Erwartungsverletzung reagieren, die rezeptiv unauffälligen Kinder diese Reaktion jedoch häufiger (85,7 %) zeigen als die rezeptiv auffälligen (46,2 %). Die Kinder mit Verstehensdefiziten zeigen etwa gleich oft (45,3 %) keine Reaktion, während die Kinder ohne Verstehensdefizite nur in 11,4 % der Fälle nicht reagieren. In beiden Gruppen kommt der isolierte Widerspruch relativ selten vor.

Tab. 4: Vergleich der Häufigkeitsverteilungen der Reaktionen auf die Erwartungsverletzungen

	Keine Reaktion	Widerspruch (Nein)	Spezifizierung/Korrektur
Rezeptiv unauffällig	11,4 %	2,9 %	85,7 %
Rezeptiv auffällig	45,3 %	8,5 %	46,2 %

### ■ Fragestellung 5: Itemanalyse

Zur Bestimmung und Verbesserung der psychometrischen Güte des entwickelten Verfahrens wurde eine Itemanalyse durchgeführt. Dabei sollen Items mit zu geringer oder hoher Schwierigkeit und zu geringer Trennschärfe ermittelt werden, um nicht genügend geeignete Items zu eliminieren oder zu verbessern.

#### *Schwierigkeitsindizes*

Der Schwierigkeitsindex einer Aufgabe ist nach Lienert und Raatz (1998) der prozentuale Anteil  $P$  der auf diese Aufgabe entfallenden richtigen Antworten in einer Analysestichprobe von der Größe  $N$ . Der Schwierigkeitsindex liegt somit bei schwierigen Aufgaben niedrig und bei leichten hoch. Ein Wert zwischen 20 und 80 % gilt dabei als akzeptabel (Lienert & Raatz, 1998). Allerdings soll das Verfahren *Halt* rezeptiv auffällige Kinder herausfiltern und erhebt nicht den Anspruch, zwischen guten Verstehern zu differenzieren. Die Schwierigkeitsindizes sollten daher in einem Bereich liegen, der diese beiden Gruppen besonders gut differenziert, was in diesem Fall einem hohen Schwierigkeitsindex entspricht. Deswegen sind auch Werte von ca. 85 % erwünscht (Gebhard, 2008).

Im Untertest Verständnisfragen erfolgt die Bestimmung der Itemschwierigkeiten über eine zufallskorrigierte Formel, da es sich hierbei um Mehrfachwahlaufgaben mit zwei Distraktoren handelt. Sieben Items in diesem Untertest zeigen zu geringe Schwierigkeiten ( $P > 85$ ), eine globale Inferenzfrage weist einen negativen Schwierigkeitsindex und damit eine zu hohe Schwierigkeit auf. Die durchschnittliche Itemschwierigkeit liegt unter Ausschluss des negativen Schwierigkeitsindex dieses Items bei  $P=77$ . Es sind sowohl relativ schwierige Aufgaben als auch mittlere Schwierigkeiten und leichte Items enthalten. Im Hinblick auf den Verwendungszweck dieses Verfahrens, die rezeptiv schwächsten Kinder der Gesamtgruppe herauszufiltern, sind auch Items mit Schwierigkeiten über  $P=80$  integriert.

Im Untertest Erwartungsverletzungen geschieht die Berechnung der Itemschwierigkeiten nicht über die Zufallskorrektur, da hier Raten eine eher untergeordnete Rolle spielt. Außerdem erfolgt keine dichotome, sondern eine dreistufige Bewertung (0/1/2 Punkte). Die Berechnung der Itemschwierigkeit wird über das Verhältnis des Itemmittelwerts zu dem maximal erreichbaren Wert eines Items berechnet. Die neun Items des zweiten Untertests weisen einen geeigneten Schwierigkeitsgrad zwischen 60-89 % auf. Die durchschnittliche Itemschwierigkeit liegt bei  $P=79$ . Damit erweist sich dieser Untertest im Gesamten als etwas leichter als der erste Untertest.

#### *Trennschärfe*

Die Trennschärfe bezeichnet die Korrelation eines Items mit dem Gesamtergebnis eines Tests. Sie gibt also an, wie gut ein einzelnes Item das gesamte Testergebnis voraussagt. Der mögliche Wertebereich liegt dabei zwischen -1 und 1. Bei einer hohen positiven Trennschärfe erfasst das Item dasselbe Kriterium wie der Gesamttest, während bei einer Trennschärfe nahe 0 das Item den Gesamttest nicht ausreichend widerspiegelt. Die Trennschärfe gibt an, wie gut ein Item zwischen Personen mit niedriger und hoher Merkmalsausprägung unterscheidet. Ein Item sollte mindestens eine Trennschärfe von 0,3 aufweisen. Zwischen Itemschwierigkeit, Trennschärfe und Reliabilität besteht ein enger Zusammenhang, da eine Variation von Itemschwierigkeiten zu einer Abnahme der Homogenität und damit zu einer Abnahme der Trennschärfe führt. Eine mittlere Itemschwierigkeit begünstigt daher eine hohe Trennschärfe und eine höhere Reliabilität (Bühner, 2011).

Sechs Items des Untertests Verständnisfragen weisen eine zu geringe Trennschärfe auf ( $< 0,3$ ), das Item mit der höchsten Trennschärfe stellt eine globale Inferenzfrage nach dem übergeordneten Ziel des Stinktiers dar.

Im Untertest Erwartungsverletzungen weisen alle Items eine hohe Trennschärfe über 0,5 auf und sind daher dazu geeignet, um zwischen Personen mit niedriger und hoher Merkmalsausprä-

gung zu trennen. Die Erwartungsverletzung mit besonders hoher Trennschärfe (0,74) beinhaltet eine Zielunterbrechung.

#### Reliabilität

Die Reliabilität bezeichnet die Messgenauigkeit eines Tests. Für beide Untertests wurden jeweils Cronbachs Alpha als Maß für die interne Konsistenz einer Skala und die Split-Half-Reliabilität berechnet. Bei der Berechnung der internen Konsistenz wird der Test in so viele Untertests zerlegt, wie er Items besitzt. Dabei erfolgt eine Überprüfung des inneren Zusammenhangs der Items unter Berücksichtigung der Testlänge. Die interne Konsistenz bezieht sich daher auf die Korrelation der Items untereinander. Es ist zu beachten, dass Cronbachs Alpha ein kombinierter Index ist, der die Höhe der Itemkorrelationen und gleichzeitig die Itemanzahl berücksichtigt (Bühner, 2011). Für die Untertests des Verfahrens *Halt* ergaben sich dabei die folgenden Werte (vgl. Tab. 5):

Tab. 5: interne Konsistenz der Untertests von Halt

Untertest	Cronbachs Alpha
Verständnisfragen	.876
Entdecken von Erwartungsverletzungen	.878

Zur Feststellung der Split-Half Reliabilität wurden die Items ihrer Reihenfolge nach in eine erste und zweite Hälfte geteilt und miteinander korreliert. Mithilfe der Spearman-Brown-Formel wird die Korrelation der beiden Hälften auf die ursprüngliche Testlänge hochgerechnet. Für den Untertest *Verständnisfragen* ergaben sich dabei folgende Werte (vgl. Tab. 6):

Tab. 6: Split-Half-Reliabilität Untertest Verständnisfragen

Reliabilitätsstatistiken Untertest VF		
Cronbachs Alpha	Teil 1	.834
	Teil 2	.724
Korrelation zwischen Formen		.744
Spearman-Brown-Koeffizient		.853

Die Ergebnisse für den Untertest *Verständnisfragen* zeigen, dass beide Testhälften jeweils bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha) aufweisen. Allerdings sind die beiden Testhälften nicht exakt parallel: Die erste Testhälfte weist eine höhere interne Konsistenz (.834) auf als die zweite Testhälfte (.724).

Die Korrelation zwischen den Testhälften ohne Spearman-Brown-Korrektur ist bereits zufriedenstellend (.744). Die Schätzung der Reliabilität der Gesamtskala unter Verwendung der Spearman-Brown-Korrektur, die eine Aufwertung der Reliabilität der Testhälften auf die Reliabilität des Gesamttests vornimmt, zeigt mit einem Wert von .853 einen etwas geringeren Wert an als die Schätzung der Reliabilität der Gesamtskala über Cronbachs Alpha.

Für den Untertest *Entdecken von Erwartungsverletzungen* wurden folgende Werte erzielt (vgl. Tab. 7):

Tab. 7: Split-Half-Reliabilität Untertest Entdecken von Erwartungsverletzungen

Reliabilitätsstatistiken Untertest Erwartungsverletzungen		
Cronbachs Alpha	Teil 1	.800
	Teil 2	.759
Korrelation zwischen Formen		.787
Spearman-Brown-Koeffizient		.881

Die beiden Testhälften zeigen jeweils bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha). Wiederum sind die beiden Testhälften nicht exakt parallel: Die erste Testhälfte weist eine etwas höhere interne Konsistenz (.800) auf als die zweite Testhälfte (.759). Die Korrelation zwischen den Testhälften ohne Spearman-Brown-Korrektur ist bereits zufriedenstellend (.787).

Die Schätzung der Reliabilität der Gesamtskala unter Verwendung der Spearman-Brown-Korrektur zeigt mit einem Wert von .881 einen etwas höheren Wert an als die Schätzung der Reliabilität der Gesamtskala über Cronbachs Alpha.

## 6 Diskussion

### 6.1 Untertest Verständnisfragen

Da das Ziel des Verfahrens *Halt* darin liegt, Kinder mit Verstehensproblemen auf Textebene herauszufiltern, wurde untersucht, ob rezeptiv auffällige Kinder beim Beantworten der Verständnisfragen schlechtere Leistungen zeigen als rezeptiv unauffällige. Dazu wurden die Mittelwerte der beiden Gruppen statistisch verglichen. Die Mittelwerte der rezeptiv auffälligen und unauffälligen Kinder unterschieden sich signifikant, sodass angenommen werden kann, dass Kinder mit Verstehensdefiziten bei der Bearbeitung des Untertests *Verständnisfragen* besondere Schwierigkeiten aufweisen. In anderen Geschichtentests konnten ebenfalls unterdurchschnittliche Leistungen von sprachgestörten Kindern beim Beantworten von Verständnisfragen zu Geschichten festgestellt werden (u. a. Heel & Janda, 2011, Skarakis-Doyle et al., 2008a, Gebhard, 2008).

In mehreren Studien wurde gezeigt, dass Kinder mit Sprachstörungen bei Aufgaben zur Bildung von Inferenzen schlechter abschneiden als gleichaltrige Kinder mit normaler Sprachentwicklung (u. a. Dodwell & Bavin, 2008, Karasinski & Ellis Weismer, 2010). Es lassen sich jedoch keine ganz übereinstimmenden Ergebnisse darüber finden, ob sprachlich auffällige Kinder ein isoliertes Defizit beim Beantworten von schlussfolgernden Fragen zu Geschichten aufweisen oder ob ebenso Schwierigkeiten beim Beantworten von Informationsfragen bestehen. Im Verfahren *Halt* schnitten sowohl die rezeptiv unauffälligen als auch die rezeptiv auffälligen Kinder besser bei den Informationsfragen als bei den Inferenzfragen ab, was für eine allgemein erhöhte Schwierigkeit der Inferenzfragen spricht. Beim statistischen Vergleich der beiden Gruppen konnte allerdings festgestellt werden, dass die rezeptiv unauffälligen Kinder bei beiden Fragearten signifikant höhere Mittelwerte erzielten. Diese Ergebnisse sprechen gegen isolierte Schwierigkeiten beim Beantworten von schlussfolgernden Fragen, die in einigen Studien berichtet werden (Dodwell & Bavin, 2008, Cain & Oakhill, 1999). Vielmehr zeigen die in dieser Studie getesteten rezeptiv auffälligen Kinder neben unterdurchschnittlichen Leistungen im Bilden von Inferenzen auch schlechtere Leistungen beim Beantworten von Informationsfragen. Verminderte Leistungen beim Lösen von Informationsfragen sprachauffälliger Kinder konnten ebenso in weiteren Studien festgestellt werden (Ellis Weismer, 1985, Bishop & Adams, 1992). Nach Bishop und Adams (1992) könnten u. a. Defizite, der Geschichte eine Struktur zu verleihen, die Schwierigkeiten bei Informations- und schlussfolgernden Fragen der rezeptiv auffälligen Kinder erklären. Ebenso könnten Probleme im phonologischen Arbeitsgedächtnis für das schlechte Beantworten von Informationsfragen verantwortlich sein, denn insbesondere die Kinder mit rezeptiven Auffälligkeiten hatten hier geringere Fähigkeiten.

Da im Verfahren *Halt* die unterschiedliche Verstehenstiefe einbezogen wird und deswegen auch die schlussfolgernden Fragen hinsichtlich ihrer Komplexität unterschieden werden, sollen diese getrennt voneinander betrachtet werden. In Anlehnung an das Konstruktions-Integrations-Modell nach van Dijk und Kintsch (1983) wird zwischen eher hierarchieniedrigen Verarbeitungsprozessen, die für die Herstellung lokaler Kohärenz verantwortlich sind, und eher hierarchiehöheren Verarbeitungsprozessen, die die Herstellung globaler Kohärenz ermöglichen, differenziert. Diese unterscheiden sich dadurch in der Distanz der zu integrierenden Informationen. Der Vergleich der beiden Gruppen ergab, dass sich signifikante Gruppenunterschiede sowohl bei den lokalen als auch den globalen Inferenzen feststellen ließen, was sich mit den Ergebnissen von Karasinski und Ellis Weismer (2010) für sprachauffällige Jugendliche deckt. Daher stellt nicht nur die globale Kohärenzbildung, die von höheren Anforderungen an das Arbeitsgedächtnis abhängt, eine besondere Schwierigkeit für sprachverstehensgestörte Kinder dar. Das Ziehen von lokalen Schlussfolgerungen ist zwar den hierarchieniedrigen Verarbeitungsprozessen zuzuordnen, bereitet den rezeptiv auffälligen Kindern ebenfalls Probleme.

Allerdings maskiert der Vergleich von Gruppenmittelwerten mögliche Variationen innerhalb der Gruppen. Weitere Faktoren wie ToM oder personbezogene Faktoren wie Motivation und Erfahrung könnten Einfluss nehmen. Denn es handelt sich beim Textverstehen um interagierende Verarbeitungsprozesse. Somit können auch inhaltliche Präferenzen eine Rolle bei der Fragenbeantwortung spielen. Beispielsweise fiel beim Vorlesen der Geschichte auf, dass besonders das Loslassen der Stinkbombe häufig Lachen oder Kommentare der Kinder hervorrief. Diese Stelle der Geschichte scheint also für die Kinder besonders interessant bzw. witzig zu sein und sollte daher auch besser verstanden und erinnert werden. Daher ist es wichtig, die vorgenommene Einteilung der Fragen als hilfreich für die qualitative Auswertung, aber nicht als einzige Betrachtungsmöglichkeit anzusehen und in weiteren Studien zusätzliche Faktoren mit zu berücksichtigen.

## 6.2 Untertest Entdecken von Erwartungsverletzungen

Der zweite Untertest dient der Erfassung der Leistungen im Monitoring des Sprachverstehens auf Textebene. Hierbei wurde eine dreistufige Bewertung gewählt, welche sich von Skarakis-Doyle (2002) und Heel und Janda (2011, 2014) unterscheidet, die die Methode allerdings bei jüngeren Kindern einsetzten. Im Einklang mit Skarakis-Doyle und Dempsey (2008b) sowie Heel und Janda (2011, 2014) wiesen die rezeptiv auffälligen Kinder signifikant geringere Mittelwerte auf als die rezeptiv unauffälligen Kinder. Bei den rezeptiv unauffälligen Kindern gab es jedoch einige Ausreißer mit geringen Werten im Monitoring des Sprachverstehens. Diese könnten möglicherweise durch geringe Fähigkeiten im phonologischen Arbeitsgedächtnis oder geringe expressive Fähigkeiten für die Korrektur der Erwartungsverletzung, aber auch durch geringe Aufmerksamkeitsleistungen erklärt werden. Studien belegen, dass Kinder mit Aufmerksamkeitsstörungen besondere Probleme im Monitoring des Sprachverstehens haben (Berthiaume, Lorch & Milich, 2010). In zukünftigen Studien sollte dieser Aspekt deshalb zusätzlich erfasst werden.

Insgesamt ist für beide Gruppen feststellbar, dass ein isolierter Widerspruch („Nein“) relativ selten vorkam. Die rezeptiv unauffälligen Kinder reagierten eindeutig am häufigsten mit einer zusätzlichen Spezifizierung bzw. einer Korrektur (85,7%). Nur in 11,4% der Fälle reagierten sie gar nicht. Die Kinder mit Verstehensdefiziten hingegen zeigten in den Bewertungskategorien *keine Reaktion* und *Spezifizierung bzw. Korrektur* etwa gleich häufige Verteilungen. Expressive Sprachprobleme könnten für eine fehlende Korrektur verantwortlich sein. Da allerdings bei den rezeptiv auffälligen Kindern häufig gar nicht reagiert wurde, spricht dies für ein Nicht-Erkennen von Unstimmigkeiten, möglicherweise aufgrund von Problemen in der Aufmerksamkeit oder aufgrund einer generellen Nichtevaluation von Inhalten in Bezug auf Erwartungsverletzungen oder Inkonsistenzen und damit einem mangelhaften Monitoring des Sprachverstehens. Dies bleibt in weiteren Studien zu klären.

## 6.3 Itemanalyse

Die meisten Items des ersten Untertests *Verständnisfragen* weisen einen geeigneten Schwierigkeitsgrad auf. Es sind sowohl relativ schwierige Aufgaben als auch mittlere Schwierigkeiten und leichte Items enthalten. Die trennscharfen Items bestehen ungefähr zu gleichen Teilen aus Informations- und schlussfolgernden Fragen. Die Itemanalyse lieferte wichtige Hinweise für die Überarbeitung des Untertests *Verständnisfragen*. Da die Methode *Entdecken von Erwartungsverletzungen* zuvor noch nicht an siebenjährigen Kindern durchgeführt wurde, stellte sich bei der Konzeption des Untertests die Frage nach einem geeigneten Bekanntheitsgrad der Geschichte und Schwierigkeitsgrad für Kinder dieses Alters. Durch den Ausschluss des zweiten Untertests *Gemeinsames Nacherzählen* und das dreistufige Bewertungssystem ohne Einbezug nonverbaler Reaktionen sollten die Anforderungen den Fähigkeiten von Schulkindern angepasst werden. Die Itemschwierigkeiten weisen darauf hin, dass dies gelungen ist. Somit stellte sich die Geschichte als ausreichend bekannt für die Durchführung der Überprüfung des Monitoring des Sprachverstehens heraus. Alle Items dieses Untertests weisen eine hohe Trennschärfe über 0,5 auf und sind daher dazu geeignet, um zwischen Personen mit niedriger und hoher Merkmalsausprägung zu trennen. Das trennschärfste Item stellte dabei eine Zielunterbrechung dar.

Für beide Untertests konnten sowohl hinsichtlich der internen Konsistenz als auch der Split-Half Reliabilität zufriedenstellende Werte festgestellt werden.

## 6.4 Methodenkritik

Für die Untersuchung wurden die Kinder anhand einer Auffälligkeit im TROG-D und/oder der Mäusegeschichte in rezeptiv auffällig vs. unauffällig eingeteilt. Diese Einteilung birgt Probleme, da die Mehrzahl der Kinder nur in einem Testverfahren und nicht in beiden auffällig war. Auch wenn das grammatische Verstehen mit dem Textverstehen korreliert, könnten einige der Kinder isolierte grammatische Schwierigkeiten aufweisen, was die durchaus auch guten Verstehensleistungen bei der Halt-Geschichte einiger Kinder der rezeptiv auffälligen Gruppe belegt. Es ist auch fraglich, ob der geforderte Wert von 7 in der Mäusegeschichte eindeutig im Hinblick auf die Gruppen differenziert, da in der gewählten Altersgruppe nur fünf Kinder unterhalb der geforderten Punkte in der Mäusegeschichte lagen. Bei 58 Kindern wären bei der Grenze von 15 % der Schlechtesten jedoch acht bis neun rezeptiv auffällige Kinder zu erwarten bzw. durch die hohe Anzahl von Kindern aus Sprachheilschulen eher mehr. In der Gruppe der rezeptiv nicht auffälligen Kinder waren u. a. Kinder, die exakt den geforderten Wert von 7 erreichten, so dass auch dieser Wert noch Schwierigkeiten belegen könnte. Dafür sprechen auch die geringen Leistungen



in den Subtests der Haltgeschichte einiger Kinder der rezeptiv nicht auffälligen Gruppe. In zukünftigen Untersuchungen sind deshalb weitere Parameter mit in die Gruppenauswahl miteinzubeziehen. Beispielsweise könnten expressive Sprachentwicklungsstörungen, Auffälligkeiten im phonologischen Arbeitsgedächtnis und in der ToM als weitere Marker oder für die Einteilung in weitere Subgruppen herangezogen werden.

Für die Überprüfung der ToM wurde das Stofftier einer Ratte herangezogen. Einige Kinder reagierten hier verunsichert oder nannten bei der Frage nach dem Inhalt der Smartiesschachtel eine ihrer Meinung nach typische Nahrung einer Ratte oder antworteten „Weiß ich nicht!“. Diese Reaktionen hätten möglicherweise umgangen werden können, wenn eine zweite Person bei der Testung anwesend gewesen wäre, die erst nach dem Besprechen des tatsächlichen Inhalts der Smartiesschachtel den Raum betreten hätte. Da vermutlich für alle Menschen eine Smarties-Schachtel den Inhalt „Smarties“ impliziert, während Tieren Smarties-Schachteln normalerweise unbekannt sind, wären die Kinder somit aufgrund der Aufgabenstellung wahrscheinlich nicht gleichermaßen verunsichert gewesen. Zudem wird diese Aufgabe normalerweise mit jüngeren Kindern durchgeführt, sodass eine Anpassung an das Alter der im Rahmen dieser Arbeit getesteten Kinder eventuell von Vorteil gewesen wäre. Ebenso würde eine Differenzierung der Überprüfung mit mehreren Fragen wie bei Kim (2016) zu genaueren Ergebnissen führen.

## 7 Fazit/Ausblick

Die Ergebnisse dieser ersten Erprobung des Verfahrens *Halt* zeigen geeignete Schwierigkeiten sowie Trennschärfen der Items und gute Reliabilitäten und sprechen somit dafür, dass mit *Halt* eine mögliche Alternative bzw. Ergänzung zu den wenigen vorhandenen Testverfahren gefunden werden konnte. Der Untertest *Verständnisfragen* sollte nach einer Kürzung und Überarbeitung der Items erneut an einer größeren Stichprobe überprüft werden. Die Wahl der Geschichte im Hinblick auf die Einordbarkeit in das gängige Geschichtenschema und die Konzeption der Verständnisfragen unter Berücksichtigung von Ziel-Plan-Strukturen der Protagonisten und dem Verstehen von Täuschungen stellte sich als wichtig zur Unterscheidung zwischen Kindern mit und ohne Defiziten im Textverstehen heraus. Durch eine qualitative Betrachtung der Verständnisfragen kann außerdem zwischen verschiedenen inhaltlichen Aspekten des Textverstehens sowie der unterschiedlichen Verstehentiefe differenziert werden. Eine quantitative Erfassung der Verstehentiefe wäre allerdings nach einer weiteren Erprobung erstrebenswert. Für den Untertest *Entdecken von Erwartungsverletzungen* konnte gezeigt werden, dass die vorgenommenen Änderungen der ursprünglichen Methode nach Skarakis-Doyle (2002) für siebenjährige Kinder geeignet sind. Neben einer umfangreicheren Stichprobe zur Evaluation wäre es wichtig, das Verfahren auch an Kindern in anderen Altersstufen zu erproben, um eine Entwicklungssensitivität der Überprüfungsmethoden feststellen und eine mögliche Erweiterung des Altersbereichs vornehmen zu können. Bei jüngeren Kindern könnte die Geschichte inhaltlich die Merkfähigkeit übersteigen. Deshalb wird derzeit für diese Kinder der Einsatz des Verfahrens mit zusätzlicher Bildunterstützung durch Metacom-Symbole überprüft. Generell wäre das Verfahren durch die geringe Durchführungs- und Auswertungszeit im sprachtherapeutischen und inklusiven Alltag praktikabel und könnte als wichtige Ergänzung zum TROG-D und zu vorhandenen Geschichtenüberprüfungen herangezogen werden, um eine umfassendere Betrachtung der Leistungen im primären Textverstehen und auch im Monitoring des Sprachverstehens zu erhalten.

## 8 Literatur

- Behrens, U. (2010): Aspekte eines Kompetenzmodells zum Zuhören und Möglichkeiten ihrer Testung. In V. Bernius & M. Imhof (Hrsg.). *Zuhörkompetenz in Unterricht und Schule. Beiträge aus Wissenschaft und Praxis* (S. 31-51). Göttingen: Vandenhoeck & Ruprecht.
- Berthiaume, K.S., Lorch, E.P. & Milich R. (2010). Inferential processing and comprehension monitoring in boys with ADHD. *Journal of Attention Disorders*, 14 (1), 31-42.
- Bishop, D.V.M. & Adams, C. (1992). Comprehension problems in children with specific language impairment: *Literal and inferential meaning*. *Journal of Speech & Hearing Research*, 35 (1), 119-129.
- Buhl, H. M., Möller, F., Oebser, M., Stein, F. & Noack, P. (2009). Zusammenhänge zwischen der Fähigkeit zur Perspektivenübernahme und dem Textverstehen. *Diskurs Kindheits- und Jugendforschung*, 1, 75-90.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. 3. aktualisierte Aufl. München: Pearson.



- Cain, K. & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, 11, 489-503.
- Clegg, J., Hollis, C., Mawhood, L. & Rutter, M. (2005). Developmental language disorders – a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, 46 (2), 128-149.
- Dodwell, K. & Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *International Journal of Language and Communication Disorders*, 43 (2), 201-218.
- Dore, R.A., Amendum, S.J., Golinkoff, R.M. & Hirsh-Pasek, K. (2018). Theory of mind: a hidden factor in reading comprehension? *Educational Psychology Review* 30 (3), 1067-1089.
- Ellis Weismer, S. (1985). Constructive comprehension abilities exhibited by languagedisordered children. *Journal of Speech and Hearing Research*, 28 (2), 175-184.
- Fox, A. V. (2006). *Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. Idstein: Schulz-Kirchner.
- Gebhard, W. (2008). *Entwicklungsbedingte Sprachverständnisstörungen bei Kindern im Grundschulalter. Status und Diagnostik im klinischen Kontext*. 2. aktualisierte Aufl. München: Herbert Utz.
- Gopnik, A. & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59 (1), 26-37.
- Graesser, A. C., Olde, B. A. & Klettke, B. (2002). How does the mind construct and represent stories? In M.C. Green, J.J. Strange & T.C. Brock (Hrsg.). *Narrative impact: Social and cognitive foundations* (S. 213-263). Mahwah, NJ: Lawrence Erlbaum.
- Hachul, C. & Schönauer-Schneider, W. (2016). *Sprachverstehen bei Kindern. Grundlagen Diagnostik Therapie*. München: Elsevier.
- Hagemeister, F. (2011). *Monitoring des Sprachverstehens: Vergleich von Kindern mit spezifischen Sprachentwicklungsstörungen und Kindern mit unterdurchschnittlichen kognitiven Fähigkeiten*. Unveröffentlichte Bachelorarbeit. Ludwig-Maximilians-Universität München.
- Hall, C. & Barnes, M.A. (2017). Inference instruction to support reading comprehension for elementary students with learning disabilities. *Intervention in School and Clinic*, 52 (5), 279-286.
- Heel, M. & Janda, S. (2011). *Diagnostik von Sprachverständnis im Vorschulalter. Explorative Untersuchung zum Textverständnis*. Unveröffentlichte Masterarbeit an der LMU München.
- Heel, M. & Janda, S. (2014). Mehrdimensionale Diagnostik von Textverständnis im Vorschulalter mit der Timogesichte – eine explorative Untersuchung. *Forschung Sprache*, 2 (1), 14-32.
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Karasinski, C. & Ellis Weismer, S. (2010). Comprehension of inferences in discourse processing by adolescents with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 53, 1268-1279.
- Kim, Y.-S. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, 141, 101-120.
- Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, 86, 128-144.
- Kintsch, W. & Rawson, K.A. (2007). Comprehension. In M.J. Snowling & Ch. Hulme (Eds.). *The Science of Reading: A Handbook* (p. 209-226). Malden, Oxford & Victoria: Blackwell Publishing.
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse*. 6. Aufl. Weinheim: Psychologie Verlags Union
- Linder, M. & Grisseemann, H. (2000). *Zürcher Lesetest*. 6. Aufl. Bern: Huber
- Magliano, J. P., Taylor, H. A. & Kim, H.-J. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory & Cognition*, 33 (8), 1357-1367.
- Mai, M. (1995). *44 Zweiminutengeschichten*. Ravensburg: Ravensburger Buchverlag.
- Marks, D., Mayer, A. & Schönauer-Schneider, W. (2018). Vermittlung von Verstehensstrategien zur Förderung des Textverständnisses. In: T. Jungmann, B. Gierschner, M. Meindl & S. Sallat (Hrsg.). *Sprach- und Bildungshorizonte. Wahrnehmen – Beschreiben – Erweitern* (S. 41-53). Idstein: Schulz-Kirchner Verlag.
- Mathieu, S. (1995). Sprachverständnis im Kindergarten – Vergleichende Untersuchung. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 64, 36-52.
- Mathieu, S. (1998). Entwicklung und Abklärung des Sprachverständnisses. In B. Zollinger (Hrsg.). *Kinder im Vorschulalter. Erkenntnisse, Beobachtungen und Ideen zur Welt der Drei- bis Siebenjährigen* (S. 83-137). Bern, Stuttgart, Wien: Haupt.
- Oakhill, J. & Cain, K. (2007). Issues of causality in children's reading comprehension. In D. Mc Namara (Hrsg.). *Reading comprehension strategies: Theories, interventions, and technologies* (S. 47-72). New York: Erlbaum.
- Petermann, F., Metz, D. & Fröhlich, M. P. (2010). *Sprachstandserhebungstest für Fünf- bis Zehnjährige (SET 5-10)*. Göttingen: Hogrefe.
- Richter, T. & Schnotz, W. (2018). Textverstehen. In S. Buch, D. Rost & J. Sparfeldt (Hrsg.). *Handwörterbuch Pädagogische Psychologie* (4. Aufl. S. 826-837). Weinheim: Beltz.
- Rißling, J.-K., Ronniger, P., Petermann, F. & Melzer, J. (2016). Psychosoziale Belastungen bei Sprachentwicklungsstörungen. *Kindheit und Entwicklung*, 25, 145-152.

- Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schulze, M. (2010). *Monitoring des Sprachverstehens im Kindergartenalter – Untersuchungen zum MSV auf Textebene bei Kindern mit und ohne spezifische Sprachentwicklungsstörungen*. Unveröffentlichte Zulassungsarbeit: Ludwig-Maximilians-Universität München
- Skarakis-Doyle, E. (2002). Young children's detection of violations in familiar stories and emerging comprehension monitoring. *Discourse Processes*, 33 (2), 175-197.
- Skarakis-Doyle, E., Dempsey, L. & Lee, C. (2008a). Identifying language comprehension impairment in preschool children. *Language, Speech, and Hearing Services in Schools*, 39, 54-65
- Skarakis-Doyle, E. & Dempsey, L. (2008). The detection and monitoring of comprehension errors by preschool children with and without language impairment. *Journal of Speech, Language and Hearing Research*, 51, 1227-1243.
- Snyder, L. & Caccamise, D. (2010). Comprehension processes for expository text: Building meaning and making sense. In M. Nippold & C. Scott (Hrsg.). *Expository discourse in children, adolescents, and adults. Development and disorders* (S. 13-39). New York, NY: Psychology Press.
- van den Broek, P.W., White, M. J., Kendeou, P. & Carlson, S. (2009). Reading between the lines. Developmental and individual differences in cognitive processes in reading comprehension. In R.K. Wagner, C. Schatschneider & C. Pythian-Sence, (Hrsg.). *Beyond Decoding. The Behavioral and Biological Foundations of Reading Comprehension* (S. 107-123). New York: The Guilford Press.
- van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wassenburg, S.I., Beker, K., van den Broek, P. & van der Schoot, M. (2015). Children's comprehension monitoring relates differently to the multiple dimensions of narrative. *Reading and Writing* 28 (8), 1203-1232.
- Weyda, C. (2012). *Entwicklung und Erprobung eines Verfahrens zur Erfassung von Textverständnis im Grundschulalter*. Unveröffentlichte Masterarbeit im Studiengang Sprachtherapie an der LMU München.
- Zollinger, B. (2010). *Die Entdeckung der Sprache*. Bern: Haupt.

## Zu den Autorinnen

Corinne Weyda (M.A.) beendete 2012 mit der Masterarbeit zum beschriebenen Diagnoseverfahren „Halt“ ihr Studium der Sprachtherapie an der Ludwig-Maximilians-Universität München. Nach Tätigkeit in sprachtherapeutischen Praxen sowie einem entwicklungstherapeutischen Kinderheim arbeitet sie seit 2016 als Sprachtherapeutin in der Spezialambulanz für Entwicklungsstörungen des Heckscher-Klinikums in München.

Wilma Schönaauer-Schneider ist Professorin für Sprachbehindertenpädagogik an der Pädagogischen Hochschule Heidelberg mit den Schwerpunkten Sprachverstehen, Monitoring des Sprachverstehens und sprachheilpädagogischer Unterricht.

## Korrespondenzadresse

Corinne Weyda  
kbo-Heckscher-Klinikum Haar  
Max-Isserlin-Straße 23  
85540 Haar  
corinne.veyda@kbo.de

Prof. Dr. Wilma Schönaauer-Schneider  
Pädagogische Hochschule Heidelberg  
Institut für Sonderpädagogik  
Keplerstr. 87  
69102 Heidelberg  
schoenauer@ph-heidelberg.de



„Forschung Sprache“ ist ein fachwissenschaftliches Organ der Deutschen Gesellschaft für Sprachheilpädagogik e.V. (dgs).

Anträge auf *Neumitgliedschaft* richten Sie bitte an die Bundesgeschäftsstelle:

Deutsche Gesellschaft für Sprachheilpädagogik e.V. (dgs)  
Bundesgeschäftsstelle  
Werderstr. 12  
D-12105 Berlin  
Telefon +49 (0) 30 661-6004  
Telefax +49 (0) 30 661-6024  
info@dgs-ev.de, www.dgs-ev.de

*Ermäßigte Mitgliedsbeiträge* gelten teilweise für Studenten, Lehramtsanwärter und Pensionäre. Details finden Sie unter [www.dgs-ev.de](http://www.dgs-ev.de) → Landesgruppen.

Bei *Adress- und Namensänderungen, Änderungen der Kontodaten* oder *Landesgruppenwechsel* durch Umzug wenden sich dgs-Mitglieder bitte an die dgs-Bundesmitgliederverwaltung unter [bundesmgv@dgs-ev.de](mailto:bundesmgv@dgs-ev.de).

*Kündigungen* richten Sie bitte schriftlich direkt an Ihre zuständige Landesgruppe.

#### Landesgruppen der dgs

Bundesland	dgs-Vertreter/-in
Baden-Württemberg	Dr. Anja Theisel Bad Friedrichshall <a href="mailto:anja@theisel.de">anja@theisel.de</a>
Bayern	Dr. Franziska Schlamp-Diekmann München <a href="mailto:franziska.schlamp@gmx.net">franziska.schlamp@gmx.net</a>
Berlin	Helmut Beek Berlin <a href="mailto:beek@dgs-ev-berlin.de">beek@dgs-ev-berlin.de</a>
Brandenburg	Grit Hentschel Cottbus <a href="mailto:schwteufel69@aol.com">schwteufel69@aol.com</a>
Bremen	Dr. Uta Lürßen, Bremen <a href="mailto:praxis@sprache-kommunikation.de">praxis@sprache-kommunikation.de</a>
Hamburg	Kristine Leites Reinbek <a href="mailto:leites@dgs-ev.de">leites@dgs-ev.de</a>
Hessen	Karin Borgwald Offenbach <a href="mailto:k.borgwald@dgs-hessen.de">k.borgwald@dgs-hessen.de</a>
Mecklenburg-Vorpommern	Beate Westphal <a href="mailto:beate.westphal@t-online.de">beate.westphal@t-online.de</a>
Niedersachsen	Susanne Fischer Celle <a href="mailto:dgs-niedersachsen@dgs-ev.de">dgs-niedersachsen@dgs-ev.de</a>
Rheinland	Heinz-Theo Schaus Essen <a href="mailto:schaus@dgs-rheinland.de">schaus@dgs-rheinland.de</a>
Rheinland-Pfalz	Birgitt Braun Wörth am Rhein <a href="mailto:birgitt_braun@t-online.de">birgitt_braun@t-online.de</a>
Saarland	Michael Monz <a href="mailto:michael.monz@hotmail.de">michael.monz@hotmail.de</a>
Sachsen	Antje Leisner Dresden <a href="mailto:dgs.sachsen@t-online.de">dgs.sachsen@t-online.de</a>
Sachsen-Anhalt	Antje Thielebein Plößnitz <a href="mailto:antjethielebein@web.de">antjethielebein@web.de</a>
Schleswig-Holstein	Regine Voß-Bremer <a href="mailto:dgs.sh@web.de">dgs.sh@web.de</a>
Thüringen	Susann Gröschel-Henkel <a href="mailto:sprachtherapie-groeschel@gmx.de">sprachtherapie-groeschel@gmx.de</a>
Westfalen-Lippe	Uta Kröger Steinfurt <a href="mailto:u.kroeger@dgs-westfalen-lippe.de">u.kroeger@dgs-westfalen-lippe.de</a>

#### Forschung Sprache E-Journal für Sprachheilpädagogik, Sprachtherapie und Sprachförderung

7. Jahrgang 2019 | ISSN 2196-6818

##### Herausgeberin

Deutsche Gesellschaft für Sprachheilpädagogik e.V. (dgs)  
Werderstr. 12 | D-12105 Berlin  
Telefon +49 (0) 30 661-6004  
Telefax +49 (0) 30 661-6024  
[info@dgs-ev.de](mailto:info@dgs-ev.de) | [www.dgs-ev.de](http://www.dgs-ev.de)

##### Redaktion

- Andreas Pohl, Dollbergen | [pohl@dgs-ev.de](mailto:pohl@dgs-ev.de)
- Prof. Dr. Wilma Schöner-Schneider, Wettstetten | [schoenauer@dgs-ev.de](mailto:schoenauer@dgs-ev.de)
- Irina Ruppert-Guglhör, Rosenheim | [ruppert-guglhoer@dgs-ev.de](mailto:ruppert-guglhoer@dgs-ev.de)
- Prof. Dr. Susanne van Minnen, Altenstadt | [van-minnen@dgs-ev.de](mailto:van-minnen@dgs-ev.de)
- Hiltrud von Kannen, Karlstadt | [von.kannen@dgs-ev.de](mailto:von.kannen@dgs-ev.de)
- Downloadredaktion: Kerstin Rimpau, München | [rimpau@dgs-ev.de](mailto:rimpau@dgs-ev.de)
- [redaktion@sprachheilarbeit.eu](mailto:redaktion@sprachheilarbeit.eu)

##### Manuskripte/Mitteilung der Redaktion

Forschung Sprache ist ein Publikationsorgan für Wissenschaftlerinnen und Wissenschaftler, um zeitnah aktuelle Studien veröffentlichen und rezipieren zu können. Es richtet sich an an Wissenschaft interessierte Leserinnen und Leser aus der Praxis, die sich für aktuelle Erkenntnisse aus der Forschung interessieren.

Manuskripte sind unter Beachtung der in den Manuskriptrichtlinien festgelegten Standards in digitaler Form an [redaktion@sprachheilarbeit.eu](mailto:redaktion@sprachheilarbeit.eu) zu senden. Für eingesandte Artikel, Fotos, Zeichnungen etc. kann keine Haftung übernommen werden.

Die Veröffentlichung von Manuskripten erfolgt als Hauptbeitrag mit eventuellem Zusatzmaterial (z.B. Fragebögen, Ergebnisse etc.). Die Beiträge werden von Beiratsmitgliedern peer-reviewed.

Aus Copyrightgründen werden grundsätzlich nur solche Arbeiten angenommen, die vorher weder im Inland noch im Ausland veröffentlicht worden sind. Die Manuskripte dürfen auch nicht gleichzeitig an anderer Stelle zur Veröffentlichung angeboten werden. Die Einsender erklären sich mit der Bearbeitung ihrer Manuskripte einverstanden.

Die in Forschung Sprache veröffentlichten und mit dem Namen der Autoren gekennzeichneten Artikel stellen deren unabhängige Meinung und Auffassung dar und stimmen nicht unbedingt mit den Ansichten der Herausgeberin, der Redaktion oder des Beirates überein.

Geschützte Warennamen (Warenzeichen) werden nicht besonders kenntlich gemacht. Aus dem Fehlen eines solchen Hinweises kann nicht geschlossen werden, dass es sich um einen freien Warennamen handelt.

Die Informationen in diesem E-Journal sind sorgfältig erwogen und geprüft, dennoch kann keine Garantie übernommen werden. Eine Haftung der Autoren, der Herausgeberin und ihrer Beauftragten inkl. des Verlages für Personen-, Sach- und Vermögensschäden ist ausgeschlossen.

**Leserbriefe** bitte per E-Mail an die Redaktion der Zeitschrift; die Redaktion behält sich eine Veröffentlichung (ganz oder in Teilen) vor.

##### Copyright

Deutsche Gesellschaft für Sprachheilpädagogik e.V. (dgs)  
Werderstr. 12 | D-12105 Berlin

##### Wissenschaftlicher Beirat

Prof. Dr. Margit Berg, Ludwigsburg; Prof. Dr. Anja Blechschmidt, Basel; Prof. Dr. Solveig Chilla, Flensburg; Dr. Uwe Förster, Hess. Oldendorf; Prof. Dr. Christian Glück, Leipzig; Dr. Bernd Hansen, Flensburg; Prof. Dr. Erich Hartmann, Fribourg; Prof. Dr. Barbara Höhle, Potsdam; Vertr. Prof. Dr. phil. Vanessa Hoffmann, Hamburg; Prof. Dr. Tanja Jungmann, Siegen; Prof. Dr. Ulrich von Knebel, Hamburg; Prof. Dr. Anette Kracht, Landau; Hannah Kreft, Gießen; Dr. Ulla Licandro, Hannover; Prof. Dr. Kathrin Mahlau, Greifswald; Prof. Dr. Andreas Mayer, München; Prof. Dr. Christiane Miosga, Hannover; Dr. Antje Orgassa, Nijmegen; Prof. Dr. Claudia Osburg, Hamburg; Prof. Dr. Monika Rothweiler, Bremen; Prof. Dr. Stephan Sallat, Halle/Saale; Prof. Dr. Christof Schreiber, Gießen; PD Dr. Katja Subellok, Dortmund

##### Datenbanken

Forschung Sprache ist in den Datenbanken EBSCO/CINAHL, ZPID/Psyndex und FIS gelistet.

##### Erscheinungsweise

2 Ausgaben 2019: 15. Mai, 15. November

##### Satz und Gestaltung

Schulz-Kirchner Verlag GmbH  
Mollweg 2 | D-65510 Idstein  
Telefon +49 (0) 6126 9320-0 | Telefax +49 (0) 6126 9320-50  
[info@schulz-kirchner.de](mailto:info@schulz-kirchner.de) | [www.schulz-kirchner.de](http://www.schulz-kirchner.de)  
Susanne Koch, Telefon +49 (0) 6126 9320-24  
[s.koch@schulz-kirchner.de](mailto:s.koch@schulz-kirchner.de)

Vertretungsberechtigte Geschäftsführer:

Dr. Ullrich Schulz-Kirchner, Nicole Eitel, Martina Schulz-Kirchner