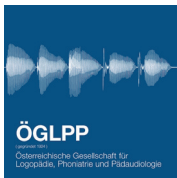


dgs

Sprachheilarbeit

# Forschung Sprache

E-Journal für Sprachheilpädagogik, Sprachtherapie  
und Sprachförderung



DLV

Wilma Schönauer-Schneider, Susanne van Minnen

Editorial . . . . . 2

Miriam Grunwald, Sandra Neumann

Phonetische und/oder phonologische Interventionen in der Sprachtherapie bei  
Kindern mit LKGS-Fehlbildungen: Systematisches Review und Risk of Bias Analyse

Phonetic and/or Phonological Speech Therapy Interventions for Children with  
Cleft Palate: A Systematic Review and Risk of Bias Analysis . . . . . 3

Erich Hartmann, Verena Hofmann, Caroline Villiger,  
Annette Tettenborn, Silke Hauri

Prävalenz und Prädiktion von Nichtresponsivität im Lichte unterschiedlicher  
RTI-Kriterien: Sekundäranalyse und Befunde aus einer präventiven  
Interventionsstudie zur Leseflüssigkeit

Prevalence and Prediction of Non-Responsiveness in the Light of  
Different RTI Criteria. Secondary Analysis and Findings from a Preventive Intervention  
Study on Reading Fluency . . . . . 25

Juliane Hinnerichs

Das Nachsprechen von Nichtwörtern zur Testung bilingualer Kinder  
Zum Einsatz der einzelsprachunabhängigen Items des LITMUS-CL-NWRT zur  
Differenzierung von Kindern mit und ohne Sprachentwicklungsstörungen

Testing bilingual children with a Nonword Repetition Task  
Using the language-independent items of LITMUS-CL-NWRT to distinguish between  
children with and without Developmental Language Disorders . . . . . 43

Christoph Till, Jasmin Kolb

Die Unterstützung von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten  
im Unterricht durch multiprofessionelle Teams. Wer hilft mit und wenn ja, wie viel?

The support of children with language, speech, and communication difficulties in regular  
classrooms through multiprofessional teams. Who helps and if so, to what extent? . . . . . 61

Impressum . . . . . 85





## Editorial

Wilma Schönauer-Schneider, Susanne van Minnen

Vom Laut bis zum Lesen. Von Fragebögen bis zu Metaanalysen. Von Therapie bis zum Team-Teaching.

Die Artikel der vorliegenden Ausgabe der *Forschung Sprache* spiegeln sowohl bezüglich des Untersuchungsgegenstandes und der Lehr-Lernkontexte als auch mit Blick auf das Forschungsdesign die ganze inhaltliche Bandbreite des Faches und die ganze Vielfalt der empirischen Methodik wider.

So nehmen Miriam Grunwald und Sandra Neumann in ihrem Artikel *Phonetische und/oder phonologische Interventionen in der Sprachtherapie bei Kindern mit LKGS-Fehlbildungen: Systematisches Review und Risk of Bias Analyse* die Effektivität verschiedener Interventionsverfahren in der Therapie mit Kindern mit LKGS-Fehlbildungen in den Blick. Ihre Metaanalyse, in die sie 19 Studien einbeziehen konnten, ergab, dass sich in fast allen dieser Studien phonologische, phonetische sowie kombinierte phonetisch-phonologische Therapieansätze als effektiv erwiesen.

Erich Hartmann, Verena Hofmann, Caroline Villiger, Annette Tettenborn und Silke Hauri stellen in ihrem Artikel *Prävalenz und Prädiktion von Nichtresponsivität im Lichte unterschiedlicher RTI-Kriterien. Sekundäranalyse und Befunde aus einer präventiven Interventionsstudie zur Leseflüssigkeit* entsprechende Studienergebnisse vor. Im Mittelpunkt stehen Kinder, die trotz präventiver Leseintervention einen nur geringen Zuwachs in ihrer Lesekompetenz erfahren. Unter anderem untersucht die Studie verschiedene Kriterien, anhand derer Bedingungen benannt werden können, die trotz präventiver Förderung zu einer geringen Leseleistung führen. Neben weiterführenden Überlegungen zur grundlegenden Methodologie werden auch Ergebnisse für das multiprofessionelle Team im (Schweizer) Schulsystem reflektiert.

In ihrem Artikel *Das Nachsprechen von Nichtwörtern zur Testung bilingualer Kinder. Zum Einsatz der einzel-sprachunabhängigen Items des LITMUS-CL-NWRT zur Differenzierung von Kindern mit und ohne Sprachentwicklungsstörungen* stellt Juliane Hinnerichs Diagnostikmöglichkeiten mit dem Nachsprechen von Nichtwörtern im Allgemeinen sowie ein neu entwickeltes Testverfahren vor. Sie lenkt hiermit den Fokus auf verbesserte diagnostische Möglichkeiten bei mehrsprachigen Kindern, deren geringe Sprachkompetenzen auf die unterschiedlich ausgeprägten Spracherwerbsbedingungen oder aber auf eine Sprachentwicklungsstörung zurückzuführen sind.

Christoph Till und Jasmin Kolb gehen in ihrem Artikel *Die Unterstützung von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten im Unterricht durch multiprofessionelle Teams. Wer hilft mit und wenn ja, wie viel?* auf strukturelle und unterrichtsorganisatorische Aspekte, wiederum aus dem Blickwinkel des Schweizer Schulsystems, ein. Es zeigte sich, dass die unterrichtsimmanente Umsetzung sprachunterstützender Maßnahmen quantitativ und qualitativ besser gelingt, wenn Kolleg:innen mit der entsprechenden sprachheilpädagogischen bzw. im Schweizer Schulsystem mit entsprechender sprachtherapeutischer Fachexpertise aktiv in das Team-Teaching eingebunden sind.

Wir wünschen Ihnen viele Anregungen für Ihren professionellen Alltag in Forschung und Praxis.

*Prof. Dr. Susanne van Minnen*

Redaktion *Forschung Sprache*, Professorin für Erziehungswissenschaft mit dem Schwerpunkt Pädagogik bei Sprachbeeinträchtigungen an Justus-Liebig-Universität Gießen

*Prof. Dr. Wilma Schönauer-Schneider*

Redaktion *Forschung Sprache*, Professorin für Sprachbehindertenpädagogik an der Pädagogischen Hochschule Heidelberg



## Phonetische und/oder phonologische Interventionen in der Sprachtherapie bei Kindern mit LKGS-Fehlbildungen: Systematisches Review und Risk of Bias Analyse\*

### Phonetic and/or Phonological Speech Therapy Interventions for Children with Cleft Palate: A Systematic Review and Risk of Bias Analysis

Miriam Grunwald, Sandra Neumann

#### Zusammenfassung

**Hintergrund:** In Europa wird aktuell jedes 700. Kind mit einer Lippen-, Kiefer-, Gaumen-, Segel-Fehlbildung (LKGSF) geboren, wovon die Hälfte der Kinder nach den Primäroperationen eine Sprachtherapie benötigt.

**Ziele:** Die Erstellung eines systematischen Reviews von prospektiven Originalstudien zur Effektivität von phonetischen, phonologischen oder daraus kombinierten Therapieansätzen bei Kindern und Jugendlichen (1;0-19;11 J.) mit LKGSF und cleft type characteristics (CTCs) plus cluster-geleitete Analyse des Risk of Bias (RoB).

**Methodik:** Es wurden die Datenbanken APA PsycInfo, Cochrane Library, HTA Database, PubMed und Web of Science ohne Zeitlimit systematisch nach Originalstudien zur Effektivität phonetischer und/oder phonologischer Therapieansätze (ohne technische Hilfsmittel) durchsucht. In die Analyse eingeschlossen wurden nur prospektive Studien, deren junge Proband:innen (1;0-19;11 J.) zu  $\geq 90\%$  nicht-syndromale LKGSF und CTCs aufwiesen.

**Ergebnisse:** Insgesamt wurden 19 Studien in die Analyse einbezogen. Fast alle Studien mit phonetischen ( $n = 3$ ), phonologischen ( $n = 11$ ), phonetisch-phonologisch kombinierten ( $n = 3$ ) oder vergleichenden Ansätzen ( $n = 2$ ) konnten signifikante Ausspracheverbesserungen belegen. Das RoB schränkt jedoch teilweise die Aussagekraft der Studien ein. Der Einbezug von phonologischen Interventionen sowie hochfrequent dargebotene TE/Intensivtherapien sind als besonders effektiv herauszustellen.

**Schlussfolgerungen:** Für eine uneingeschränkte Evidenz und RoB sollten zukünftige Studiendesigns an der Cochrane-Leitlinie bzw. PRISMA-Statement orientiert werden.

#### Schlüsselwörter

Systematisches Review, Phonetische Therapie, Phonologische Therapie, Lippen-Kiefer-Gaumen-Segel-Fehlbildung

#### Abstract

**Background:** In Europe, one in 700 children is currently born with a cleft palate and/or lip (CP±L), half of whom require speech therapy after their primary surgery.

**Aims:** To conduct a systematic review of prospective original studies on the effectiveness of phonetic, phonological, or combined therapy approaches in children and adolescents (1;0-19;11 yrs.) with CP±L and cleft type characteristics (CTCs) using cluster-led Risk of Bias (RoB) analysis.

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

**Methods:** APA PsycInfo, Cochrane Library, HTA, PubMed, and Web of Science databases were systematically searched for original studies on the effectiveness of phonetic and/or phonological therapy approaches (without technical devices) without time limit. Only prospective studies with  $\geq 90\%$  participants (1;0-19;11 yrs.) with nonsyndromic CP $\pm$ L and CTCs were comprised in the analysis.

**Results:** The review included a total of 19 studies. Almost all trials using phonetic ( $n = 3$ ), phonological ( $n = 11$ ), combined phonetic-phonological ( $n = 3$ ) or comparative approaches ( $n = 2$ ) showed significant speech improvements. However, RoB partially limits the validity of the studies. The inclusion of phonological interventions and high-frequency therapy as well as intensive therapy stand out as particularly effective.

**Conclusions:** For unconstrained evidence and RoB, future study designs should follow the Cochrane Guideline and PRISMA-Statement.

#### Keywords

systematic review, phonetic therapy, phonological therapy, cleft palate

## 1 Einleitung

In Europa wird aktuell etwa jedes 700. Kind mit einer Lippen-, Kiefer-, Gaumen-, Segel-Fehlbildung (LKGsf) geboren (Mossey & Modell, 2012). Die Hälfte der Kinder benötigt nach den Primäroperationen von Lippe, Gaumen und Velum eine sprachtherapeutische Versorgung in wohnortnahen, sprachtherapeutischen Praxen (Peterson-Falzone, Hardin-Jones, & Karnell, 2010). Bei den betroffenen Kindern können Sprechauffälligkeiten wie Resonanzstörungen (Hypernasalität), nasaler Durchschlag, nasale Turbulenzen sowie LKGsf-typische Artikulationsverlagerungen (cleft type characteristics/CTCs) auftreten. Grunwell & Harding (1993) beschreiben typische Artikulationsauffälligkeiten bei LKGsf-Fehlbildungen als artikulatorische Rückverlagerungen, die sich oral, nasal, glottal oder pharyngeal zeigen können. Hierbei hat das Velum eine Schlüsselfunktion inne. Ein postoperativ zu kurzes oder zu unbewegliches Velum kann eine velopharyngeale Insuffizienz (VPI) auslösen, die die Bildung aller oralen Laute negativ beeinträchtigt (Neumann, 2017).

Internationale Expert:innen präferieren eine phonologische Betrachtungsweise der LKGsf-typischen Sprechauffälligkeiten, da diese nicht ausschließlich phonetischer Natur sind. Vielmehr handelt es sich um „phonetisch motivierte phonologische Störungen“, bei denen phonetische und phonologische Störungsanteile vernetzt sind, indem phonetische Fehlentwicklungen in das phonologische System habituiert integriert werden (Golding-Kushner, 2001; Kummer, 2014; Neumann, 2010; Wohlleben, 2004). So behalten z. B. Kinder nach den Primäroperationen und voll funktionstüchtigem Velum bzw. nach Restloch-OP im harten Gaumen Rückverlagerungen bei, obwohl auf phonetischer Ebene die korrekte Artikulationsrealisation möglich ist. Harding & Grunwell (1996) konstatierten Einflüsse auf die phonologische Entwicklung, wenn mehr als ein Ziellaut verlagert bzw. substituiert wurde. Es erfolgt eine phonologische Integration des (geringen) phonetischen Inventars.

Daher kommen (inter-)national in der sprachtherapeutischen Betreuung von Kindern mit LKGsf ab ca. 2;6 Jahren insbesondere phonetische und/oder phonologische Methoden zum Einsatz (Howard & Lohmander, 2011).

Die aktuelle akademische Sprachtherapie verfolgt den Anspruch einer evidenzbasierten Praxis. Nach wie vor mangelt es jedoch an systematischen Reviews über die sprachtherapeutische Therapieeffektivität bei Kindern mit LKGsf. Die vorliegenden internationalen Übersichtsarbeiten von Bessell et al. (2013) sowie Sand, Hagberg & Lohmander (2022) liefern erste Daten zur Effektivität und Evidenz diverser sprachtherapeutischer Konzepte bei Menschen mit LKGsf. In beiden Reviews werden allerdings die Einschränkungen der Ergebnisse durch ein Verzerrungsrisiko/Risk of Bias (RoB) nur grob betrachtet. Das Vorliegen eines oder mehrerer RoB schränkt jedoch die Aussagekraft der Studien ggf. stark ein. Für eine evidenzbasierte Sprachtherapie von Kindern und Jugendlichen mit LKGsf ab 1;0 Jahren bedarf es hingegen der klaren Orientierung an internationalen Evidenzen.

Das vorliegende systematische Review verfolgt daher das Ziel, die Evidenz für die Wirksamkeit ausschließlich phonologischer, phonetischer bzw. kombinierter phonetisch-phonologischer Interventionen bei Kindern und Jugendlichen mit LKGsf herauszustellen und zu bewerten. Hierbei liegt der Schwerpunkt auf der ausführlichen, clustergeleiteten Analyse der Art und Häufigkeit

des RoB in den Bereichen Selection Bias (Stichprobenverzerrung), Performance Bias (Durchführungsverzerrung) und Outcome Bias (Ergebnisverzerrung) aller eingeschlossenen Studien (s. Kap. 4 & Anhang III).

## 2 Methode

Diese systematische Übersichtsarbeit ist bei PROSPERO registriert ([https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=244429](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=244429)). Ein hochwertiges systematisches Review sollte möglichst viele Kriterien der ‚Preferred Reporting Items for Systematic Reviews and Meta-Analysis‘/PRISMA-IPD Checklist (Page et al., 2021) erfüllen. Diese Checkliste definiert Vorgaben zur konkreten Durchführung, sowie Methoden- und Ergebnisdarstellung eines systematischen Reviews. Somit kann eine gründliche und transparente systematische Übersicht erstellt werden, die es den Leser:innen ermöglicht, die Stärken und Schwächen der Überprüfungsergebnisse leicht zu beurteilen.

In dem vorliegenden Artikel werden 20 von 27 Kriterien der PRISMA-IPD Checklist (Page et al., 2021) vollständig, drei Kriterien teilweise und vier Kriterien nicht erfüllt (s. Anhang I). Ein Kriterium wurde als erfüllt angesehen, wenn mindestens ein Unterkriterium bestätigt wurde. Informationen zu verwendeten Analysematerialien sowie das Review Protokoll können bei der Erstautorin erfragt werden. An letzterem wurden keine Änderungen vorgenommen.

### 2.1 Datenbankrecherche und Suchstrategie

Die Datenbanken APA PsycInfo, Cochrane Library, HTA Database, PubMed und Web of Science wurden von Beginn der Aufzeichnungen bis Januar 2022 in den Kategorien Titel und Zusammenfassung durchsucht. Die Suchstrategie/Syntax umfasste fünf Kategorien von Suchbegriffen: Sprechstörungen (z. B. ‚cleft palat\*‘, ‚speech sound disorder\*‘), Proband:innen (< 20 Jahre) (z. B. ‚pre school\*‘, ‚young adult\*‘), Beschreibung der Intervention (z. B. ‚phonolog\*‘, ‚therap\*‘) und Ergebnisse (z. B. ‚effect\*‘, ‚eviden\*‘) (s. Anhang II).

### 2.2 Inklusionskriterien

Es wurden nur veröffentlichte Arbeiten von empirischen, prospektiven Studien (inkl. Studienprotokolle) berücksichtigt. Hierbei gab es keine Einschränkungen hinsichtlich des Veröffentlichungsdatums, des geografischen Standorts der Autor:innen, der Sprache oder des Studiendesigns. Die folgenden Auswahlkriterien entsprechend des PICO-Schemas (Page et al., 2021) wurden beachtet:  $\geq 90\%$  Kinder und Jugendliche (1;0-19;11 Jahre) mit nicht-syndromaler LKGSF (Ausschluss: isolierte Lippen-Fehlbildung), CTCs oder mit Syndromen ohne bekannte Entwicklungsverzögerung; phonologische und/oder phonetische sprachtherapeutische Interventionen ohne technische Hilfsmittel, ausschließlich von Sprachtherapeut:innen (ST) durchgeführt; irrelevant, ob eine Vergleichsgruppe existiert oder nicht; Angabe von Ergebnismaßen für sprachliche Beurteilungen. Gründe für den Ausschluss von Referenzen sind im Flowchart (s. Abb. 1) aufgeführt.

### 2.3 Studienauswahlprozess

Alle Studien wurden anhand der oben genannten Inklusionskriterien in einem zweistufigen Verfahren gescreent. Nach dem obligatorischen Entfernen der Duplikate wurden die verbleibenden Referenzen in Covidence (Covidence systematic review software, o. J.) übertragen und anhand des Titels und des Abstracts sowie anschließend anhand des Volltexts auf Relevanz geprüft. Eine Autorin (M. G.) überprüfte alle Titel und Abstracts auf ihre Zulässigkeit. Die andere Autorin (S. N.) bewertete zudem bei einer Stichprobe von 50 zufällig ausgewählten Referenzen unabhängig und verblindete die Titel und Zusammenfassungen. Die Interrater-Reliabilität, berechnet als Cohen’s Kappa-Koeffizient, zeigte in allen Fällen eine Übereinstimmung ( $K = 1,0$ ).

### 2.4 Strategie der Datenextraktion

Zur Datenanalyse kam ein Datenextraktionsformular in Anlehnung an das PICO-Schema (Page et al., 2021) sowie in Anlehnung an Bessell et al. (2013) zum Einsatz. Es wurden die folgenden Informationen extrahiert: Studiendesign und Evidenzlevel, Ort, Proband:innen (Stichprobengröße, Geschlechterverteilung, Alter, Art der Fehlbildung nach Angaben der Autor:innen), die Intervention und Vergleichsgruppe, die Art der Durchführung (Dauer pro Therapieeinheit/TE, Frequenz, Gesamtdauer) sowie die primären Ergebnismaße und Ergebnisse.

Um die Vollständigkeit der Analysekriterien zu gewährleisten, wurde das Datenextraktionsformular von beiden analysierenden Autorinnen an drei zufällig ausgewählten Studien getestet und ergänzt.

### 3 Ergebnisse

#### 3.1 Eingeschlossene Studien

Bei zwei systematischen Suchläufen am 10.01.2022 wurden 958 Studien für die Überprüfung identifiziert. Es wurden fünf Referenzen ergänzt, die über die Quellen anderer Artikel gefunden wurden. Nach Entfernung der Duplikate verblieben 637 Studien für ein erstes Screening nach Titel sowie Abstract. Hiernach wurden 558 Studien exkludiert. Die Volltexte der verbleibenden 79 Referenzen wurden eingehend im Sinne der Inklusionskriterien geprüft. Bei Studien ohne Zugriffsmöglichkeit wurden die Autor:innen persönlich kontaktiert. Nach dem weiteren Ausschluss von 60 Studien gemäß der Inklusionskriterien, konnten insgesamt 19 Studien in die detaillierte Analyse eingeschlossen werden (s. Abb. 1). Einige Referenzen, die Bessell et al. (2013) und/oder Sand et al. (2022) integrierten, wurden begründet ausgeschlossen (s. Anhang V).

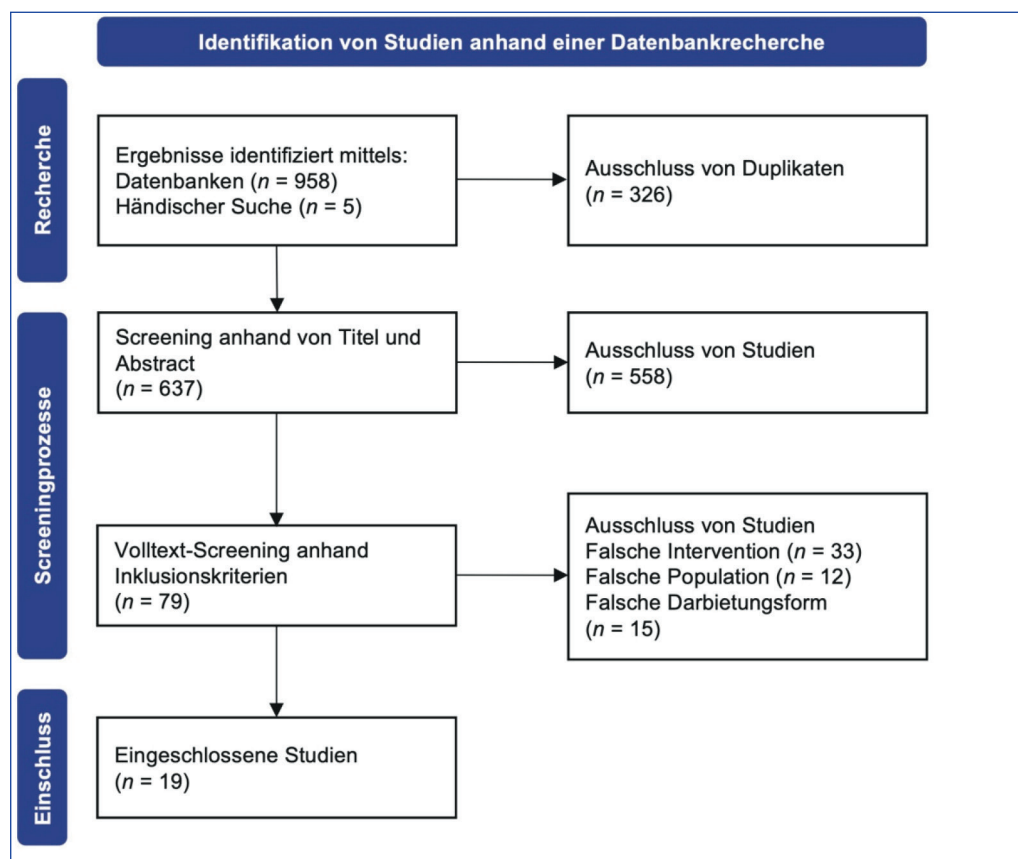


Abb. 1: PRISMA-Flowchart (Page et al., 2021)

#### 3.2 Studiencharakteristika<sup>1</sup>

Von den 19 eingeschlossenen Studien lag deren Publikation zwischen 1971 und 1999 ( $n = 4$ ), zwischen 2000 und 2009 ( $n = 2$ ), zwischen 2010 und 2019 ( $n = 9$ ) sowie zwischen 2020 und 2021 ( $n = 4$ ).

Das Evidenzlevel der identifizierten Therapiestudien erwies sich als sehr heterogen. Alle Referenzen wurden mittels der Evidenz-Skala des Joanna Briggs Institute for Evidence Based Practice (2014) kategorisiert. Das Ranking reicht von der Evidenzstufe I (höchste Evidenzstufe) bis zur Evidenzstufe V (niedrigste Evidenzstufe). Die Studiendesigns umfassten randomisierte Kontroll-

<sup>1</sup> Hinweis: Einige Studien lieferten zu einzelnen Analysekriterien (z.B. Art der Durchführung, Frequenz, Gesamtdauer) keine Angaben. Die fehlenden Informationen werden in Tabelle 1 konkret benannt, jedoch in den Kapiteln 3.3.1 - 3.3.4 nicht detailliert aufgeführt. Die tabellarische Darstellung dient zudem der schnelleren Vergleichbarkeit der Studiendetails.

studien (RCT;  $n=9$ ), Fallserienstudien ohne Vergleichsgruppen ( $n=8$ ) sowie je eine Kohortenstudie/Interventionsvergleichsstudie ( $n=1$ ) und randomisierte Studie ( $n=1$ ). Es wurden über alle Studien hinweg 569 Proband:innen (inkl. Vergleichsgruppen) untersucht, davon pro Studie drei bis 90 Teilnehmende mit einer Altersrange von 1;3 bis 19;6 Jahren. Die Gesamtzahl der Proband:innen ist im Vergleich zu anderen systematischen Übersichtsarbeiten im Bereich der Sprachtherapie bei Kindern, z. B. Crowe, Cuervo, Guiberson, & Washington (2021) relativ gering, insbesondere vor dem Hintergrund der betrachteten Zeitspanne. Die Geschlechterverteilung war, sofern angegeben, in den meisten Studien nicht homogen mit einem größeren Anteil männlicher Probanden.

Rein phonetische Sprachtherapieansätze ( $n=3$ ) wurden neben phonologischen Ansätzen ( $n=11$ ), einer Kombination aus beiden ( $n=3$ ), sowie einem Methodenvergleich ( $n=2$ ) eingesetzt. Hierbei wurden konkrete Therapiekonzepte benannt, z. B. phonologische Therapie im Rahmen des Whole Language Models (WLM,  $n=6$ ) (Pamplona, Ysunza, & Ramírez, 2004; Pamplona et al., 2005; Pamplona et al., 2012; Pamplona, Ysunza, & Morales, 2014; Pamplona, Ysunza, & Morales, 2017; Pamplona & Ysunza, 2018), Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT,  $n=1$ ) (Herrerias Mercado, Simpson, & Bellom-Rohrbacher, 2019), Artikulationstherapie nach Van Riper ( $n=2$ ) (Alighieri et al., 2020; Luyten et al., 2016), der METAPHON-Ansatz ( $n=2$ ) (Alighieri et al., 2020, 2021) oder das Enhanced Milieu Teaching with phonological emphasis (EMT+PE,  $n=3$ ) (Kaiser, Scherer, Frey, & Roberts, 2017; Scherer et al., 2020, 2021).

Die Dauer der Studiendurchführung betrug bei einer Studie nur drei Tage, während andere Studien zahlreiche Sitzungen über mehrere Monate umfassten ( $n=10$ ). Bezüglich der Art der Durchführung fand in den Studien Intensivtherapie ( $n=7$ ), hochfrequente Therapie ( $n=7$ ) wöchentliche Sprachtherapie ( $n=1$ ) oder ein Vergleich von wöchentlicher bzw. hochfrequenter und Intensivtherapie ( $n=2$ ) statt. Die Dauer der einzelnen Therapiesitzungen pro Tag variierte stark zwischen 30 Minuten ( $n=3$ ) und 240 Minuten ( $n=4$ ). Auch die verwendeten Messungen bzw. Diagnostika zur Analyse der Aussprache erwiesen sich in ihrer Art und Weise als stark heterogen. Folgende Verfahren kamen zum Einsatz: Quantitative Zählung der phonologischen Prozesse/CTCs, Einschätzung des Schweregrades der Artikulationsstörung (z. B. Prozentsatz korrekt realisierter Konsonanten ‚PCC‘) anhand von Tonbandaufnahmen der Spontansprache bzw. konstruierten Spielsituationen, Artikulationsüberprüfung mittels flexibler transnasaler Endoskopie, Analyse mittels der Software PRAAT sowie eine Verständlichkeitsmessung anhand der Skala zur Verständlichkeit im Kontext (ICS). Additive Messungen über die Aussprache/Verständlichkeit hinaus wurden in dieser Arbeit nicht berücksichtigt.

Tab. 1: Charakteristika der Studien

Autor:innen (Land/Staat)	Studien-design (Evidenz-level)	N	Alter	Fehlbildungs-art	Interven-tion	Vergleichs-gruppe	Durch-führung	Frequenz, Dauer pro TE	Gesamt-dauer	Primäre Ergebnismaße	Ergebnisse
<b>Phonetische Interventionen</b>											
Van Denmark (1971) (USA)	Fallserie, Vergleich Prä-Post (IV)	11 (2w, 9m)	k. A.	GF	Klassische Artikulations-therapie	-	intensiv	täglich: 150 Min. ET + 90 Min. GT	6 Wochen	Sprachprobe: Schweregrad-einschätzung Artikulation 105-Items-Artikulationstest (= Templin-Darley Screening Test of Articulation & Iowa Pressure Articulation Test (IPAT)) Satz wiederholung Lautdiskriminierung Lautstimulationstest	Schweregradeinschätzung: sign. Reduktion Iowa Pressure Articulation Test: sign. Verbess. (n = 10) 105-Items-Artikulationstest: sign. Verbess. (n = 10), Ø korrigierter Laute: n = 13 Templin-Darley Screening Test of Articulation: nicht sign. Verbess. Satz wiederholungstest: Ø korrigierter Laute: n = 37, sign. Verbess. Gesamtzahl der Fehler Stimulationstest: Reduktion von fehlerhafter Lautrealisation von 66% (Prä) auf 53 % (Post), Verbess. um 61 Laute
Jahanbin et al. (2014) (Iran)	Fallserie, Vergleich Prä-Post (IV)	8 (5w, 3m)	4;0-11;0 J. (M = 6.95)	LKGSF	Klassische Artikulations-therapie	-	k. A.	k. A.	13 Mo.	Software Praat Formantenanalyse (F1, F2, F3), Laute /b/, /p/, /f/, /v/, /k/ und /g/	Formantenanalyse: keine sign. Verbess. bei /f/, /v/, /k/ & /g/, aber sign. Verbess. der Hypemasalität /b/ und /p/
Herreras Mercado et al. (2019) (USA)	Fallserie, Vergleich Prä-Post (IV)	3	4;4, 4;4 & 12;8 J.	BLKGSF	PROMPT	-	hoch-frequent	3x 45 Min./Wo.	1 Mo.	Audioaufnahmen: Verständlichkeit & Transkription Goldman-Fristoe Test of Articulation (GFTA)	Proband: in 1: sign. Verbess. Anzahl korrekter Konsonanten Proband: in 2: sign. Verbess. Anzahl korrekter Konsonanten & bei beiden sign. Verständlichkeitsverbess. Proband: in 3: geringe Verbess.
<b>Phonologische Interventionen</b>											
Van Denmark & Hardin (1986) (USA)	Fallserie, Vergleich Prä-Post (IV)	13 (7w, 9m)	6;8-12;0 J.	LKGSF (n = 4) BLKGSF (n = 4) GF (n = 4) SF (n = 1)	Multiple-Sound-Ansatz	-	intensiv	4x 60 Min. an 5 Tagen/Wo.	26 Tage	Spontansprachprobe: Schweregrad-einschätzung Artikulation Iowa Cleft Palate Articulation Test Satz wiederholungstest	An T2 (nicht signifikante) Verbesserung der Artikulationsfähigkeit in Konversationen (n = 8) sign. Reduktion des Schweregrades der Artikulationseinschränkung von T1 nach T3 (Follow-up) Satz wiederholungstest: bessere Artikulationsfähigkeiten als bei Produktion von Einzelwörtern
Pamplona et al. (2004) (Mexiko)	Prospektive Randomisierte Studie (RT) (II)	30	3;0-7;1 J.	LKGSF	Phonol. Therapie im Rahmen des WLM (n = 15)	Traditionelle phonologische Therapie (n = 15)	hoch-frequent	2x 60 Min./Wo.	1; 4-27 Mo. VG: 6-22 Mo.	Sprachprobe: Transkription & phononet.-phonol. Analyse	Vergleichsgruppe: Abbau von CTCs nach kürzerem Zeitraum (M = 14.53 Mo.) Versuchsgruppe: Abbau von CTCs nach M = 16.27 Mo. → kein sign. Unterschied
Pamplona et al. (2005) (Mexiko)	Interventionsvergleich (III.2)	90	3;0-10;0 J.	LKGSF GF	Phonol. Therapie im Rahmen des WLM (n = 45)	Konventionelle Therapie (n = 45)	l: intensiv VG: hoch-frequent	l: 4x 240 Min. an 5 Tagen/Wo. VG: 2x 60 Min./Wo.	1 Jahr (104 Std.)	Sprachprobe (freies Spiel und Erzählen einer Bildgeschichte): Transkription & phonet.-phonol. Analyse + Schweregrad Artikulationseinschränkung	Versuchsgruppe & Vergleichsgruppe: sign. Reduktion des Schweregrades kompensatorischer Artikulation → kein sign. Unterschied



Autor:innen (Land/Staat)	Studien- design (Evidenz- level)	N	Alter	Fehlbildungs- art	Interven- tion	Vergleichs- gruppe	Durch- führung	Frequenz, Dauer pro TE	Gesamt- dauer	Primäre Ergebnismaße	Ergebnisse
<b>Phonologische Interventionen (Fortsetzung)</b>											
Pamplona et al. (2012) (Mexiko)	Fallserie, Vergleich Prä-Post (IV)	50	4;4-10;11 J.	LKGSF	Phonol. Therapie im Rahmen des WLM mit phonol. Techniken	-	intensiv	5x 240 Min./Wo.	4 Wo.	Sprachprobe (15 Min. freies Spiel, Geschichten erzählen): Transkription & phonet.-phonol. Analyse + Schweregrad Artikulationseinschränkung	Sign. Korrelation zwischen der Effektivität aller Strategien und dem Schweregrad der kompensatorischen Artikulation; 3 Interventionstechniken mit direkten Anweisungen (Einsatz von phonologischen Cues, Lautanbahnung & Bewusstmachung phonematischer Merkmale) → bei schwer betroffenen Proband:innen sign. Verbess. d. Artikulation
Pamplona et al. (2014) (Mexiko)	RCT (II)	90	3;0-6;8 J.	LKGSF	Phonol. Therapie im Rahmen des WLM mit phonol. Techniken je CTCs- Level (n = 45)	Phonol. Therapie im Rahmen des WLM mit phonol. Techniken (n = 45)	intensiv	5x 240 Min./Wo.	4 Wo.	Sprachprobe (20 Min.): Transkription & phonet.-phonol. Analyse + Schweregrad Artikulationseinschränkung	Versuchsgruppe & Vergleichsgruppe: sign. Verbess. des Artikulationsniveaus; Versuchsgruppe: sign. stärkere Abnahme der CTCs
Kaiser et al. (2017) (USA)	RCT (II)	19	1;3-3;0 J.	LKGSF BLKGSF (n = 5) GF (n = 3)	EMT+PE (n = 8)	Business- as-usual (BAU) (n = 11)	hoch- frequent	2x 30 Min./ Wo.	6 Mo. (48 TE)	Profiles of Early Expressive Phonological Skills (PEEPS)	Versuchsgruppe: sign. Verbesserung des PCC gegenüber der Vergleichsgruppe
Pamplona et al. (2017) (Mexiko)	RCT (II)	82	3;0-6;4 J.	LKGSF	Phonol. Therapie im Rahmen des WLM + audiovisuelles Material (n = 41)	Phonol. Therapie im Rahmen des WLM (n = 41)	intensiv	5x 240 Min./Wo.	3 Wo.	Sprachprobe (20 Min.): Transkription & phonet.-phonol. Analyse + Schweregrad Artikulationseinschränkung	Beide Gruppen: sign. Verbess. der kompensatorischen Artikulationseinschränkung; Versuchsgruppe: sign. bessere Artikulationslevel gegenüber der Vergleichsgruppe
Pamplona & Ysunza (2018) (Mexiko)	RCT (II)	32	3;5-6;7 J.	LKGSF	Phonol. Therapie im Rahmen des WLM + Reflektierte Praxis (Deliberate Practice) (n = 16)	Phonol. Therapie im Rahmen des WLM (n = 16)	k. A.	45 Min.	15 TE	Goldman-Fristoe Test of Articulation (GFTA)	Beide Gruppen: sign. Verbesserung der Artikulation; Versuchsgruppe: sign. höhere Verbesserungen im Artikulationslevel gegenüber der Vergleichsgruppe
Scherer et al. (2020) (USA)	RCT (II)	30 (12w, 18m)	1;3-3;0 J.	LKGSF (n = 16) GF (n = 8) BLKGSF (n = 6)	EMT+PE (n = 15)	Business- as-usual (BAU) (n = 15)	I: wöchentlich VG: (n = 6) monatliche TE	I: 2x 30 Min./ Wo VG: 1x/Mo.	I: 6 Mo. (48 TE) VG: 4-6 Mo.	Analyse Konsonanteninventar, PCC, Nasalität	Beide Gruppen: keine sign. Effekte der Intervention auf den PCC bzw. das Konsonanteninventar

Autoren:innen (Land/Staat)	Studien-design (Evidenz-level)	N	Alter	Fehlbildungs-art	Interven-tion	Vergleichs-gruppe	Durch-führung	Frequenz, Dauer pro TE	Gesamt-dauer	Primäre Ergebnismaße	Ergebnisse
Scherer et al. (2021) (Brasilien)	RCT (II)	24 (15w, 9m)	1;6-2;8 J.	LKGSF (n = 14) GF (n = 7) BLKGSF (n = 3)	EMT+PE (n = 12)	Business-as-usual (BAU): keine TE (n = 12)	hoch-frequent	I: 2x 30-45 Min./Wo.	I: 3 Mo.	Profiles of Early Expressive Phonological Skills (PEEPS) Transkription, phonet.-phonol. und Analyse, PCC Skala zur Verständlichkeit im Kontext (ICS)	Beide Gruppen: kein sign. Unterschied bei kompensatorischer Artikulation; Versuchsgruppe: keine sign. Verbess. der Aussprache an T1, bei T2 (Follow-up) aber sign. Verbess. im PCC; der PCC von Plosiven und Verständlichkeit im Kontext
Alighieri et al. (2021) (Belgien)	RCT (II)	12 (7w, 5m)	4;0-12;0 J.	LKGSF/BLKGSF (n = 8) GF (n = 2) SMGF (n = 2)	Modifizier-tes META-PHON (n = 6)	Modifizier-tes META-PHON (n = 6)	I: intensiv VG: wöchentlich	I: 5x 60 Min./Wo. VG: 60 Min./Wo.	I: 2 Wo. VG: 10 Wo.	Spontansprachprobe + Zahlen, Wörter & Sätze aufsagen: Transkription phonet.-phonol. Analyse, PCC Cleft Audit Protocol for Speech-Augmented (CAPS-A)	Beide Gruppen: kein sign. Zeit x Gruppe-Effekt; Versuchsgruppe: An T5 (unmittelbar nach Therapie) sign. Verringerungen im Auftreten anteriorer oraler CTCs, nicht-oraler CTCs und dem Prozentsatz korrekter Artikulationsstellen (PCP) auf Satzebene + sowie sign. Verbess. von Sprechverstehen & Sprechakzeptanz
<b>Kombiniert phonetisch-phonologische Interventionen</b>											
Sell & Grunwell (1990) (Sri Lanka)	Fallserie, Vergleich Prä-Post (IV)	18	11;0-19;0 J.	Gruppe 1: LKGSF/BLKGSF (n = 8) Gruppe 2: GF/SF (n = 6) Gruppe 3: Gruppen-Vergleich (n = 4)	Individuelle phonetisch-phonologische Therapie: Gruppen-Vergleich	-	k. A.	k. A.	k. A.	Sprachprobe (Imitation: Dreisilber, Pseudo-Silben; Zählen von 1-20, Aufzählen von Tagen, Wochen und Monaten; Wortliste, Wiederholung von drei Sätzen, Spontansprachprobe, 2 Min.); phonet.-phonol. Analyse	Alle 3 Gruppen: Beibehaltung von kompensatorischer Artikulation; Verbess. in Imitation von Dreisilbern und Korrektur der Artikulationsart und -stelle
Derakhshandeh et al. (2016) (Iran)	Fallserie, Vergleich Prä-Post (IV)	5 (2w, 3m)	4;6-9;0 J.	LKGSF (n = 2) GF (n = 1) SMGF (n = 1) SF (n = 1)	phonetisch-phonologische Therapie (n = 5)	-	hoch-frequent	4x 45 Min./Wo.	2,5 Monate (40 TE)	Nachsprechen: Analyse korrekter Konsonanten bei Wortimitation bzw. Benennen: Anzahl korrekter Konsonanten CAPS-A (Cleft Audit Protocol for Speech)	Hohe Effektivität (PND = 80%) bei der Reduzierung nicht-oraler CTCs + Verbess. stimulierbarer Konsonanten (PND = 80%) bei n = 4; Erhöhung des PCC beim Benennen (PND = 80%) bei n = 3
Luyten et al. (2016) (Uganda)	Fallserie, Vergleich Prä-Post (IV)	5 (4w, 1m)	7;3-19;6 J.	LKGSF (n = 3) GF (n = 2)	phonetisch-phonologische Therapie (Van Riper & Distinctive Feature Approach)	-	intensiv	2x 60 Min./Tag	3-4 Tage	Sprachprobe (Artikulation /l/; Aufzählen der Zahlen 1-10 und 60-70 sowie der Wochentage + Photo Articulation Test: Third Edition (PAT-3): phonet.-phonol. Analyse Cleft Audit Protocol for Speech-Augmented (CAPS-A)	Klinisch relevante Verbesserungen in der Artikulation + erhebliche Verbesserungen im PCC auf Wortebene bei n = 4

Autor:innen (Land/Staat)	Studien-design (Evidenz-level)	N	Alter	Fehlbildungs-art	Interven-tion	Vergleichs-gruppe	Durch-führung	Frequenz, Dauer pro TE	Gesamt-dauer	Primäre Ergebnismaße	Ergebnisse
<b>Vergleichende Studien zu phonetischer vs. phonologischer Intervention</b>											
Pamplona et al. (1999) (Mexiko)	Randomisierte Studie (RT) (II)	29	3;0-9;4 J.	LKGFSF	Klassische Artikulationstherapie (Van Riper) (n = 15)	Traditionelle phonologische Therapie (n = 14)	hochfrequent	2x 60 Min./Wo.	I: 14-46 Mo. VG: 6-22 Mo.	Sprachprobe (10 Min.): Transkription & phonet.-phonol. Analyse	Versuchsgruppe: benötigte 14-46 Mo. (M = 30,07 Mo.) Therapie zur Korrektur kompensatorischer Artikulation Vergleichsgruppe: sign. kürzere Therapiedauer → 6-22 Mo. (M = 14,50 Mo.)
Alighieri et al. (2020) (Belgien)	RCT (II)	14 (7w, 7m)	4;0-12;0 J.	LKGFSF/BLKGFSF (n = 9) GF (n = 3) SMGF (n = 2)	Modifiziertes METAPHON (n = 7)	Artikulationstherapie (Van Riper) (n = 7)	intensiv	5x 60 Min./Wo.	2 Wochen (10 Tage)	Spontansprachprobe + Zahlen, Wörter & Sätze auftragen: Transkription phonet.-phonol. Analyse, PCC Cleft Audit Protocol for Speech-Augmented (CAPS-A)	Versuchsgruppe: an T5 (unmittelbar nach der Intervention) sign. Verbess. beim PCC-R & Prozent korrekter Artikulationsstellen (PCP) auf Wort- und Satzebene + sign. Verringerungen von anterioren oralen CTCs, nicht-orale CTCs + sign. Verbess. im Sprechverhalten & Sprechakzeptanz Vergleichsgruppe: Verbess. in Artikulationsleistungen, aber nicht sign.

Anmerkung: BAU = Business-as-usual; BLKGFSF = beidseitige Lippen-, Kiefer-, Gaumen-, Segelfehlbildung; CTCs = cleft type characteristics; EMT+PE = Enhanced Milieu Teaching with phonological emphasis; ET = Einzeltherapie; GF = Gaumenfehlbildung; GT = Gruppentherapie; I = Intervention; J = Jahre; k. A. = keine Angabe; LKGFSF = Lippen-, Kiefer-, Gaumen-, Segel-Fehlbildung; M = Mittelwert; m = männlich; Min. = Minuten; Mo. = Monate; PCC = Prozentsatz korrekt realisierter Konsonanten; PCC-R = Prozentsatz korrekt realisierter Konsonanten - überprüft; PCP = Prozentsatz korrekter Artikulationsstellen; phonol. = phonologisch; phomet. = phonetisch; PND = percentage of non-overlapping data; PROMPT = Prompts for Restructuring Oral Muscular Phonetic Targets; RCT = Randomisiert kontrollierte Studie; SF = Segel-Fehlbildung; sign. = signifikant; SMGF = submuköse Gaumenfehlbildung; TE = Therapieeinheit; Verbess. = Verbesserung; VG = Vergleichsgruppe; w = weiblich; WLM = Whole Language Model; Wo. = Woche

### 3.3 Therapieeffektivität

Die in den Studien eingesetzten therapeutischen Ansätze lassen sich in vier Gruppen einteilen: 1. phonetische Ansätze/Artikulationstherapie, 2. phonologisch-linguistische Ansätze, 3. kombiniert phonetisch-phonologische Interventionen sowie 4. Studien zum direkten Vergleich phonetischer vs. phonologischer Therapieansätze. Im jeweiligen Überblick wird nachfolgend durchgängig das PICO-Schema (Page et al., 2021) für die zusammenfassende Darstellung der vier Interventionsformen verwendet. So erhält der/die Leser:in eine strukturierte Übersicht über die Anzahl und Charakteristika der Proband:innen (P), eingesetzte Interventionen (I), ob eine und ggf. welche Vergleichsgruppe vorlag (C) und welche Ergebnisse (O) konstatiert werden konnten. Daran anschließend folgt jeweils eine kurze Darlegung der einzelnen Studien, die teilweise durch Inhalte in Tabelle 1 ergänzt wird. Dies ist notwendig, da die Studien sich in ihrem Design sehr heterogen und kaum zusammenfassbar darstellen lassen. Die Ergebnisdarstellung ist ausschließlich begrenzt auf phonetische bzw. phonologische Ergebnismessungen. Es werden (signifikante) Gruppenunterschiede bzw. messbare Ergebnisse wie Effektstärken oder klinisch bedeutsame Veränderungen dargelegt.

#### 3.3.1 Phonetische Ansätze

##### 3.3.1.1 Überblick

Die traditionelle Artikulationstherapie wurde bei drei Studien eingesetzt und auf deren Effektivität überprüft (Herreras Mercado et al., 2019; Jahanbin, Pahlavannezhad, Savadi, & Hasanzadeh, 2014; Van Demark, 1971). Alle Studien gestalteten sich als Fallserien (Evidenzlevel IV).

**(P) Proband:innen:** In den drei Referenzen wurden insgesamt 26 Kinder (Gruppengröße:  $n = 3-12$ ) zwischen 4;0 und 12;8 Jahren eingeschlossen. Die Mehrheit der Proband:innen wies eine einseitige Lippen-, Kiefer-, Gaumen-, Segel-Fehlbildung (LKGsf) (46,15%) oder eine Gaumenfehlbildung (GF) (42,31%) auf.

**(I) Intervention:** Zwei verschiedene phonetisch-motorische Ansätze wurden verwendet. Es kamen die klassische Artikulationstherapie ( $n = 2$ ), sowie das Konzept PROMPT ( $n = 1$ ) zum Einsatz.

Die Dauer und Frequenz der therapeutischen Interventionen unterschieden sich stark. Es wurden sowohl eine stationäre, tägliche Intensivtherapie über sechs Wochen ( $n = 1$ ), als auch eine hochfrequente wohnortnahe Therapie ( $n = 1$ , 3x/Wo.) durchgeführt.

**(C) Vergleich:** Keine der Studien integrierte eine Vergleichsgruppe.

**(O) Ergebnisse:** Die Ergebnisse zeigen bei allen drei Studien signifikante Verbesserungen der Artikulationsleistung, in einer Studie jedoch nur auf Ebene der lautbegleitenden Hypernasalitätsreduktion.

##### 3.3.1.2 Detaillierte Studienergebnisse

###### Van Demark (1971)

Van Demark (1971) führte eine sechswöchige Intensivtherapie mit elf Kindern (keine Altersangabe) nach traditioneller Artikulationstherapie ohne Vergleichsgruppe durch. Die Kinder erhielten pro Tag 90 Min. Gruppen-, sowie 150 Min. Einzeltherapie. Es konnte mit diversen Artikulationstests eine signifikante Reduktion der Anzahl von Artikulationsfehlern als auch der inkorrekten Realisationen von Plosiven und Frikativen sowie in der übergeordneten Schweregradmessung der Artikulationsstörung gezeigt werden.

###### Jahanbin et al. (2014)

Die Studie von Jahanbin et al. (2014) verfolgte das Ziel, die Effektivität einer Artikulationstherapie der Laute /b/, /p/, /f/, /v/, /k/ und /g/ über einen Zeitraum von 13 Monaten bei acht Kindern (4;0-11;0 J.) mittels PRAAT ohne Vergleichsgruppe zu überprüfen. Es liegen keine Informationen zur Dauer oder Frequenz der TE vor. Die Ergebnisse zeigten keine signifikante Verbesserung der Konsonantenrealisation in den Formanten (F1, F2 & F3) nach der Therapie. Nur bei den Bilabialen /b/ und /p/ war eine signifikante Reduktion der Hypernasalität festzustellen.

###### Herreras Mercado et al. (2019)

Bei der Studie von Herreras Mercado et al. (2019) handelt es sich um eine Fallserie von drei Kindern (4;4, 4;4 & 12;8 J.), bei denen dreimal pro Woche à 45 Min. über einen Monat eine Sprachtherapie mit dem Konzept PROMPT (Hayden, 2015) durchgeführt wurde. Die Ergebnisse

zeigten Verbesserungen in der Realisation der Ziellaute auf Silben-, Wort- und Satzebene bei allen Proband:innen, bei zwei Proband:innen sogar eine Korrektheit über 80 %. Bei diesen konnte zudem eine signifikante Verbesserung bei der Anzahl korrekt realisierter (und verstandener) Konsonanten herausgestellt werden, was sich auch in der Verständlichkeitseinschätzung durch verblindete Personen widerspiegelte.

### 3.3.2 Phonologische Interventionen

#### 3.3.2.1 Überblick

Die Wirksamkeit phonologischer Interventionen wurde von der Mehrheit der Studien ( $n = 11$ ) untersucht (Alighieri et al., 2021; Kaiser et al., 2017; Pamplona et al., 2004, 2005, 2012, 2014, 2017; Pamplona & Ysunza, 2018; Scherer et al., 2020, 2021; Van Demark & Hardin, 1986). Positiv hervorzuheben ist, dass die Mehrzahl der Studien ( $n = 8$ ) als randomisierte Kontrollstudien (RCT) durchgeführt wurden. Zudem lagen zwei Fallserienstudien sowie eine Interventions-Vergleichsstudie vor. Das mexikanische Team um Pamplona führte mehr als die Hälfte der besagten Studien ( $n = 6$ ) durch (Pamplona et al., 2004, 2005, 2012, 2014, 2017; Pamplona & Ysunza, 2018).

**(P) Proband:innen:** In den Referenzen wurden insgesamt 472 Kinder (Gruppengröße:  $n = 12-90$ ) zwischen 1;3 und 12;0 Jahren eingeschlossen. Die Mehrheit der Proband:innen wies eine einseitige Lippen-, Kiefer-, Gaumen-, Segel-Fehlbildung (LKGSF) (71,40 %) oder eine Gaumenfehlbildung (GF) (5,08 %) auf.

**(I) Intervention:** Vier verschiedene phonologisch-linguistische Ansätze kamen zum Einsatz. Es wurden die phonologische Therapie im Rahmen des Whole-Language-Models (WLM) ( $n = 6$ ), die naturalistische Intervention ‚Enhanced Milieu Teaching with Phonological Emphasis‘ (EMT+PE,  $n = 3$ ), der Multiple-Sound-Ansatz ( $n = 1$ ) sowie METAPHON ( $n = 1$ ) verwendet. Auch Dauer und Frequenz der therapeutischen Interventionen variierten stark. Neben Intensivtherapien ( $n = 4$ ) wurden hochfrequente Sprachtherapien (2-3x/Wo.,  $n = 4$ ) sowie eine wöchentliche Therapie durchgeführt. Zwei weitere Studien verglichen wöchentliche bzw. hochfrequente Sprachtherapie mit Intensivtherapie.

**(C) Vergleich:** Drei Studien nahmen eine Gruppe von Kindern mit LKGSF als Vergleich, die als ‚Business-as-usual‘ (BAU) bezeichnet wurde. Unter BAU fielen z. B. Untergruppen, die keine Therapie, Frühförderung oder lediglich Check-ups erhielten. Zwei Studien integrierten keine Vergleichsgruppe, eine führte eine konventionelle Sprachtherapie durch. In den Vergleichsgruppen von drei Studien wurden bei gleichem Therapiekonzept als Unterschied keine zusätzlichen Therapietechniken eingesetzt. Eine Studie nutzte in der Vergleichsgruppe traditionelle phonologische Therapie, während eine andere wöchentliche Sprachtherapie (Vergleichsgruppe) mit Intensivtherapie verglich.

**(O) Ergebnisse:** In neun der elf Studien konnten bei der Versuchsgruppe signifikante Verbesserungen der Artikulationsleistungen in der Reduktion der CTCs bzw. Erhöhung des PCC nachgewiesen werden. Nur zwei Studien, die den Ansatz EMT+PE evaluierten, stellten keine signifikanten Verbesserungen in der Artikulation ihrer Proband:innen nach Therapie fest.

#### 3.3.2.2 Detaillierte Studienergebnisse

##### Van Demark & Hardin (1986)

Die Studie untersuchte die Effektivität einer Intensivtherapie (Sommercamp; 4 Stunden Therapie pro Tag für 26 Tage) mit 13 Kindern im Alter von 6;8-12;0 Jahren ( $M = 8;7$  J.). Hierfür wurde der Multiple-Sound-Ansatz (Mc Cabe & Bradley, 1975) gewählt. Acht Proband:innen zeigten unmittelbar nach der Therapie (T2) eine (nicht signifikante) Verbesserung der Artikulationsfähigkeit in Konversationen. Nach einer darauf aufbauenden ambulanten Therapie (9 Mo.) konnte eine signifikante Reduktion ( $p < .005$ ) des durchschnittlichen Schweregrades der Artikulationsbeeinträchtigung zwischen T1 (Prä-Therapie) und T3 (Follow-up nach 9 Monaten) belegt werden.

##### Pamplona et al. (2004)

In dieser prospektiven randomisierten Studie ( $N = 30$ , 3;0-7;1 J.) wurde eine phonologische Intervention im Rahmen des WLM (Versuchsgruppe,  $n = 15$ ) mit einer traditionellen phonologischen Therapie (Vergleichsgruppe,  $n = 15$ ) verglichen, um zu untersuchen, welcher Ansatz am schnellsten zu einer vollständigen Reduktion der CTCs führt. Alle Proband:innen erhielten zweimal wöchentlich 60 Min. Sprachtherapie. Die Daten zeigten, dass die Vergleichsgruppe nach einem etwas kürzeren Zeitraum von 6-22 Monaten ( $M = 14.53$  Monate), die Versuchsgruppe nach

4-27 Monaten ( $M = 16.27$  Monate) Therapie zum erwünschten Ergebnis kam, was jedoch keinen signifikanten Unterschied darstellte ( $p = .331$ ).

#### **Pamplona et al. (2005)**

Diese Studie stellte eine phonologische Therapie im Rahmen des WLM als Intensivtherapie einer wöchentlichen, konventionellen Sprachtherapie gegenüber. Die Proband:innen der Versuchsgruppe ( $n = 45$ , 3;0-10;0 J.) erhielten im Rahmen eines dreiwöchigen Sommercamps insgesamt 60 Stunden Sprachtherapie, während die Vergleichsgruppe ( $n = 45$ , 3;0-10;0 J.) über einen Zeitraum von einem Jahr (zweimal wöchentlich à 60 Minuten) insgesamt 104 Stunden Therapie bekam. Bei beiden Gruppen konnte eine signifikante Reduktion der kompensatorischen Artikulation verzeichnet werden ( $p < .05$ ), die sich jedoch nicht signifikant voneinander unterschied.

#### **Pamplona et al. (2012)**

Die Proband:innen ( $N = 50$ , 4;4-10;11 J.) dieser Studie wurden im Rahmen des WLM mit phonologischen Techniken zur Reduktion kompensatorischer Artikulation therapiert. Hierbei wurden fünf verschiedene Therapietechniken (z. B. Modellieren) verwendet. Alle Proband:innen erhielten eine vierwöchige Intensivtherapie (240 Minuten pro Tag an fünf Tagen pro Woche). Die einzelne Methodeneffektivität korrelierte signifikant ( $p < .05$ ) mit dem Level der kompensatorischen Artikulationsbeeinträchtigung der Kinder. Die Technik des Modellierens mit oder ohne Betonung zeigte sich nur effektiv bei Kindern mit einer leichten Artikulationsbeeinträchtigung. Besonders die drei Interventionstechniken, die direkte Anweisungen beinhalteten (Einsatz von phonologischen Cues, Lautanbahnung, Bewusstmachung phonematischer Merkmale), zeigten vor allem bei schwer betroffenen Proband:innen einen signifikant höheren Prozentsatz an Artikulationsverbesserungen ( $p < .001$ ).

#### **Pamplona et al. (2014)**

In dieser Studie (RCT) wurde in einem Gruppenvergleich mit 90 Proband:innen (3;0-6;8 J.) im Rahmen einer vierwöchigen Intensivtherapie (240 Minuten pro Tag an fünf Tagen pro Woche) untersucht, ob der Einsatz phonologischer Techniken (Modellieren, Modellieren mit Betonung des Zieltargets nach vorheriger Pause, Einsatz von phonologischen Cues, Lautanbahnung, Bewusstmachung phonematischer Merkmale) wirksamer ist, wenn diese gezielt auf dem Level der kompensatorischen Artikulationsbeeinträchtigung (CTCs) der Kinder eingesetzt werden. In der Versuchsgruppe ( $n = 45$ ) wurden Strategien eingesetzt, die speziell auf das Level der CTCs der Kinder abgestimmt waren, während in der anderen Gruppe alle Strategien unterschiedslos eingesetzt wurden. Beide Gruppen zeigten nach der Intervention eine signifikante Verbesserung ihres Artikulationsniveaus ( $p < .0001$ ). Im Gruppenvergleich wies die Versuchsgruppe jedoch eine signifikant stärkere Abnahme der CTCs auf ( $z$  sub  $t=5.845$ ,  $p < .0001$ ).

#### **Kaiser et al. (2017)**

In dieser Pilotstudie (RCT) erfolgte eine Effektivitätsprüfung der naturalistischen Intervention EMT+PE u.a. auf Artikulationsleistungen von sehr jungen Proband:innen ( $n = 8$ , 1;3-3;0 J.). Die Kinder erhielten zweimal wöchentlich 30 Minuten Therapie (Gesamtdauer: sechs Monate = 48 Sitzungen). Die Vergleichsgruppe (BAU,  $n = 11$ ) erhielt eine typische Sprachtherapie ( $n = 6$ ) bzw. keine Intervention ( $n = 5$ ). Die Versuchsgruppe zeigte eine signifikante Verbesserung ( $p = .015$ ) des PCC gegenüber der Vergleichsgruppe.

#### **Pamplona et al. (2017)**

Diese Studie (RCT) untersuchte die Effektivität des zusätzlichen Einsatzes von audiovisuellem Material parallel zur phonologischen Intervention im Rahmen des WLM in einer dreiwöchigen Intensivtherapie (240 Minuten pro Tag an fünf Tagen pro Woche) bei Proband:innen mit LKGSF ( $N = 82$ , 3;0-6;4 J.). Die Versuchsgruppe ( $n = 41$ ) erhielt neben der phonologischen Therapie zusätzlich audiovisuelles Material (CD mit Liedern und Geschichten mit hohem Input an Ziel-lauten + passendem Ausmalbuch) für die Verwendung im häuslichen Setting zusammen mit den Sorgeberechtigten. Die Vergleichsgruppe ( $n = 41$ ) erhielt ausschließlich phonologische Therapie. Die Ergebnisse belegten in beiden Gruppen eine signifikante Verbesserung in der kompensatorischen Artikulationsbeeinträchtigung ( $p < .01$ ). Die Kinder der Versuchsgruppe zeigten zudem signifikant bessere Artikulationslevel gegenüber der Vergleichsgruppe ( $p < .001$ ).

**Pamplona & Ysunza (2018)**

Die Studie (RCT) evaluierte die Effektivität einer phonologischen Intervention bei Kindern mit LKGSF ( $N = 32$ , 3;5-6;7 J.) anhand von fünfzehn 45-minütigen Therapieeinheiten auf Basis des WLM (Vergleichsgruppe,  $n = 16$ ) versus WLM plus Einsatz des Konzepts der Reflektierten Praxis (Deliberate Practice) (Versuchsgruppe,  $n = 16$ ). Die Methode der Reflektierten Praxis kommt aus dem Höchstleistungssport und ist durch höchst strukturiertes, stetig angepasstes, lang andauerndes Üben gekennzeichnet. Beide Gruppen erzielten eine signifikante Verbesserung der Artikulation ( $p < .05$ ). Die Versuchsgruppe wies jedoch signifikant höhere Verbesserungen im Artikulationslevel gegenüber der Vergleichsgruppe auf ( $p < .05$ ).

**Scherer et al. (2020)**

In dieser Studie (stratifiziertes RCT) wurde wie bei Kaiser et al. (2017) bei kleinen Kindern mit LKGSF ( $N = 30$ , 1;3-3;0 J.) untersucht, inwieweit die naturalistische Intervention EMT+PE die Aussprache der Proband:innen verbessert. Die Versuchsgruppe ( $n = 15$ ) erhielt EMT+PE (48 Therapieeinheiten, zweimal 30 Minuten Therapieeinheit pro Woche, über 6 Monate). Aus der BAU-Vergleichsgruppe ( $n = 15$ ) erhielten Kinder gemeindegestützte Frühförderungsangebote ( $n = 6$ ) bzw. einmal monatlich gemeindegestützte Therapieangebote im häuslichen Setting ( $n = 6$ ).

Insgesamt konnten keine signifikanten Effekte der Intervention auf den PCC bzw. das Konsonanteninventar für die beiden Gruppen herausgestellt werden. Kinder der Versuchsgruppe, die die größten Fortschritte in der Sprachproduktion erzielten, waren 23-33 Monate alt, sprachen mit einer Geschwindigkeit von mindestens 7-10 Wörter pro Minute (WPM) und hatten einen berichteten aktiven Wortschatz von 50 Wörtern. Ein Prädiktor zur Verringerung der kompensatorischen Artikulation nach jeglicher Intervention stellte die Sprechgeschwindigkeit dar. Hier profitierten Kinder, die über eine höhere Sprechgeschwindigkeit ( $> 10$  WPM) verfügten.

**Scherer et al. (2021)**

Eine weitere Studie zur Effektivität der naturalistischen Intervention EMT+PE verglich diesen Ansatz mit einer BAU-Vergleichsgruppe bei sehr jungen Proband:innen ( $N = 24$ , 1;6-2;8 J.,  $M = 2$  J.). Die Kinder der Versuchsgruppe ( $n = 12$ ) erhielten EMT+PE (2x/Wo. zweimal 30-45 Min. TE, 3 Mo.). Die BAU-Vergleichsgruppe ( $n = 12$ ) erhielt keine Sprachtherapie. Es wurden nach der Intervention keine signifikanten Unterschiede kompensatorischer Artikulation zwischen beiden Gruppen belegt. In direktem Anschluss an die Therapie (T1) konnte bei der Versuchsgruppe keine signifikante Verbesserung der Aussprache herausgestellt werden. Der PCC, der PCC von Plosiven und die Verständlichkeit im Kontext zeigten jedoch eine signifikante Verbesserung beim Follow-up (T2) nach 3 Monaten.

**Alighieri et al. (2021)**

Die Studie (RCT) stellte bei Kindern mit LKGSF im Vorschul- bzw. Schulalter ( $N = 12$ , 4;0-12;0 J.,  $M = 8$  J.) eine wöchentliche Therapie (Vergleichsgruppe, 1x 60 Min./Woche, über 10 Wochen) einer Intensiv-Therapie (Versuchsgruppe; 5x 60 Min./Woche, über 2 Wochen) mittels METAPHON (inklusive spezieller Anpassungen für LKGSF) gegenüber. Beide Gruppen erhielten demnach nur eine unterschiedliche Frequenz der TE, während die Anzahl (10 TE) gleich blieb. Es zeigte sich kein signifikanter Zeit x Gruppe-Effekt. Unmittelbar nach der Intervention (T5) konnten in der Versuchsgruppe signifikante Verringerungen im Auftreten anteriorer oraler CTCs ( $p \leq .001$ ), nicht-oraler CTCs und dem Prozentsatz korrekter Artikulationsstellen (PCP) auf Satzebene ( $p < .001$ ), sowie signifikante Verbesserungen des Sprechverstehens ( $p = .002$ ) und der Sprechakzeptanz ( $p = .003$ ) herausgestellt werden. Diese Erfolge blieben auch nach drei Wochen (T6) sowie nach drei Monaten (T7) bestehen.

**3.3.3 Phonetisch-phonologisch kombinierte Interventionen****3.3.3.1 Überblick**

Drei Studien evaluierten die Effektivität von phonetisch-phonologisch kombinierten Interventionen (Derakhshandeh et al., 2016; Luyten et al., 2016; Sell & Grunwell, 1990) non-oral and passive cleft speech characteristics.  
 Methods  
 Five children with cleft palate (+/-cleft lip). Alle eingeschlossenen Studien sind Fallserienstudien, wodurch das Evidenzlevel sehr niedrig (IV) ist.

**(P) Proband:innen:** In den drei Referenzen wurden insgesamt 28 Kinder bzw. Jugendliche ( $n = 5-18$ ) zwischen 4;6 und 19;6 Jahren eingeschlossen. Die Mehrheit der Proband:innen wies eine

einseitige Lippen-, Kiefer-, Gaumen-, Segel-Fehlbildung (LKGSF) (46,43 %) oder eine Gaumenfehlbildung (GF) (25 %) auf.

**(I) Intervention:** In den implizierten Arbeiten führten europäische LKGSF-Spezialist:innen binnen kurzer Zeit im Iran, in Uganda sowie in Sri Lanka Sprachtherapie durch. Zwei Studien nutzten individuelle phonetisch-phonologische Therapieansätze, die dritte Studie setzte eine Intervention nach Van Riper in Kombination mit dem Distinctive Feature Approach ein. Es wurden sowohl eine Intensivtherapie (6-8 TE in 3-4 Tagen), eine Therapie über 4 Monate (Frequenz unbekannt), als auch eine hochfrequente wöchentliche Therapie (40 TE in 2.5 Monaten) durchgeführt.

**(C) Vergleich:** Keine der Studien bezog eine Vergleichsgruppe mit ein.

**(O) Ergebnisse:** Alle drei Studien konnten bei ihren Proband:innen klinisch bedeutsame Verbesserungen in der Artikulationsleistungen durch Reduktionen der CTCs aufzeigen. Eine Intensivtherapiestudie bewies hohe Effektivitätsstärken in den Artikulationsverbesserungen, jedoch nicht bei allen Proband:innen.

### 3.3.3.2 Detaillierte Studien- und Ergebnisdarstellung

#### **Sell & Grunwell (1990)**

Diese Studie untersuchte bei 18 Proband:innen (11;0-19;0 J., M = 14 J.) aus Sri Lanka, die erst acht Monate zuvor operiert wurden, ob eine kombinierte phonetisch-phonologische Therapie zu einer Verbesserung der Artikulationsleistungen führt. Die Proband:innen wurden in drei Gruppen entsprechend der Art ihrer Fehlbildung eingeteilt: Gruppe 1: durchgehende LKGSF/BLKGSF ( $n = 8$ ), Gruppe 2: Gaumen- und Segel-F. ( $n = 4$ ) und Gruppe 3: Isolierte Segel-F. (teils mit G-Anteil,  $n = 6$ ). Nach Abschluss der Therapie zeigten alle Gruppen noch immer kompensatorische Artikulation. Es kam allerdings zu einer Verbesserung bei der Imitation von Dreisilbern, bei der Korrektur der Artikulationsart und -stelle in allen drei Gruppen.

#### **Derakhshandeh et al. (2016)**

In dieser Studie wurde mit fünf Kindern mit LKGSF ( $n = 5$ , 4;6-9;0 J.) eine hochfrequente kombinierte phonetisch-phonologische Therapie (4x 45 Minuten pro Woche über 2,5 Monate = 40 TE) im Iran durchgeführt. Es zeigte sich eine hohe Effektivität (PND = 80 %) bei der Reduzierung nicht-oraler CTCs sowie bei der effektiven Verbesserung stimulierbarer Konsonanten (PND = 80 %) bei jeweils vier Proband:innen. Weiterhin konnte eine Erhöhung des Prozentsatzes korrekter Konsonanten beim Benennen (PND = 80 %) bei drei Proband:innen erzielt werden. Die Korrektur der passiven Lautveränderungen gestaltete sich weniger effektiv. Alle Proband:innen behielten in der 4-wöchigen Follow-up-Phase die Ergebnisse bei.

#### **Luyten et al. (2016)**

In dieser Kurz-Intensivtherapie-Studie (2x 60 Min./Tag, 3-4 Tage) wurde anhand von Proband:innen mit LKGSF in Uganda ( $N = 5$ , 7;3-19;6 J.) eine kombinierte phonetisch-phonologische Therapie nach Van Riper und dem Distinctive Feature Approach (Costello & Onstine, 1976) durchgeführt. Klinisch relevante Verbesserungen in der Artikulation wurden von vier Proband:innen erzielt. Dieselben Proband:innen zeigten auf Wortebene erhebliche Verbesserungen im PCC.

### 3.3.4 Vergleichende Studien zu phonetischer vs. phonologischer Intervention

Zwei Studien verglichen die Effektivität eines phonetischen versus eines phonologischen Therapieansatzes bei Kindern mit LKGSF (Alighieri et al., 2020; Pamplona, Ysunza, & Espinosa, 1999). Im Folgenden werden beide Studien direkt beschrieben.

#### **Pamplona et al. (1999)**

Diese Studie (RCT) verglich die benötigte Dauer in Monaten einer hochfrequenten Sprachtherapie (2 x 60 Min./Woche) zur Eliminierung kompensatorischer Artikulation bei Kindern mit LKGSF ( $N = 29$ ; 3;0-9;4 J.) bei einem rein phonetischen (Versuchsgruppe,  $n = 15$ ) vs. phonologischen Ansatz (Vergleichsgruppe,  $n = 14$ ). Die Versuchsgruppe benötigte 14-46 Monate ( $M = 30,07$  Mo.) Therapie, um das Ziel zu erreichen. Die Vergleichsgruppe erreichte dieses bereits nach 6-22 Monaten ( $M = 14,50$  Mo.), was einen signifikanten Unterschied darstellte ( $p < .001$ ).



**Alighieri et al. (2020)**

Diese Intensivtherapie-Studie (RCT) wurde mit Kindern mit LKGSF ( $N = 14$ , 4;0-12;0 J.) innerhalb von 2 Wochen (1x 60 Min./Tag = 10 TE) durchgeführt und verglich den phonetischen Ansatz von Van Riper (Vergleichsgruppe,  $n = 7$ ) mit dem phonologischen Ansatz METAPHON für LKGSF modifiziert (Dean, Howell, Waters, & Reid, 1995) (Versuchsgruppe,  $n = 7$ ).

Unmittelbar nach der Intervention (T5) konnten in der Versuchsgruppe signifikante Verbesserungen beim Prozentsatz korrekter Konsonanten – überprüft (PCC-R) und Prozentsatz korrekter Artikulationsstellen (PCP) auf Wort- und Satzebene, signifikante Verringerungen im Auftreten von anterioren oralen CTCs ( $p \leq .001$ ), nicht-oralen CTCs ( $p \leq .001$ ) sowie signifikante Verbesserungen des Sprechverstehens ( $p = .006$ ) und der Sprechakzeptanz ( $p < .001$ ) herausgestellt werden. Diese Erfolge blieben auch nach zwei Wochen (T7) sowie nach drei Monaten (T8) bestehen. Die Vergleichsgruppe erreichte ebenfalls Verbesserungen in den Artikulationsleistungen, die jedoch nicht signifikant waren.

#### 4 Einschätzung des Risk of Bias (RoB) in den analysierten Studien

Die beiden Autorinnen (M. G. und S. N.) bewerteten die Studien unabhängig voneinander anhand einer Analysetabelle in Anlehnung an Higgins et al. (2022) und Meusch (2010), wobei jedes der folgenden neun Kriterien entweder als hohes, unklares oder geringes RoB eingestuft wurde. Die genauen Analysedefinitionen pro Kriterium finden sich in Anhang III. Unstimmigkeiten wurden diskutiert und im Konsens gelöst. Ein unklares RoB ist ebenso negativ wie ein hohes RoB zu beurteilen, da es bedeutet, dass die benötigten Informationen nicht erwähnt wurden bzw. nicht ausreichend waren, um das RoB zu bewerten. Das RoB wurde in Anlehnung an Cochrane (Higgins et al., 2022) in drei Cluster (Selection Bias, Performance Bias und Outcome Bias) unterteilt (s. Tab. 2).

Tab. 2: RoB-Cluster nach Cochrane (Higgins et al., 2022)

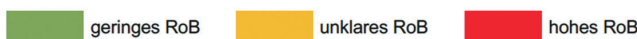
Cluster 1 (Selection Bias)	Cluster 2 (Performance Bias)	Cluster 3 (Outcome Bias)
Beschreibung der Testgruppe(n)	Diagnostikverfahren	Verblindung
Vergleichbarkeit der Testgruppen	Transkription	Vollständigkeit der Ergebnisdaten
Inklusions- und Exklusionskriterien	Datenerhebung	Endpunkte

Die Bewertung der Studien ergab Einschränkungen in allen drei Clustern (s. Tab. 3). Die prozentuale RoB-Verteilung entsprechend der einzelnen Kriterien ist Anhang IV zu entnehmen.

In der Gesamtanalyse ist erkennbar, dass 83 % der eingeschätzten Kriterien ein geringes RoB aufweisen (s. Tab. 3, grüne Kästen). Die verbleibenden 17 % der Kriterien sind kritisch einzuschätzen. Davon belegen 11 % ein unklares RoB (gelbe Kästen) und 6 % ein hohes RoB (rote Kästen). Nur fünf Studien kann für alle untersuchten Kriterien ein geringes RoB bescheinigt werden (Alighieri et al., 2020, 2021; Derakhshandeh et al., 2016; Luyten et al., 2016; Scherer et al., 2020). Weitere vier Studien (Pamplona et al., 2004; Pamplona & Ysunza, 2018; Scherer et al., 2021; Van Demark & Hardin, 1986) zeigen lediglich in jeweils einem Kriterium ein eingeschränktes RoB. Die restlichen zehn Studien sind durch zwei ( $n = 6$ ) bis drei ( $n = 4$ ) Kriterien pro Studie durch ein RoB eingeschränkt. Keine der phonetischen Studien ist ohne eingeschränktes RoB. Insgesamt weisen die kombinierten phonetisch-phonologischen Interventionen unter anderem das geringste RoB auf. Hier sind nur zwei von 26 Kriterien durch RoB eingeschränkt. Die meisten Studien mit hohem RoB befinden sich unter den phonetischen Interventionen, mit einem hohen bzw. unklaren RoB-Anteil von ca.  $\frac{1}{3}$  der Kriterien.

Tab. 3: Analyse des RoB pro Studie

	Cluster 1 (Selection Bias)			Cluster 2 (Performance Bias)			Cluster 3 (Outcome Bias)		
	Beschreibung der Testgruppe(n)	Vergleichbarkeit der Testgruppen	Inklusions- und Exklusionskriterien	Diagnostikverfahren	Transkription	Datenerhebung	Verblindung	Vollständigkeit der Ergebnisdaten	Endpunkte
<b>Phonetische Interventionen</b>									
Van Demark (1971)	gering	gering	hoch	gering	gering	gering	gering	gering	gering
Jahanbin et al. (2014)	gering	gering	hoch	gering	gering	gering	gering	gering	gering
Herreras Mercado et al. (2019)	gering	gering	gering	gering	gering	gering	gering	gering	gering
<b>Phonologische Interventionen</b>									
Van Demark & Hardin (1986)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Pamplona et al. (2004)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Pamplona et al. (2005)	gering	gering	gering	hoch	gering	gering	gering	gering	gering
Pamplona et al. (2012)	gering	gering	gering	hoch	gering	gering	gering	gering	gering
Pamplona et al. (2014)	gering	gering	gering	hoch	gering	gering	gering	gering	gering
Kaiser et al. (2017)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Pamplona et al. (2017)	gering	gering	gering	hoch	gering	gering	gering	gering	gering
Pamplona & Ysunza (2018)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Scherer et al. (2020)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Scherer et al. (2021)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Alighieri et al. (2021)	gering	gering	gering	gering	gering	gering	gering	gering	gering
<b>Kombiniert phonetisch-phonologische Interventionen</b>									
Sell & Grunwell (1990)	gering	gering	hoch	gering	gering	gering	gering	gering	gering
Derakhshandeh et al. (2016)	gering	gering	gering	gering	gering	gering	gering	gering	gering
Luyten et al. (2016)	gering	gering	gering	gering	gering	gering	gering	gering	gering
<b>Vergleichende Studien zu phonetischer vs. phonologischer Intervention</b>									
Pamplona et al. (1999)	gering	gering	gering	hoch	gering	gering	gering	gering	gering
Alighieri et al. (2020)	gering	gering	gering	gering	gering	gering	gering	gering	gering



Im ersten Cluster (Selection Bias) weisen zwölf von 19 Studien insgesamt ein geringes RoB auf. Das verbleibende analysierte Risiko verteilt sich über die drei Kriterien die Testgruppenbeschreibung betreffend, z. B. weil unzureichende Informationen über Inklusions- und Exklusionskriterien genannt wurden. Auffällig ist das erhöhte RoB bei fünf von neun Kriterien der phonetischen Studien.

Im zweiten Cluster (Performance Bias) sind acht der 19 Studien durch ein geringes RoB charakterisiert. Das Cluster ist mit 24% der Kriterien am deutlichsten von hohem bzw. unklarem RoB betroffen. Hier zeigen sich Risiken bei acht von 16 Studien in der Durchführung der Transkription, z. B. keine mit statistischen Werten belegte Intra- und/oder Interrater-Reliabilität. Bei fünf Studien liegt das hohe RoB-Problem im Bereich der Diagnostikverfahren, z. B. hauptsächlich, weil die Verfahren nicht standardisiert und/oder veröffentlicht waren (s. Tab. 3).

Das dritte Cluster (Outcome Bias) demonstriert die wenigsten Einschränkungen, da 89% der analysierten Kriterien mit geringem RoB eingestuft werden konnten. Hierin weisen zwölf der 19 Studien ein geringes RoB auf. Die beiden Kriterien ‚Vollständigkeit der Ergebnisse‘ und ‚Endpunkte‘ zeigen für alle implizierten Studien ein geringes RoB. Sechs der Studien charakterisierten

sich durch hohes bzw. unklares RoB im Bereich der Verblindung der Ergebnisdaten, da entweder keine Verblindung stattfand oder Unklarheit über diese bestand. Im Folgenden werden die Studien je nach Interventionsform eingeschätzt und pro Studie genaue Gründe für das RoB genannt.

#### 4.1 Phonetische Interventionen

Ein unklares bzw. hohes RoB konnte bei 32 % aller Kriterien bei Studien, die phonetische Interventionen verwendeten ( $n = 3$ ), herausgestellt werden.

Die größte Schwierigkeit ist im ersten Cluster (Selection Bias) erkennbar. Hier sind fünf von neun Kriterien durch hohes bzw. unklares RoB belastet. Jahanbin et al. (2014) und Van Demark (1971) erwähnen oder verwenden keine Inklusions- und/oder Exklusionskriterien, während Herreras Mercado et al. (2019) diese nicht im Detail beschreibt. Die Beschreibung der Testgruppe(n) erfolgte in einer Studie unzureichend (Van Demark, 1971). Jahanbin et al. (2014) lieferten keine detaillierten Angaben zu den Proband:innen.

Im zweiten Cluster (Performance Bias) zeigte sich nur ein Kriterium mit unklarem RoB eingeschränkt: Herreras Mercado et al. (2019) führten eine Transkription durch, dokumentierten jedoch die Intra- und/oder Interrater-Reliabilität nicht mit statistischen Werten.

Das dritte Cluster (Outcome Bias) weist bei 22 % der Kriterien ein unklares RoB auf, da zwei Studien keine Angaben lieferten, ob eine Verblindung der Ergebnisse für die Analyse erfolgte (Jahanbin et al., 2014; Van Demark, 1971).

#### 4.2 Phonologische Interventionen

Die Analyse der Studien mit phonologischer Intervention ( $n = 11$ ) zeigt insgesamt nur bei 15 % aller Kriterien ein unklares bzw. hohes RoB.

Das erste Cluster (Selection Bias) weist im Vergleich zu den anderen beiden Clustern das geringste RoB auf. Es zeigte nur bei zwei Studien (Kaiser et al., 2017; Scherer et al., 2021) ein RoB unklarer Art, da heterogene Testgruppen (Alter, Fehlbildungsart) miteinander verglichen wurden. Das zweite Cluster (Performance Bias) ist vollständig unauffällig im Kriterium der ‚Datenerhebung‘. Es ist jedoch durch ein hohes RoB im Kriterium ‚Diagnostikverfahren‘ bei vier Studien charakterisiert, da die verwendeten Diagnostikverfahren nicht standardisiert und/oder veröffentlicht waren (Pamplona et al., 2005, 2012, 2014, 2017). Weiterhin zeigten sechs von elf Studien ein unklares RoB im Kriterium der Transkription, weil keine Daten zu statistischen Werten bzgl. der Intra- und/oder Interraterreliabilität angegeben wurden (Pamplona et al., 2004, 2005, 2012, 2014, 2017; Pamplona & Ysunza, 2018). Das dritte Cluster (Outcome Bias) weist einzig im Bereich ‚Verblindung‘ Auffälligkeiten im RoB auf: Ein hohes RoB ist bei Kaiser et al. (2017) wegen fehlender Verblindung zu verzeichnen. In den Studien von Pamplona et al. (2005) und Van Demark & Hardin (1986) ist unklar, ob eine Verblindung stattfand (unklares RoB).

#### 4.3 Kombiniert phonetisch-phonologische Interventionen

Die Studien ( $n = 3$ ), die kombiniert phonetisch-phonologische Interventionen einsetzen, zeigen insgesamt ein unklares bzw. hohes RoB nur bei 8 % aller Kriterien.

Das erste Cluster (Selection Bias) weist nur bei einer Studie (Sell & Grunwell, 1990) eine Auffälligkeit (hohes RoB) auf, da hierin keine Inklusions- und Exklusionskriterien benannt wurden. Das zweite Cluster (Performance Bias) zeigt für alle Studien in allen Kriterien ein geringes RoB. Das dritte Cluster (Outcome Bias) weist für eine Studie bei einem Kriterium ein unklares RoB auf: Bei Sell & Grunwell (1990) bleibt unklar, ob eine Verblindung der Ergebnisse für die Auswertung stattfand.

#### 4.4 Vergleichende Studien zu phonetischer vs. phonologischer Intervention

Die beiden vergleichenden Studien zeigen insgesamt ein unklares bzw. hohes RoB bei 17 % aller Kriterien. Im dritten Cluster (Outcome Bias) liegen keinerlei Einschränkungen im RoB vor (geringes RoB = 100 %). Darauf folgt das erste Cluster (Selection Bias) mit einem unklaren RoB bei einer Studie, die bezüglich der Zusammensetzung der Testgruppen ungenaue Angaben machte (Pamplona et al., 1999). Die Einschränkung (hohes und unklares RoB) im zweiten Cluster (Performance Bias) liegt bei 33 % der Kriterien, da die Studie von Pamplona et al. (1999) nicht veröffentlichte/standardisierte Diagnostikverfahren integrierte und keine Reliabilitätsmessungen dokumentierte.

## 5 Diskussion

Die vorliegende systematische Übersichtsarbeit verfolgte das Ziel, Evidenzen für die Effektivität phonetischer und/oder phonologischer sprachtherapeutischer Interventionen bei Kindern bzw. Jugendlichen mit LKGSF und LKGSF-typischen Artikulationsauffälligkeiten (CTCs) herauszustellen. In diesem Rahmen wurden die eingeschlossenen Studien auch in Hinblick auf deren RoB analysiert. Es konnten 19 Referenzen in die Analyse eingeschlossen werden, die phonetische ( $n = 3$ ), phonologische ( $n = 11$ ), kombinierte phonetisch-phonologische Interventionen ( $n = 3$ ) als auch interventionsvergleichende Ansätze ( $n = 2$ ) umfassten. Es wurde gezeigt, dass sich sowohl die phonologischen als auch die phonetischen und kombinierten phonetisch-phonologischen Therapiekonzepte als effektiv erweisen, da fast alle gefundenen Referenzen Verbesserungen nach der Therapie mit der jeweiligen Intervention belegten. Fünf der 19 Studien untersuchten die Effektivität phonetischer Verfahren, sowohl als Versuchsgruppe ( $n = 3$ ), als auch als Vergleichsgruppe zur phonologischen Intervention ( $n = 2$ ). Insgesamt wurden jedoch nur 48 Proband:innen einbezogen, was die Aussagekraft deutlich mindert. Zwei Studien setzten das Konzept nach Van Riper ein, eine Studie arbeitete nach PROMPT und zwei weitere benannten kein konkretes Konzept. Interessant ist, dass in vier der fünf Studien eine hohe Therapiefrequenz gewählt wurde: zwei Studien nutzten eine Intensivtherapie, zwei weitere eine hochfrequente wöchentliche Therapie. Drei der fünf Studien berichteten signifikante Verbesserungen in den Artikulationsleistungen ihrer Proband:innen, eine Studie kam nach einem längeren Zeitraum (14-46 Mo.) zum Ergebnis des vollständigen Abbaus kompensatorischer Artikulation. Nur eine Studie konnte keine klinisch bedeutsamen Fortschritte herausstellen. Erfolge durch phonetische Therapieansätze stehen im Einklang mit der internationalen Grundlagenliteratur zur sprachtherapeutischen Herangehensweise bei Kindern mit LKGSF (Howard & Lohmander, 2011; Peterson-Falzone et al., 2010). Die Mehrzahl der analysierten Studien ( $n = 13$ ) evaluierten jedoch phonologische Ansätze entweder als Versuchsgruppe ( $n = 11$ ) oder im Vergleich zu phonetischen Ansätzen ( $n = 2$ ). In diesen Studien wurden 493 Proband:innen (von insgesamt 569) eingeschlossen. Vier verschiedene phonologisch-linguistische Ansätze wurden verwendet: der Multiple-Sound-Ansatz ( $n = 1$ ), die naturalistische Intervention EMT+PE ( $n = 3$ ), METAPHON ( $n = 2$ ) sowie die phonologische Intervention im Rahmen des WLM ( $n = 6$ ). Eine Studie gab kein genaues Konzept an. Auch in diesem Studienbereich wurde verstärkt ein hochfrequentes Setting gewählt. Als Intensivtherapie gestalteten sich fünf Studien, vier Studien setzten ein hochfrequentes Therapieangebot (2-3x/Wo.) ein, eine Studie bot wöchentliche Therapie an, und zwei Studien verglichen die Effektivität von wöchentlicher bzw. hochfrequenter vs. intensiver Therapie.

Nahezu alle Referenzen konnten signifikante Ergebnisse im Sinne von Artikulationsverbesserungen erzielen. Nur zwei von drei Studien, die den Einsatz von EMT+PE evaluierten, konnten bei den Proband:innen keine signifikanten Verbesserungen verzeichnen. Es wurde dargelegt, dass phonologisch ausgerichtete Intensivtherapien (täglich über eine Gesamtdauer von 2-4 Wochen) eine sehr gute Effektivität erreichen. Pamplona et al. (2005) belegten, dass sich eine Intensivtherapie über drei Wochen genauso effektiv zeigt wie eine wöchentliche Sprachtherapie über die Dauer von einem Jahr. Darüber hinaus wurde eine signifikant kürzere Therapiedauer bei wöchentlicher Therapie unter dem Einsatz phonologischer Ansätze vs. rein phonetischer belegt. Hierbei verkürzte sich der Zeitraum um die Hälfte, bis die CTCs bei den Proband:innen abgebaut wurden.

Bei Betrachtung der eingesetzten Therapiemethoden kam unter anderem METAPHON zum Einsatz und demonstrierte gute Ergebnisse. Hier ist jedoch die geringe Anzahl der Studien ( $n = 2$ ) sehr kritisch zu betrachten, die die Aussagekraft der Evidenz einschränkt.

Das Team um Pamplona und Kolleg:innen evaluierte in sechs Studien phonologische Interventionen im Rahmen des WLM. Dieses Therapiekonzept verfolgt die Annahme, dass bessere und schnellere Ergebnisse in der Kommunikationsfertigkeit bei Proband:innen erzielt werden können, wenn simultan an semantisch-lexikalischen, grammatischen und phonetisch-phonologischen Items im Kontext sinnvoller Kommunikation gearbeitet wird. Die Studien zeigten nahezu ausschließlich signifikante Verbesserungen. Es wurden zudem jedoch auch additive Methoden impliziert (z. B. audiovisuelles Material). Der Einsatz von audiovisuellem Material, wie Hörgeschichten mit passenden Ausmalbüchern, konnte zu signifikant besseren Ergebnissen führen. Dies betonte bereits Golding-Kushner (2001) in ihrem Standard-Therapiewerk, indem sie die Wichtigkeit des Einsatzes von visuellen und taktilen Stimuli bei der Aussprachetherapie von Kindern mit LKGSF hervorhebt. Für den deutschsprachigen Bereich liegen derzeit jedoch keine ex-

plizit auf LKGSF ausgerichteten audiovisuellen Therapiematerialien vor. Dies kann zukünftig ein sinnvolles Forschungsdesiderat darstellen. Es kam weiterhin das Konzept der Reflektierten Praxis (Deliberate Practice) zum Einsatz, was mit seinem sehr hohem Übungsfaktor pro Zielitem zu guten Ergebnissen führte. Diese Methode ähnelt dem ‚drill‘ mit hoher Inputspezifizierung und maximaler Wiederholung, was in den USA zum Standard in der Sprachtherapie gehört. Hierzu gibt es im Bereich der Sprachentwicklungsstörungen (SES) international belegte Evidenzen (Cleave, Becker, Curran, Van Horne, & Fey, 2015; Plante, Mettler, Tucci, & Vance, 2019). Auch die gezielte Abstimmung von Therapietechniken (wie Modellieren, Modellieren mit Betonung des Zieltargets mit vorheriger Pause, Einsatz von phonologischen Cues) auf das betroffene Kind zeigte signifikante Verbesserungen. Da sowohl rein phonetische als auch phonologische Ansätze zu klinisch bedeutsamen Verbesserungen in der Aussprache der einbezogenen Proband:innen beitrugen, erwiesen sich Studien mit kombinierter phonologisch-phonetischer Intervention genauso effektiv. Es zeigten alle Proband:innen ( $n = 28$ ) Verbesserungen ihrer Artikulationsleistungen.

Die Qualität und Aussagekraft der analysierten Studien erweisen sich als heterogen. In Bezug auf die Analyse der Evidenzlevel können acht Studien als RCT im Bereich II eingeordnet werden, was als Goldstandard bei prospektiven Studien gilt. Hinsichtlich der Einschätzung der Effektivität muss das Evidenzlevel jedoch in Bezug auf das RoB betrachtet werden. Hier fällt auf, dass nur zwei der RCT-Studien ein komplett unauffälliges (geringes) RoB zeigen, was die Aussagekraft der übrigen RCTs einschränkt. Rein phonetisch ausgelegte Studien erwiesen sich hierbei am stärksten eingeschränkt. Ihr Evidenzlevel ist niedrig (IV) und das RoB hoch. Bei den kombiniert phonetisch-phonologischen Studien ist das Evidenzlevel zwar als niedrig einzuschätzen (IV), das RoB über die Studien hinweg jedoch als gering zu betrachten (s. Kapitel 4.3).

Zukünftige Studien zur Sprachtherapieeffektivität bei Kindern und Jugendlichen mit LKSGF und CTCs sollten ein hohes Evidenzlevel anstreben und keinerlei RoB aufweisen. Hierbei sollte sich beim Studiendesign und bei der Publikation der Ergebnisse an den aktuellen Cochrane-Leitlinien (Higgins et al., 2022) sowie dem PRISMA-Statement (Page et al., 2021) orientiert werden. Nur mittels zuverlässiger Studiendesigns ohne RoB, wie z. B. vollständig vorliegende Informationen zum Studiendesign, der Studiendurchführung und der Datenauswertung, einer größtmöglichen Verblindung und einer großen Anzahl von Proband:innen, können aussagekräftige Evidenzen gewonnen werden. Unter Berücksichtigung des PRISMA-Statements sind in Cluster 1 (Selection Bias) die Verwendung von Inklusions- und Exklusionskriterien ebenso wie die Dokumentation studienrelevanter Charakteristika der Proband:innen unabdingbar (Page et al., 2021). Im zweiten Cluster (Performance Bias) könnte das RoB im vorliegenden Forschungsbereich zukünftig minimiert werden, wenn ausschließlich standardisierte, publizierte Ergebnismessungen zum Einsatz kämen, alle Sprachkorpi nach IPA transkribiert, und dazu die Intra- und/oder Interrater-Reliabilität mit Angabe statistischer Werte dokumentiert werden würden. Das RoB im dritten Cluster (Outcome Bias) könnte weiterhin minimiert werden, würden zukünftige Studien ihre Ergebnisdaten verblindet auswerten und über diese Verblindung bei der Veröffentlichung berichten.

## 5.1 Limitationen

Schlosser, Wendt & Sigafos (2007) empfehlen den Einbezug von Büchern, Buchkapiteln und unveröffentlichter Literatur in die Analyse, um eine Verzerrung der Ergebnisse zu vermeiden. Dies wurde in dem vorliegenden Review nicht umgesetzt, was die Vollständigkeit der Ergebnisse möglicherweise begrenzt. Zudem konnten mehrere Datenbanken nicht einbezogen werden, da kein Zugriff auf deren Inhalte bestand.

## 6 Fazit

Zusammenfassend konnte das vorliegende systematische Review zeigen, dass sowohl rein phonetische, phonologische und kombinierte Interventionen bei Kindern und Jugendlichen mit LKGSF und CTCs zu signifikanten Ergebnissen führten. Die Studienlage bzw. Studienqualität (RoB) reicht jedoch aktuell noch nicht aus, um eindeutige Schlüsse ziehen zu können. Phonologische Therapieansätze erwiesen sich bisher als besonders effektiv und im Vergleich zu phonetischen Ansätzen als wirksamer. Sie führten bei gleicher Dauer und Frequenz der TE in kürzerer Zeit zum angestrebten Ziel. Hier wären zukünftig jedoch vermehrt vergleichende Studiendesigns vonnöten, um diese (noch schwache) Evidenz ggf. stärken zu können.

Darüber hinaus erwiesen sich auch die Studien, die beide Ansätze miteinander kombinierten, als effektiv. Traditionelle phonetische Lautanbahnung kann Kindern mit LKGSF helfen, einzelne Laute zielgerichtet zu erlernen, wobei sich der Einbezug mehrerer Sinne (visuell, taktil-kinästhetisch, auditiv) als vorteilhaft zeigt. Auch hochfrequent dargebotene Therapieeinheiten und Intensivtherapien mit phonologischer Beteiligung sind als effektiv herauszustellen. Hier ist die Studienlage aber auch auszubauen, um diese Schlussfolgerung zu bestätigen. Besonders wichtig erscheint außerdem die phonologische Herangehensweise der Kontrastierung distinktiver Merkmale von Phonemen, die durch den Kontrast nasal-oral bei Kindern mit LKGSF und CTCs zu ergänzen wäre (Howard & Lohmander, 2011). Im deutschsprachigen Bereich werden bei Kindern mit LKGSF nur wenige Intensivtherapien angeboten. Neu ist die in Salzburg entwickelte Kurzzeit-Intensiv-Therapie nach König, Specht-Moser & Neumann (2019), die zuletzt erfolgreich evaluiert wurde (König, 2022). In dieser werden als Gruppen- und Einzeltherapie mit täglichen phonetisch-phonologischen Therapieeinheiten innerhalb von einer Woche signifikante Verbesserungen in der Aussprache von Kindern erzielt, die auch bei Follow-up-Messungen bestehen bleiben. Für die wohnortnahe Sprachtherapie wäre es anzustreben, (wo notwendig) vermehrt phonologische Anteile in die Therapie zu integrieren und insbesondere die Frequenz der Therapieeinheiten auf 2-5x wöchentlich zu erhöhen, um den betroffenen Kindern eine möglichst kurze, aber maximal effiziente Sprachtherapie zu ermöglichen.

Die als Anhang I-V betitelten Zusatzmaterialien können unter dem folgenden Link eingesehen werden: <https://kurzelinks.de/sRLKGSF>

## 7 Literaturverzeichnis

- Alighieri, C., Bettens, K., Bruneel, L., D'haeseleer, E., Gaever, E. V., & Lierde, K. V. (2020). Effectiveness of Speech Intervention in Patients With a Cleft Palate: Comparison of Motor-Phonetic Versus Linguistic-Phonological Speech Approaches. *Journal of Speech, Language, and Hearing Research*, 63(12), 3909–3933. [https://doi.org/10.1044/2020\\_JSLHR-20-00129](https://doi.org/10.1044/2020_JSLHR-20-00129)
- Alighieri, C., Van, L. K., De, C. A.-S., Demuyneck, K., Bruneel, L., D, 'haeseleer Evelien, & Bettens, K. (2021). Is High-Intensity Speech Intervention Better? A Comparison of High-Intensity Intervention Versus Low-Intensity Intervention in Children With a Cleft Palate. *Journal of Speech, Language, and Hearing Research*, 64(9), 3398–3415. [https://doi.org/10.1044/2021\\_JSLHR-21-00189](https://doi.org/10.1044/2021_JSLHR-21-00189)
- Bessell, A., Sell, D., Whiting, P., Roulstone, S., Albery, L., Persson, M., ... Ness, A. R. (2013). Speech and Language Therapy Interventions for Children with Cleft Palate: A Systematic Review. *The Cleft Palate Craniofacial Journal*, 50(1), 1–17. <https://doi.org/10.1597/11-202>
- Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O., & Fey, M. E. (2015). The Efficacy of Recasts in Language Intervention: A Systematic Review and Meta-Analysis. *American Journal of Speech-Language Pathology*, 24(2), 237–255. [https://doi.org/10.1044/2015\\_AJSLP-14-0105](https://doi.org/10.1044/2015_AJSLP-14-0105)
- Costello, J., & Onstine, J. M. (1976). The Modification of Multiple Articulation Errors Based on Distinctive Feature Theory. *Journal of Speech and Hearing Disorders*, 41(2), 199–215. <https://doi.org/10.1044/jshd.4102.199>
- Covidence systematic review software. (o. J.). Melbourne, Australia: Veritas Health Innovation. Abgerufen von [www.covidence.org](http://www.covidence.org)
- Crowe, K., Cuervo, S., Guiberson, M., & Washington, K. N. (2021). A Systematic Review of Interventions for Multilingual Preschoolers With Speech and Language Difficulties. *Journal of Speech, Language, and Hearing Research*, 64(11), 4413–4438. [https://doi.org/10.1044/2021\\_JSLHR-21-00073](https://doi.org/10.1044/2021_JSLHR-21-00073)
- Dean, E. C., Howell, J., Waters, D., & Reid, J. (1995). Metaphon: A metalinguistic approach to the treatment of phonological disorder in children. *Clinical Linguistics & Phonetics*, 9(1), 1–19. <https://doi.org/10.3109/02699209508985318>
- Derakhshandeh, F., Nikmaram, M., Hosseinabad, H. H., Memarzadeh, M., Taheri, M., Omrani, M., ... Sell, D. (2016). Speech characteristics after articulation therapy in children with cleft palate and velopharyngeal dysfunction – A single case experimental design. *International Journal of Pediatric Otorhinolaryngology*, 86, 104–113. <https://doi.org/10.1016/j.ijporl.2016.04.025>
- Golding-Kushner, K. J. (2001). *Therapy Techniques for Cleft Palate Speech & Related Disorders*. San Diego: Singular.
- Grunwell, P., & Harding, A. (1993). The Relationship Between Speech and Timing of Palatal Repair. In *Analyzing Cleft Palate Speech* (pp. 48–82). London: Whurr.
- Harding, A., & Grunwell, P. (1996). Characteristics of cleft palate speech. *European Journal of Disorders of Communication: The Journal of the College of Speech and Language Therapists, London*, 31(4), 331–357. <https://doi.org/10.3109/13682829609031326>
- Hayden, D. (2015). *PROMPT Introduction to Technique Manual* (Revised edition). Santa Fe, New Mexico: PROMPT Institute.
- Herrerias Mercado, R., Simpson, K., & Bellom-Rohrbacher, K. H. (2019). Effect of Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT) on Compensatory Articulation in Children With Cleft Palate/Lip. *Global Pediatric Health*, 6. <https://doi.org/10.1177/2333794X19851417>
- Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2022). *Cochrane Handbook for Systematic Reviews of Interventions*. Abgerufen 6. Januar 2023, von [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook)
- Howard, S., & Lohmander, A. (2011). *Cleft Palate Speech: Assessment and Intervention*. West Sussex, UK: Wiley-Blackwell.
- Jahanbin, A., Pahlavannezhad, M. R., Savadi, M., & Hasanzadeh, N. (2014). The effect of speech therapy on acoustic speech characteristics of cleft lip and palate patients: A preliminary study. *Special Care in Dentistry*, 34(2), 84–87. <https://doi.org/10.1111/scd.12031>

- Kaiser, A. P., Scherer, N. J., Frey, J. R., & Roberts, M. Y. (2017). The Effects of Enhanced Milieu Teaching With Phonological Emphasis on the Speech and Language Skills of Young Children With Cleft Palate: A Pilot Study. *American Journal of Speech-Language Pathology*, 26(3), 806–818. [https://doi.org/10.1044/2016\\_AJSLP-16-0008](https://doi.org/10.1044/2016_AJSLP-16-0008)
- König, C. (2022). *Entwicklung und Evaluation einer 6-Tages-Intensivtherapie bei Kindern (5-10 Jahre) mit Lippen-Kiefer-Gaumen-Segel-Fehlbildung*. Universität Erfurt, Erfurt.
- König, C., Specht-Moser, B., & Neumann, S. (2019). 6-Tage-Intensiv-Camp für Kinder mit LKGS-Fehlbildung. Grundlagen, Aufbau und erste Evidenzen. *LogoTHEMA*, 16(1), 7–13.
- Kummer, A. W. (2014). *Cleft Palate and Craniofacial Anomalies. The Effects on Speech and Resonance* (3. Aufl.). San Diego: Thomson.
- Luyten, A., Bettens, K., D'haeseleer, E., Hodges, A., Galiwango, G., Vermeersch, H., & Van Lierde, K. (2016). Short-term effect of short, intensive speech therapy on articulation and resonance in Ugandan patients with cleft (lip and) palate. *Journal of Communication Disorders*, 61, 71–82. <https://doi.org/10.1016/j.jcomdis.2016.03.006> therapy. *Acta Symbolica*, (6:1).
- Meinusch, M. (2010). *Orofaziale Dysfunktionen bei LKGS-Fehlbildungen: Ein systematisches Review* (Unveröffentlichte Diplomarbeit an der Universität zu Köln, Humanwissenschaftliche Fakultät)
- Mossey, P. A., & Modell, B. (2012). Epidemiology of Oral Clefts 2012: An International Perspective. In M. T. Coubourne (Hrsg.), *Cleft Lip and Palate: Epidemiology, Aetiology and Treatment* (Bd. 16). S.Karger AG. <https://doi.org/10.1159/000337464>
- Neumann, S. (2010). *Sprachtherapeutische Diagnostik bei Menschen mit Lippen-Kiefer Gaumen-Segel-Fehlbildung*. Hamburg: Verlag Dr. Kovač.
- Neumann, S. (2017). Lippen-Kiefer-Gaumen-Segel-Fehlbildungen (LKGSF). In *Kompandium der akademischen Sprachtherapie und Logopädie* (Bd. 3, S. 273–287). München: Kohlhammer.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. <https://doi.org/10.1136/bmj.n71>
- Pamplona, M. del C., Ysunza, A., Chavelas, K., Arámburu, E., Patiño, C., Martí, F., & Morales, S. (2012). A Study of Strategies for Treating Compensatory Articulation in Patients with Cleft Palate. *Journal of Maxillofacial and Oral Surgery*, 11(2), 144–151. <https://doi.org/10.1007/s12663-011-0314-x>
- Pamplona, M. del C., Ysunza, A., & Espinosa, J. (1999). A comparative trial of two modalities of speech intervention for compensatory articulation in cleft palate children, phonologic approach versus articulatory approach. *International Journal of Pediatric Otorhinolaryngology*, 49(1), 21–26. [https://doi.org/10.1016/S0165-5876\(99\)00040-3](https://doi.org/10.1016/S0165-5876(99)00040-3)
- Pamplona, M. del C., Ysunza, A., & Morales, S. (2014). Strategies for Treating Compensatory Articulation in Patients with Cleft Palate. *International Journal of Biomedical Science: IJBS*, 10(1), 43–51.
- Pamplona, M. del C., Ysunza, A., Patiño, C., Ramírez, E., Drucker, M., & Mazón, J. J. (2005). Speech summer camp for treating articulation disorders in cleft palate patients. *International Journal of Pediatric Otorhinolaryngology*, 69(3), 351–359. <https://doi.org/10.1016/j.ijporl.2004.10.012>
- Pamplona, M. del C., Ysunza, A., & Ramírez, P. (2004). Naturalistic intervention in cleft palate children. *International Journal of Pediatric Otorhinolaryngology*, 68(1), 75–81. <https://doi.org/10.1016/j.ijporl.2003.09.007>
- Pamplona, M. del C., Ysunza, A., & Urióstegui, C. (1996). Linguistic interaction: The active role of parents in speech therapy for cleft palate patients. *International Journal of Pediatric Otorhinolaryngology*, 37(1), 17–27. [https://doi.org/10.1016/0165-5876\(96\)01362-6](https://doi.org/10.1016/0165-5876(96)01362-6)
- Pamplona, M. del C., & Ysunza, P. A. (2018). Deliberate Practice: Preliminary Results of a Useful Strategy for Correcting Articulation in Children With Cleft Palate. *Journal of Craniofacial Surgery*, 29(6), 1490–1494. <https://doi.org/10.1097/SCS.00000000000004707>
- Pamplona, M. del C., Ysunza, P. A., & Morales, S. (2017). Audiovisual materials are effective for enhancing the correction of articulation disorders in children with cleft palate. *International Journal of Pediatric Otorhinolaryngology*, 93, 17–23. <https://doi.org/10.1016/j.ijporl.2016.12.011>
- Peterson-Falzone, S. J., Hardin-Jones, M. A., & Karnell, M. P. (2010). *Cleft Palate Speech*. Mosby/Elsevier.
- Plante, E., Mettler, H. M., Tucci, A., & Vance. (2019). Maximizing Treatment Efficiency in Developmental Language Disorder: Positive Effects in Half the Time. *American Journal of Speech-Language Pathology*, 28(3), 1233–1247. [https://doi.org/10.1044/2019\\_AJSLP-18-0285](https://doi.org/10.1044/2019_AJSLP-18-0285)
- Sand, A., Hagberg, E., & Lohmander, A. (2022). On the Benefits of Speech-Language Therapy for Individuals Born With Cleft Palate: A Systematic Review and Meta-Analysis of Individual Participant Data. *Journal of Speech, Language, and Hearing Research*, 65(2), 555–573. [https://doi.org/10.1044/2021\\_JSLHR-21-00367](https://doi.org/10.1044/2021_JSLHR-21-00367)
- Scherer, N. J., Kaiser, A. P., Frey, J. R., Lancaster, H. S., Lien, K., & Roberts, M. Y. (2020). Effects of a naturalistic intervention on the speech outcomes of young children with cleft palate. *International Journal of Speech-Language Pathology*, 22(5), 549–558. <https://doi.org/10.1080/17549507.2019.1702719>
- Scherer, N. J., Yamashita, R., de Oliveira, D. N., DiLallo, J., Trindade, I., Fukushiro, A. P., & Richards, K. (2021). Early speech and language intervention in Brazilian-Portuguese toddlers with cleft lip and/or palate. *Clinical Linguistics & Phonetics*, 36(1), 34–53. <https://doi.org/10.1080/02699206.2021.1912187>
- Schlosser, R. W., Wendt, O., & Sigafos, J. (2007). Not all systematic reviews are created equal: Considerations for appraisal. *Evidence-Based Communication Assessment and Intervention*, 1(3), 138–150. <https://doi.org/10.1080/17489530701560831>
- Sell, D. A., & Grunwell, P. (1990). Speech results following late palatal surgery in previously unoperated Sri Lankan adolescents with cleft palate. *The Cleft Palate Journal*, 27(2), 162–168; discussion 174–175. [https://doi.org/10.1597/1545-1569\(1990\)027<0162:srflps>2.3.co;2](https://doi.org/10.1597/1545-1569(1990)027<0162:srflps>2.3.co;2)
- The Joanna Briggs Institute. (2014). The Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party: Supporting Document for the Joanna Briggs Institute Levels of Evidence and Grades of Recommendation. Abgerufen von <https://jbi.global/sites/default/files/2019-05/JBI%20Levels%20of%20Evidence%20Supporting%20Documents-v2.pdf>
- Van Demark, D. R. (1971). Articulatory changes in the therapeutic process. *The Cleft Palate Journal*, 8, 159–166.
- Van Demark, D. R., & Hardin, M. A. (1986). Effectiveness of intensive articulation therapy for children with cleft palate. *The Cleft Palate Journal*, 23(3), 215–224.
- Wohlleben, U. (2004). *Die Verständlichkeitsentwicklung von Kindern mit Lippen-Kiefer-Gaumen-Segel-Spalten. Eine Längsschnittstudie über spalttypische Charakteristika und deren Veränderung*. Idstein: Schulz-Kirchner.

## Zu den Autorinnen

*Miriam Grunwald*, M.A. (geb. Platte) ist wissenschaftliche Mitarbeiterin im Arbeitsbereich „Inklusive Bildungsprozesse bei Beeinträchtigungen von Sprache und Kommunikation“ an der Universität Erfurt und promoviert im Bereich der LKGS-Fehlbildungen.

*Prof. Dr. Sandra Neumann* leitet den Arbeitsbereich „Inklusive Bildungsprozesse bei Beeinträchtigungen von Sprache und Kommunikation“ an der Universität Erfurt. Ihre Arbeits- und Forschungsschwerpunkte liegen im Bereich der LKGS-Fehlbildungen, der ICF-CY, kommunikativer Partizipation und Verständlichkeit im Kontext.

## Korrespondenzadresse

Miriam Grunwald  
Inklusive Bildungsprozesse bei Beeinträchtigungen von Sprache und Kommunikation  
Fachgebiet Sonder- und Sozialpädagogik  
Erziehungswissenschaftliche Fakultät  
Universität Erfurt  
Nordhäuser Str. 63  
99089 Erfurt  
E-Mail: [miriam.grunwald@uni-erfurt.de](mailto:miriam.grunwald@uni-erfurt.de)

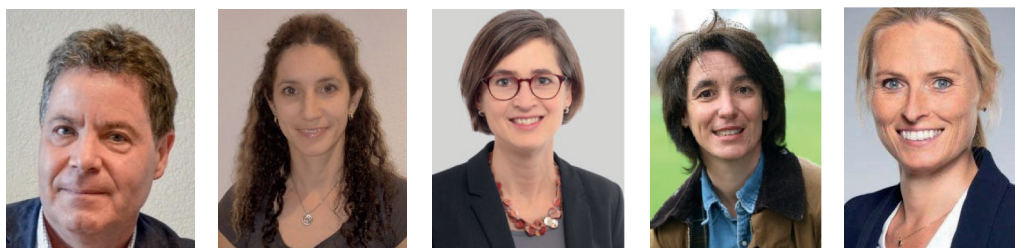
### **Erklärung zu Interessenkonflikten**

Die Autorinnen erklären, dass es keine potenziellen Interessenkonflikte sowie finanzielle Unterstützung in Bezug auf die Forschung, Autorenschaft und/oder Veröffentlichung dieses Artikels gibt.

### **ORCID iDs**

Miriam Grunwald <https://orcid.org/0000-0003-1645-1386>  
Sandra Neumann, PhD <https://orcid.org/0000-0003-1220-9787>





## Prävalenz und Prädiktion von Nichtresponsivität im Lichte unterschiedlicher RTI-Kriterien\*

Sekundäranalyse und Befunde aus einer präventiven Interventionsstudie zur Leseflüssigkeit

## Prevalence and Prediction of Non-Responsiveness in the Light of Different RTI Criteria

Secondary Analysis and Findings from a Preventive Intervention Study on Reading Fluency

Erich Hartmann, Verena Hofmann, Caroline Villiger, Annette Tettenborn, Silke Hauri

### Zusammenfassung

**Hintergrund:** Kinder mit inadäquatem Fortschritt in präventiven Leseinterventionen bedürfen der vermehrten Aufmerksamkeit von Forschung und Praxis. Die Identifikation von solchen „Nonrespondern“ (NR) erfolgt bislang anhand uneinheitlicher Kriterien, deren Einfluss auf die Prävalenz und die Vorhersage von Nichtresponsivität noch wenig erforscht ist.

**Ziel:** Die Studie untersucht, wie verschiedene NR-Kriterien die Prävalenz und die Prädiktion von Nichtresponsivität bei präventiv geförderten Kindern mit Leseschwierigkeiten beeinflussen.

**Methode:** Es werden sechs NR-Kriterien definiert und auf die Daten von 130 Drittklässler:innen angewendet, die ein präventives Training der Leseflüssigkeit erhielten. Die explorierten Kriterien differieren hinsichtlich der Methode (Endstatus, Zuwachs, duale Diskrepanz) bzw. der Referenzgruppe (normbasiert vs. stichprobenbasiert) zur Bestimmung von NR.

**Ergebnisse:** Die Studie liefert Evidenz für einen deutlichen Einfluss des NR-Kriteriums auf die Prävalenz von unresponsiven Kindern. Die NR-Anteile liegen zwischen 11 % und 67 %. Normbasierte Kriterien führen zu höheren Prävalenzraten als stichprobenbasierte. Variierende NR-Kriterien beeinflussen auch die Vorhersage eines inadäquaten Outcomes, wobei sich nur zwei Prädiktoren (Leseleistung zum Interventionsbeginn, sozio-ökonomischer Status) über mehrere Modelle hinweg als bedeutsam erweisen.

**Diskussion/Ausblick:** Die Ergebnisse bestätigen und ergänzen Befunde aus angloamerikanischen Studien und verweisen auf Probleme von heterogenen Kriterien zur Beurteilung des Interventionsoutcomes von Kindern mit Leseschwierigkeiten. Weiterführende Forschung zur Identifikation von nicht responsiven Lernenden anhand von zuverlässigen und praktikablen RTI-Kriterien und evidenzbasierten Instrumenten ist notwendig.

### Schlüsselwörter

Prävention, Leseflüssigkeit, Response-to-Intervention, Nonresponder

### Abstract

**Background:** Children with inadequate progress in preventive reading interventions require increased attention from researchers and practitioners. To date, identification of such “non-responders” (NR) has been based on inconsistent criteria whose influence on the prevalence and prediction of non-responsiveness remains poorly understood.

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

**Purpose:** This study examines how different NR criteria influence the prevalence and prediction of non-responsiveness in preventively tutored children with reading difficulties.

**Method:** Six NR criteria are defined and applied to data from 130 third-grade children who received preventive reading fluency training. The explored criteria differ in terms of method (final status, growth, dual discrepancy) or reference group (norm-based vs. sample-based) for determining NR.

**Results:** The study provides evidence for a significant influence of the NR criterion on the prevalence of unresponsive children. NR proportions range from 11 % to 67 %. Norm-based criteria result in higher prevalence rates than sample-based criteria. Varying NR criteria also influence the prediction of an inadequate outcome, with only two predictors (reading achievement at intervention onset, socioeconomic status) proving significant across multiple models.

**Discussion/Outlook:** The results confirm and complement findings from Anglo-American studies and highlight problems of heterogeneous criteria for evaluating intervention outcomes of children with reading difficulties. Further research is needed to identify non-responsive learners using reliable and feasible RTI criteria and evidence-based instruments.

#### Keywords

Prevention, reading fluency, response-to-intervention, non-responder

## 1 Hintergrund und Zielsetzung

Aufgrund der in den letzten Dekaden vorangetriebenen Leseinterventionsforschung besteht ein Konsens über den Nutzen von problemspezifischen Trainings oder Förderprogrammen sowohl bei Risikolernenden als auch bei Schulkindern mit einer Lese-Rechtschreibstörung (LRS). Die empirische Befundlage spricht für die Wirksamkeit von verschiedenen Methoden für basale Lesefertigkeiten und das Leseverständnis, wobei die in Sekundärstudien berichteten Effektstärken im mittleren bis hohen Bereich liegen (z. B. Edmonds et al., 2009; Galuschka et al., 2014; Mayer & Marks, 2019; Scammacca et al., 2007; Wanzek et al., 2010). Diese positive Bilanz darf nicht darüber hinwegtäuschen, dass längst nicht alle Schüler:innen mit Leseschwierigkeiten bedeutsam von evidenzbasierten Interventionen profitieren. In der Literatur finden sich Termini wie non-responder, low/poor responder oder treatment resister, um auf solche Kinder zu verweisen (z. B. Al Otaiba & Fuchs, 2002). Nach Vaughn und Mitarbeitende (2009, S. 167; Orig. Engl.) sind „low responders (...) von besonderem Interesse, weil sie eine einzigartige Stichprobe für die Untersuchung konzeptioneller Fragen im Zusammenhang mit der Leseentwicklung, dem Unterricht und der Sonderpädagogik darstellen“.

So ist „Nichtresponsivität“ im mehrstufigen *Response-to-Intervention-Modell* zur Identifikation und Prävention von schulischen Lernstörungen das zentrale Kriterium für die Intensivierung von Fördermaßnahmen für Risikolernende auf Ebene II (sekundäre Prävention) und schließlich für eine vertiefte Diagnostik und sonderpädagogische Interventionen für Kinder mit nachhaltigen Schriftsprachdefiziten auf Ebene III (tertiäre Prävention) (vgl. Barth et al., 2008; Hartmann, 2008; McMaster et al., 2005; Voß et al., 2016). Als Fachleute für die Diagnostik, Prävention und Therapie von sprachlichen und schriftsprachlichen Entwicklungsbeeinträchtigungen spielen Logopäd:innen und Sprachheilpädagog:innen in inklusiven Schulen wichtige Rollen bei der Identifikation von Kindern mit Lernschwierigkeiten sowie bei der Implementierung und Evaluation von effektiven Fördermaßnahmen für diese heterogene Zielgruppe (Hartmann, 2021; Justice, 2006; Nellis et al., 2014). Die im Mittelpunkt dieses Beitrags stehende Frage, wie Schüler:innen zuverlässig erfasst werden können, die auf gezielte Schriftsprachinterventionen nicht erfolgreich ansprechen, ist somit (auch) für Sprachtherapeut:innen von hoher Relevanz.

Die bisherige angloamerikanische Nonresponder-Forschung hat sich vornehmlich mit der Frage beschäftigt: „Who are the young children for whom best practices in reading are ineffective?“ (Al Otaiba & Fuchs, 2006, S. 414). Von deren Beantwortung erhofft man sich ein besseres Verständnis von kindlichen Lernproblemen, zuverlässige Verfahren zur Identifikation von Lernenden mit Risiken bzw. LRS sowie Informationen für optimierende Anpassungen von Interventionen für nicht erfolgreich lernende Kinder. Dabei ist die wissenschaftliche Untersuchung von individuellen Merkmalen zur Erklärung einer unzureichenden Response-to-Intervention (RTI) mit dem Sachverhalt konfrontiert, dass eine einheitliche, allgemein anerkannte Operationalisierung von Nichtresponsivität aussteht; stattdessen gelangen in Forschung und Praxis unterschiedliche Kriterien zur Anwendung, um geförderte Schulkinder als unresponsiv bzw. responsiv zu klassifizieren (vgl. Al Otaiba & Fuchs, 2002; Fletcher et al., 2011; McMaster et al., 2005). Wie ein-

zelne bisherige angloamerikanische Studien nahelegen, beeinflusst die Art des Kriteriums nicht nur die Prävalenzraten von NR in Leseinterventionsstudien, sondern auch die Bedeutung von potenziellen individuellen Prädiktoren für ein inadäquates Outcome wie beispielsweise eine unveränderte oder verschlechterte geförderte Leistung (Barth et al., 2008; Frijters et al., 2013; Hendricks et al., 2020). Zur weiteren Klärung dieses Einflusses ist zusätzliche Forschung erforderlich, und zwar vorteilhaft auch in nicht-englischsprachigen Ländern und mit anderen Zielgruppen und Interventionsinhalten.

Vor diesem Hintergrund verfolgt dieser Beitrag das Ziel, anhand einer Sekundäranalyse einer präventiven Interventionsstudie zur Leseflüssigkeit bei Schweizer Drittklässler:innen mit Leseschwierigkeiten (Villiger et al., 2019) zu explorieren, inwieweit sich unterschiedliche NR-Kriterien auf die Prävalenz und die Vorhersage eines inadäquaten Interventionsoutcomes auswirken. Ausgehend vom Forschungsstand zum NR-Konstrukt (Kap. 2) und zu Prädiktoren der RTI (Kap. 3) sowie einem Überblick zur eigenen Interventionsstudie (Kap. 4) werden die Fragestellungen (Kap. 5), die Methodik (Kap. 6) und die Befunde präsentiert (Kap. 7), diskutiert (Kap. 8) und kritisch gewürdigt (Kap. 9). Ein Ausblick rundet den Beitrag ab (Kap. 10).

## 2 Zur Heterogenität des NR-Kriteriums in der Leseinterventionsforschung

In einer Übersicht von 23 Interventionsstudien mit Risikolernenden (Kindergarten bis 1. Klasse) fanden Al Otaiba und Fuchs (2002), dass sich der NR-Anteil je nach Primärarbeit zwischen 8 % und 80 % bewegt. Zu dieser hohen Variabilität können inhaltliche und methodische Studienmerkmale beitragen. Neben Unterschieden in Bezug auf die Stichproben, die implementierten Interventionen (u. a. Inhalt, Intensität, Dauer, Implementationstreue) dürften vor allem differierende Kriterien für NR in den ausgewerteten Untersuchungen ausschlaggebend sein (Al Otaiba & Fuchs, 2002; 2006; Nelson et al., 2003). Barth et al. (2008, S. 2) führen in diesem Zusammenhang an, dass die Klassifikation von Risikolernenden als unresponsiv oder responsiv ein operationales Kriterium benötigt, welches die auf einem Kontinuum existierende individuelle RTI dichotomisiert bzw. als binäres Outcome betrachtet. Wie eingangs erwähnt, wird dieses – letztlich arbiträre – Kriterium in Forschung und Praxis uneinheitlich definiert. Die zur Anwendung gelangenden Operationalisierungen lassen sich nach Barth et al. (2008) anhand verschiedener Dimensionen beschreiben und vergleichen, wie in Tabelle 1 aufgeführt und charakterisiert (vgl. Frijters et al., 2013).

Tab. 1: Dimensionen von RTI-Kriterien in der Leseinterventionsforschung

Dimensionen	Beschreibung
Methode	Zur Bestimmung der RTI werden folgende Methoden genutzt, die alle gewisse Vorteile und Nachteile aufweisen: <ol style="list-style-type: none"> <li>1. Endstatus-Methode (final status, normalization): Zur Beurteilung des Outcomes wird der Leistungsstand am Ende der Intervention herangezogen. Die finale Leistung wird anhand normbasierter Referenzwerte (Prozentrang, T-Wert u. a.) oder kriterienbasierten Benchmarks (z. B. 30 Richtige Wörter pro Minute) beurteilt. Risikolernende, deren Endleistung das definierte Kriterium nicht erreichen, gelten als nicht responsiv, die restlichen als responsiv.</li> <li>2. Zuwachs-Methode (Growth, slope-discrepancy): Beurteilt wird der Lernfortschritt (Lernrate) im Vergleich zur Referenzgruppe, wobei hierfür u. a. CBM-Messungen oder Post-Prä-Differenz ausgewertet werden. Kinder mit unzureichendem Lerngewinn gemäß verwendetem Kriterium sind NR.</li> <li>3. Duale Diskrepanz-Methode: Kombination von Endstatus- und Zuwachs-Methoden. Als NR gelten Interventionskinder, die im Lernfortschritt und in der Endleistung diskrepanz sind zur Referenzgruppe.</li> </ol>
Referenzgruppe	Mögliche Referenzgruppen zur Beurteilung der RTI sind: <ol style="list-style-type: none"> <li>a) Eine normative Referenzgruppe (normative sample) repräsentiert den gesamten Bereich von Fähigkeiten in normierten Tests;</li> <li>b) Eine beschränkt normative Referenzgruppe (limited norm sample) umfasst den Leistungsbereich der Kinder, die an derselben Intervention teilgenommen haben (Interventionsstichprobe).</li> </ol>
Zielvariable/ Messung	Die RTI wird anhand einer einzelnen Zielvariable (z. B. Wortlesen) oder einer Kombination von mehreren Variablen (z. B. Wortlesen, Leseverstehen) bestimmt, die mittels standardisierter oder kriterienbasierter Tests oder CBM-Verfahren gemessen werden.
Trennwerte	Diverse Cut-points werden in Form von negativen Standardabweichungen (z. B. -0.5, -1), Perzentilen/Prozenträngen (z. B. PR < 16, 33), T-Werten oder anderen Indikatoren wie Median Split, kritische Differenz, Reliable Change Index (RCI) angegeben. Kinder, deren Endleistung oder Lernzuwachs den definierten Wert nicht erreichen, sind NR.

Bisherige Forschungen zur Frage, wie sich differierende Operationalisierungen auf die Prävalenz von NR auswirken, berichten mehr oder weniger stark divergierende Anteile (und Profile) von unresponsiven Lernenden je nach gewählten Kriterien. So bewegen sich die NR-Prävalenzraten in der Studie von Barth et al. (2018) zwischen 2 % und 70 %, bei Frijters et al. (2013) zwischen 14 % und 75 % und bei Hendricks et al. (2020) zwischen 20 % und 81 %.

In der bisher umfassendsten Arbeit zu dieser Fragestellung untersuchten Barth et al. (2008), wie sich systematische Unterschiede des RTI-Kriteriums hinsichtlich der Methode (Endstatus, Zuwachs, duale Diskrepanz), der Cut-Points (-0.5, -1.0, -1.5 SD) und Art der Messungen (growth assessment vs. end-of-year assessment) auf die Beurteilung des Interventionsoutcomes auswirken. Hierzu wurden zahlreiche Merkmalskombinationen auf die Datensätze von 400 Schulkindern angewendet, die in den ersten zwei Schuljahren präventive Leseinterventionen erhalten hatten. Gemäß den Ergebnissen tendieren verschiedene Operationalisierungen dazu, unterschiedliche Kinder als unresponsiv bzw. responsiv zu identifizieren. Während die Übereinstimmung der NR-Klassifikationen allgemein gering war, fiel sie bei Respondern etwas höher aus. Von den 808 analysierten Kombinationen ergaben die meisten keine Kappa-Werte mit einem Mindestmaß an Übereinstimmung (0.40). Kappa-Werte, die diesen Schwellenwert erreichten, zeigten selten eine bedeutende Übereinstimmung der Identifizierung von NR. Eine klassifikatorische Überlegenheit der Zuwachs- gegenüber der Endstatus-Methode oder der dualen Diskrepanzmethode gegenüber anderen Methoden war nicht ersichtlich. Der Cut-Point erwies sich als die wichtigste Determinante für den RTI-Status. Nach Barth et al. (2008) sollten zur Identifikation von NR bzw. von Kindern mit LRS vorteilhaft mehrere Indikatoren und Quellen (z. B. Testergebnisse, Lehrpersonenurteile) herangezogen werden, um so eine rein formale Entscheidungsfindung anhand von letztlich willkürlichen Trennwerten zu vermeiden.

Während es Hinweise gibt für den Vorteil der dualen Diskrepanzmethode, die weniger NR identifiziert als die Endstatus- und die Zuwachsmethode (McMaster et al., 2005), bleibt beim gegenwärtigen Forschungsstand ungeklärt, wie kindliche Nichtresponsivität bzw. Responsivität am besten zuverlässig, effizient und kosteneffektiv erfasst werden kann (Frijters et al., 2013; Hendricks & Fuchs, 2020; Lam & McMasters, 2014). Weiterführende Forschung zu dieser Frage ist nicht nur von wissenschaftlichem Interesse, sondern auch grundlegend für die Praxis. Damit kooperierende Lehrpersonen und sonderpädagogische Fachleute die ihnen obliegenden förderdiagnostischen Aufgaben erfolgreich bewältigen können, benötigen sie entsprechendes evidenzbasiertes Wissen und geeignete Instrumente, die es erlauben, (Risiko-)Kinder mit Lernschwierigkeiten gezielt zu erkennen, um deren Förderung intensivieren und stärker individualisieren zu können. Aus Gründen der Qualitätssicherung und der Chancengerechtigkeit ist zu fordern, dass sich die Schulen dabei auf anerkannte und einheitliche bzw. standardisierte Vorgehensweisen und Kriterien zur Beurteilung von kindlichen Lernprozessen stützen (Al Otaiba & Fuchs, 2002; Barth et al., 2008; Hendricks & Fuchs, 2020).

### 3 Prädiktoren für Nichtresponsivität

Um zu klären, welche individuellen Merkmale mit NR in präventiven Interventionen assoziiert sind, werteten Al Otaiba und Fuchs (2002) 23 Studien mit Risikolernenden (Kindergarten bis 1. Klasse) aus (vote counting Ansatz). Die in den einzelnen Arbeiten untersuchten individuellen Variablen wurden in deskriptiven Analysen als prädiktiv, nicht prädiktiv oder nicht schlüssig / gemischt klassifiziert. Die Ergebnisse sind insgesamt konform mit der linguistischen Defizithypothese, wonach Leselernproblemen bzw. LRS primär linguistisch-kognitive Defizite (phonologische Verarbeitung u. a.) zugrunde liegen (Vellutino, 1979). Der mit NR am stärksten assoziierte Faktor war ein Defizit der phonologischen Bewusstheit. Weitere prädiktive Zusammenhänge fanden sich für Schwächen in den Bereichen phonologisches Gedächtnis, schnelles Benennen, Wortschatz, (verbale) Intelligenz, orthographische Verarbeitung, Aufmerksamkeit sowie für behaviorale Schwierigkeiten und einzelne demographische Merkmale (SES, Mehrsprachigkeit) (Al Otaiba & Fuchs, 2002).

Nelson et al. (2003) wählten einen metaanalytischen Ansatz zur Identifikation von Prädiktoren für NR. In ihre Sekundäranalyse gingen 30 Leseinterventionsstudien mit Risikolernenden (Kindergarten bis 3. Klasse) ein. Den Ergebnissen entsprechend ordnen die Autoren die untersuchten Variablen nach ihrer Vorhersagekraft (Fisher's Z-transformed effect sizes) wie folgt an: Schnelles Benennen ( $Z_r = 0.51$ ), Problemverhalten ( $Z_r = 0.46$ ), phonologische Bewusstheit ( $Z_r = 0.42$ ), alphabetisches Prinzip ( $Z_r = 0.35$ ), Gedächtnis ( $Z_r = 0.31$ ), IQ ( $Z_r = 0.26$ ) und demografische

Merkmale ( $Z_r = 0.07$ ). Mit Ausnahme des letzten, nicht signifikanten Faktors korrespondieren diese Befunde mit denjenigen von Al Otaiba und Fuchs (2002).

In einer weiteren Sekundärarbeit (Lam et al., 2014) zur Prädiktion der RTI bei Risikolernenden (Kindergarten bis 3. Klasse) ergab die Analyse von 14 experimentellen Studien das Folgende: Worterkennen, alphabetisches Prinzip, Leseflüssigkeit und Phonembewusstheit waren konsistent prädiktiv, andere Variablen hingegen inkonsistent (Intelligenz, Gedächtnis) oder nicht prädiktiv (demographische Merkmale, Sprachkompetenzen).

Trotz teilweise divergierender Befunde, die in Zusammenhang mit konzeptuellen und methodischen Unterschieden von Primär- und Sekundärstudien zu sehen sind, verweist der Forschungsstand insgesamt auf linguistisch-kognitive und schriftsprachliche Faktoren, die nicht responsive Schulkinder bereits vor der Intervention von erfolgreichen Lernenden unterscheiden. Die angeführten Arbeiten lassen indes offen, inwieweit sich Unterschiede von RTI-Kriterien auf die Prädiktion von Nichtresponsivität auswirken. Zur Klärung dieser Frage analysierten Frijters und Mitarbeitende (2013) Daten von 270 Schulkindern mit Lesedefiziten (6.-8. Klasse). Die über ein Jahr implementierte Intervention (125h) adressierte Wortlesen, Leseflüssigkeit und Leseverstehen. Die Forschenden definierten vier RTI-Kriterien: a) Normalisierung relativ zur Altersnorm, b) Post-Prä-Differenz (Reliable Change Index), c) individuelle Zuwachskurve und d) Fortschritt (Post-Prä-Differenz) über mehrere Outcomemaße („within-individual gains, replicated over tests“/WIGROT), und klassifizierten die Interventionskinder dann dementsprechend. Zur Vorhersage der RTI wurden binär-logistische Regressionsanalysen mit demographischen und individuellen Prädiktoren gerechnet (Alter, Geschlecht, IQ, phonologische Bewusstheit, schnelles Benennen). Je nach Kriterium fand sich eine erhebliche Variabilität der resultierten signifikanten Prädiktoren, wobei die Varianzaufklärung (Nagelkerke  $R^2$ ) in den verschiedenen Modellen zwischen .29 und .09 liegt. Die Normalisierungsmethode produzierte die größte Variabilität der Prädiktoren und ergab, dass die Wahrscheinlichkeit für ein positives Outcome mit besseren kognitiv-linguistischen Prätest-Fähigkeiten (verbaler IQ, schnelles Benennen, phonologische Bewusstheit) zunahm. Die Methode der Post-Prä-Differenz und die Methode der Zuwachskurve waren vergleichsweise geeigneter für die Identifikation von mehreren relevanten Prädiktoren. Die WIGROT-Methode identifizierte über alle Messungen hinweg nur einen einzigen bedeutsamen Prädiktor (Alter).

Hendricks und Fuchs (2020) analysierten Daten von 229 schwachen Lesern (4.-5. Klasse), die an einem 15-wöchigen Leseverständnistraining teilgenommen hatten (3 x 45 Min./Woche). Mittels logistischer Regressionen wurde die prädiktive Bedeutung von sieben Prätest-Variablen (Klassenstufe, Wortschatz, Aufmerksamkeit, Arbeitsgedächtnis, nonverbaler IQ, Wortlesen, Leseverstehen) für eine Responder-Klassifikation in Abhängigkeit von zehn verschiedenen Kombinationen von Methoden (Endstatus vs. Zuwachs-Methode) und Messungen (standardisierte Tests, informelle Verfahren) geschätzt. Für die Prädiktoren Wortschatz, nonverbaler IQ, Aufmerksamkeit und Leseverstehen fanden sich über alle Modelle hinweg signifikante Zusammenhänge. Bei Kindern mit besseren sprachlich-kognitiven und schriftsprachlichen Voraussetzungen erhöht sich die Wahrscheinlichkeit für ein positives Outcome, während schwächere Prätestwerte mit NR assoziiert sind, was mit Befunden von Frijters et al. (2013) korrespondiert. Die verschiedenen RTI-Operationalisierungen zeigten mehr oder weniger deutliche Unterschiede hinsichtlich der Anzahl der signifikanten Prädiktoren und ihrer Bedeutung. Risikolernende mit stärkeren kognitiven, linguistischen und schriftsprachlichen Ausgangsleistungen wurden mit der Endstatus-Methode eher als responsiv klassifiziert als mit der Zuwachs-Methode.

Zusammenfassend fokussierte die bisherige Forschung vornehmlich sprachlich-kognitive Prädiktoren der RTI. Noch kaum untersucht und geklärt ist hingegen die prädiktive Bedeutung von motivationalen Voraussetzungen und vor allem von Implementationsmerkmalen der Interventionen. Die Implementationsqualität umfasst nicht nur die konsistente Umsetzung der Intervention (Inhalte, Methoden, Modalitäten, Intensität u. a.), sondern auch die Qualität der Schulung der Trainierenden und von Unterstützungsangeboten während des Trainings. Obwohl die Wichtigkeit der Implementationsqualität für die interne Validität (Nachweis der Wirksamkeit) und die externe Validität (Generalisierung der Befunde) einer Intervention unbestritten ist (McElvany & Van Steensel, 2009), wurde solchen Aspekten in der NR-Forschung kaum Beachtung geschenkt. So finden sich in den Übersichten von Al Otaiba und Fuchs (2002) und Nelsen et al. (2003) nur vereinzelt Studien, die Angaben über die Trainierenden, ihre Schulung und die Implementationsqualität machen. Ohne solche Informationen kann jedoch nicht ausgeschlossen werden, dass „(...) variation in responsiveness was not merely a reflection of differences in the

accuracy of treatment implementation“ (Al Otaiba & Fuchs, 2002, S. 313). Forschende sind daher aufgefordert, Indikatoren der Implementationsqualität in Interventionsstudien zu erfassen und diese in Modelle zur Analyse von Prädiktoren der RTI aufzunehmen und zu überprüfen (Al Otaiba & Fuchs, 2006).

## 4 LiT-Studie im Überblick

Die zu präsentierende Studie verortet sich im Forschungsprojekt „Lesen im Tandem“ (LiT). In dessen Mittelpunkt steht die Frage nach der differenziellen Wirksamkeit eines präventiven Trainings der Leseflüssigkeit bei Drittklässler:innen durch Eltern oder Coaches. Für diese RCT-Studie mit drei Messzeitpunkten wurden 198 Kinder zufällig der Elterngruppe (n=67), der Coachgruppe (n=64) oder der Kontrollgruppe (n=67) zugeordnet. Die teilnehmenden Kinder bekundeten aufgrund eines Screenings (Metze, 2009), bestätigt durch das Lehrpersonenurteil, erhebliche Lese Probleme, hatten jedoch weder eine LRS-Diagnose noch sonderpädagogische Maßnahmen erhalten („Risikokinder“). Zwischen den drei Gruppen bestanden zu Beginn der Studie keine relevanten Unterschiede hinsichtlich Schriftsprachleistungen und verschiedener Kontrollvariablen. Die Intervention erstreckte sich über 20 Wochen bei zwei bis drei wöchentlichen Förderheiten von 20 Minuten. Sie basiert auf der empirisch validierten *Paired Reading-Methode* (PR) von Topping (1987; 1988), die sich auf das simultane laute Textlesen konzentriert und folgende Kernmerkmale beinhaltet: a) Synchronität des Lesens hinsichtlich Tempo und Lautstärke, b) Steuerung des Lesens mit dem Finger, c) Fehlerkorrektur und d) Lob (Näpflin et al., 2020). Die Schulung von Eltern und Coaches umfasste zwei Sitzungen (je 1,5 Stunden) teilweise unter Einbezug der Trainingskinder. Dabei wurde vor allem auf die korrekte Umsetzung der PR-Kernmerkmale geachtet. Wie Näpflin et al. (2020) anhand von Lesetagebüchern und Videoaufnahmen der Tandems aufzeigen, wurden diese Merkmale in beiden Gruppen weitgehend wie intendiert im Training umgesetzt. Bezüglich der Wirksamkeit der Intervention ergaben die Ergebnisse, dass die Coach-Gruppe kurzfristig stärker davon profitierte als die Eltern-Gruppe. Kinder mit höherer Lesefertigkeit im Vortest erzielten größere Fortschritte als schwache Leser. Der Effekt auf die Leseflüssigkeit ließ sich im Follow-up nicht bestätigen (vgl. Villiger et al., 2019).

## 5 Fragestellungen

Anknüpfend an den referierten Forschungsstand werden anhand einer Sekundäranalyse von Daten aus der LiT-Studie folgende Fragen beantwortet:

1. Wie wirken sich verschiedene NR-Kriterien auf die Prävalenz von Nichtresponsivität bzw. Responsivität innerhalb der Interventionsstichprobe aus?
2. Welche Zusammenhänge bestehen bei den einzelnen Kriterien zwischen individuellen, familiären und interventionsbezogenen Einzelprädiktoren und einer NR-Klassifikation?
3. Wie beeinflussen die NR-Kriterien die Vorhersage eines inadäquaten Outcomes unter Kontrolle von mehreren signifikanten Prädiktoren?

Ziel des Beitrags ist es nicht, den Einfluss von zahlreichen potenziellen NR-Kriterien (vgl. Kap. 2) auf die Prävalenz und die Vorhersage des Interventionsoutcomes bei Risikolernenden systematisch zu untersuchen. Vielmehr werden obige Fragen unter Berücksichtigung methodischer Möglichkeiten exemplarisch anhand von sechs NR-Kriterien (Kap. 6.2) und einer Reihe von Prädiktoren (Kap. 6.3) bearbeitet. Dabei wird angenommen, dass unterschiedliche Kriterien die Prävalenz und die Prädiktion von NR verändern können.

## 6 Methode

### 6.1 Stichprobe

Nachfolgende Analysen stützen sich auf die Datensätze von 130 Drittklässler:innen, die am LiT-Lesetraining teilgenommen haben und für die Prä- und Posttest-Werte für die Variable *Leseflüssigkeit* (erfasst mittels LDL, Walter, 2009) vorliegen als Voraussetzung für die Bestimmung der RTI. Tabelle 2 vermittelt Angaben zur gesamten Interventionsstichprobe, die 67 Kinder der Eltern-Gruppe und 63 Kinder der Coach-Gruppe umfasst.

Tab. 2: Merkmale der Interventionsstichprobe (N = 130)

Variablen	Deskriptive Statistiken
Alter in Jahren: M (SD)	8.87 (0.54)
Geschlecht: m (%)	63.10
Nonverbal-kognitive Fähigkeiten (RW CFT 1-R): M (SD)	62.08 (10.16)
Wortschatz (RW SET 5-10): M (SD)	31.69 (4.96)
HISEI: M (SD)	50.89 (15.61)
Erstsprache Deutsch (%)	89.80
Leseverstehen T <sub>1</sub> (RW AVS): M (SD)	6.59 (3.82)
Leseflüssigkeit T <sub>1</sub> (RW LDL): M (SD)	33.08 (10.22)

Anmerkung: Informationen zu den Instrumenten finden sich unter 6.3.

Die Trainingsgruppe zeigt durchschnittliche nonverbal-kognitive Fähigkeiten im CFT 1-R (Weiss & Osterland, 2013), der expressive Wortschatz (Subtest Bildbenennung aus *SET 5-10*, Petermann, 2012) liegt ebenfalls im Normbereich (T-Wert 48). Der Gruppenwert für den sozio-ökonomischen Hintergrund (HISEI) korrespondiert mit demjenigen eines Technikers bzw. einer Sekretärin (Ganzeboom & Treiman, 1996). Das mit dem VSL (Walter, 2013) gemessene Leseverständnis liegt vor dem Training im unteren Normalbereich (T-Wert: 42.5). Demgegenüber verzeichnet das Interventionssample im Prätest eine klar unterdurchschnittliche Leseleistung im LDL (T-Wert: 36.14; Walter, 2009). Gemäß den LDL-Werten für den Posttest konnte sich die Stichprobe in der Leseleistung verbessern, der erzielte RWM (T-Wert: 41.71) liegt jedoch noch deutlich unter demjenigen der Eichstichprobe (RWM 71.8).

## 6.2 Nonresponder-Kriterien

Die in dieser Studie definierten und angewandten sechs NR-Kriterien (vgl. Tab. 3) basieren auf Messungen der Leseleistung mittels LDL (Walter, 2013). Anlehnend an die erläuterten Dimensionen von RTI-Kriterien (Kap. 2) handelt es sich um drei normbasierte (normative sample) und drei stichprobenbasierte (limited norm sample) Operationalisierungen. Für die normbasierten Kriterien wurden Indizes bzw. Trennwerte (Prozentrang, kritische Differenz für die 3. Klasse) aus dem LDL-Testhandbuch herangezogen (Walter, 2009, S. 27 und 45). Bei den stichprobenbasierten Operationalisierungen kamen analoge Statistiken zur Anwendung, die aufgrund des eigenen Datensatzes berechnet worden waren (LDL Posttest, LDL Post-Prä-Differenz). Innerhalb dieser beiden Kategorien sind drei Methoden zu unterscheiden: Bei der Endstatus-Methode wird das Trainingskind als NR klassifiziert, wenn der LDL-Wert im Posttest unter Prozentrang 16 liegt (RWM < 44). Bei der Zuwachs-Methode wurde die Post-Prä-Differenz herangezogen und anhand der kritischen Differenz beurteilt (vgl. Nachtigall & Suhl, 2005). Aufgrund der im LDL-Manual angegebenen kritischen Differenz (3. Klasse) wurde bestimmt, ob der Unterschied zwischen den beiden Messwerten bei einer Wahrscheinlichkeit von 95 % ( $z_{krit} = 1.96$ ) und einer Reliabilität von  $r_{tt} = .90$  (Grundschule) nicht zufallsbedingt ausfällt (Walter, 2013, S. 27). Kinder, die den kritischen Wert (RWM = 18) nicht erreichen, sind NR. Für das entsprechende stichprobenbasierte Kriterium wurde anhand der bei Walter (2013, S. 27) angegebenen Formel eine eigene kritische Differenz berechnet, wobei die bei Walter (2013) angegebene Reliabilität übernommen wurde. Kinder mit einem geringeren Zuwachs gelten als NR. Das duale Diskrepanz-Kriterium kombiniert die beiden Methoden; demnach werden Kinder als NR eingestuft, wenn sie sowohl mit der Endstatus- als auch mit der Zuwachsmethode als nicht responsiv beurteilt werden.

$$(1) Diff_{krit} = z_{krit} \times s_e$$

( $Diff_{krit}$  = kritische Differenz;  $z_{krit}$  = z-Wert der kritischen Differenz;  $s_e$  = Standardmessfehler)

$$(2) s_e = s_x \times \sqrt{1 - r_{tt}} \quad (s_x = \text{Standardabweichung von } x)$$

Tab. 3: Angewandte Kriterien für NR

Normbasiert (LDL Eichstichprobe, N = 207)	
Endstatus:	LDL Posttest < PR 16 (RWM < 44)
Zuwachs:	LDL Post-Prä-Differenz < kritische Differenz (95 %, $r_{tt} = .90$ , RWM < 18)
Duale Diskrepanz:	Kombination von Endstatus und Zuwachs
Stichprobenbasiert (LiT-Interventionsgruppe, N = 130)	
Endstatus:	LDL Posttest < PR 16 (RW < 33)
Zuwachs:	LDL Post-Prä-Differenz < kritische Differenz (95 %, $r_{tt} = .90$ , RWM < 7)
Duale Diskrepanz:	Kombination von Endstatus und Zuwachs

### 6.3 Prädiktoren und Instrumente

In der LiT-Studie wurden neben der zentralen Variable Leseflüssigkeit verschiedene andere Outcome-, Kontroll- und Prädiktorvariablen mittels Tests, Fragebögen, Tagebüchern und Videoanalysen erhoben (Näpflin et al., 2020; Villiger et al., 2019). Die in dieser Arbeit berücksichtigten individuellen (1-7), familiären (8) und interventionsbezogenen (9-15) Prädiktoren für NR und die entsprechenden Instrumente werden unten charakterisiert. Die Auswahl dieser Variablen erfolgte aufgrund des in Kapitel 3 präsentierten Forschungsstands, wonach kognitive, linguistische und schriftsprachliche (Ausgangs-)Leistungen sowie der sozio-ökonomische Hintergrund von Risikolernenden einen prädiktiven Zusammenhang mit der RTI aufweisen können. Ein weiteres kindspezifisches Merkmal ist hier die Lesemotivation, die für den Lernerfolg ebenfalls als relevant angesehen wird. Die zusätzlichen interventionsbezogenen Prädiktoren umfassen die Trainingsbedingung (Eltern vs. Coaches), die effektive Gesamtzeit des Lesetrainings sowie fünf Variablen zur Implementationsqualität, deren Bedeutung für die Vorhersage einer NR-Klassifikation ebenfalls angenommen wird.

- 1. Kognitive Fähigkeiten T1:** Die nonverbale Intelligenz wurde mit dem CFT 1-R gemessen (Weiss & Osterland, 2013), dessen Reliabilität zufriedenstellend ist (Kuder-Richardson 20 = .90). Für die Analysen wurde der Gesamtrohwert verwendet.
- 2. Wortschatz T1:** Der Wortschatz wurde mittels Untertest Bildbenennung des SET 5-10 erfasst (Petermann, 2012). Dabei sind 40 Abbildungen (Gegenständen, Handlungen) zu benennen. Die interne Konsistenz ist zufriedenstellend (Kuder-Richardson 20 = .83). In die Analysen ging der erzielte Rohwert der Anzahl der richtigen Benennungen ein.
- 3. Konzentration T1:** Vor der Intervention wurden die Probanden mit dem D2-r untersucht (Brickenkamp et al., 2010), der die Konzentrationsfähigkeit sowie Schnelligkeit und Genauigkeit bei der Unterscheidung visueller Reize misst. Der Test verfügt über eine hohe Messgenauigkeit (Cronbachs  $\alpha$  zwischen .89 und .95, Retest-Reliabilität nach 1 bzw. 10 Tagen = .94 und .91 bzw. .85 und .92). Der Kennwert für die Genauigkeit ist befriedigend (Cronbachs  $\alpha$  .80 bis .91, Retest-Reliabilität: .84 bzw. .47). Der erzielte Rohwert floss in die Analysen ein.
- 4. Leseflüssigkeit T1:** Zur Erfassung der Leseflüssigkeit wurde die Lernfortschrittsdiagnostik Lesen verwendet (LDL, Walter, 2009). Das Verfahren besteht aus Texten, die vom Schulkind eine Minute lang laut gelesen werden. Mittels Auswertungsbogen wird die Anzahl der richtig gelesenen Wörter pro Minute (RWM) ermittelt. Die Paralleltestreliabilität liegt für Grundschüler:innen bei  $r_{tt} = .90$ . Statusdiagnostische Normen (T-Wert, Prozentrang) existieren u. a. für die 1. Klasse (Ende), die 2. Klasse (Anfang) sowie für die Schuljahresmitte der Klassenstufen 2-4. Für die Klassenstufen 1-8 werden kritische Differenzen zur Beurteilung von zwei Messwerten ausgewiesen. In nachfolgende Analysen ging der RWM im Prätest ein (Rohwert).
- 5. Leseverstehen T1:** Das Leseverständnis wurde mit der Verlaufsdiagnostik sinnerfassenden Lesens erfasst (VSL, Walter, 2013). Die Kinder lesen einen Text, bei dem bei jedem siebten Wort eine Klammer mit drei Auswahlwörtern (2 Distraktoren, passendes Wort) steht. Die interne Konsistenz beträgt .93, die Konstruktreliabilität .94, die Paralleltestreliabilität .77-.86. Der im Prätest erzielte Rohwert wurde verwendet.
- 6. Lesemotivation T1:** Die Lesemotivation vor der Intervention wurde anhand von zwei Items aus einem Fragebogen von Stutz et al. (2017) gemessen: 1. Ich lese, weil ich möglichst gut lesen möchte, 2. Ich lese, weil ich im Lesen besser werden möchte. Die Kinder beurteilen die



Aussagen anhand einer 4-Punkt-Likert-Skala (1. stimmt gar nicht, 2. stimmt eher nicht, 3. stimmt eher, 4. stimmt genau.). Die für die Gesamtstichprobe (inkl. Kontrollgruppe; N=194) ermittelte interne Konsistenz ist adäquat (Cronbachs  $\alpha = .80$ ).

7. **Rechtschreiben T1:** Orthographische Kompetenzen wurden mit dem SLRT II erhoben (Moll & Landerl, 2014). Der Rechtschreibteil misst die Fähigkeit, diktierete Wörter orthographisch korrekt in Sätze einzufügen. Die Retest-Reliabilität liegt zwischen .80 bis .97. Als Variable wurde die Anzahl der richtig geschriebenen Wörter verwendet (Rohwert).
8. **Sozio-ökonomischer Status/HISEI:** Die Eltern der Trainingskinder bearbeiteten einen Fragebogen zum familiären Hintergrund und machten darin u. a. Angaben zum Beruf der Eltern. Jedem Elternteil wurde nach einer standardisierten Klassifizierung von Berufen (Ganzeboom & Treiman, 1996) ein Index zugewiesen. Der höhere Index (HISEI) ging in die Analysen ein.
9. **Trainingsbedingung:** Diese Variable gibt an, ob das Training durch Eltern oder Coaches durchgeführt wurde (Referenzgruppe: Coach).
10. **Lesegesamtzeit Training:** Häufigkeit und Dauer des Trainings wurden anhand eines Lesetagebuchs erhoben, welches die Trainierenden im Anschluss an jede Sitzung ausfüllten. Anhand der Anzahl der Sitzungen und der jeweiligen Trainingsdauer wurde die Lesegesamtzeit (Minuten) ermittelt (Näpflin et al., 2020).
11. **Kernmerkmal Synchronität Tempo:** Die Übereinstimmung des Tempos beim gemeinsamen Lesen wurde anhand eines hoch inferenten Ratings einer Videoaufnahme einer Trainingssitzung erfasst (N=113). Zwei Rater beurteilten die Umsetzung auf einer vierstufigen Skala von (1) sehr geringe Ausprägung (Coach liest während des Trainings größtenteils schneller oder langsamer als das Kind) bis (4) sehr hohe Ausprägung (Coach passt sich dem Tempo des Kindes über das ganze Training an). Die relative Konsistenz (Generalisierbarkeitskoeffizient) fiel für diese Variable wie auch für die Kernmerkmale 12 und 13 sehr zufriedenstellend aus ( $> .92$ ). Für die Analyse wurde jeweils der Mittelwert der beiden Ratings verwendet (Näpflin et al., 2020).
12. **Kernmerkmal Synchronität Lautstärke:** Die Übereinstimmung der Lautstärke beim gemeinsamen Lesen wurde hoch inferent geratet. Die Beurteilung erfolgte auf einer vierstufigen Skala von (1) sehr geringe Ausprägung (Coach oder Kind ist fast immer besser hörbar) bis (4) sehr hohe Ausprägung (Coach und Kind sind gleich gut hörbar).
13. **Kernmerkmal Fingerlenkung:** Die Steuerung des Leseprozesses mit dem Finger durch die Trainierenden wurde hoch inferent auf einer vierstufigen Skala von (1) sehr geringe Ausprägung (Coach gebraucht den Finger unsystematisch) bis (4) sehr hohe Ausprägung (Coach steuert konstant) beurteilt.
14. **Kernmerkmal Fehlerkultur:** Dieses Merkmal wurde niedrig inferent geratet anhand eines Rasters, das die Reaktionen der Trainierenden bei Lesefehlern erfasst. Die Codes richten sich danach, ob Fehler übergangen oder aufgegriffen und bearbeitet werden. Dieses Merkmal wurde auf einer Skala von (1) Fehler wird übergangen bis (4) Fehlerhinweis mit ungenutzter Selbstkorrekturmöglichkeit beurteilt. Die Interrater-Reliabilität (prozentuale Übereinstimmung) beträgt  $> 90\%$  (Näpflin et al., 2020).
15. **Kernmerkmal Lob:** Der Einsatz von Lob während des Trainings (z. B. „gut“, „bravo“) wurde niedrig inferent codiert: (1) kommt vor, (0) kommt nicht vor. Die prozentuale Übereinstimmung beider Ratings liegt bei  $97\%$  (Näpflin et al., 2020).

#### 6.4 Datenanalysen

Zur Beantwortung von Fragestellung 1 wurden die prozentualen Anteile der als nicht responsiv (NR) bzw. responsiv (R) klassifizierten Kinder mittels Cochran Q-Test auf signifikante Unterschiede geprüft. Dies erfolgte separat für die norm- und stichprobenbasierten Klassifikationen, um zu ermitteln, ob die Methoden Endstatus, Zuwachs und duale Diskrepanz zu bedeutsamen Unterschieden in den Prävalenzraten führen.

Zur Beantwortung der Fragestellungen 2 und 3 wurden binär-logistische Regressionsmodelle zur Vorhersage einer NR-Klassifikation durchgeführt. Die logistische Regression erfordert keine speziellen Verteilungsannahmen, allerdings sollte zwischen den Prädiktoren keine Multikollinearität vorherrschen (Fromm, 2010). Deshalb wurden vor den Hauptanalysen zu Fragestellung 3 alle Korrelationen (nach Pearson) zwischen den Prädiktoren berechnet, welche gemeinsam in die Modelle aufgenommen werden sollten. Des Weiteren beträgt das empfohlene Minimum an Beobachtungen bei der logistischen Regression  $N = 50$  bzw.  $N = 10$  pro Prädiktor (Pavlou et al.,

2015). Die Stichprobengröße umfasst je nach Variable 102 bis 130 Fälle, da für gewisse Variablen (u. a. Videoanalyse) fehlende Werte vorlagen. Weil relativ viele Prädiktoren überprüft werden sollten, wurden zunächst Modelle mit einzelnen Prädiktoren gerechnet, um deren Beitrag zur Erklärung der Wahrscheinlichkeit für eine NR-Klassifikation zu bestimmen. Da sich diese Prädiktion je nach Operationalisierung unterscheiden kann, wurden die Analysen für alle normbasierten und alle stichprobenbasierten Kriterien einzeln durchgeführt. Schließlich wurden Gesamtmodelle mit all jenen Prädiktoren gerechnet, die in mindestens einer der Einzelanalysen signifikant ausgefallen waren. Diese sechs Prädiktoren wurden simultan in das Modell aufgenommen. Auch diese Analysen erfolgten separat für alle Kriterien. Die Vorhersagekraft der verschiedenen Modelle wurde anhand Nagelkerkes R-Quadrat und der korrekten Klassifikation von NR und R verglichen (Fromm, 2010).

## 7 Ergebnisse

### 7.1 Fragestellung 1: Einfluss des NR-Kriteriums auf die Prävalenz von unresponsiven und responsiven Kindern

Bei der Klassifikation in unresponsive und responsive Trainingskinder zeigten sich deskriptiv große Unterschiede zwischen den Operationalisierungen. Insbesondere werden bei stichprobenbasierten Kriterien deutlich weniger Probanden als NR eingestuft (zwischen 11.5 % und 18.5 %) als bei normbasierten (zwischen 43.8 % und 66.9 %; Tab. 4). Weiter spielt es eine Rolle, ob die Endstatus-, die Zuwachs- oder die duale Diskrepanz-Methode gewählt wird. So führt sowohl innerhalb der normbasierten als auch der stichprobenbasierten Kriterien die Zuwachs-Methode zum größten NR-Anteil und die duale Diskrepanz zum geringsten. Die Unterschiede zwischen den NR-Kriterien innerhalb der jeweiligen Referenzgruppe sind überzufällig ( $p < .01$  resp.  $p < .05$ ).

Tab. 4: Absolute und prozentuale Anteile an nicht responsiven und responsiven Kindern (N=130) bei Anwendung unterschiedlicher NR-Kriterien

NR-Kriterien	Absolute und prozentuale Anteile	
	NR	R
<b>Normbasiert</b>		
Endstatus-N	59 (45.4 %)	71 (54.6 %)
Zuwachs-N	87 (66.9 %)	43 (33.1 %)
Duale Diskrepanz-N	57 (43.8 %)	73 (56.2 %)
	$Q = 52.75; p < .001$	
<b>Stichprobenbasiert</b>		
Endstatus-S	20 (15.4 %)	110 (84.6 %)
Zuwachs-S	24 (18.5 %)	106 (81.5 %)
Duale Diskrepanz-S	15 (11.5 %)	115 (88.5 %)
	$Q = 8.71; p = .013$	

Anmerkung: Mittels Cochran Q-Tests wurden die Anteile NR / R über alle normbasierten bzw. stichprobenbasierten Kriterien verglichen.

Abbildung 1 präsentiert die Prä- und Posttest-Werte bzw. die Leistungszuwächse in der Leseflüssigkeit für die jeweiligen Gruppen von unresponsiven und responsiven Kindern. Wie die visuelle Datenanalyse erkennen lässt, zeigen die NR-Gruppen über alle Kriterien hinweg schwächere Ausgangs- und Endleistungen als die R-Gruppen, und die Lernzuwächse der NR fallen konsistent geringer aus als die Fortschritte der responsiven Schüler:innen.

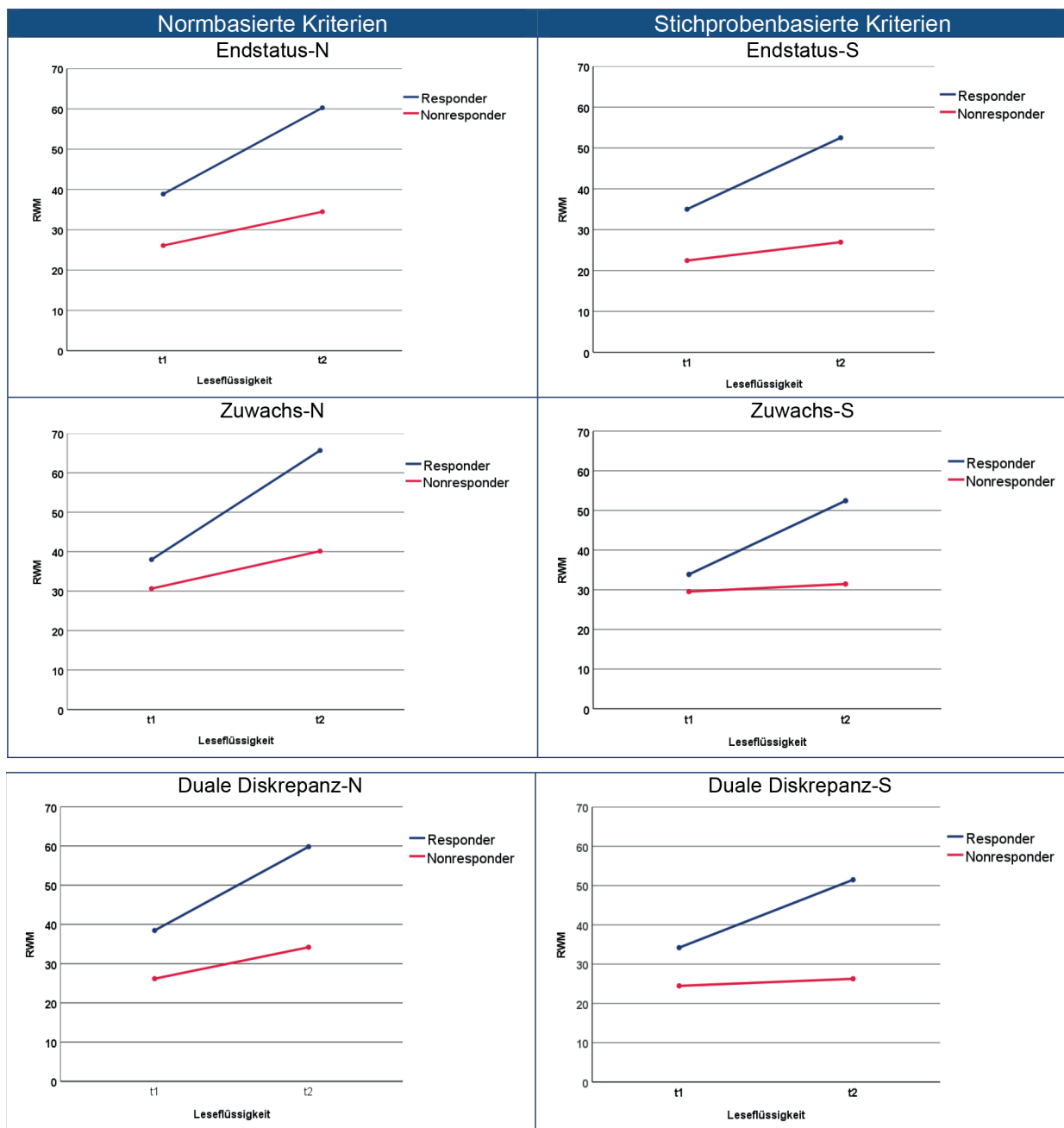


Abb. 1: Prä- und Posttestwerte für Leseflüssigkeit (RWM) der NR- und R-Gruppen bei sechs Kriterien im Vergleich

### 7.2 Fragestellung 2: Einfluss des NR-Kriteriums auf Einzelprädiktoren für NR

Hinsichtlich relevanter Prädiktoren zur Vorhersage des NR-Status erwiesen sich unterschiedliche Merkmale bei einem oder bei mehreren Kriterien als statistisch bedeutsam. Dazu gehören alle erhobenen schriftsprachlichen Ausgangsleistungen (Leseflüssigkeit, Leseverstehen, Rechtschreiben), der HISEI und vereinzelt auch die interventionsbezogenen Variablen Trainingsbedingung und Synchronität des Lesetempos. Ein höheres Ausgangsniveau schriftsprachlicher Fähigkeiten (Leseflüssigkeit, Leseverstehen, Rechtschreiben) verringerte bei (fast) allen Operationalisierungen signifikant das Risiko einer NR-Klassifikation (siehe Tab. 5, ein tendenziell signifikantes Ergebnis für Leseflüssigkeit). Dasselbe zeigte sich für einen höheren HISEI, wobei dessen Effekt nur bei den normbasierten Kriterien deutlich ausfällt. Hier war für das Zuwachs-Kriterium zudem relevant, ob es sich bei den Tutoren um die Eltern oder um einen Coach handelt, wobei sich bei den Eltern das Risiko des NR-Status signifikant erhöhte. Bei allen anderen Kriterien spielte die

Trainingsbedingung keine Rolle. Bei den drei stichprobenbasierten Operationalisierungen hatte dagegen das Kernmerkmal Synchronität Tempo einen klaren Einfluss: Je mehr beim gemeinsamen Lesen auf ein übereinstimmendes Tempo geachtet wurde, desto geringer war die Wahrscheinlichkeit einer NR-Klassifikation.

Tab. 5: Effekte der einzelnen Prädiktoren auf die Klassifikation als NR im logistischen Regressionsmodell bei sechs Kriterien im Vergleich

Variablen/ Stichprobengröße (N)	NR-Kriterien											
	Endstatus Norm		Zuwachs Norm		Duale Diskrepanz Norm		Endstatus Stichprobe		Zuwachs Stichprobe		Duale Diskrepanz Stichprobe	
	OR	p	OR	p	OR	p	OR	p	OR	p	OR	p
Kognitive Fähigkeiten (130)	1.00	.917	0.98	.253	1.00	.989	0.99	.620	0.99	.592	0.99	.621
Wortschatz T1 (130)	0.96	.307	0.95	.244	0.97	.445	0.94	.150	0.95	.197	0.94	.184
Konzentration T1 (129)	1.60	.395	0.98	.126	0.99	.501	0.99	.382	0.99	.645	0.99	.398
Leseflüssigkeit T1 (130)	<b>0.77</b>	<b>&lt;.001</b>	<b>0.93</b>	<b>&lt;.001</b>	<b>0.79</b>	<b>&lt;.001</b>	<b>0.79</b>	<b>&lt;.001</b>	0.95	.063	<b>0.87</b>	<b>&lt;.001</b>
Leseverstehen T1 (130)	<b>0.78</b>	<b>&lt;.001</b>	<b>0.80</b>	<b>&lt;.001</b>	<b>0.78</b>	<b>&lt;.001</b>	<b>0.79</b>	<b>.008</b>	<b>0.84</b>	<b>.023</b>	<b>0.79</b>	<b>.021</b>
Lesemotivation T1 (129)	0.96	.879	1.11	.691	0.89	.652	0.97	.935	0.72	.277	0.78	.497
Rechtschreiben T1 (130)	<b>0.85</b>	<b>&lt;.001</b>	<b>0.83</b>	<b>&lt;.001</b>	<b>0.84</b>	<b>&lt;.001</b>	<b>0.83</b>	<b>.001</b>	<b>0.87</b>	<b>.024</b>	<b>0.83</b>	<b>.015</b>
HISEI (128)	<b>0.97</b>	<b>.029</b>	<b>0.97</b>	<b>.030</b>	<b>0.98</b>	<b>.033</b>	0.99	.367	0.99	.397	0.98	.305
Trainingsbedingung (130)	1.38	.361	<b>2.39</b>	<b>.023</b>	1.39	.354	1.18	.736	1.40	.462	1.47	.488
Lesegesamtzeit (128)	1.00	.539	1.00	.97	1.00	.601	1.00	.396	1.00	.453	1.00	.230
Synchronität Tempo (120)	0.70	.192	0.84	.537	0.64	.105	<b>0.47</b>	<b>.023</b>	<b>0.52</b>	<b>.041</b>	<b>0.46</b>	<b>.036</b>
Synchronität Lautstärke (102)	0.72	.273	0.84	.574	0.69	.215	0.64	.219	0.85	.668	0.89	.792
Fingerlenkung (112)	0.87	.510	0.93	.766	0.83	.395	0.67	.119	0.86	.602	0.64	.129
Fehlerkultur (112)	1.19	.522	1.15	.620	1.14	.626	0.90	.770	1.23	.584	0.79	.582
Lobb (112)	0.87	.742	0.72	.460	0.79	.581	1.53	.535	1.10	.879	1.56	.585

Anmerkung: Geringere *N*s bei Tempo und Lautstärke wegen fehlender gemeinsamer Lesephasen; OR = Odds Ratio; <sup>a</sup> Referenzkategorie = Coach; <sup>b</sup> Referenzkategorie = kein Lob; Fett = signifikanter Einzeleffekt bei mindestens einem NR-Kriterium.

### 7.3 Fragestellung 3: Einfluss von NR-Kriterien auf die Vorhersage von NR unter Berücksichtigung von mehreren signifikanten Prädiktoren

Bevor die Gesamtmodelle mit allen (mindestens einmal) signifikanten Prädiktoren gerechnet wurden, wurde überprüft, ob zwischen diesen Variablen Multikollinearität vorherrschen könnte. Wie in Tabelle 6 ersichtlich ist, resultierten keine hohen Korrelationen. Zwischen den Schriftsprachkompetenzen (Leseflüssigkeit, Leseverstehen, orthographische Fähigkeiten) bestanden mittlere Effekte zwischen  $r = .38$  und  $r = .46$ , zwischen dem Kernmerkmal Tempo und den orthographischen Fähigkeiten ebenfalls eine mittlere Korrelation ( $r = .33$ ). Alle übrigen Korrelationen bewegten sich im tiefen Bereich. Es wurde deshalb nicht als problematisch erachtet, die Prädiktoren gemeinsam in das logistische Modell aufzunehmen.

Tab. 6: Korrelationen zwischen den sechs signifikanten Prädiktoren

	1	2	3	4	5	6
1 Leseflüssigkeit T1	--					
2 Leseverstehen T1	.417**	--				
3 Rechtschreiben T1	.383**	.464**	--			
4 HISEI	-.073	.159	.194*	--		
5 Trainingsbedingung <sup>a</sup>	.001	-.128	-.103	.056	--	
6 Synchronität Tempo	.235*	.084	.331**	.100	-.002	--

Anmerkung: \*\* $p < .01$  (zweiseitig getestet); \* $p < .05$  (zweiseitig getestet); <sup>a</sup>Referenzkategorie = Coach.

In Tabelle 7 sind die Gesamtmodelle für jedes NR-Kriterium dargestellt, welche alle Prädiktoren enthalten, die mindestens in einer der Einzelanalysen signifikant ausfielen. Um die Modelle in ihrer Vorhersagekraft vergleichen zu können, wurden überall dieselben Prädiktoren eingefügt (selbst wenn diese nicht in allen Einzelmodellen signifikant waren). Die Stichprobengröße ist aufgrund fehlender Werte bei einigen Variablen auf 102 reduziert, was bei sechs gleichzeitig berücksichtigten Prädiktoren ausreichend ist (Pavlou et al., 2015).

Von den Prädiktoren erwiesen sich, unter Kontrolle der anderen Faktoren, nur noch die Leseflüssigkeit T1 in fast allen Modellen (außer beim normbasierten Zuwachs-Kriterium) sowie der HISEI bei allen normbasierten Kriterien als signifikant. Das Nagelkerke R-Quadrat als Maß für die Güte des Gesamtmodells war höher bei den normbasierten Kriterien (.34 - .71) als bei den stichprobenbasierten (.21 - .53). Gemäß Backhaus (2006) kann ein Wert ab .2 als akzeptabel (kleiner Effekt), ab .4 als gut (mittlerer Effekt) und ab .5 als sehr gut (großer Effekt) bezeichnet werden, womit aus vorliegenden Modellen akzeptable bis sehr gute Werte resultierten. Der Trennwert für die Klassifikation in NR und R durch das logistische Modell wurde anhand des jeweiligen NR-Anteils festgelegt (Fromm, 2010), also beispielsweise .45 bei 45 % nicht responsiven Kindern. Die korrekte NR-Klassifikation war bei den normbasierten Kriterien (68.2 % - 86.7 %) etwas höher als bei den stichprobenbasierten (58.8 % - 81.8 %), während die Werte bei der im Sinne des gewählten Kriteriums korrekten Klassifikation von R zwischen 68.6 % und 85.5 % (normbasiert) bzw. zwischen 70.6 % und 81.4 % (stichprobenbasiert) lagen. Die korrekte Gesamtklassifikation bewegte sich zwischen 69.4 % und 86.3 % (normbasiert) bzw. zwischen 68.6 % und 81.4 % (stichprobenbasiert).

Tab. 7: Gesamtmodelle aller Prädiktoren, die mindestens in einem Modell einen signifikanten Einzeleffekt auf die NR-Klassifikation aufwiesen, bei sechs Kriterien im Vergleich

Variablen (N = 102)	NR-Kriterien											
	Endstatus Norm		Zuwachs Norm		Duale Diskrepanz Norm		Endstatus Stichprobe		Zuwachs Stichprobe		Duale Diskrepanz Stichprobe	
	OR	p	OR	p	OR	p	OR	p	OR	p	OR	p
Leseflüssigkeit T1	<b>0.73</b>	<b>&lt;.001</b>	<b>0.92</b>	<b>.005</b>	<b>0.78</b>	<b>&lt;.001</b>	<b>0.75</b>	<b>.001</b>	0.98	.532	<b>0.89</b>	<b>.035</b>
Leseverstehen T1	0.82	.068	0.89	.123	0.85	.092	0.85	.375	0.81	.063	0.78	.160
Rechtschreiben T1	0.99	.888	0.94	.411	0.96	.616	1.08	.444	0.97	.689	1.03	.767
HISEI	<b>0.94</b>	<b>.009</b>	<b>0.96</b>	<b>.027</b>	<b>0.95</b>	<b>.021</b>	0.98	.525	0.99	.528	0.99	.788
Trainingsbedingung <sup>a</sup>	4.04	.050	1.99	.163	3.24	.069	2.89	.208	1.53	.475	1.99	.378
Synchronität Tempo	1.14	.781	1.30	.481	3.24	.819	0.60	.282	0.63	.197	0.62	.282
Nagelkerke R <sup>2</sup> Klassifikation	.71		.34		.65		.53		.21		.33	
% korrekt gesamt	86.3		69.4		84.3		81.4		68.6		78.4	
% korrekt NR	87.2		68.2		86.7		81.3		58.8		81.8	
% korrekt R	85.5		68.6		82.5		81.4		70.6		78.0	

Anmerkung: OR = Odds Ratio; <sup>a</sup> Referenzkategorie = Coach; Fett = signifikante Effekte.

## 8 Diskussion

Diese Studie untersuchte erstmals im deutschsprachigen Raum den Einfluss von verschiedenen RTI-Kriterien auf die Prävalenz und die Vorhersage eines inadäquaten Interventionsoutcomes bei Schulkindern mit Leseschwierigkeiten. Die berücksichtigten Operationalisierungen differieren hinsichtlich der Methode (Endstatus, Zuwachs, duale Diskrepanz) bzw. der Referenzgruppe (normbasiert versus stichprobenbasiert).

Mit Blick auf Fragestellung 1 dokumentieren die Ergebnisse einen klaren Einfluss des Kriteriums, wobei die resultierenden NR-Anteile zwischen 11 % und 67 % liegen. Diese beachtliche Spannweite überlappt sich größtenteils mit Angaben in kriterienvergleichenden angloamerikanischen Untersuchungen, die je nach Operationalisierung NR-Prävalenzen zwischen 2 % und 70 % (Barth et al., 2008), 14 % und 75 % (Frijters et al., 2013) oder 20 % und 81 % (Hendricks et al., 2020) berichten. Dabei fällt auf, dass in unserer Studie die höchste Prävalenzrate (67 %) durchwegs (etwas) geringer ausfällt.

Die in dieser Studie analysierten normbasierten Operationalisierungen generieren generell deutlich höhere NR-Prävalenzraten (43 % - 67%;  $M = 52\%$ ) als die stichprobenbasierten (11.5 % - 18.5%;  $M = 15\%$ ), was in Anbetracht des insgesamt eingeschränkten Leistungsbereichs der Interventionsgruppe (limited norm sample) kaum erstaunt. Unabhängig von der Referenzgruppe identifiziert die Zuwachsmethode am meisten NR (66.9 %, 18.5 %), gefolgt von der Endstatus-Methode (45.4 %, 15.4 %) und der dualen Diskrepanz-Methode (43.8 %, 11.5 %). Dass letztgenannte Methode bei der NR-Klassifikation insgesamt konservativer ausfällt als die anderen Methoden, weist in dieselbe Richtung wie Befunde von McMaster et al. (2005), wonach die duale Diskrepanz weniger NR identifiziert als die Endstatus- und die Zuwachsmethode für sich.

Die bearbeiteten Fragestellungen 2 und 3 sind insofern innovativ, als neben individuellen Variablen erstmals auch Implementationsmerkmale als Prädiktoren der RTI berücksichtigt wurden. Wie die separaten Analysen zunächst zeigen, findet sich nur bei sechs von fünfzehn Prädiktoren ein signifikanter Zusammenhang mit einer NR-Klassifikation. Dazu gehören schriftsprachliche Ausgangsleistungen, die bei allen (Leseverstehen T1, Rechtschreiben T1) oder fast allen (Leseflüssigkeit T1) Operationalisierungen bedeutsam ausfallen. Dabei erhöht sich bei schwacher Prätest-Leistung die Wahrscheinlichkeit eines negativen Outcomes. Dies konvergiert mit Ergebnissen von Frijters et al. (2013) und Hendricks und Fuchs (2020) und speziell auch mit Befunden von Villiger et al. (2019), wonach Kinder mit schwacher Leseflüssigkeit im Prätest weniger von der PR-Methode profitierten als Kinder mit besserer Ausgangsleistung. Weitere signifikante Zusammenhänge ergeben sich in den Einzelanalysen für den sozio-ökonomischen Hintergrund (je tiefer, desto höher das NR-Risiko) (vgl. Al Otaiba & Fuchs, 2002) sowie für die Trainingsbedingung (Förderung durch Eltern erhöht das NR-Risiko) und Synchronität Tempo (je synchroner das gemeinsame Lesen, umso geringer die Wahrscheinlichkeit einer NR-Klassifikation). Alle anderen Implementationsvariablen fallen unbedeutsam aus. Doch nicht nur für den HISEI (nur signifikant bei normbasierten Kriterien), sondern auch für die Trainingsbedingung (nur signifikant beim Zuwachs-Norm-Kriterium) und die Synchronität Tempo (nur signifikant bei den stichprobenbasierten Kriterien) resultiert eine geringere Konsistenz der prädiktiven Relevanz über die sechs Operationalisierungen hinweg als bei den schriftsprachlichen Voraussetzungen. Die Befunde zu den Implementationsvariablen sprechen insgesamt dafür, dass (gewisse) Unterschiede in der Umsetzung der PR-Methode keinen bedeutsamen Einfluss auf die RTI der trainierten Drittklässler:innen hatten.

Nach den Resultaten zu Fragestellung 3 resultieren in den Regressionen mit mehreren Prädiktoren nur noch für die Leseflüssigkeit T1 und den HISEI signifikante Zusammenhänge mit einer NR-Klassifikation. Während die prädiktive Bedeutung der initialen Leseflüssigkeit auch hier weitgehend konsistent ist (signifikant bei 5 von 6 Kriterien), wird der HISEI lediglich bei den normbasierten Kriterien signifikant, was möglicherweise methodische Gründe hat (kleine NR-Subgruppen bei stichprobenbasierten Operationalisierungen). Mit Blick auf den prädiktiven Zusammenhang der lesetechnischen Ausgangsleistung zeichnet es sich auch in den Gesamtmodellen ab, dass die eingesetzte Fördermethode – zumindest bei Anwendung durch nicht professionelle Trainierende – für Schulkinder mit geringer Leseleistung nicht zielführend ist. Ein vorausgehendes Training von spezifischen Voraussetzungen für Leseflüssigkeit wie insbesondere der direkten und indirekten Worterkennung dürfte für solche Risikokinder wichtig und hilfreich sein.

In der Gesamtschau korrespondieren unsere Ergebnisse mit bisherigen Erkenntnissen der angloamerikanischen Forschung, wonach die Art des Kriteriums nicht nur die NR-Prävalenzraten beeinflusst, sondern – zumindest teilweise – auch die Prädiktoren einer inadäquaten RTI. Diese Befunde bleiben vor dem Hintergrund von Stärken und Limitationen der Studie kritisch zu würdigen.

## 9 Stärken und Limitationen

Neben der Analyse von Daten aus einer qualitativ hochstehenden Interventionsstudie besteht eine Stärke dieses Beitrags darin, dass für die Prädiktion des NR-Status neben einer Reihe von individuellen und familiären Variablen auch Indikatoren der Implementationsqualität berücksichtigt wurden, was frühere Studien vermissen ließen (Frijters et al., 2013; Hendricks et al., 2020). Außerdem ergänzt unsere Arbeit den Kenntnisstand der angloamerikanischen NR-Forschung um eine Untersuchung mit leseschwachen Primarschulkindern aus dem deutschsprachigen Raum in einem spezifischen Interventionssetting (außerschulisches Training durch Eltern oder Coaches).

Nichtsdestotrotz weist die Studie auch Limitationen auf. Ein erster Punkt betrifft die Stichprobe, die mit  $N = 130$  zwar für die Anzahl inkludierter Prädiktoren im Modell als genügend betrachtet werden kann, allerdings besteht je nach NR-Kriterium eine kleine NR-Subgruppe bzw. teilweise stark unausbalancierte Gruppengrößen. Obschon das Verhältnis zwischen unresponsiven und responsiven Kindern für die Klassifikation berücksichtigt wurde, könnten künftige Studien mit umfangreicheren Stichproben möglicherweise aufgrund einer höheren Teststärke Effekte aufdecken, welche hier nicht gefunden werden konnten.

Eine weitere Limitation kann darin gesehen werden, dass eine geringere Anzahl an NR-Operationalisierungen untersucht wurde als in früheren kriterienvergleichenden Studien (z. B. Frijters et al., 2013; Hendricks & Fuchs, 2020). Bezüglich der Zuwachsmethode, die auch in das duale Diskrepanzkriterium einfließt, ist anzumerken, dass für die Bestimmung des Lernzuwachses nur zwei Messzeitpunkte vorlagen. Es wäre wünschenswert gewesen, während der Intervention mehrere Messungen im Sinne von curriculumbasierten Erhebungen durchzuführen, um die kindlichen Fortschritte auch im zeitlichen Verlauf zu erfassen und den Einfluss eines solchen zusätzlichen Indikators auf die Prävalenz und Prädiktion von Nichtresponsivität schätzen zu können. Bezüglich der Vorhersage des RTI-Status blieben in unserer Studie einige potenziell relevante Faktoren wie Arbeitsgedächtnis oder schnelles Benennen unberücksichtigt (vgl. Al Otaiba & Fuchs, 2002; Nelson et al., 2003). Aufgrund der spezifischen Bedingungen der implementierten Intervention – außerschulisches Lesetraining durch Coaches oder Eltern – sollten die berichteten Befunde nicht ungeprüft auf andere Interventionskontexte und Zielgruppen übertragen werden.

Last but not least kann auch diese Studie mit einem retrospektiven Ansatz die Frage nach der „besten“ Vorgehensweise zur Identifikation von unresponsiven Schulkindern nicht abschließend klären. Für belastbare und generalisierbare Antworten sind weitere diagnostische Studien bzw. prospektive Längsschnittuntersuchungen mit umfangreichen, heterogenen und repräsentativen Stichproben erforderlich, um die psychometrische Validität und Effizienz von verschiedenen NR-Kriterien systematisch bestimmen und evaluieren zu können (Barth et al., 2008; Hendricks & Fuchs, 2020).

## 10 Ausblick

Die Identifikation von Kindern mit einer ungenügenden RTI auf präventive (und therapeutische) Schriftsprachinterventionen ist eine wichtige wie herausfordernde Aufgabe von Schulen. Gleichzeitig besteht in der Literatur ein Konsens, dass die Heterogenität und Kontextabhängigkeit der in Forschung und Praxis angewandten NR-Kriterien ein relevantes Problem darstellt. Hendricks et al. (2020, S. 14) weisen zurecht darauf hin, dass ohne einen evidenzbasierten Konsens darüber, wie nicht responsive Schulkinder am besten erfasst werden sollten, der RTI-Ansatz – ähnlich wie das traditionelle IQ-Diskrepanzmodell der LRS-Diagnostik – arbiträr erscheint und eine Reihe von praktischen und bildungspolitischen Problemen aufwirft. Im Hinblick auf die Beurteilung der Zuverlässigkeit und Nützlichkeit von verschiedenen RTI-Operationalisierungen schlagen Barth et al. (2008, S. 13) vor, sich in der Forschung zunächst auf eine „Art von Goldstandard“ zu einigen, anhand dessen andere Kriterien verglichen werden können. Sodann ließen sich mittels geeigneter Methoden zur Vorhersage des Goldstandards optimale Grenzwerte bestimmen, die entlang der kontinuierlichen Verteilung der individuellen RTI liegen. Zudem könnten latente Klassenmodelle, die mehrere Indikatoren für den Interventionserfolg berücksichtigen, nützlich sein. Dadurch ließe sich klären, ob der Goldstandard praktikabel ist, was die Entscheidungsfindung bezüglich des RTI-Status erleichtern sollte.

Der für solche wissenschaftlichen Bemühungen erforderliche Aufwand ist beachtlich, in Anbetracht der auch in diesem Beitrag aufgezeigten Schwierigkeiten von heterogenen RTI-Kriterien jedoch sehr relevant und lohnenswert. Während sich aus dem derzeit unvollständigen Forschungsstand und unserer Grundlagenstudie in erster Linie Implikationen für die zukünftige Forschung ergeben, ist der Erkenntnisgewinn des Beitrags auch für die schulische Praxis bedeutsam. Neben theoretischen Informationen zum Konstrukt der Nichtresponsivität und einem Überblick über verschiedene Möglichkeiten der Operationalisierung der RTI vermittelt der Artikel begründete Argumente für die Wichtigkeit einer konsequenten Evaluation des kindlichen Lernerfolgs zwecks optimierender Anpassung von Interventionen für nicht erfolgreiche Lernende. Trotz offener Fragen spricht aktuell einiges für die Beurteilung des Interventionsoutcomes von Risikolernenden sowohl anhand des Leistungszuwachses als auch des erreichten Leistungsstandes (im Sinne des dualen Diskrepanz-Kriteriums) (Hendricks & Fuchs, 2020; McMaster et

al., 2005). Zur Erfassung dieser Parameter des Interventionserfolgs sollten vorteilhaft validierte status- und verlaufdiagnostische Instrumente genutzt werden, die auf adäquaten normativen Referenzgruppen basieren, zumal solche Verfahren die Gefahr von falsch-negativen NR-Klassifikationen – im Vergleich zu eingeschränkten schul- oder samplebasierten Referenzdaten – im Interesse von Kindern mit Lernschwierigkeiten reduzieren (Barth et al., 2008; Frijters et al., 2013).

Als Mitglieder von Schulteams können Sprachtherapeut:innen aufgrund ihrer spezifischen professionellen Kompetenzen nicht nur zur Implementierung von schulweiten Präventionsmaßnahmen aktiv beitragen, sondern auch Lehrpersonen bei der Auswahl und beim Einsatz von geeigneten Kriterien und praktikablen Instrumenten zur Beurteilung der schriftsprachlichen Entwicklung (nicht nur) von Risikolernenden unterstützen.

## Finanzierung/Danksagung

Das LiT-Projekt wurde vom Schweizerischen Nationalfonds (Projekt Nr. 149560) und den Kantonen Luzern und Freiburg finanziert. Die Autorenschaft dankt den Tutor:innen sowie den Schulkindern für ihre Teilnahme an der Studie.

## Literatur

- Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special education, 23*(5), 300-316. doi.org/10.1177/07419325020230050501
- Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39*(5), 414-431. doi.org/10.1177/0022219406039005040
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2006). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (11. Aufl.). Berlin: Springer.
- Barth, A. E., Stuebing, K. K., Anthony, J. L., Denton, C. A., Mathes, P. G., Fletcher, J. M., & Francis, D. J. (2008). Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences, 18*(3), 296-307. doi.org/10.1016/j.lindif.2008.04.004
- Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of educational research, 79*(1), 262-300. doi.org/10.3102/0034654308325998
- Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review, 40*(1), 3-22.
- Frijters, J. C., Lovett, M. W., Sevcik, R. A., & Morris, R. D. (2013). Four methods of identifying change in the context of a multiple component reading intervention for struggling middle school readers. *Reading and writing, 26*(4), 539-563. doi.org/10.1007/s11145-012-9418-z
- Fromm, S. (2010). *Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten* (Vol. 2). Wiesbaden: Springer Fachmedien.
- 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*(1), 1-12. doi.org/10.1371/journal.pone.0089900
- Ganzeboom, H. B., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25*(3), 201-239.
- Hartmann, E. (2004). Konzeption und Diagnostik von schriftsprachlichen Lernstörungen im Responsiveness-to-Intervention-Modell: eine kritische Würdigung. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 2*(77), 123-137.
- Hartmann, E. (2021). Aussichtsreicher inklusiver Schriftsprachunterricht für Kinder mit erhöhtem Förderbedarf im Lesen- und Schreibenlernen. In A. Kunz, R. Luder, & C. Müller Bösch (Hrsg.), *Inklusive Pädagogik und Didaktik* (S. 178-197). Bern: hep Verlag AG.
- Hendricks, E. L., & Fuchs, D. (2020). Are individual differences in response to intervention influenced by the methods and measures used to define response? Implications for identifying children with learning disabilities. *Journal of Learning Disabilities, 53*(6), 428-443. doi.org/10.1177/0022219420920
- Justice, L. M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools, 37*(4), 284-297. https://doi.org/10.1044/0161-1461(2006/033)
- Lam, E. A., & McMaster, K. L. (2014). Predictors of responsiveness to early literacy intervention: A 10-year update. *Learning Disability Quarterly, 37*(3), 134-147. doi.org/10.1177/0731948714529772
- Mayer, A., & Marks, D. K. (2019). Förderung des Textverständnisses durch die Vermittlung von Verstehensstrategien – Eine Metaanalyse zur Effektivität. *Forschung Sprache, 1*, 4-36.
- McElvany, N., & Van Steensel, R. (2009). Potentials and challenges of family literacy interventions: The question of implementation quality. *European Educational Research Journal, 8*(3), 418-433. doi.org/10.2304/eej.2009.8.3.418
- McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*(4), 445-463. doi.org/10.1177/001440290507100404
- Metze, W. (2009). *Stolperwörter-Lesetest*. Abgerufen von [http://www.wilfriedmetze.de/Handanweisung\\_2009.pdf](http://www.wilfriedmetze.de/Handanweisung_2009.pdf) [10.9.13]
- Nachtigall, C., & Suhl, U. (2005). Evaluation intraindividuelle Veränderung: ein Vergleich verschiedener Veränderungskennwerte. *Zeitschrift für Klinische Psychologie und Psychotherapie, 34*(4), 241-247. doi.org/10.1026/1616-3443.34.4.241
- Näpflin, C., Frommelt, M., Hugener, L., Tettenborn, A., Krammer, K., Villiger, C., Hauri, S., & Hartmann, E. (2020). Implementationsqualität unter der Lupe: Unterscheiden sich Eltern und Lesecoachs in der Umsetzung eines Trainings zur Förderung der Leseflüssigkeit? *Psychologie in Erziehung und Unterricht, 67*(2), 95-111.



- Nellis, L. M., Sickman, L. S., Newman, D. S., & Harman, D. R. (2014). Schoolwide collaboration to prevent and address reading difficulties: Opportunities for school psychologists and speech-language pathologists. *Journal of Educational and Psychological Consultation*, 24(2), 110-127. doi.org/10.1080/10474412.2014.903187
- Nelson, R. J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, 18(4), 255-267. doi.org/10.1111/1540-5826.00080
- Pavlou, M., Ambler, G., Seaman, S., De Iorio, M., & Omar, R. Z. (2016). Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine*, 35, 1159-1177. doi.org/10.1002/sim.6782
- Petermann, F. (2012). *Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren: SET 5-10*. Göttingen: Hogrefe.
- Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: Center on Instruction. Abgerufen von <https://eric.ed.gov/?id=ED521837> [12.05.2021]
- Stutz, F., Schaffner, E., & Schiefele, U. (2017). Measurement invariance and validity of a brief questionnaire on reading motivation in elementary students. *Journal of Research in Reading* 40(4), 439-461. doi.org/10.1111/1467-9817.12085
- Topping, K. J. (1987). Paired reading: A powerful technique for parent use. *The Reading Teacher*, 40, 608-614.
- Topping, K. J. (1988). *Paired reading training pack*. Kirklees Paired Learning Project.
- Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention examining higher and lower responders. *Exceptional Children*, 75(2), 165-183.
- Vellutino, F. (1979). *Dyslexia: Theory and Research*. Cambridge/MA: MIT Press.
- Villiger, C., Hauri, S., Tettenborn, A., Hartmann, E., Näpflin, C., Hugener, I., & Niggli, A. (2019). Effectiveness of an extracurricular program for struggling readers: A comparative study with parent tutors and volunteer tutors. *Learning & Instruction*, 60, 54-65. doi.org/10.1016/j.learninstruc.2018.11.004
- Voß, St., Blumenthal, Y., Mahlau, K., Marten, K., Diehl, K., Sikora, S., & Hartke, B. (2016). *Der Response-to-Intervention-Ansatz in der Praxis. Evaluationsergebnisse zum Rügener Inklusionsmodell*. Münster: Waxmann.
- Walter, J. (2009). *LDL: Lernfortschrittsdiagnostik Lesen. Ein curriculumbasiertes Verfahren*. Göttingen: Hogrefe.
- Walter, J. (2013). *VSL: Verlaufsdagnostik sinnerfassenden Lesens*. Göttingen: Hogrefe.
- Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*, 23(8), 889-912. doi:10.1007/s11145-009-9179-5
- Weiss, R.H., & Osterland, J. (2013). *CFT-1R. Grundintelligenztest Skala 1 (Revision)*. Göttingen: Hogrefe.

## Zu den Autor:innen

*Prof. Dr. Erich Hartmann* ist Professor für Logopädie am Departement für Sonderpädagogik der Universität Freiburg/CH. Sein Arbeitsgebiet umfasst die evidenzbasierte Diagnostik und Intervention bei kindlichen Sprach- und Schriftsprachstörungen.

*Dr. Verena Hofmann* ist Lektorin am Departement für Sonderpädagogik der Universität Freiburg/CH. Sie forscht im Bereich der sprachlich-kommunikativen Fähigkeiten sowie im Bereich Peerbeziehungen und Verhaltensentwicklung bei Kindern und Jugendlichen mit besonderen Bedürfnissen.

*Prof. Dr. Caroline Villiger* ist Leiterin des Forschungsschwerpunkts «Familie – Bildung – Schule» und Dozentin an der Pädagogischen Hochschule Bern/CH. Ihre Arbeitsthemen sind: Familiäre und schulische Leseförderung, Erziehungs- und Bildungskoooperation Familie – Schule, elterliche Lernbegleitung.

*Prof. Dr. Annette Tettenborn* ist Leiterin des Instituts für Professions- und Unterrichtsforschung an der Pädagogischen Hochschule Luzern/CH. Ihre Expertise liegt in den Bereichen der Lehr-Lernforschung und der Professionalisierungsforschung.

*Silke Hauri*, Msc. ist Bereichsleiterin «Berufspraktische Ausbildung Primarstufe» an der Pädagogischen Hochschule Zürich/CH.

## Korrespondenzadresse

Prof. Dr. Erich Hartmann  
Departement für Sonderpädagogik, Abteilung Logopädie  
Petrus-Kanisius-Gasse 21  
CH-1700 Freiburg  
erich.hartmann@unifr.ch



4.,  
überarbeitete  
Auflage

4.,  
überarbeitete  
Auflage

Jutta Fleckenstein, Sabine Jankuhn,  
Susanne Meiering, Holger Scholz,  
4., überarbeitete Auflage 2023,  
kartoniert: ISBN 978-3-8248-1149-6,  
72 Seiten,  
E-Book: ISBN 978-38248-9965-4, PDF,  
EUR 24,00 [D]

## Diagnostischer Leitfaden zur Feststellung des sonderpädagogischen Unterstützungsbedarfs

Beobachtungskriterien – diagnostische Fragestellungen – Formulierungshilfen – Fördermaßnahmen zu den Förderschwerpunkten

- Lernen (L)
- emotionale und soziale Entwicklung (esE)
- körperliche und motorische Entwicklung (kmE)
- und in Bezug auf Lernende aus dem Autismus-Spektrum

Inklusion bedeutet für die Schule, dass sich die Unterrichtsangebote an den Entwicklungsmöglichkeiten jeder Schülerin und jedes Schülers orientieren und deren Bedürfnissen – unabhängig von ihren individuellen Voraussetzungen – Rechnung tragen.

Dieser Leitfaden bietet die Möglichkeiten zur differenzierten Wahrnehmung und Beschreibung von Schülerinnen und Schülern. Speziell für den Unterstützungsbedarf im Bereich des Lernens, der körperlichen und motorischen Entwicklung sowie der sozialen und emotionalen Entwicklung werden ei-

nerseits Formulierungsangebote gemacht und andererseits auch mögliche Maßnahmen der Unterstützung im Lernprozess dargestellt.

In der Praxis bewährt sich der Leitfaden zur Erstellung differenzierter und aussagekräftiger pädagogischer Gutachten bzw. Stellungnahmen.

Für die vierte Neuauflage wurden insbesondere mit Blick auf die Schülerschaft aus dem Autismus-Spektrum Ergänzungen vorgenommen.



Tel.: +49 6126 9320-13 | Fax: +49 6126 9320-50  
bestellung@schulz-kirchner.de | www.skvshop.de

Lieferung versandkostenfrei innerhalb Deutschlands





## Das Nachsprechen von Nichtwörtern zur Testung bilingualer Kinder\*

Zum Einsatz der einzelsprachunabhängigen Items des LITMUS-CL-NWRT zur Differenzierung von Kindern mit und ohne Sprachentwicklungsstörungen

### Testing bilingual children with a Nonword Repetition Task

Using the language-independent items of LITMUS-CL-NWRT to distinguish between children with and without Developmental Language Disorders

Juliane Hinnerichs

#### Zusammenfassung

**Hintergrund:** Die Differenzierung zwischen bilingualen Kindern, bei denen eine Sprachentwicklungsstörung (SES) vorliegt und denen, deren geringe Deutschkenntnisse lediglich auf unzureichenden Input zurückzuführen sind, gilt als Herausforderung. Die Überprüfung beider/aller erworbener Sprachen ist oft nicht möglich, deshalb scheint das Nachsprechen von Nichtwörtern (NWRT) als einzelsprachunabhängige Methode vielversprechend. Der LITMUS-CL-NWRT wurde gezielt zur Testung bilingualer Kindern mit diversen sprachlichen Hintergründen konstruiert.

**Ziele:** Nach einer ersten Studie zur Auswirkung verschiedener Auswertungsvorgehensweisen auf die Ergebnisse des LITMUS-CL-NWRT soll dessen Eignung zur Differenzierung von bilingualen Kindern mit und ohne SES überprüft werden. Anschließend soll untersucht werden, ob der LITMUS-CL-NWRT dabei eine bessere diagnostische Validität aufweist als ein einzelsprachspezifischer NWRT (LS-NWRT).

**Methode:** 103 monolinguale und bilinguale Kinder zwischen 3;7 und 13;9 Jahren wurden im Rahmen des BiliSAT-Projekts getestet und daraufhin jeweils in eine Gruppe mit und ohne SES aufgeteilt. Die NWRT-Ergebnisse wurden mit einer Kovarianzanalyse verglichen, anschließend wurde die Sensitivität und Spezifität der Testungen berechnet.

**Ergebnisse:** Beide NWRTs differenzieren Kinder mit und ohne Sprachentwicklungsstörungen voneinander unabhängig davon, ob diese bilingual oder monolingual aufwachsen. Der LITMUS-CL-NWRT zeigt bessere Ergebnisse zur diagnostischen Validität als der einzelsprachspezifische NWRT.

**Schlussfolgerungen:** Sowohl die Itemkonstruktion als auch die Auswertungsvorgehensweise beeinflussen die Ergebnisse von NWRTs. Einzelsprachunabhängig konstruierte Nichtwörter eignen sich besser als LS-NWRTs zur Testung bilingualer Kinder.

#### Schlüsselwörter

Bilingualismus, Sprachentwicklungsstörungen, Nachsprechen von Nichtwörtern

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

### Abstract

**Background:** The distinction between children with Developmental Language Disorders (DLD) and those with low German language skills simply due to insufficient input is regarded as a challenge. The assessment of both/all acquired languages is often impossible, thus making approaches based on repetition of nonwords a promising language-independent tool. The LITMUS-CL-NWRT was specifically designed to test bilingual children with various language backgrounds.

**Aims:** After a first study on the impact of different scoring procedures on the results of the LITMUS-CL-NWRT, its suitability to distinguish between bilingual children with or without DLD is to be assessed. Subsequently, it has to be checked if the diagnostic accuracy of the LITMUS-CL-NWRT is higher than that of a language-specific NWRT (LS-NWRT).

**Methods:** 103 monolingual and bilingual children, aged 3;7 - 13;9, were tested within the BiliSAT-project and divided up into two groups with and without DLD. The results of the NWRTs were compared with an analysis of covariance, afterwards test sensitivity and specificity were determined.

**Results:** Independent of bilingualism as an influencing factor, with both NWRTs, children with and without DLD can be distinguished. The LITMUS-CL-NWRT shows a higher diagnostic accuracy than the LS-NWRT.

**Conclusion:** The item construction as well as the scoring procedure affect the results of NWRTs. Language-independent constructed nonwords are more suitable than LS-NWRTs to test bilingual children.

### Keywords

Bilingualism, Developmental Language Disorders, Nonword Repetition Test

## 1 Einleitung – Zur Testung von bilingualen Kindern

Der Umgang mit bilingualen<sup>1</sup> Kindern im sprachtherapeutischen Kontext ist eine Herausforderung. Ein Drittel der befragten Sprachtherapeut:innen geben an, bei der Diagnostik mehrsprachiger Kinder überfordert zu sein und mehr als die Hälfte fühlt sich bei der Auswahl geeigneter Testverfahren unsicher (Marxen, Lüke & Cholin 2022, S.103). Lange Zeit wurden Testungen und Therapieansätze in der Annahme entwickelt, dass Kinder lediglich eine Sprache erwerben. In den letzten Jahren entstand zunehmend ein Bewusstsein für die steigende Anzahl bilingualer Kinder in Deutschland, wobei „praxistaugliche sprachwissenschaftliche fundierte Verfahren für mehrsprachig aufwachsende Kinder in Deutschland“ noch in der Entwicklung sind (Chilla 2022, S.125). Die korrekte Einschätzung der sprachlichen Fähigkeiten bilingualer Kinder ist schwierig. Ein Grund hierfür ist die Überschneidung der Symptome einer SES mit den Phänomenen des unauffälligen Zweitspracherwerbs (Chilla 2022; Scherger 2022). Ist der Wortschatz im Deutschen aufgrund des begrenzten Inputs so klein oder liegt eine Störung im semantisch-lexikalischen Bereich vor? Wurde die korrekte Verbstellung aufgrund der kurzen Sprachkontaktdauer noch nicht erworben oder ist der Erwerb grammatischer Strukturen generell beeinträchtigt? Die Berichte von Fehldiagnosen zeigen, dass es in der Praxis schwierig ist, die sprachliche Entwicklung bilingualer Kinder korrekt einzuschätzen (Chilla 2022; Grimm & Schulz 2014; Rothweiler 2007).

Die Prävalenzrate für eine SES<sup>2</sup>, die „nicht auf sensorische, organische, mentale oder gravierende sozial-emotionale Defizite zurückgeführt“ werden kann, beträgt circa 5-8 % (De Langen-Müller, Kauschke, Kiese-Himmel, Neumann & Noterdaeme 2011, S.32), unabhängig davon, ob eine oder mehrere Sprachen erworben werden. Hierbei wird immer wieder betont, dass bei mehrsprachigen Kindern alle Sprachen von der Störung betroffen sein müssen (Chilla 2022). Sobald sich in einer Sprache eine unauffällige Entwicklung abzeichnet, können keine angeborenen Spracherwerbsdefizite vorliegen. Die Auffälligkeiten in der bzw. den anderen Sprachen müssen dann auf die Inputsituation zurückzuführen sein.

Idealerweise müssten also die Fähigkeiten in allen erworbenen Sprachen überprüft werden, dies ist jedoch oft nicht möglich. So mangelt es einerseits an normierten Testverfahren in der Erstsprache sowie andererseits auch an Sprachtherapeut:innen, die diese beherrschen. Des Weiteren gilt es als problematisch, die Normwerte monolingualer Kinder für den bilingualen Erwerbskontext zu verwenden, unabhängig davon, ob es sich um die Erst- oder Zweitsprache handelt (Chilla 2022). Für die meisten Sprachtestungen im Deutschen liegen keine gesonderten

1 Der Begriff bilingual wird hier für Kinder verwendet, die mindestens zwei, gegebenenfalls aber auch drei oder mehr Sprachen erwerben.

2 Andere Bezeichnungen: Umschriebene SES (USES), Spezifische SES (SSES); im Englischen nach aktuellem Konsens Developmental Language Disorder (DLD; Bishop, Snowling, Thompson & Greenhalgh 2017)

Normwerte für bilinguale Kinder vor. Wenn bilinguale Normen erhoben wurden, so scheint es fast unmöglich mit diesen die vielfältigen bilingualen Spracherwerbsbiografien zufriedenstellend abzudecken, wie beispielsweise den simultanen, frühen oder späten sukzessiven Erwerb und die darauf folgenden Variationen der Sprachkontaktdauer. So enthält beispielsweise die Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ; Schulz & Tracy 2011) für bilinguale Kinder mehrere Normwerttabellen pro Altersgruppe, die je nach Kontaktdauer zum Deutschen ausgewählt werden können. Dennoch gibt es natürlich immer wieder Kinder, für deren Alter und Kontaktdauer keine entsprechende Normwerttabelle vorliegt.

Somit wird offensichtlich, dass bei bilingualen Kindern die Einordnung der sprachlichen Fähigkeiten in der Einzelsprache als „auffällig“ und „unauffällig“ oft schwierig ist. Eine sinnvolle Unterstützung bei der Differentialdiagnose von bilingualen Kindern wäre demnach ein Verfahren, das zugrundeliegende Sprachverarbeitungsfähigkeiten unabhängig von den Einzelsprachen erfasst.

### 1.1 Nachsprechen von Nichtwörtern

Das Nachsprechen von Nichtwörtern wird in diesem Kontext als vielversprechende Methode angesehen, da hiermit einzelsprachunabhängige Fähigkeiten überprüft werden sollen. Mit diesem Verfahren sollen also nicht bereits erworbene sprachliche Fähigkeiten (zum Beispiel im Bereich Lexikon oder Grammatik) untersucht werden, sondern Prozesse, von denen angenommen wird, dass sie der Sprachverarbeitung zugrunde liegen. Das Ziel ist hierbei, die Funktionsweise zugrundeliegender Spracherwerbsmechanismen unabhängig von einzelsprachspezifischem Wissen zu testen und hiermit den diagnostischen Prozess bei bilingualen Kindern mit Verdacht auf SES zu erleichtern.

Die Annahme, dass beim Nachsprechen von Nichtwörtern einzelsprachunabhängige Fähigkeiten überprüft werden, wird damit begründet, dass die Items neu konstruierte Lautfolgen sind, die keine realen Wörter repräsentieren und somit für alle Kinder gleichermaßen unbekannt sind. Mit dem Nachsprechen von Nichtwörtern soll die Leistungsfähigkeit des phonologischen Kurzzeitgedächtnisses<sup>3</sup> gemessen werden, da die kurzfristige Speicherung von neuem verbalen Input als essentielle Voraussetzung für den Spracherwerb gilt (Gathercole & Baddeley 1990; Hasselhorn & Körner 1997; Pierce, Genesee, Delcenserie & Morgan 2017). Der NWRT ist ein bereits seit langem im sprachtherapeutischen Alltag eingesetztes Testinstrument, einer aktuellen Umfrage nach setzen ihn zwei Drittel der Sprachtherapeut:innen ein (Heeg et al. 2022). Die Idee, unbekannte Lautfolgen nachzusprechen, geht auf Grete Mottier (1951) zurück und wurde von ihr zur Testung von Kindern mit Lese-Rechtschreib-Störungen (LRS) entwickelt (Gamper, Keller, Messerli, Moser & Wüst, 2012). Heutzutage werden NWRTs in diversen Testungen und Sprachscreenings zur Erkennung von LRS und auch SES eingesetzt, sowie auch in einigen Sprachstandsverfahren<sup>4</sup>.

Jedoch wurde in Bezug auf die bislang eingesetzten Nichtwörter mehrfach betont, dass diese bei weitem nicht so einzelsprachunabhängig sind wie teilweise angenommen (Ennemoser, Poupou & Hartung 2012 u.a.). So wird an dem Beispiel des Items „Glösterkeit“ aus dem SETK 3-5 (Grimm 2015) deutlich, dass hier sowohl das deutsche Morphem „keit“ verwendet wird als auch die Konsonantenverbindung /gl/, welche nicht unbedingt in der entsprechenden Erstsprache eines bilingualen Kindes vorkommt. Ebenso kommt der Laut /œ/ zwar nicht nur im Deutschen vor, jedoch gibt es viele Sprachen dieser Welt, die diesen Vokal nicht in ihrem Inventar haben. Somit wird offensichtlich, dass dieses Item für ein monolinguales Kind in Deutschland wesentlich vertrauter klingt – selbst wenn es dieses noch nie zuvor gehört hat – als für ein Kind mit wenig Kontakt zum Deutschen.

Die Nachsprecheleistungen werden durch Vertrautheit mit der Sprache, nach der die Nichtwörter konzipiert wurden, beeinflusst: So zeigen zahlreiche Studien, dass bilinguale Kinder durch NWRTs benachteiligt werden, wenn die Items nach den Wortbildungsregeln der weniger dominanten Sprache konstruiert wurden (u. a. Messer et al. 2010; Parra, Hoff & Core 2011; Sorenson Duncan & Paradis 2016).

3 Dieser zugrundeliegende Mechanismus wird unter anderem auch als verbales Kurzzeitgedächtnis, verbales oder phonologisches Arbeitsgedächtnis oder phonologische Schleife bezeichnet (Rescorla 2014; Schöler & Brunner 2008; Schwob et al. 2021). Die meisten Autor:innen beziehen sich dabei auf das Arbeitsgedächtnismodell nach Baddeley & Hitch (1994).

4 Siehe unter anderem BISC von Jansen, Mannhaupt, Marx & Skowronek (2002); Delfin 4 von Fried, Briedigkeit, Isele & Schunder (2009); HASE von Schöler & Brunner (2008); KISS von Holler-Zittlau, Euler & Neumann (2011); PhoMo-Kids von Stadie & Schöppe (2013); SETK 3-5 von Grimm (2015) und SSV von Grimm, Aktas & Kießig (2003)

Grimm (2015, S.66/75) sieht in dieser Benachteiligung kein Problem, indem sie darauf verweist, dass bei bilingualen Kinder mit den Nichtwörtern des SETK 3-5 die schon erworbenen Sprachfähigkeiten des Deutschen untersucht werden und nicht – wie bei den monolingualen Kindern – das Vorliegen einer SES. Des Weiteren führt sie an, dass „man gar nicht erst ernsthaft darüber nachdenken“ möchte, wie Nichtwörter konstruiert werden müssten, um möglichst viele unterschiedliche Sprachen zu repräsentieren (Grimm 2015, S.75).

## 1.2 LITMUS-CL-NWRT

Im Rahmen des internationalen Forschungsnetzwerkes COST Action IS0804 stellte sich ein Team von Wissenschaftler:innen genau dieser Herausforderung. Es wurde ein NWRT entwickelt, dessen Items in ihrem Aufbau mit möglichst vielen Sprachen kompatibel sind, der LITMUS-CL-NWRT (Language Impairment Testing in Multilingual Settings – Crosslinguistic Nonword-Repetition-Test; Chiat 2015). Zugleich betont Chiat (2015, S.133), dass selbst diese Items nicht komplett neutral, also uneingeschränkt einzelsprachunabhängig sind: “However we construct non-words, they will be more or less like real word phonology in different languages with respect to length, prosody, segmental constituents and their frequency and phonetic realisation of segments”. So kann eine bestimmte Wortlänge oder Silbenstruktur in einer Sprache als typisch gelten, während sie in einer anderen Sprache wiederum eher selten oder gar nicht vorkommt (Chiat 2015, S.133).

Leonard (2014, S.3) erklärt beispielsweise die Tatsache, dass italienische Vorschulkinder drei- und viersilbige Nichtwörter besser nachsprechen können als englische Vorschulkinder, damit, dass italienische Wörter im Schnitt länger sind und die italienischen Kinder somit besser mit der Aussprache langer Wörter vertraut sind. Des Weiteren finden Szewczyk, Marecka, Chiat & Wodniecka (2018, S.33) in ihren Daten, dass polnische Kinder im Gegensatz zu englischsprachigen Kindern keine zusätzlichen Schwierigkeiten durch Konsonantenverbindungen innerhalb der ihnen präsentierten Nichtwörter haben und erklären dies damit, dass im Polnischen vergleichsweise viele Konsonantenverbindungen vorkommen.

Obwohl Nichtwörter also niemals gleichermaßen neutral in Bezug auf alle Sprachen der Welt konstruiert werden können, so zeigen bisherige Studien zum LITMUS-CL-NWRT vielversprechende Ergebnisse in Bezug auf dessen möglichst einzelsprachunabhängige Itemkonstruktion. So wurde die Testung bereits in über 15 Ländern erprobt und die hierbei erhaltenen Daten von mehr als 1800 monolingualen und bilingualen Proband:innen zeigen, dass die bilingualen Kinder, unabhängig von ihrer Kontaktdauer, beim LITMUS-CL-NWRT nicht schlechter als monolinguale Kinder abschneiden (Polišenská et al. 2022). Gleichzeitig finden sich in diesen Daten bedeutsame Unterschiede zwischen Kindern mit sprachlichen Auffälligkeiten und sprachlich unauffälligen Kindern (im Folgenden TD-Kinder für engl. Typically Developing). Des Weiteren haben Untersuchungen aus den Niederlanden und Irland gezeigt, dass sich der LITMUS-CL-NWRT besser zur Testung bilingualer Kinder eignet als ein einzelsprachspezifisch konstruierter NWRT (im Folgenden LS-NWRT für engl. Language Specific Nonword Repetition Test) (Antonićević-Elliott et al. 2020; Boerma et al. 2015; Verhagen & Andringa 2021). Die Autor:innen zweier Metaanalysen zum Einsatz von NWRTs bei der Identifikation von Kindern mit SES kommen ebenfalls zu dem Schluss, dass sich einzelsprachunabhängige Nichtwörter besser zur Testung bilingualer Kinder eignen als einzelsprachspezifische Items (Ortiz 2021, S. 2275; Schwob et al. 2021, S.3589).

Der LITMUS-CL-NWRT wurde auch für Testungen im deutschsprachigen Raum adaptiert (Hinnerichs, Polišenská & Chiat 2016; siehe auch Hinnerichs 2016) und dort zunächst anhand einer kleinen Stichprobe (n=24) erprobt. Um die Testung mit einem LS-NWRT zu vergleichen wurde der Untertest Nachsprechen von Kunstwörtern des Heidelberger auditiven Screenings in der Einschulungsuntersuchung (HASE-NK; Schöler & Brunner 2008; siehe auch 3.2) herangezogen. Es zeigte sich hierbei, dass der LITMUS-CL-NWRT die monolingualen und bilingualen Kinder mit und ohne Sprachentwicklungsstörungen gut voneinander unterscheiden kann, jedoch nicht besser als der HASE-NK. Die Ergebnisse von Hinnerichs (2016) sollten jedoch aufgrund der kleinen Stichprobe mit Vorsicht betrachtet werden. In einer groß angelegten Untersuchung von Zaretsky, Lange & Hey (2022; n=1441) zeigte sich hingegen, dass die bilingualen Kinder in Deutschland durch die zwei eingesetzten LS-NWRTs benachteiligt werden, während sie im LITMUS-CL-NWRT sogar besser abschneiden als die monolingualen Kinder. Zaretsky et al. (2022) setzten jedoch nicht die Items des HASE-NK sondern die des KiSS (Holler-Zittlau, Euler & Neumann 2011) und des SSV (Grimm, Aktas & Kießig 2003) zum Vergleich mit dem LIT-

MUS-CL-NWRT ein. Da die Items dieser beiden LS-NWRTs im Vergleich zum HASE-NK insofern komplexer und einzelsprachspezifischer sind, dass deren Items Konsonantenverbindungen (Beispiel aus KiSS: Verklasenaft) und Morpheme des Deutschen (Beispiel aus SSV: Pristobierlichkeit) enthalten, können die Ergebnisse von Zaretsky et al. (2022) nur bedingt mit denen von Hinnerichs (2016) verglichen werden. Beim Vergleich von NWRTs sollte also nicht nur berücksichtigt werden, ob diese einzelsprachspezifisch oder einzelsprachunabhängig konstruiert wurden, die Itemkonstruktion muss in jedem Fall im Detail betrachtet werden. In Bezug auf die Itemkonstruktion von Nichtwörtern gibt es zahlreiche Faktoren – hier seien beispielhaft die Länge, die phonologische Komplexität und die Nähe zu realen Wörtern genannt – die einen Einfluss auf die Nachsprecheleistungen der Proband:innen ausüben können (für einen Überblick siehe Szewczyk et al. 2018).

### 1.3 Zum Auswertungsvorgehen bei NWRTs

Ein weiterer Punkt, der Unterschiede zwischen NWRT-Studien hervorrufen kann, ist das Auswertungsvorgehen. Während das Vorgehen bei Elisionen einheitlich ist – die Tilgung von Lauten oder Silben wird immer als Fehler gewertet –, so variiert die Vorgehensweise bei Additionen und Substitutionen zwischen den Studien. Laut- und Silbenadditionen werden zwar in den meisten Studien als Fehler gewertet, jedoch schließen Boerma et al. (2015, S.1753) diese mit der Begründung, dass dabei lediglich etwas hinzugefügt wird, aber kein Informationsverlust verzeichnet werden kann, von der Fehlerwertung aus.

In Bezug auf Substitutionen wird meist zwischen Vokalen und Konsonanten unterschieden. Während minimale Vokalabweichungen oft nicht als Fehler gewertet werden (beispielsweise [u/o] oder [i/ə]; Chiat, Polišíenská, Szewczyk & Kapalková 2016 und Tuller et al. 2018), so schließen Abed Ibrahim & Fekete (2019) jegliche Vokalabweichungen bei der Fehlerwertung aus. Bei Konsonanten werden oft Abweichungen bei der Stimmgebung toleriert, da bilinguale Kinder hierdurch benachteiligt werden können. So wird Stimmhaftigkeit je nach Sprache unterschiedlich realisiert (Dos Santos & Ferré 2016, S.5) und zudem liegt nicht in allen Sprachen die stimmhafte und stimmlose Variante eines Lautes vor. Im Standardarabischen gibt es beispielsweise nur den stimmhaften Labial /b/, nicht aber das stimmlose /p/ (Watson 2002, S.14). Des Weiteren werden oft konstante Substitutionen von Konsonanten bei der Fehlerwertung ausgeschlossen, da diese als „unreife“ Sprache oder Artikulationsschwierigkeit eingestuft werden und nicht als Symptom einer Sprachentwicklungsstörung (Abed Ibrahim & Fekete 2019, S.11; Boerma et al. 2015, S.1753; Chiat et al. 2016). Hierbei wird nicht berücksichtigt, dass konstante Substitutionen nicht nur ein Symptom phonetischer, also artikulatorischer Defizite sein können, sondern auch auf phonologische Schwierigkeiten hinweisen können, welche im deutschsprachigen Raum als Symptom einer Sprachentwicklungsstörung eingestuft werden (Kauschke & Vogt 2019, S. 179). Ebenso sprechen die Beobachtungen, dass Aussprachestörungen oft mit einer Sprachentwicklungsstörung einhergehen können (Vuolo & Goffmann 2020, S.2) und phonetische und phonologische Defizite ohnehin oft nicht klar voneinander abgrenzbar sind (Weinrich & Zehner 2017, S.37/38), dafür, konstante Substitutionen im Rahmen eines NWRTs als Fehler zu werten (Hinnerichs in Vorb.).

## 2 Zielsetzung und Forschungsfragen

Das Ziel der vorliegenden Untersuchung ist, zunächst anhand der Manipulation verschiedener Auswertungskriterien eine Auswertungsvorgehensweise festzulegen, die bestmöglich zwischen Kindern mit und ohne SES differenziert ohne dabei bilinguale Kinder zu benachteiligen (Studie 1). Daraufhin soll unter Anwendung dieser Auswertungsvorgehensweise untersucht werden, ob sich der LITMUS-CL-NWRT als Testinstrument im bilingualen Kontext eignet und dabei bessere Ergebnisse als ein bereits etablierter LS-NWRT aufweist (Studie 2).

Hierfür sollen folgende Forschungsfragen beantwortet werden:

- Inwiefern wird die Differenzierungsfähigkeit des LITMUS-CL-NWRT beeinflusst, wenn Additionen, Vokalsubstitutionen und konstante Substitutionen als Fehler gewertet werden? (Studie 1)
- Mit welcher Auswertungsvorgehensweise differenziert der LITMUS-CL-NWRT Kinder mit und ohne Sprachentwicklungsstörungen am besten voneinander? (Studie 1)

- Unterscheidet der LITMUS-CL-NWRT monolinguale und bilinguale Kinder mit und ohne Sprachentwicklungsstörungen? (Studie 2)
- Gibt es zwischen dem LITMUS-CL-NWRT und dem HASE-NK Unterschiede bezüglich der diagnostischen Validität? (Studie 2)

### 3 Methoden

Die hier beschriebenen Daten wurden im Rahmen des BiliSAT-Projekts (BILInguale Entwicklung von Schüler:innen mit Arabisch und Türkisch als Erstsprache; DFG-CH-1112/4-1; Prof. Dr. Chilla, Universität Flensburg/Prof. Dr. Hamann, Universität Oldenburg) erhoben. Abgesehen von den beiden NWRTs wurden mit den Proband:innen des Projekts zahlreiche weitere Testungen in der Erst- und Zweitsprache durchgeführt, um ihren Sprachstand adäquat einschätzen zu können. Für die Erhebungen waren 3-4 Testsitzungen notwendig, die meisten Kinder wurden zuhause getestet, einige auch in der Schule, im Kindergarten oder in einer logopädischen Praxis.

Die Proband:innen wurden entweder von der Autorin selbst oder von wissenschaftlichen Mitarbeiter:innen und Hilfskräften getestet. Die Auswertungen des LITMUS-CL-NWRT und HASE-NK wurden von der Autorin selbst vorgenommen und mit den Auswertungen einer anderen Wissenschaftlerin (Interraterin) abgeglichen.

#### 3.1 Proband:innen

Insgesamt wurde der LITMUS-CL-NWRT mit 103 Kindern durchgeführt, davon wurden bis auf drei Kinder auch alle mit dem HASE-NK getestet. Die Kinder wurden in vier Gruppen aufgeteilt, jeweils zwei Gruppen mit sprachlich unauffälligen Kindern (monolinguale Gruppe: MoTD; bilinguale Gruppe: BiTD) und zwei Gruppen mit Kindern, bei denen eine Sprachentwicklungsstörung (engl. DLD) diagnostiziert wurde (MoDLD und BiDLD).

##### 3.1.1 Gruppenzuordnung

Um festzustellen, bei welchen Kindern eine Sprachentwicklungsstörung vorliegt, wurden deren sprachliche Fähigkeiten im Deutschen und Türkischen bzw. Arabischen getestet. Für das Türkische wurde dabei der TELD-3-T (Turkish Test of Early Language Development; Form A; Topbaş & Güven 2011) durchgeführt, im Arabischen der ELO-L (Batterie d'Evaluation du Langage Oral chez l'enfant libanais; Zebib et al. 2017). Das Deutsche wurde mit der PLAKSS-II (Psycholinguistische Analyse kindlicher Aussprachestörungen; Fox-Boyer 2014), dem WWT (Wortschatz- und Wortfindungstest für 6 bis 10-Jährige; Glück 2011), dem TROG-D (Test zur Überprüfung des Grammatikverständnisses; Fox 2009) und der LiSe-DaZ (Linguistische Sprachstandserhebung – Deutsch als Zweitsprache; Schulz & Tracy 2011) getestet. Ein monolinguales Kind wurde der Gruppe mit Sprachentwicklungsstörungen zugeordnet, wenn es auf mindestens zwei sprachlichen Ebenen (Phonologie, Lexikon rezeptiv, Grammatik) im auffälligen Bereich lag, das heißt unter  $-1,25$  SD (Kriterium nach Tomblin, Records & Zhang 1996; siehe auch Hamann & Abed Ibrahim 2017). Zum Vorgehen bei Kindern, die zum Testzeitpunkt jünger bzw. älter als die Normgruppen einzelner Testungen waren, sei auf die ausführlichen Beschreibungen bei Hinnerichs (in Vorb.) verwiesen. Bei den bilingualen Kindern wurden die Normabweichungen zur Definition einer Auffälligkeit je nach Sprachdominanz angepasst (mit Ausnahme der LiSe-DaZ, die bereits Normwerte für bilinguale Kinder enthält).

Die Sprachdominanz wurde auf der Basis eines ausführlichen Elternfragebogens (PaBiQ, Tuller 2015) berechnet (Dominante Sprache:  $-1,5$  SD; nicht-dominante Sprache:  $-2,25$  SD; ausgeglichener „balanced“ Sprachstatus:  $-1,75$  SD; Thordardottir 2015 und Tuller et al. 2018). Es wurden hierbei Informationen zum Sprachkontakt in den ersten vier Lebensjahren, sowie auch zum aktuellen Sprachgebrauch erfragt (für eine ausführliche Beschreibung hierzu siehe Abed Ibrahim & Fekete 2019, S.8). Durch die Elternbefragung wurde ebenfalls das Vorliegen von Hörstörungen und anderen Beeinträchtigungen ausgeschlossen, mit der Raven's Coloured Progressive Matrices (CPM; Bulheller & Häcker 2002) wurde überprüft, ob der nonverbale IQ im Normbereich liegt.

##### 3.1.2 Gruppenbeschreibung

Eine Übersicht der Proband:innen findet sich in Tabelle 1. Hier wird deutlich, dass die Kinder eine große Altersspanne umfassen (3;7 - 13;9) und auch die Kontaktdauer zum Deutschen, sowie das Alter zu Erwerbsbeginn (im Folgenden: AoO, nach dem engl. Age of Onset) in den bilin-



gualen Gruppen stark variieren. Die BiTD-Gruppe hat mehr Kontaktdauer ( $M = 5;0$  Jahre) als die BiDLD-Gruppe ( $M = 4;6$  Jahre), der Unterschied ist jedoch nicht signifikant ( $p > .05$ ;  $U = 347.5$ ). Die Unterschiede bezüglich des Alters zu Erwerbsbeginn des Deutschen sind hingegen bedeutsam ( $p < .01$ ;  $U = 181.5$ ). Die BiTD-Kinder hatten mit durchschnittlich 4;10 Jahren deutlich später Kontakt zum Deutschen als die Kinder der BiDLD-Gruppe ( $M = 2;2$  Jahre). Dieser Unterschied scheint widersprüchlich zu den Angaben der Kontaktdauer und lässt sich dadurch erklären, dass die Kinder der BiTD-Gruppe auch deutlich älter sind ( $M = 9;8$  Jahre) als die der BiDLD-Gruppe ( $M = 6;8$ ), wodurch sie trotz eines höheren AoO eine etwas längere Kontaktdauer aufweisen. Die MoDLD- und Mo-TD-Gruppe liegen mit ihren Altersdurchschnittswerten zwischen den Gruppen der bilingualen Kinder (MoDLD:  $M = 8;0$  Jahre; MoTD:  $M = 8;7$ ), die Altersunterschiede sind nur zwischen der BiDLD- und BiTD-Gruppe signifikant ( $p < .01$ ). Dies wird bei den statistischen Berechnungen berücksichtigt.

Tab. 1: Übersicht zu Alter und Sprachkontaktdaten aller Teilnehmer:innen (nach Hinnerichs in Vorb.)

Gruppe	n	Alter (in Monaten)		Kontaktdauer zum Deutschen (in Monaten)		AoO des Deutschen (in Monaten)	
		M (SD)	Spanne	M (SD)	Spanne	M (SD)	Spanne
MoTD	18	102.7 (26.0)	51-145	-			
MoDLD	14	96.0 (34.3)	43-153	-			
BiTD	58	115.7 (28.1)	49-165	59.9 (37.0)	13-135	57.6 (37.6)	0-126
BiDLD	13	80.2 (12.4)	67-113	54.3 (20.4)	17-83	25.9 (19.0)	0-59
Gesamt	103	106.3 (29.6)	43-165	58.7 (34.6)	13-135	51.5 (36.9)	0-126

Anhand der Angaben zum AoO wird deutlich, dass manche Kinder schon von Geburt an Kontakt zum Deutschen hatten, also simultan bilingual aufwachsen, während andere Kinder sogar erst mit über 10 Jahren Kontakt zum Deutschen hatten. Insgesamt wachsen zwei Drittel der bilingualen Kinder dieser Untersuchung sukzessiv bilingual auf, während ein Drittel dem simultanen Erwerb (AoO < 36 Monate) zugeordnet werden kann. Der sozio-ökonomische Status wurde durch den Bildungsgrad der Mutter festgelegt (Hoff 2006), es fanden sich keine signifikanten Unterschiede zwischen den Gruppen (Hinnerichs in Vorb.).

Für die Berechnungen von Studie 2 wurden die Ergebnisse aller in Tabelle 1 aufgeführten Kinder miteinbezogen. Für Studie 1 zum Auswertungsvorgehen wurde nur ein Teil der Proband:innen berücksichtigt, da zum Zeitpunkt der Analysen noch nicht alle Daten erhoben und ausgewertet waren. Hierfür wurden nur Kinder der BiTD- und MoDLD-Gruppe ausgewählt und verglichen, da die Unterscheidung dieser Gruppen meist als größte Herausforderung gilt: Wenn die Nachteile, die eventuell durch den geringeren Sprachkontakt bilingualer Kinder entstehen zu denselben, vergleichsweise niedrigen Testergebnissen führen wie die Symptome einer Sprachentwicklungsstörung, dann werden die Kinder der BiTD-Gruppe fälschlicherweise als sprachlich auffällig diagnostiziert. Ein Auswertungsvorgehen, durch welches die BiTD- und MoDLD-Gruppe bestmöglich voneinander differenziert werden, kann also als sensitiv für die Defizite einer Sprachentwicklungsstörung gelten ohne dabei bilinguale Kinder zu benachteiligen. Für Studie 2 wurden 22 Kinder der BiTD-Gruppe (Alter Ø 10;2 Jahre) mit 10 Kindern der MoDLD-Gruppe (Alter Ø 9;0 Jahre) verglichen.

### 3.2 Testungen

Der LITMUS-CL-NWRT enthält 16 Items, wobei zu jedem Item vier bis sechs Alternativen konstruiert wurden, aus denen das passendste für die jeweilige Sprache ausgesucht werden kann. Für das Deutsche wurde die Testung nach Hinnerichs (2016) adaptiert und enthält nur Items, die weder im Deutschen, noch in den Erstsprachen der hier getesteten Kinder als reale Wörter vorkommen (Hinnerichs, Polišenská & Chiat 2016). Bei der Lautauswahl der Items wurde darauf geachtet, dass diese mit möglichst vielen Sprachen kompatibel sind. Die Nichtwörter enthalten die drei Vokale /a, i, u/, und die elf Konsonanten /p, b, t, d, k, g, s, z, l, m, n/ (siehe Tabelle 2). Die Testung besteht aus jeweils vier 2-, 3-, 4- und 5-Silbern, alle Silben sind offen und haben nur einen Konsonanten im Anlaut (CV-Struktur). Die Items werden mit neutraler Prosodie präsentiert, das heißt, dass alle Silben gleichermaßen betont werden.

Die Items des HASE-NK entstammen dem Untertest Nachsprechen von Kunstwörtern aus dem HASE (Schöler & Brunner 2008; siehe auch 1.2). Sie bestehen ebenfalls nur aus CV-Silben,

die Lautauswahl ist jedoch etwas größer als beim LITMUS-CL-NWRT. So enthält der HASE-NK sechs Vokale (/a, e, ə, i, o, u/) und elf Konsonanten (/p, b, d, k, g, f, v, ʁ, l, m, n/), also drei Vokale mehr als der LITMUS-CL-NWRT (siehe Tabelle 2). Das Betonungsmuster entspricht dem Deutschen, Zweisilber werden beispielsweise auf der ersten Silbe betont, Dreisilber auf der zweiten Silbe.

Tab. 2: Nichtwörter des LITMUS-CL-NWRT und HASE-NK

Testung	2-Silber	3-Silber	4-Silber	5-Silber
LITMUS-CL-NWRT	/naki/ - 1	/nalidu/ - 11	/zipalita/ - 13	/litapimudi/ - 9
	/zipu/ - 6	/zipula/ - 2	/gazulumi/ - 8	/sipunakila/ - 10
	/tula/ - 7	/panuti/ - 3	/nugitala/ - 5	/tulikazimu/ - 15
	/lumi/ - 16	/luniga/ - 14	/litisagu/ - 12	/maluzikupa/ - 4
HASE-NK	/mali/	/lakemo/	/pukavobə/	
	/lufa/	/vunobə/	/fodəkina/	
	/bidə/	/godepi/	/Ɂibanelu/	

Die Nichtwörter des HASE-NK wurden, wie im Manual vorgegeben, in der hier aufgeführten Reihenfolge präsentiert, also von oben nach unten, beginnend in der linken Spalte. Die Nichtwörter des LITMUS-CL-NWRTs wurden im Gegensatz zu denen des HASE-NK nicht mit steigender Silbenanzahl präsentiert, sondern in der durch die Zahlen in Tabelle 2 angegebenen Reihenfolge.

Die beiden Testungen wurden meistens direkt hintereinander durchgeführt. Die Items des LITMUS-CL-NWRT wurden dabei im Rahmen einer Power Point-Präsentation abgespielt, die Nichtwörter werden im Kontext einer Geschichte „Zauberwörter“ genannt, mit denen eine kaputte Perlenkette wieder „ganz gezaubert“ werden kann (Polišenská & Kapalková 2014; siehe Abbildung 1). Die Items des HASE-NK wurden ebenfalls als Zauberwörter eingeführt, jedoch nur als Audioaufnahme, ohne Power-Point-Präsentation, abgespielt, um möglichst nahe an den Vorgaben des Manuals zu bleiben (Schöler & Brunner 2008).

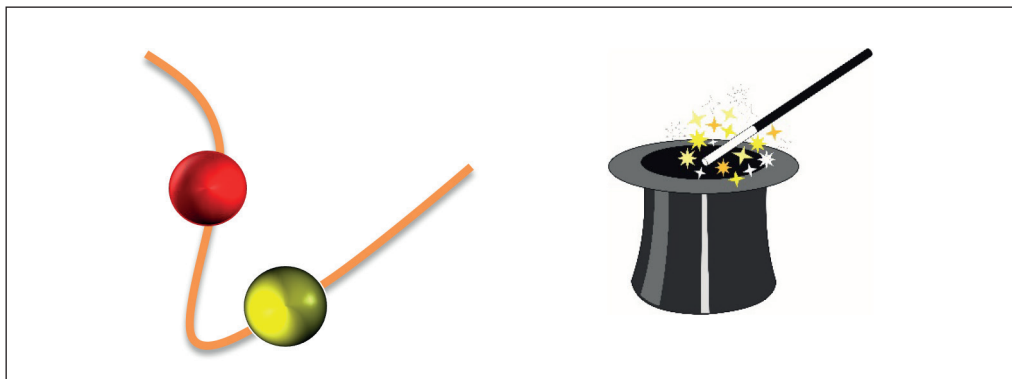


Abb. 1: Power-Point-Präsentation zur Durchführung des LITMUS-CL-NWRT: „Perlen auf eine Kette zaubern“ (Polišenská & Kapalková 2014)

### 3.3 Auswertungsvorgehensweisen

Um den Einfluss verschiedener Auswertungsvorgehensweisen zu vergleichen, wurde zunächst festgelegt, welche Abweichungen grundsätzlich als Fehler gelten und welche Abweichungen untersucht werden sollten, da sie in bisherigen Studien nicht einheitlich als Fehler bzw. zu tolerierende Abweichungen eingestuft wurden. So werden Laut- und Silbenelisionen und Laut- und Silbenumstellungen als Fehler gewertet, ebenso auch Nullreaktionen oder Itemwiederholungen, die keinerlei Ähnlichkeit mehr mit dem vorgegebenen Nichtwort haben. Konsonantensubstitutionen werden ebenfalls als Fehler gewertet, mit Ausnahme von Abweichungen bei der Stimmgebung, da hier die bilinguale Erwerbssituation unabhängig vom Vorliegen einer Sprachentwicklungsstörung zu Abweichungen führen kann (siehe 1.3).

Des Weiteren wurde in Studie 1 untersucht, inwiefern sich die Wertung von Laut- und Silbenadditionen, konstanten Substitutionen und Vokalsubstitutionen auf die Differenzierungsfähigkeit des LITMUS-CL-NWRT auswirkt. In Bezug auf Vokalsubstitutionen werden diese bei minimalen Abweichungen grundsätzlich nicht als Fehler gewertet, um ebenso wie in Bezug auf

die Stimmhaftigkeit von Konsonanten die bilinguale Erwerbssituation zu berücksichtigen. Ob es jedoch empfehlenswert ist, jegliche Vokalsubstitution von der Fehlerwertung auszuschließen (siehe 1.3), soll mit den vorliegenden Daten überprüft werden.

## 4 Ergebnisse

Das Auswertungsvorgehen, das durch die Ergebnisse von Studie 1 festgelegt wurde, diente als Basis für die Auswertungen zu Studie 2. Es wurde mit einem Interraterabgleich überprüft, die Interrater-Reliabilität ergab eine fast vollkommene Übereinstimmung ( $k = .997$ ; Hinnerichs in Vorb.).

Alle statistischen Berechnungen wurden mit SPSS 27 durchgeführt. Die Ergebnisse sind Auszüge des Dissertationsprojekts der Autorin (Hinnerichs in Vorb.).

### 4.1 Ergebnisse von Studie 1

Um der Frage nachzugehen, welchen Einfluss verschiedene Auswertungsvorgehensweisen auf die Unterscheidung der BiTD- und MoDLD-Gruppe haben, wurde zunächst verglichen, wie oft Additionen, Vokalsubstitutionen und konstante Substitutionen in den beiden Gruppen vorkommen. Aufgrund der kleinen Stichprobe wurde der Mann-Whitney-U-Test herangezogen. Dieser ergab, dass sowohl Additionen als auch Vokalsubstitutionen signifikant häufiger in der MoDLD- als in der BiTD-Gruppe vorkommen (Additionen:  $p < .05$ ;  $U = 58.5$ ; Vokalsubstitutionen:  $p < .01$ ;  $U = 44.5$ ). Konstante Substitutionen kamen in keiner Gruppe vor und werden deshalb für die weiteren Berechnungen nicht berücksichtigt.

Die deskriptive Auswertung zeigt, dass Additionen zwar bei sieben der zehn Kinder der MoDLD-Gruppe vorkommen, jedoch nicht alle diese Kinder durch den Ausschluss von Additionen einen höheren Gesamtpunktwert erreichen (siehe Tabelle 3). Der Grund hierfür liegt darin, dass viele Items, bei deren Nachsprechen Additionen verzeichnet werden konnten, noch weitere Abweichungen enthalten - beispielsweise Elisionen - und somit ohnehin als falsch gewertet würden (Hinnerichs in Vorb.). Ebenso führen die 22 Vokalsubstitutionen, die in der MoDLD-Gruppe beobachtet und laut Kriterium 2 nicht als Fehler gewertet wurden, nur bei drei Kindern zu einem Anstieg von jeweils einem Punkt (siehe Tabelle 3).

Tab. 3: Übersicht zum Auftreten der Fehlerarten (Studie 1)

Fehlerarten	MoDLD	BiTD
<b>Kriterium 1: Additionen</b>		
Anzahl der Kinder, bei denen diese Abweichung vorkommt:	7/10 Kinder (70 %)	5/22 Kinder (23 %)
Anzahl der Additionen:	10	6
Auswirkung auf das Ergebnis:	3 mal + 1 Punkt	2 mal + 1 Punkt
<b>Kriterium 2: Vokalsubstitutionen</b>		
Anzahl der Kinder, bei denen diese Abweichung vorkommt:	8/10 Kinder (80 %)	6/22 Kinder (27 %)
Anzahl der Vokalsubstitutionen:	22	10
Auswirkung auf das Ergebnis:	3 mal + 1 Punkt	3 mal + 1 Punkt

Deshalb wurde überprüft, ob das häufigere Auftreten von Additionen und Vokalsubstitutionen in der MoDLD-Gruppe letztendlich dazu führt, dass diese hierdurch tatsächlich schlechter als die BiTD-Gruppe im LITMUS-CL-NWRT abschneidet.

Zu diesem Zweck wurde verglichen, wie hoch die Differenz der Mittelwerte bei Anlegung verschiedener Auswertungsvorgehensweisen ist. In Tabelle 4 befindet sich eine Beschreibung der verschiedenen Auswertungsvorgehensweisen.

Tab. 4: Übersicht und Beispiele zu den Auswertungsvorgehensweisen in Studie 1 (nach Hinnerichs in Vorb.)

Auswertungsvorgehensweisen	Beispiele
<b>Strenge Auswertung</b> Kriterien 1 und 2 werden als Fehler gewertet (minimale Vokalabweichungen werden toleriert)	/nugikatala/ statt /nugitala/ > falsch /zipulita/ statt /zipalita/ > falsch /panoti/ statt /panuti/ > korrekt
Kriterium 1: <b>Laut- und Silbenadditionen</b> zählen NICHT als Fehler	/nugikatala/ statt /nugitala/ > korrekt
Kriterium 2: <b>Vokalsubstitutionen</b> zählen NICHT als Fehler (alle Vokalabweichungen werden toleriert)	/zipulita/ statt /zipalita/ > korrekt
<b>Milde Auswertung</b> Kriterien 1 und 2 werden NICHT als Fehler gewertet	/nugikatala/ statt /nugitala/ > korrekt /zipulita/ statt /zipalita/ > korrekt /panoti/ statt /panuti/ > korrekt

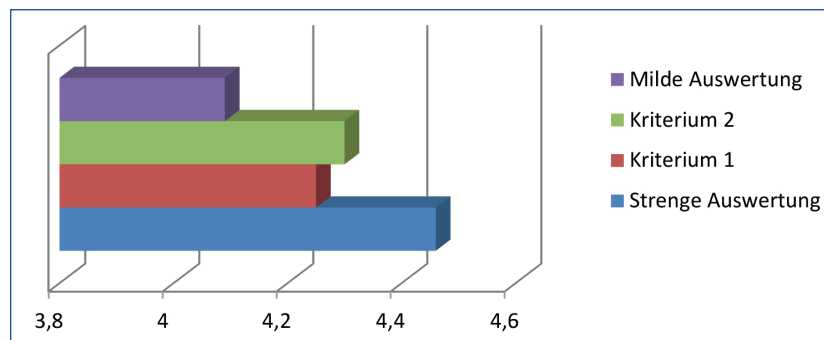
So werden, wie unter 3.3 beschrieben, einige Abweichungen, wie Elisionen und Umstellungen, immer als Fehler gewertet, während andere Abweichungen, wie Veränderungen der Stimmhaftigkeit oder minimale Vokalabweichungen, grundsätzlich toleriert werden. In Bezug auf Additionen (Kriterium 1) und Vokalsubstitutionen (Kriterium 2) wird untersucht, wie sich deren Wertung auf die Endergebnisse des LITMUS-CL-NWRTs auswirkt. Vergleicht man zunächst die Ergebnisse des Mann-Whitney-U-Tests, der nach jedem dieser Auswertungsvorgehensweisen durchgeführt wurde, so zeigt sich, dass alle Auswertungsvorgehensweisen die BiTD- und MoDLD-Gruppe signifikant voneinander unterscheiden (strenge Auswertung:  $p < .01$ ;  $U = 13.5$ ; Kriterium 1:  $p < .01$ ;  $U = 15.00$ ; Kriterium 2:  $p < .01$ ;  $U = 11.5$ ; milde Auswertung:  $p < .01$ ;  $U = 11.5$ ). Die BiTD-Gruppe schneidet also unabhängig von der Auswertungsvorgehensweise immer besser beim LITMUS-CL-NWRT ab.

In einem nächsten Schritt wurden die Ergebnisse für jede Gruppe im Mittel berechnet, woraufhin die Differenz der beiden Mittelwerte gebildet wurde. Tabelle 5 fasst diese Werte für die vier Auswertungsvorgehensweisen zusammen.

Tab. 5: Statistische Werte zu den einzelnen Auswertungsvorgehen im Gruppenvergleich

Statistische Werte	Auswertungsvorgehensweisen			
	Strenge Auswertung	Kriterium 1	Kriterium 2	Milde Auswertung
<b>Mittelwert (SD)</b>	BiTD: 14.36 (1.29) MDLD: 9.9 (2.88)	BiTD: 14.45 (1.3) MDLD: 10.2 (2.9)	BiTD: 14.5 (1.19) MDLD: 10.2 (2.9)	BiTD: 14.64 (1.09) MDLD: 10.5 (2.92)
<b>Minimum-Maximum (Spannweite)</b>	BiTD: 11-16 (5) MDLD: 6-14 (8)	BiTD: 11-16 (5) MDLD: 6-14 (8)	BiTD: 12-16 (4) MDLD: 6-14 (8)	BiTD: 12-16 (4) MDLD: 6-14 (8)
<b>Mittelwertsdifferenz</b>	<b>4.46</b>	<b>4.25</b>	<b>4.3</b>	<b>4.14</b>

Mit Abbildung 2 wird verdeutlicht, dass durch die strenge Auswertung mit 4.46 Punkten die größte Mittelwertdifferenz zwischen der MoDLD- und BiTD-Gruppe zustande kommt. Die geringste Differenz erfolgt mit 4.14 Punkten durch die milde Auswertung, die Anwendungen von Kriterium 1 und 2 liegen dazwischen. Insgesamt liegen die Mittelwertdifferenzen nicht weit auseinander, der maximale Unterschied beträgt 0.32 Punkte.



Werte von oben nach unten: 4.14; 4.3; 4.25; 4.46

Abb. 2: Mittelwertdifferenzen zwischen der BiTD- und MoDLD-Gruppe bei den verschiedenen Auswertungsvorgehensweisen (aus Hinnerichs in Vorb.)

#### 4.2 Ergebnisse von Studie 2

Mit Studie 2 soll zunächst die Frage beantwortet werden, ob der LITMUS-CL-NWRT monolinguale und bilinguale Kinder mit und ohne Sprachentwicklungsstörungen voneinander unterscheiden kann. Vergleicht man die Mittelwerte der vier Gruppen, so wird deutlich, dass die beiden TD-Gruppen beim LITMUS-CL-NWRT besser abschneiden als die DLD-Gruppen. So erreicht die MoTD-Gruppe durchschnittlich 13.61 Punkte ( $SD = 1.65$ ) und die BiTD-Gruppe 13.2 Punkte ( $SD = 2.29$ ), während es bei der MoDLD-Gruppe im Mittel 9.36 Punkte ( $SD = 2.92$ ) und bei der BiDLD 9.08 Punkte ( $SD = 2.99$ ) sind.

Mittels einer Kovarianzanalyse wurde überprüft, ob diese Unterschiede signifikant sind. Als Kovariaten wurden hierbei das Alter und der sozio-ökonomische Status berücksichtigt. Im Gegensatz zum Alter unterscheiden sich die Gruppen zwar hinsichtlich des sozio-ökonomischen Status nicht signifikant (siehe 3.1.2.), jedoch wurde dieser als Kovariate miteinbezogen, da sich schwache Korrelationen zwischen dem LITMUS-CL-NWRT und dem sozio-ökonomischen Status der Teilnehmer:innen zeigten ( $r_s = .227, p < .05, n = 100$ ). Zwischen dem Alter und dem LITMUS-CL-NWRT zeigte sich ein mittlerer Zusammenhang ( $r = .479, p < .01, n = 103$ ). Die Kontaktdauer und das AoO wurden nicht als Kovariaten in die Berechnungen miteinbezogen, da für beide gezeigt werden konnte, dass sie keinen Zusammenhang zu den Ergebnissen des LITMUS-CL-NWRT aufweisen, wenn der Einfluss des Alters mit einer partiellen Korrelation kontrolliert wird (Kontaktdauer:  $r_p = .061$ ; AoO:  $r_p = -.061$ ).

Die Ergebnisse der Kovarianzanalyse zeigen, dass sich die Gruppen hinsichtlich ihrer Ergebnisse beim LITMUS-CL-NWRT mit einer großen Effektstärke signifikant voneinander unterscheiden ( $F(3, 94) = 12.5, p < .01, \eta_p^2 = .29$ ). Durch die Post-hoc-Analysen mit Bonferroni-Korrektur wird deutlich, dass beide DLD-Gruppen signifikant weniger Punkte beim LITMUS-CL-NWRT erreichen als die TD-Gruppen ( $p < .01$ ), während sich weder zwischen der MoTD- und BiTD-, noch zwischen der BiDLD- und MoDLD-Gruppe signifikante Unterschiede zeigen ( $p > .05$ ). Bezogen auf die Kovariaten zeigt sich, dass das Alter einen signifikanten Einfluss auf die Ergebnisse des LITMUS-CL-NWRT hat ( $p < .01$ ), während dies beim sozio-ökonomischen Status nicht der Fall ist ( $p > .05$ ).

Abbildung 3 veranschaulicht die Gruppenunterschiede und zeigt, dass diese mit zunehmender Itemlänge größer werden.

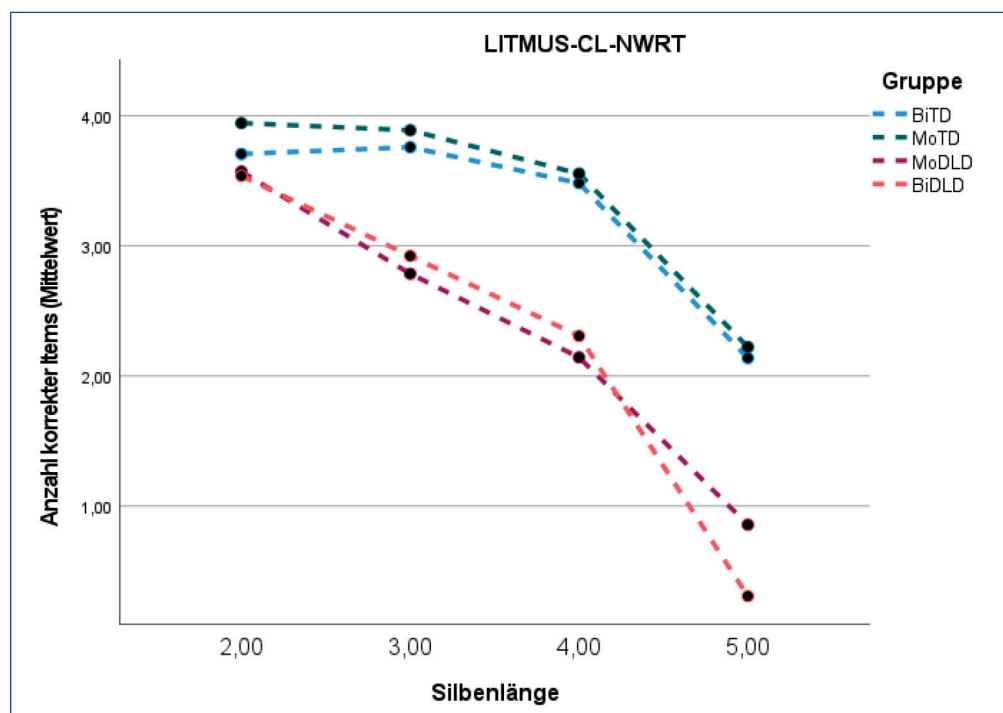


Abb. 3: Diagramm zum Längeneffekt beim LITMUS-CL-NWRT (aus Hinnerichs in Vorb.)

Während beim Nachsprechen von Zweisilbern keine bedeutsamen Unterschiede zwischen den TD- und DLD-Gruppen zu verzeichnen sind, sprechen die Kinder der TD-Gruppen Drei-, Vier-

und Fünfsilber signifikant besser nach als die Kinder der DLD-Gruppen, wobei die Effektstärke bei den Fünfsilbern den höchsten Wert erreicht (Dreisilber  $\eta_p^2 = .28$ ; Viersilber  $\eta_p^2 = .24$ ; Fünfsilber  $\eta_p^2 = .30$ ). Ebenso wie in Bezug auf die Gesamtergebnisse zeigt sich auch für die einzelnen Itemlängen, dass kein bedeutsamer Unterschied zwischen den bilingualen und monolingualen Kindern zu verzeichnen ist.

Zu den Ergebnissen des HASE-NK wurde ebenfalls eine Kovarianzanalyse berechnet und es zeigt sich, dass analog zu den Berechnungen beim LITMUS-CL-NWRT auch hier signifikante Unterschiede zwischen den TD- und DLD-Gruppen zu verzeichnen sind, während der Faktor Bilingualismus keinen Einfluss auf die Ergebnisse hat (Näheres unter Hinnerichs in Vorb.).

Da signifikante Gruppenunterschiede nicht ohne Weiteres mit einer zufriedenstellenden diagnostischen Validität gleichzusetzen sind, wurde in einem nächsten Schritt berechnet, wie viele Kinder die beiden NWRTs korrekt als sprachlich unauffällig bzw. auffällig einordnen (siehe Tabelle 6). Mittels einer ROC-Kurve wurde der ideale Cut-off-Wert berechnet, jeweils für die monolingualen und bilingualen Kinder getrennt. Alle Kinder, die auf oder unter diesem Wert liegen, werden als sprachlich auffällig klassifiziert. Die Sensitivität gibt an, wieviel Prozent der Kinder mit DLD korrekt als sprachlich auffällig erkannt wurden, die Spezifität bezieht sich auf die korrekte Klassifizierung der TD-Kinder. Der AUC-Wert gibt die Vorhersagegenauigkeit einer Testung an. Liegt er unter .60, gilt die Testung als wertlos, bis zu .70 als schwach, bis zu .80 als akzeptabel, bis zu .90 als gut und ein Wert zwischen .90 und 1.0 gilt als sehr gut (Armon-Lotem & Meir 2016).

Tab. 6: Übersicht der Werte zur diagnostischen Validität des LITMUS-CL-NWRT und HASE-NK (nach Hinnerichs in Vorb.)

		LITMUS-CL-NWRT	HASE-NK
<b>Bilinguale Kinder</b>	N	71	68
	AUC	.873	.774
	Cut-off-Wert	12	6
	Sensitivität	<b>85 %</b>	67%
	Spezifität	71 %	73 %
<b>Monolinguale Kinder</b>	N	32	32
	AUC	.903	.873
	Cut-off-Wert	12	7
	Sensitivität	<b>86 %</b>	<b>93 %</b>
	Spezifität	78 %	61 %

Vergleicht man die AUC-Werte der beiden NWRTs, so wird deutlich, dass der LITMUS-CL-NWRT sowohl die monolingualen als auch die bilingualen Kinder besser unterscheidet. So ist die Vorhersagegenauigkeit des HASE-NK bei den bilingualen Kindern im akzeptablen Bereich, während sie beim LITMUS-CL-NWRT im guten Bereich liegt. Bei den monolingualen Kinder liegt der HASE-NK zwar auch im guten Bereich, der LITMUS-CL-NWRT weist jedoch eine sehr gute Vorhersagegenauigkeit auf. In Bezug auf die bilingualen Kinder weist der LITMUS-CL-NWRT eine deutlich höhere Sensitivität auf, wobei der HASE-NK eine etwas höhere Spezifität als der LITMUS-CL-NWRT zeigt. Bei den monolingualen Kindern ist es genau umgekehrt, hier zeigt der HASE-NK eine höhere Sensitivität und der LITMUS-CL-NWRT eine höhere Spezifität.

Legt man diesen Werten die Einschätzung von Plante & Vance (1994) zugrunde, wonach Sensitivitäts- und Spezifitätswerte ab 90 % für eine gute diagnostische Genauigkeit stehen, ab 80 % akzeptabel sind, aber unter 80 % eine zu hohe Rate von Fehldiagnosen anzeigen, so liegen einige der Werte aus Tabelle 6 unterhalb des zufriedenstellenden Bereichs (die Werte, die im zufriedenstellenden Bereich liegen, wurden in Tabelle 6 fett markiert). Beim HASE-NK ist lediglich die Sensitivität bezüglich monolingualer Kinder im guten Bereich, alle anderen Werte liegen deutlich unter 80 %. Die Sensitivität des LITMUS-CL-NWRT liegt immer im akzeptablen Bereich, sowohl bei den monolingualen als auch bei den bilingualen Kindern. Die Spezifitätswerte liegen beide unter 80 %, wobei der Wert der monolingualen Kinder mit 78 % nahe am akzeptablen Bereich liegt.

## 5 Diskussion und Schlussfolgerungen

Die vorliegende Studie begründet sich durch die Herausforderung, die sprachlichen Fähigkeiten bilingualer Kinder im Hinblick auf das Vorliegen einer Sprachentwicklungsstörung einzuschätzen. Da die Testung einzelner Sprachen viele Hindernisse mit sich bringt, wird das Nachsprechen von Nichtwörtern aufgrund seiner Neutralität in Bezug auf Einzelsprachen als vielversprechendes Testinstrument angesehen.

Jedoch zeigen bisherige Studien vielfach, dass auch NWRTs nicht so einzelsprachunabhängig sind wie angenommen.

Der LITMUS-CL-NWRT wurde im Rahmen eines internationalen Forschungsprojekts entwickelt, wobei bei der Itemkonstruktion auf die Kompatibilität mit möglichst vielen Sprachen Wert gelegt wurde.

### 5.1 Diskussion zu Studie 1

Bevor die Eignung des LITMUS-CL-NWRTs anhand von vier Gruppen mit monolingualen und bilingualen Kindern mit und ohne Sprachentwicklungsstörung untersucht wurde, sollten durch eine kleiner angelegte Studie einzelne Auswertungskriterien überprüft werden. So war es das Ziel von Studie 1 herauszufinden, inwiefern sich die Wertung von Additionen, Vokalsubstitutionen und konstanten Konsonantensubstitutionen auf die Differenzierung der BiTD- und MoDLD-Gruppe auswirkt. Die Ergebnisse zeigen, dass es keinen großen Unterschied macht, wenn Additionen und/oder Vokalsubstitutionen von der Fehlerwertung ausgenommen werden. Dennoch sprechen die vorliegenden Daten dafür, diese beiden Abweichungsarten als Fehler zu werten, da sich hierdurch die Differenz zwischen den Kindern mit und ohne Sprachentwicklungsstörung leicht erhöht, ohne dass dabei die bilingualen Kinder benachteiligt werden.

Konstante Substitutionen kamen hingegen weder in der BiTD- noch in der MoDLD-Gruppe vor. Dies ist wohl vor allem durch das Alter der Proband:innen zu erklären. Alle Kinder in Studie 1 sind mindestens 6 Jahre alt, die meisten sogar älter. Wenn konstante Substitutionen im NWRT durch das Vorliegen konstanter phonologischer Prozesse begründet sind, so ist davon auszugehen, dass diese in dieser Altersgruppe nicht mehr beobachtet werden, da phonologische Prozesse im unauffälligen Spracherwerb spätestens am Ende des vierten Lebensjahres überwunden werden (Kannengieser 2019). Sollten also bei Kindern über fünf Jahren im Rahmen eines NWRTs konstante Substitutionen beobachtet werden, so ist es ohnehin nicht ratsam, diese aus der Fehlerwertung auszuschließen, da sie das Symptom einer phonologischen Verzögerung oder Störung sein können (Weinrich & Zehner 2017, S.41f.).

Wie verhält es sich nun aber mit jüngeren Kindern? Sollten bei diesen, wie es von einigen Wissenschaftler:innen gehandhabt wird, konstante Substitutionen von der Fehlerwertung ausgeschlossen werden, da diese durch altersgemäße, physiologische phonologische Prozesse zustande kommen könnten?

Es gibt einige Gründe, die dagegen sprechen: Vor allem bei groß angelegten Screeninguntersuchungen, in deren Rahmen NWRTs oft eingesetzt werden, kann im Einzelfall wohl kaum eingeschätzt werden, ob der konkrete Substitutionsprozess physiologisch und noch altersadäquat ist. So müsste beim Ausschluss konstanter Substitutionen aus der Fehlerwertung beispielsweise auch die konstante Rückverlagerung von /t/ zu /k/ toleriert werden, obwohl dies als ungewöhnlicher und therapiebedürftiger Prozess gilt.

Ein anderer Aspekt hinsichtlich konstanter Substitutionen ist das Argument, dass diese auch durch phonetische Störungen hervorgerufen werden können. Letztere gelten im Gegensatz zu phonologischen Auffälligkeiten nicht als Symptom einer Sprachentwicklungsstörung, sondern werden zu den Aussprachestörungen gezählt (Kauschke & Vogt 2019). Bei der Durchführung eines Screenings sollten jedoch auch Kinder mit phonetischen Störungen von einem Screeninginstrument herausgefiltert werden, da sie ebenso wie Kinder mit einer Sprachentwicklungsstörung einen Therapiebedarf haben. Auch unter diesem Aspekt scheint es also nicht ratsam, konstante Substitutionen aus der Fehlerwertung auszuschließen. Im Hinblick auf Studien, die einen NWRT gezielt bezüglich seiner Eignung als Testung bei Sprachentwicklungsstörungen untersuchen und somit Kinder mit Aussprachestörungen nicht erfassen wollen, sei angemerkt, dass phonetische und phonologische Störungen oft in Kombination auftreten und in vielen Fällen keine klare Zuordnung zu einem Störungsschwerpunkt möglich ist (Weinrich & Zehner 2017).

Ein weiteres Argument, das für die strengere Auswertung spricht, also für den Einbezug von Additionen, Vokalsubstitutionen und konstanten Substitutionen in die Fehlerwertung, ist die

Praktikabilität. Je mehr Abweichungen aus der Fehlerwertung ausgeschlossen werden, desto fehleranfälliger wird diese in der praktischen Anwendung. Die Abweichungen, die nicht als Fehler gewertet werden dürfen, müssen erst als solche erkannt werden. Im Berufsalltag ist oft wenig Zeit und eine Liste von Abweichungsarten, die nicht als Fehler gezählt werden dürfen, macht die Auswertung von NWRTs komplizierter und langwieriger. Eine Ausnahme sollten also lediglich die Abweichungen bilden, die durch die bilinguale Erwerbssituation erklärt werden können. In diesem Fall lohnt es sich die Auswertungsvorgehensweise etwas komplexer zu gestalten, um Fehldiagnosen bezüglich bilingualer Kinder möglichst zu vermeiden. Dies betrifft nach aktuellem Kenntnisstand Abweichungen bei der Stimmgebung und minimale Unterschiede bei der Realisierung von Vokalen. Für Substitutionen durch Vokale, die im Spektrum weiter auseinander liegen, ebenso wie auch für Additionen konnte mit den Daten aus Studie 1 gezeigt werden, dass durch deren Einbezug in die Fehlerwertung keine Benachteiligung bilingualer Kinder entsteht. Da die vorliegende Stichprobe sehr klein ist, sollten diese Erkenntnisse durch weitere Untersuchungen überprüft werden.

## 5.2 Diskussion zu Studie 2

Die hier als „streng“ definierte Auswertungsvorgehensweise (siehe Tabelle 4) wurde der Gesamtauswertung des LITMUS-CL-NWRTs zugrunde gelegt. Beim Vergleich der vier untersuchten Gruppen zeigen die Ergebnisse, dass mit dem LITMUS-CL-NWRT ein signifikanter Unterschied zwischen Kindern mit und ohne Sprachentwicklungsstörung gemessen werden kann, ohne dass dabei die bilinguale Erwerbssituation einen nennenswerten Einfluss auf die Nachsprecheleistungen hat. Vor allem der Einsatz längerer Items erhöht die Differenzierungsfähigkeit der Testung. Während das Nachsprechen von Zweisilbern keinen nennenswerten Unterschied zwischen den TD- und DLD-Gruppen hervorruft, können bei Drei-, Vier- und Fünfsilbern signifikante Unterschiede verzeichnet werden.

Da signifikante Gruppenunterschiede nicht automatisch bedeuten, dass ein Großteil der Kinder diagnostisch korrekt erkannt wird, wurden in einem weiteren Schritt die Spezifität und Sensitivität berechnet. Die Werte zur diagnostischen Validität des LITMUS-CL-NWRTs liegen hierbei größtenteils im akzeptablen und guten Bereich, wobei die Spezifität sowohl bei den bilingualen als auch bei den monolingualen Kindern nicht zufriedenstellend ist. Letzteres wird durch die große Altersspanne der Proband:innen erklärt. Da sich die Leistungen beim NWRT mit zunehmendem Alter verbessern (Buil-Legaz, Aguilar-Mediavilla & Adrover-Roig 2016, Boerma & Blom 2021), ist es nicht empfehlenswert, einen Cut-off-Wert für eine Stichprobe zu berechnen, die 10 Jahre umfasst. Dies wird durch eine nähere Analyse der Fehldiagnosen bestätigt: So ist die Hälfte der Kinder, die durch den LITMUS-CL-NWRT fälschlicherweise als sprachlich auffällig eingestuft wurden, obwohl sie der BiTD-Gruppe angehören, deutlich jünger als der Durchschnitt der Gesamtgruppe (Hinnerichs in Vorb.). Durch einen niedrigeren Cut-off-Wert für diese jüngeren Kinder hätte also die Spezifität deutlich verbessert werden können. Die Berechnung separater Cut-off-Werte für einzelne Altersstufen war im Rahmen der vorliegenden Studie nicht möglich, da die Proband:innenanzahl hierfür zu gering war. Die Tatsache, dass die Werte zur diagnostischen Validität trotz dieser ungünstigen Alterszusammensetzung größtenteils zufriedenstellend sind, lässt darauf schließen, dass der LITMUS-CL-NWRT ein geeignetes Testinstrument im Kontext von Bilingualismus und Sprachentwicklungsstörung sein könnte. Dies müsste jedoch mit weiteren Studien, in deren Rahmen idealerweise eine große Anzahl von gleichaltrigen Proband:innen getestet wird, überprüft werden.

Des Weiteren müssten auch die Erkenntnisse zum Vergleich des LITMUS-CL-NWRT mit dem HASE-NK überprüft werden. Im Einklang mit bisherigen Studien weisen die vorliegenden Ergebnisse ebenfalls darauf hin, dass sich der LITMUS-CL-NWRT besser zur Testung bilingualer Kinder eignet als der HASE-NK. Jedoch sind die Unterschiede bezüglich der diagnostischen Validität zwischen dem LITMUS-CL-NWRT und dem HASE-NK nicht so groß wie in der Studie aus den Niederlanden: Hier liegt die diagnostische Validität des LITMUS-CL-NWRT bei bilingualen Kindern im sehr guten Bereich, während die des LS-NWRTs lediglich akzeptabel ist (Boerma et al. 2015). Dies könnte unter anderem an der Itemkonstruktion liegen, da die Nichtwörter des LS-NWRTs bei Boerma et al. (2015) im Gegensatz zum HASE-NK komplexer sind. Die Items des HASE-NK sind in ihrer Konstruktion sehr nahe an der des LITMUS-CL-NWRT, da sie ebenfalls nur CV-Silben enthalten. Die Unterschiede zwischen den hier vorliegenden Daten und den Ergebnissen aus den Niederlanden könnten also dadurch zustande kommen, dass die bilingualen Kinder bei Boerma et al. (2015) durch den LS-NWRT noch stärker benachteiligt



werden, da dessen Items komplexer und dadurch eventuell auch einzelsprachspezifischer sind als die des HASE-NK.

Der Unterschied zwischen dem LITMUS-CL-NWRT und dem HASE-NK besteht in der Itemkonstruktion lediglich aus einer etwas größeren Lautauswahl beim HASE-NK und der Prosodie, die beim LITMUS-CL-NWRT neutral gehalten wird. Ob bilinguale Kinder bei NWRTs jedoch tatsächlich durch den Einsatz von Lauten, die nicht in ihrer Erstsprache vorkommen, benachteiligt werden, wurde bisher kaum untersucht. Erste Ergebnisse deuten darauf hin, dass sich für bestimmte bilinguale Gruppen ein Nachteil ergeben könnte (Hinnerichs in Vorb.).

Die schlechtere diagnostische Validität des HASE-NK könnte aber auch auf andere Faktoren der Testkonstruktion zurückzuführen sein. So beinhaltet der HASE-NK zwar Drei- und Viersilber, jedoch keine Fünfsilber. Da Letztere bezüglich der Unterscheidung der TD- und DLD-Gruppen die größte Effektstärke beim LITMUS-CL-NWRT aufweisen, könnte dies auch zu dessen vergleichsweise höheren diagnostischen Validität geführt haben. Ein weiterer Grund könnte die Itemanzahl sein. Der HASE-NK beinhaltet mit 9 Items fast nur halb so viele Nichtwörter wie der LITMUS-CL-NWRT (16 Items). In testtheoretischen Ansätzen wird davon ausgegangen, dass ein Merkmal besser erfasst werden kann, wenn die entsprechende Testung mehr Items enthält (Brandt & Moosbrugger 2020, S.50; Zaretsky, Lange & Hey 2022, S.11). Welche der genannten Faktoren letztendlich in welchem Ausmaß dazu geführt hat, dass der LITMUS-CL-NWRT eine bessere diagnostische Validität aufweist als der HASE-NK, kann mit den vorliegenden Daten nicht abschließend beantwortet werden. Zieht man die bisherigen Ergebnisse von NWRT-Studien mit bilingualen Kindern heran, so zeigte sich bereits mehrfach, dass der LITMUS-CL-NWRT besser zur Testung bilingualer Kinder geeignet ist als einzelsprachspezifische NWRTs (Boerma et al. 2015; Ortiz 2021; Zaretsky, Lange & Hey 2022). Die hier vorliegenden Ergebnisse legen nahe, dass der HASE-NK aufgrund seiner Itemkonstruktion bilinguale Kinder eventuell weniger benachteiligt als andere LS-NWRTs. Dies müsste mit weiteren Untersuchungen überprüft werden. Zukünftige Studien sollten also gezielt die Auswirkung der Test- und Itemkonstruktion auf die Nachsprehleistungen von monolingualen und bilingualen Kindern untersuchen. Auf der Basis dieser Erkenntnisse könnten dann NWRTs erstellt werden, die Kinder mit und ohne DLD möglichst gut voneinander unterscheiden, ohne dabei bilinguale Kinder zu benachteiligen.

Abgesehen davon, dass der LITMUS-CL-NWRT Kinder mit und ohne DLD ausreichend genau voneinander differenzieren kann, misst er ihre Fähigkeiten unabhängig von ihrem sozialen und sprachlichen Hintergrund.

Dies wird dadurch belegt, dass der sozio-ökonomische Status keinen bedeutsamen Einfluss auf die Nachsprehleistungen hat und auch die Sprachkontaktdauer und das AoO keinen nennenswerten Zusammenhang zu den Ergebnissen des LITMUS-CL-NWRT aufweisen. Diese Erkenntnisse stehen im Einklang mit bisherigen Studien (Antonijevic-Elliott et al. 2020; Chiat & Polišenská 2016; Verhagen & Andringa 2021).

Die große Altersspanne der Proband:innen und die gleichzeitig kleinen Stichproben der DLD-Gruppen schränken die Aussagekraft der vorliegenden Studie ein. Gleichzeitig ist das diagnostische Vorgehen zur Gruppenzuweisung eine Stärke dieser Untersuchung. Während einige Studien die Zuweisungen zur TD- bzw. DLD-Gruppe lediglich auf der Basis eines alleinigen Testverfahrens in der Zweitsprache vornehmen (Zaretsky, Lange & Hey 2022), wurden im Rahmen des BiliSAT-Projekts internationale Empfehlungen eingehalten: In beiden Sprachen wurden mit standardisierten und normierten Testungen mehrere sprachliche Ebenen überprüft, die Normabweichung zur Definition einer Störung wurde an die Sprachdominanz angepasst. Somit kann festgehalten werden, dass die Ergebnisse trotz der großen Altersspanne und der kleinen Stichprobe eine Aussagekraft haben, da die Kinder der DLD-Gruppen eine bestmöglich abgesicherte Diagnose erhalten haben.

## 6 Fazit und Ausblick

Das Nachsprechen von Nichtwörtern wird in der Diagnostik von Sprachentwicklungsstörungen häufig eingesetzt, sowohl im Rahmen von Screeningverfahren als auch in der sprachtherapeutischen Praxis.

Dabei gelten die eingesetzten NWRTs oft fälschlicherweise pauschal als einzelsprachunabhängig, wodurch bilinguale Kinder benachteiligt werden können.

Es ist daher essenziell, beim Einsatz von NWRs in der Praxis die Konstruktion der Items und die Spracherwerbssituation des jeweiligen Kindes zu berücksichtigen und die Ergebnisse mit Vorsicht zu interpretieren.

Der Einsatz von möglichst einzelsprachunabhängig konstruierten Testungen ist zu empfehlen, bisherige Forschungsergebnisse sind vielversprechend. Weitere Untersuchungen hierzu sind wünschenswert, ein konsistentes Auswertungsvorgehen ist dabei für eine Vergleichbarkeit der Ergebnisse unerlässlich.

Des Weiteren zeigen die Werte zur diagnostischen Validität der vorliegenden Arbeit sowie auch die früherer NWR-Studien, dass selten alle Kinder korrekt als sprachlich auffällig bzw. unauffällig erkannt werden. Deshalb wird immer wieder betont, dass der NWR niemals isoliert, sondern immer in Kombination mit weiteren Testungen zur Einschätzung sprachlicher Fähigkeiten herangezogen werden sollte (Ellis Weismer et al. 2000; Schwob et al. 2021). Dies gilt nicht nur für den sprachtherapeutischen Kontext, wo eine fundierte Therapieplanung ohnehin nur durch die ausführliche Überprüfung der einzelnen sprachlichen Ebenen stattfinden kann. Auch im Rahmen von Screeningverfahren zur Erkennung von Risikokindern sollte der NWR immer durch andere Testungen ergänzt werden (Ortiz 2021).

Die vorliegenden Daten bestärken die bisherigen internationalen Ergebnisse zum LITMUS-CL-NWR und legen nahe, dass dieser auch im deutschsprachigen Raum zu einer verbesserten Erkennung von Sprachentwicklungsstörungen bei bilingualen Kindern führen kann.

## Danksagung

Mein Dank geht an alle Familien und insbesondere an die Kinder, die durch ihre Teilnahme an der Datenerhebung diese Studie möglich gemacht haben. Ebenso danke ich Solveig Chilla und Cornelia Hamann und allen anderen Mitarbeiter:innen des BiliSAT-Projekts (DFG-CH-1112/4-1; Prof. Dr. Chilla, Universität Flensburg/Prof. Dr. Hamann, Universität Oldenburg), im Rahmen dessen die Daten erhoben und ausgewertet wurden. Großer Dank gebührt ebenso der Graduiertenakademie der Universität Heidelberg, durch deren finanzielle Unterstützung über verschiedene Stipendienprogramme die Dissertation realisiert werden konnte.

## Literatur

- Abed Ibrahim, L. & Fekete I. (2019) What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks. *Frontiers in Psychology* 9:2757. <https://doi.org/10.3389/fpsyg.2018.02757>
- Antonijevic-Elliott, S., Lyons, R., O'Malley, M. P., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M. & Fitzmaurice, Y. (2020). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics*, 34 (4), 293–311.
- Armon-Lotem, S. & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, 51 (6), 715–731.
- Baddeley, A. D. & Hitch, G. J. (1994). Developments in the Concept of Working Memory. *Neuropsychology*, 8 (4), 485–493.
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F. & Blom, E. (2015). A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research*, 58, 1747–1760.
- Boerma, T. & Blom, E. (2021). Crosslinguistic nonword repetition and narrative performance over time. In S. Armon-Lotem & K. K. Grohmann (Hrsg.), *Language Impairment in Multilingual Settings. LITMUS in action across Europe* (S.301-328). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Brandt, H. & Moosbrugger, H. (2020). Planungsaspekte und Konstruktionsphasen von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion*, 3. Auflage (S. 39–66). Berlin: Springer.
- Buil-Legaz, L., Aguilar-Mediavilla, E. & Adrover-Roig, D. (2016). Longitudinal trajectories of the representation and access to phonological information in bilingual children with specific language impairment. *International Journal of Speech-Language Pathology*, 18 (5), S. 473–482.
- Bulheller, S. & Häcker, H. O. (2002). *Raven's Coloured Progressive Matrices*. Frankfurt am Main: Pearson Assessment.
- Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, J. De Jong & N. Meir (Hrsg.), *Assessing Multilingual Children – Disentangling Bilingualism from Language Impairment* (S. 125–150). Bristol: Multilingual Matters.
- Chiat, S. & Polišenská, K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. *Journal of Speech, Language, and Hearing Research*, 59 (3), 1179–1189.
- Chiat, S., Polišenská, K., Szewczyk, J. & Kapalková, S. (2016). *Scoring Instructions – Crosslinguistic Nonword Repetition Tasks: British English version*. Unpublished material.
- Chilla, S. (2022). *Kindliche Mehrsprachigkeit*. München: Ernst Reinhardt Verlag.
- Dos Santos, C. & Ferré, S. (2016). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition*, 25, 1–14. <https://doi.org/10.1080/10489223.2016.1243692>
- Ennemoser, M., Pepouna, S. & Hartung, N. (2012). Kulturfaire und prognostisch valide Erfassung des phonologischen Arbeitsgedächtnisses bei Kindern mit Migrationshintergrund. *Unterrichtswissenschaft*, 40 (1), 26–46.

- Fox, A. (2009). *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner.
- Fox-Boyer, A. (2014). *Psycholinguistische Analyse kindlicher Aussprachestörungen (PLAKSS-II)*. Frankfurt am Main: Pearson.
- Fried, L., Briedigkeit, E., Isele, P. & Schunder, R. (2009). Delfin 4 – Sprachkompetenzmodell und Messgüte eines Instrumentariums zur Diagnose, Förderung und Elternarbeit in Bezug auf die Sprachkompetenz. *Zeitschrift für Grundschulforschung*, 2, 13-26.
- Gamper, H., Keller, U., Messerli, N., Moser, M. & Wüst, J. (2012). *Normen für den Mottier-Test bei 4- bis 12-jährigen Kindern. Praxisforschung der Erziehungsberatung des Kantons Bern, Band 12*. Zugriff am 07.09.2022 auf <https://docplayer.org/25529012-Normen-fuer-den-mottier-test.html>
- Gathercole, S. E. & Baddeley, A. D. (1990b). Phonological Memory Deficits in Language Disordered Children: Is There a Causal Connection? *Journal of Memory and Language*, 29, 336-360.
- Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6–10*. München: Elsevier.
- Grimm, A. & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, 7 (4), 821-841.
- Grimm, H. (2015). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5). Diagnose von Sprachverarbeitungs-fähigkeiten und auditiven Gedächtnisleistungen*. Göttingen: Hogrefe.
- Grimm, H., Aktas, M. & Kießig, U. (2003). *SSV. Sprachscreening für das Vorschulalter. Kurzform des SETK 3-5*. Göttingen: Hogrefe.
- Hamann, C. & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, 2 (16), 1-19.
- Hasselhorn, M. & Körner, K. (1997). Nachsprechen von Kunstwörtern: Zum Zusammenhang zwischen Arbeitsgedächtnis und syntaktischen Sprachleistungen bei Sechs- und Achtjährigen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 29, 212-224.
- Heeg, K., Bloder, T., Eikerling, M., Frey, N., Starke, A. & Lüke, C. (2022). *Der Mottier-Test in der Anwendung – Überprüfung von Parametern der Durchführung*. Vortrag am XXXIV. Workshop Klinische Linguistik.
- Hinnerichs, J. (2016). *Spezifische Sprachentwicklungsstörungen bei Kindern mit Deutsch als Zweitsprache – Zur Eignung des Nachsprechens von quasi-universalen Nichtwörtern als Differentialdiagnostik*. Unveröffentlichte Masterarbeit, Universität Heidelberg.
- Hinnerichs, J. (in Vorb.). *Ein Testverfahren zur Differenzierung von bilingualen Kindern mit und ohne Sprachentwicklungsstörungen – Untersuchungen zum Nachsprechen einzelsprachunabhängiger Nichtwörter mit dem LITMUS-CL-NWRT bei Kindern mit Deutsch als Erst- und Zweitsprache*. Dissertation: Universität Heidelberg.
- Hinnerichs, J., Polišká, K. & Chiat, S. (2016). *The Crosslinguistic Nonword Repetition Test: German version*.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55-88.
- Holler-Zittlau, I., Euler, H. A., & Neumann, K. (2011). Kindersprachscreening (KiSS) – das hessische Verfahren zur Sprachstandserfassung. *Spracheheilerarbeit*, 5-6, 263-268.
- Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Recht-schreibschwierigkeiten*. Göttingen: Hogrefe.
- Kannengieser, S. (2019). *Sprachentwicklungsstörungen – Grundlagen, Diagnostik und Therapie*. München: Elsevier.
- Kauschke, C. & Vogt, S. (2019). Positionspapier zur Terminologie und Definition von Sprachentwicklungsstörungen. *Logos*, 27 (3), 174-181.
- Marxen, R., Lüke, C. & Cholin, J. (2022). Kompetenzzempfinden bei der Sprachentwicklungsdiagnostik mehrsprachiger Kinder. *Logos*, 30 (2), 97-106.
- Meir, N. & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic Status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, 8:1442.
- Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache (Hrsg.) (2013): *Qualitätsmerkmale für Sprachstandsverfahren im Elementarbereich. Ein Bewertungsrahmen für fundierte Sprachdiagnostik in der Kita*. Köln.
- Messer, M.H., Leseman, P.P.M., Boom, J. & Mayo, A.Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual pre-schoolers. *Journal of Experimental Child Psychology*, 105, 306-323.
- Parra, M, Hoff, E. & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, 108, 113-125.
- Pierce, L. J., Genesee, J., Delcenserie, A. & Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics*, 38, 1265-1300.
- Plante, E. & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Language, Speech and Hearing Services in Schools*, 25, 15-24.
- Polišká, K. & Kapalková, S. (2014). Improving Child Compliance on a Computer-Administered Nonword Repetition Task. *Journal of Speech, Language, and Hearing Research*, 57, 1060-1068.
- Polišká, K., Chiat, S., Szewczyk, J., Antonijevic-Elliott, S., Blom, E., Boerma, T., Bohnacker, U., Nutcher, M., Chan, A., Chilla, S., Chondrogianni, V., Fitzmaurice, Y., Fu, N. C., Gatt, D., Grech, H., Haddad, R., Hamann, C., Hinnerichs, J., Holzinger, D., Jezek, M., Judge-Clayden, F., Kapalková, S., Kunnari, S., Mayer-Crittenden, C., O'Malley, M. P., Öberg, L., Oudgenoeg-Paz, O., Polatidou, A., Schwob, S., Shiyun, M., Skoruppa, K., Tabone, N., van den Berghe, R., Verhagen, J. & White, M. (2022). *Is the crosslinguistic nonword repetition test valid crosslinguistically? Evidence from performance in different language groups and countries*. Vortrag auf der vierten Bi-SLI Konferenz 2022 in Berlin/ZAS.
- Rescorla, L. (2014). Entwicklungswege von Late-Talkern. In S. Ringmann & J. Siegmüller (Hrsg.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen – Jugend- und Erwachsenenalter* (S. 57-77). München: Elsevier.
- Rothweiler, M. (2007). „Mistaken identity“ – Zum Problem der Unterscheidung typischer grammatischer Strukturen bei SSES und bei Mehrsprachigkeit. In U. de Langen-Müller & V. Maihack (Hrsg.). *Früh genug – aber wie? Sprachförderung per Erlass oder Sprachtherapie auf Rezept? Tagungsbericht vom 8. wissenschaftlichen Symposium des Deutschen Bundesverbandes der akademischen Sprachtherapeuten* (S.110-128). Köln: ProLog.
- Scherger, A.-L. (2022). The role of age and timing in bilingual assessment: non-word repetition, subject-verb agreement and case marking in L1 and eL2 children with and without SLI. *Clinical Linguistics & Phonetics*, 36 (1), 54-74, <https://doi.org/10.1080/02699206.2021.1885497>
- Schöler, H. & Brunner, M. (2008). *HASE – Heidelberger Auditives Screening in der Einschulungsuntersuchung*. Wertingen: Westra.
- Schulz, P. & Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe.

- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R. & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, 64, 3578-3593.
- Sorenson Duncan, T. & Paradis, J. (2016). English Language Learners' Nonword Repetition Performance: The Influence of Age, L2 Vocabulary Size, Length of L2 Exposure, and L1 Phonology. *Journal of Speech, Language, and Hearing Research*, 59, 39-48.
- Stadie, N. & Schöppe, D. (2013). *PhoMo-Kids. Phonologie Modellorientiert für Kinder vom Vorschulalter bis zum dritten Schuljahr*. Köln: Prolog.
- Szewczyk, J. M., Marecka, M., Chiat, S. & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23-26.
- Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Hrsg.), *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment* (S. 331-358). Bristol: Multilingual Matters.
- Tomblin, J. B., Records, N. L. & Zhang, X. (1996). A System for the Diagnosis of Specific Language Impairment in Kindergarten Children. *Journal of Speech and Hearing Research*, 39, 1284-1294.
- Topbaş, S. & Güven, S. (2011). *Türkçe Erken Dil Gelişim Testi – TEDİL-3:T*. Ankara: Detay.
- Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Hrsg.), *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment* (S.229-328). Bristol: Multilingual Matters.
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morrin, E., Prevost, P., Dos Santos, C., Abed Ibrahim, L. & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, 53, 4, S.888-904.
- Verhagen, J. & Andringa, S. J. (2021). *The role of existing language knowledge in bilingual and multilingual toddlers' repetition of cross-linguistic and language-specific nonwords*. *Linguistic Approaches to Bilingualism*. <https://doi.org/10.1075/lab.20050.ver>
- Vuolo, J. & Goffmann, L. (2020). Vowel Accuracy and Segmental Variability Differentiate Children With Developmental Language Disorder in Nonword Repetition. *Journal of Speech, Language, and Hearing Research*, 63, 3945-3960. [https://doi.org/10.1044/2020\\_JSLHR-20-00166](https://doi.org/10.1044/2020_JSLHR-20-00166)
- Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. New York: Oxford University Press.
- Weinrich, M & Zehner, H. (2017). *Phonetische und phonologische Störungen bei Kindern*. Springer: Berlin.
- Zaretsky, E., Lange, B. P. & Hey, C. (2022): Phonological Short-Term Memory: When Bilingualism Matters. *Language Learning and Development*, <https://doi.org/10.1080/15475441.2021.2014846>
- Zebib, R., Henri, G., Khoms, A., Messara, C. & Hreich, E. (2017). *Batterie d'Evaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Kerserwan: LTE.

## Zur Autorin

Juliane Hinnerichs ist akademische Sprachtherapeutin (B. Sc. Patholinguistik) und absolvierte den Master „Deutsch als Zweitsprache“ an der Universität Heidelberg. Dort forscht und lehrt sie zu Bilingualismus, Spracherwerb und Sprachstandsfeststellungen. In ihrer Dissertation beschäftigte sie sich mit Sprachentwicklungsstörungen bei bilingual aufwachsenden Kindern.

## Korrespondenzadresse

Juliane Hinnerichs, M.A.  
 Universität Heidelberg  
 Neuphilologische Fakultät  
 Institut für Deutsch als Fremdsprachenphilologie  
 Plöck 55  
 69117 Heidelberg  
[juliane.hinnerichs@posteo.de](mailto:juliane.hinnerichs@posteo.de)



## Die Unterstützung von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten im Unterricht durch multiprofessionelle Teams. Wer hilft mit und wenn ja, wie viel?\*

The support of children with language, speech, and communication difficulties in regular classrooms through multiprofessional teams. Who helps and if so, to what extent?

Christoph Till, Jasmin Kolb

### Zusammenfassung:

**Hintergrund:** Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten sind vermehrt in den Klassenzimmern der Regelschule zu finden. Mit Sprachbildungs-, Sprachförder- und Sprachtherapiemaßnahmen können diese Kinder auch inklusiv unterstützt werden. In der Schweiz sind Regellehrpersonen sowie Fachkräfte für schulische Heilpädagogik und Logopädie für diese sprachunterstützenden Maßnahmen verantwortlich.

**Fragestellung:** Es stellt sich die Frage, wie häufig die verschiedenen Berufsgruppen diese Maßnahmen im Klassenzimmer umsetzen, von welchen Bedingungen die Durchführungshäufigkeit abhängt und ob sich diese Bedingungen in den Berufsgruppen unterschiedlich auswirken.

**Methodik:** Mithilfe eines Online-Fragebogens wurden 142 Regellehrpersonen, 89 schulische Heilpädagog:innen und 86 Logopäd:innen zur Häufigkeit der unterrichtsintegrierten sprachunterstützenden Maßnahmen befragt. Zusätzlich wurden Daten zu Rahmenbedingungen sowie zur Häufigkeit der multiprofessionellen Kooperation und entsprechenden Einstellungen erhoben. In ordinalen Regressionsmodellen wurden ausgewählte Zusammenhänge überprüft.

**Ergebnisse:** Schulische Heilpädagog:innen und vor allem Logopäd:innen setzen die Maßnahmen deutlich seltener unterrichtsintegriert um. Die Häufigkeit des praktizierten Team-Teachings steigert die Durchführungswahrscheinlichkeit dieser Maßnahmen deutlich. Am stärksten zeigt sich dieser Effekt bei den Logopäd:innen. Aber auch die Anzahl Schulhäuser, für die eine Person verantwortlich ist, sowie die Intention, vermehrt unterrichtsintegriert zu arbeiten, können einen positiven Einfluss haben.

**Schlussfolgerungen:** Je häufiger Fachpersonen in Team-Teachings eingebunden sind, desto größer ist die Wahrscheinlichkeit, dass die sprachunterstützenden Maßnahmen vermehrt unterrichtsintegriert umgesetzt werden. Folglich müssen entsprechende Zusammenarbeitsformen aufgebaut werden.

### Schlüsselwörter

Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten, Inklusion, sprachunterstützende Maßnahmen, multiprofessionelle Kooperation, Fragebogen, ordinale Regressionen

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

## Abstract

**Background:** The number of children with language difficulties in regular classrooms is increasing. Measures like language instruction, language facilitation and speech-language therapy can be applied in-class. In Switzerland, classroom teachers as well as special education teachers and speech-language therapists are responsible for these measures of language support.

**Aims:** The question arises as to how often these measures are implemented in-class by the different occupational groups, what further conditions influence the frequency, and whether these conditions have different effects in the occupational groups.

**Methods:** 142 classroom teachers, 89 special education teachers, and 86 speech-language therapists were asked about the frequency of in-class measures of language support, about the determining factors, the frequency of multiprofessional cooperation and corresponding attitudes. Selected correlations were tested in ordinal regression models.

**Results:** Special education teachers and especially speech-language therapists are less likely to implement measures of language support in the classroom. The frequency of team teaching significantly increases the likelihood of these measures being implemented in-class. This effect is strongest among speech-language therapists. The number of school buildings as well as high intentions to support children in-class more often, can also have a positive influence.

**Conclusions:** Consequently, to increase the frequency of in-class measures of language support, appropriate forms of collaboration should be established.

## Keywords

children with language difficulties, inclusion, measures of language support, multiprofessional cooperation, questionnaire, ordinal regressions

# 1 Inklusion von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten durch multiprofessionelle Angebote

## 1.1 Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten im Unterricht

Unterrichtliche Aktivitäten involvieren häufig komplexe mündliche und schriftliche Sprache. Sprache – gesprochen wie geschrieben – ist Gegenstand, Medium und Steuerungsmechanismus in (fast) allen Unterrichtsfächern (Mahlau, 2016), weswegen die Sprachkompetenzen auch gemäß dem Schweizer Lehrplan 21 als „entscheidender Schlüssel zum Schulerfolg“ (Deutschschweizer Erziehungsdirektoren-Konferenz, 2016, S. 10) beschrieben werden.

Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten (Sammelbegriff für unterschiedliche Ausprägungsformen von Auffälligkeiten in den genannten Bereichen gemäß Kauschke et al., 2023) bringen oftmals nicht die notwendigen sprachlichen Kompetenzen mit, um die curricularen Anforderungen bewältigen zu können (Archibald, 2017). Sie sind daher besonders gefährdet, zusätzlich eine Lernbehinderung attestiert zu bekommen (Reber & Schönauer-Schneider, 2011). Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten werden außerdem als sozial unreif wahrgenommen und können soziale Situationen schlechter erfassen. Die Folgen und Begleiterscheinungen lassen sich bis ins Jugend- und Erwachsenenalter nachweisen (Mahlau, 2018).

Die Schwierigkeiten dieser Kinder sollten so gut wie möglich überwunden und/oder kompensiert werden. Glück, Reber, Spreer und Theisel (2013) definieren unter anderem den Unterricht, die Förderung und die Therapie als zentrale Unterstützungsangebote. Der Unterricht als „sprachsensibler Unterricht steht für unterschiedliche Unterrichtskonzepte, die Sprache bewusst als Mittel des Denkens und Kommunizierens einsetzen, um fachliches und sprachliches Lernen zu verknüpfen. Ein wesentliches Element ist die in den Unterricht integrierte, gezielte sprachliche Unterstützung“, die sich an alle Schülerinnen und Schüler richtet (Woerfel & Giesau, 2018, S. 1). Sprachförderung hingegen umfasst Massnahmen für Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten, welche mit einfachen diagnostischen Mitteln erfasst werden müssen (Mahlau & Herse, 2017). Sprachförderung kann zusätzlich oder parallel zum Unterricht oder als integrierte Form im Einzel-, Gruppen- oder Klassensetting umgesetzt werden, mit dem Ziel, sprachliche Kompetenzen in der Umgangs- und Bildungssprache aufzubauen und somit einen schulischen Lernerfolg ermöglichen (Reber & Schönauer-Schneider, 2017). Sprachtherapie meint gezielte und intensive Maßnahmen, die auf Grund einer differenzierten Diagnostik auf die individuellen Bedürfnisse eines Kindes angepasst sind, lernerschwerende Bedingungen reduzieren sowie Sprachkompetenzen aufbauen und erweitern und eingeschränkte Kompetenzen

kompensieren sollen, u. a. um die Teilhabe am Alltag und am Unterricht zu ermöglichen (Hartmann, Till & Winkes 2019). Die individuellen Bedürfnisse stehen im Vordergrund, wobei curriculare Anforderungen nach Möglichkeit berücksichtigt werden. Dementsprechend kann die Therapie auch ins Unterrichtsgeschehen verlagert werden, wenn die therapeutischen Ziele und die Arbeitsbedingungen dies rechtfertigen (Glück et al., 2013). Im Leitfaden „Integration und besondere Massnahmen in der Volksschule“ des Kantons Bern heißt es u. a., dass die sprachunterstützenden Maßnahmen „in der Regel unterrichtsintegriert“ durchgeführt werden sollen (Erziehungsdirektion des Kantons Bern, 2019, S. 27).

Die Wirksamkeit unterrichtsintegrierter Sprachförderung und -therapie wurde vor allem in den 1990er Jahren und Anfang des neuen Jahrtausends erforscht und u. a. im Review von Cirrin et al. (2010) zusammenfassend analysiert. Zusammenfassend lauten die Ergebnisse dieses Reviews zu unterrichtsintegrierter Sprachtherapie, dass diese besonders dann sinnvoll ist, wenn die Kinder schwere Sprachentwicklungsstörungen haben und wenn die Sprachtherapie im Team-Teaching mit der Regellehrperson angeboten wird. In den anderen untersuchten Fällen finden sich keine signifikanten Unterschiede in den Effekten, sodass unterrichtsintegrierte Sprachtherapie keine bedeutenden Vor- oder Nachteile zur separativen Therapie aufweist. Im Review findet sich vor allem eine Studie (Throneburg, Calvert, Sturm, Paramboukas & Paul, 2000), die einen deutlichen Nachweis liefert, dass unterrichtsintegrierte Sprachtherapie (genauer gesagt: Wortschatztherapie) wirksamer sein kann als die Therapie in separativen Settings. Die Wirksamkeit wird bedingt durch das Team-Teaching der *speech-language pathologists* mit den *classroom teachers* ( $d = 1.65$ ); arbeitet die logopädische Fachkraft unterrichtsintegriert, aber allein, findet sich nach wie vor ein signifikanter Effekt ( $d = .30$ ), dieser ist dem Effekt der separativen Wortschatztherapie aber unterlegen ( $d = .76$ ).

Gemäß Reber (2012) ist die unterrichtsintegrierte Sprachtherapie auch nicht in jedem Kontext sinnvoll: So sei der Einbezug logopädischer Methoden in den Unterricht bei Wortschatz-, Grammatik- oder auch Sprachverständnisstörungen sehr wirksam, bei Aussprachestörungen hingegen jedoch nicht. In Bezug auf die verschiedenen Fachbereiche, die unterrichtet werden, liegt nahe, Sprachtherapie mit dem Deutschunterricht zu verknüpfen, da der Erwerb der Schriftsprache sinnvoll unterstützt werden kann. Die Verknüpfung von eher sprachfernen Fachbereichen mit Sprachtherapie liegt jedoch weniger auf der Hand, ist aber ebenso sinnvoll und möglich. Auch sind nicht alle Sozialformen (Einzelarbeit, Partnerarbeit, Gruppenarbeit, Frontalunterricht, Unterrichtsgespräche; Pukas, 2008) gleichermaßen geeignet für den Einbezug sprachtherapeutischer Elemente. Und letzten Endes ist es nicht sinnvoll, alle Phasen des sprachtherapeutischen Prozesses unterrichtsintegriert durchzuführen. So kann die Anbahnung und die Einübung von Fähigkeiten weiterhin gut im separativen Setting geschehen. Vor allem, wenn es um den Alltagstransfer sprachlicher Fähigkeiten geht, sind klassenzimmerintegrierte Vorgehensweisen jedoch vorzuziehen (Archibald, 2017; Cirrin et al., 2010).

Lehrpersonen benötigen diverse Fähigkeiten, um die Herausforderungen der inklusiven<sup>1</sup> Beschulung von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten gerecht zu werden. Bauer, Glück, Jonas, Mayer, Sallat und Stitzinger (2022, S. 230) nennen u. a. die Fähigkeit zur Identifikation und Reduktion von sprachlich-kommunikativen Barrieren, die Schaffung von sprachlich-kommunikativen Voraussetzungen für erfolgreiches schulisches Lernen, die Optimierung der Vermittlung von Fach- und Bildungssprache, die Verknüpfung von sprachheilpädagogischen Unterrichtsprinzipien und fachdidaktischen Methoden u. v. m.

Es ist jedoch unangemessen, die Verantwortung für die sprachauffälligen Kinder allein den Regellehrpersonen zu übertragen. Die Fachwelt ist sich dahingehend einig, dass die erfolgreiche Inklusion sprachauffälliger Kinder von der Kooperation verschiedener Fachkräfte abhängt, da eine Fachperson allein nicht alle notwendigen Kompetenzen für diesen Auftrag mitbringen kann (Archibald, 2017; Glück et al., 2013; Sallat & Siegmüller, 2016). Der Unterricht mit sprachauffälligen Kindern muss umfassende und spezifische Kenntnisse aus der Spracherwerbsforschung, der Linguistik, der allgemeinen Pädagogik und Didaktik, sowie Sprachpathologie und Medizin berücksichtigen (Mahlau, 2016, S. 117f.) und erfordert daher „die Bereitstellung multi-professioneller Angebote [...] Dabei ist eine gute Zusammenarbeit notwendig, damit eine Stärkung in der Sprachunterstützung der betroffenen Kinder erfolgt, und keine Kompetenzdelegierung mit Verwässerung der notwendigen sprachheilpädagogischen Förderung“ (Mahlau, 2018, S. 49). Es gilt also, die Zuständigkeiten und Kompetenzen der beteiligten Fachpersonen zu klären, um einer

1 Gemäß Hinz (2004) wird der Begriff „Inklusion“ immer dann verwendet, wenn von der Vision einer inklusiven Schule gesprochen wird, der Begriff „Integration“ wird verwendet, wenn die aktuelle Praxis thematisiert wird.

wenig effizienten Umsetzung von sprachunterstützenden Maßnahmen gezielt entgegenwirken zu können. Die schulischen und sozial-emotionalen Schwierigkeiten der betroffenen Kinder – so bedeutsam diese auch sind – werden im Folgenden nicht mehr berücksichtigt, da diese die Zusammenarbeit mit weiteren Fachpersonen und eine deutliche Erweiterung des Forschungsfelds bedeuten würde. Hier soll der Fokus auf die Sprachunterstützung gelegt werden.

## 1.2 Sprachunterstützung durch Regellehrpersonen, schulische Heilpädagog:innen und Logopäd:innen

Ehren (2000) verweist auf die Schwierigkeit der Unterteilung der Zuständigkeiten in der sprachlichen Unterstützung von Kindern, da sich alle beteiligten Fachkräfte mit dem gleichen Gegenstand beschäftigen. Es ist nicht sinnvoll, eine Gruppe für die Wortschatzentwicklung und eine andere Gruppe für die Grammatikentwicklung der Kinder verantwortlich zu machen. Die Sprachentwicklung stellt ein Kontinuum dar, das sich durch alle Situationen und Kontexte zieht. Insofern müssen auch alle Fachgruppen jederzeit involviert sein.

Stattdessen teilt Ehren (2000) die genannten Maßnahmen drei verschiedenen Berufsgruppen zu: So wird der (sprachensible) Unterricht durch *classroom teachers* verantwortet, die Sprachförderung durch *special education teachers* und die Sprachtherapie durch *speech-language pathologists*. Auch Lüdtkke (2017) orientiert sich mit dem „Komplementären Unterstützungsprofil Sprache und Kommunikation“ an dieser Dreiteilung. Das Profil beschreibt, wie die beteiligten Fachpersonen ihre Kernkompetenzen und damit ihre fachliche Autonomie wahren und gleichzeitig in einen Kompetenztransfer mit den anderen Fachpersonen treten können. Lüdtkke ordnet im Profil den Grundschullehrer:innen die Kernkompetenz „allgemeine Pädagogik“, den Sonderpädagog:innen mit Förderschwerpunkt Sprache die Kompetenzen „Sprachpädagogik“ sowie „Sprachdidaktik“ und den Sprachtherapeut:innen/Logopäd:innen die Kernkompetenz „Sprachtherapie“ zu. Im Rahmen des Kompetenztransfers können bspw. Logopäd:innen den Grundschullehrer:innen Grundlagen zu Spracherwerbsprozessen und -störungen vermitteln, während sie selbst schul- und unterrichtsbezogene pädagogisch-didaktische Kompetenzen von den Grundschullehrkräften vermittelt bekommen. Voraussetzungen sind die Offenheit gegenüber den Perspektiven der anderen Fachpersonen sowie eine gewisse Experimentierfreudigkeit, sich auf neue Arbeitsformen einzulassen.

In den Schulsystemen der Deutschschweizer Kantone finden sich mit den Regellehrpersonen (RLP), den Fachkräften für Schulische Heilpädagogik (SHP) und den Fachkräften für Logopädie (LOG) die entsprechenden Pendanten. Es ist jedoch anzumerken, dass die Ausbildung in Schulischer Heilpädagogik zwar sprachheilpädagogische Inhalte umfasst, diese aber nicht mit einer Spezialisierung im Förderschwerpunkt Sprache zu vergleichen ist, wie sie von Bauer et al. (2022) beschrieben und von Theisel (2015) untersucht wurde. Dafür sind in der Schweiz als Spezialist:innen für Sprachentwicklung und Sprachstörungen die pädagogisch-therapeutisch (im Gegensatz zu den medizinisch-therapeutisch) tätigen LOG in der Regel in den Schulen integriert oder in unmittelbarer Nähe zur Schule situiert. Betroffene Kinder können zumeist weiterhin die Regelschule besuchen und erhalten die therapeutischen Lektionen während der Unterrichtszeit (Kempfe Preti, 2010). Trotz dieser günstigen Ausgangslage findet Sprachtherapie gemäß den Studienergebnissen von Blechschmidt, Besset, Läber, Reber und Allemann (2013) nach wie vor bevorzugt im separativen Einzelsetting und selten bis nie im Klassenzimmer statt. Selbst wenn es um den „Transfer sprachlicher Fähigkeiten und Fertigkeiten“ in den (Schul-)Alltag geht, wird die Sprachtherapie nur in 11 % der Fälle innerhalb des Klassenzimmers umgesetzt. Eine grundsätzliche Verknüpfung der Sprachtherapie mit Unterrichtsinhalten wird in 26 % der Fälle berichtet, vor allem bezüglich Schriftspracherwerbsstörungen (87 %), semantisch-lexikalischen (61 %) und kommunikativ-pragmatischen Auffälligkeiten (54 %).

Es genügt nicht, die für das Einzelsetting konzipierte Therapie 1:1 im Klassenzimmer durchzuführen. Die Ziele und die für deren Verfolgung notwendigen Maßnahmen müssen für die Klasse und den Unterricht neu geplant werden (Kempfe Preti, 2010, S. 10). Die Inklusion betroffener Kinder führt zu einer Neuverteilung bzw. einer Verschiebung der Aufgaben der verschiedenen Fachpersonen im multiprofessionellen Team (Glück et al., 2013). Wie bei Blechschmidt et al. (2013) zu sehen war, wurden inklusive bzw. unterrichtsintegrierte Maßnahmen kaum umgesetzt. Ob dies heutzutage anders aussieht und ob damit einhergehend auch ein neues Aufgabenprofil bzw. Rollenverständnis entstanden ist, ist einer der Fragen dieser Untersuchung, ob die genannten Berufsgruppen diese Aufgaben als multiprofessionelles Team lösen, eine weitere.



### 1.3 Unterrichtsintegrierte sprachunterstützende Maßnahmen durch multiprofessionelle Teams

Gräsel, Fußangel und Pröbstel (2006) unterscheiden drei Formen der Kooperation: den Austausch (von Informationen und Materialien), die Arbeitsteilung bzw. Synchronisation (gemeinsam abgesprochene Ziele, die individuell erarbeitet werden) und die Kokonstruktion (Zielerreichung nur durch gemeinsame Aufgaben- und Problemlösung). Team-Teaching wird der Kokonstruktion zugeordnet und gilt als eindeutigste bzw. höchste Form der Kooperation, da es hohe Ansprüche an die Zusammenarbeit stellt und unmittelbar im Unterricht stattfindet (Gräsel et al., 2006; Kolbe & Reh, 2008; Widmer-Wolf, 2018). In verschiedenen Studien zur Kooperation von RLP und SHP (z. B. Arndt & Werning 2013; Baumann, Henrich & Studer 2012; Gröhlich, Drossel & Winkelsetz, 2015; Hernandez, 2013; Luder, 2021) sowie von RLP und LOG (vgl. Archibald, 2017; Baxter, Brookes, Bianchi, Rashid & Hay, 2009; Beck & Dennis, 1997; Blechschmidt et al., 2013; Brimo & Huffman, 2023; Elksnin & Capilouto 1994; Green, Chance & Stockholm, 2019; Hartas, 2004; Pershey & Rapping, 2003; Pfeiffer, Pavelko, Hahs-Vaughn & Dudding, 2019) zeigt sich, dass die Kooperationsform Austausch am häufigsten und die Kokonstruktion (inklusive Team-Teaching) am seltensten durchgeführt wird (Baumann et al., 2012; Luder, 2021; Blechschmidt et al., 2013; Brimo & Huffman, 2023; Pershey & Rapping, 2003; Pfeiffer et al., 2019). Der Austausch ist die einzige Kooperationsform, die von allen Berufsgruppen gleichermaßen oft genutzt wird (Luder, 2021). Team-Teaching insgesamt wird zwar äußerst selten, grundsätzlich aber von allen Berufsgruppen genutzt (Arndt & Werning, 2013; Baumann et al., 2012; Luder, 2021; Beck & Dennis, 1997; Blechschmidt et al., 2013; Brimo & Huffman, 2023; Green et al., 2019). Die Team-Teaching-Formen, bei denen eine Person (i.d.R. die RLP) den Unterricht verantwortet, und die andere Person (SHP oder LOG) einzelne Kinder bei Bedarf unterstützt, werden von allen Berufsgruppen bevorzugt durchgeführt (Arndt & Werning, 2013; Beck & Dennis, 1997; Blechschmidt et al., 2013; Green et al., 2019; Luder, 2021). Dies steht zwar im Widerspruch zur Einschätzung von RLP und LOG, dass interaktives Unterrichten die pädagogisch sinnvollste Art des Team-Teachings ist (Beck & Dennis, 1997), passt aber zu den Ergebnissen von Throneburg et al. (2000), dass unterrichtsintegrierte Sprachtherapie nur dann wirklich wirksam ist, wenn sie im Team-Teaching durchgeführt wird.

Gemäß Huber und Ahlgrimm (2012) wird die erfolgreiche Kooperation in der Schule auf mehreren Ebenen bedingt: der personellen, der institutionellen sowie der organisationskulturellen Ebene. Drossel und Willems (2014) ergänzen noch die Ebene des Schulleitungshandelns, welche u. a. die Bereitstellung kooperationsförderlicher Rahmenbedingungen umfasst.

Auf der *personellen Ebene* sind u. a. das Wissen, die Fähigkeiten und Fertigkeiten einer Person vor dem Hintergrund der eigenen Berufsbiografie, der Reflexion der eigenen Rolle und Ähnliches zentral. So liegt auf der Hand, dass auf dieser Ebene vor allem die *unterschiedlichen Rollen* bzw. die Aufgaben, die von RLP, SHP und/oder LOG entsprechend ihrer Ausbildung eingenommen und übernommen werden, von entscheidender Bedeutung sind (Widmer-Wolf, 2018). Aus den Studien zur Kooperation von RLP, SHP und LOG geht hervor, dass eher eine Rollenkonfusion vorherrscht, da nicht abschließend geklärt ist, wer welche Verantwortung und Aufgaben hat. RLP haben demnach keine klare Vorstellung davon, über welche Kompetenzen LOG genau verfügen und wie diese im Unterricht von Nutzen sein könnten (Baxter et al., 2009; Brimo & Huffman, 2023; Green et al., 2019; Pershey & Rapping, 2003). Deswegen scheint das Vertrauen der RLP in logopädische Fachkräfte in inklusiven Unterrichtssettings nicht besonders groß zu sein und letztere werden höchstens als „Klassenassistenten“ im Klassenzimmer geduldet (Green et al., 2019). Dies trifft auch auf SHP zu – so äußerten sich RLP dahingehend, dass sie die Rolle der SHP vor allem darin sehen würden, die RLP zu entlasten, wohingegen SHP sich äußerten, gerne mehr Verantwortung im gemeinsamen Unterricht übernehmen zu wollen (Luder, 2021).

Die *institutionelle Ebene* steht u. a. für das Vorhandensein von Teamstrukturen, die Bereitstellung von zeitlichen Ressourcen und Ähnlichem, um kooperative Arbeitsformen zu begünstigen. *Team-Teaching* als höchste und intensivste Form der Kooperation eignet sich daher besonders als Indikator für die umgesetzte Zusammenarbeit im Rahmen integrativer Settings (Luder, 2021). Wie Throneburg et al. (2000) außerdem zeigen konnten, scheint Team-Teaching zudem ein guter Indikator für besonders wirksame unterrichtsintegrierte sprachunterstützende Maßnahmen zu sein. Reber weist bereits 2012 auf die gute Ausgangslage in Schweizer Schulen und die bereits gemachten Erfahrungen mit Team-Teaching zwischen LOG und RLP hin und sieht hier einen Vorsprung vor Deutschland bei der Entwicklung von inklusiven Unterstützungsformen für sprachauffällige Kinder in der Schule.

Auf der *organisationskulturellen Ebene* sind u. a. gemeinsam getragene Wert- und Zielvorstellungen, eine konstruktive Kultur der Kollegialität, ein Klima der Offenheit für Innovationen und der Veränderungsbereitschaft und viele weitere ähnliche Bedingungen verortet. Grundsätzlich wirken sich positive Einstellungen zur Kooperation auf die Qualität ebendieser aus (Vogt, Kunz-Heim, Baez, Safi & Zumwald, 2022). Gemäß der Theory of Planned Behavior (Ajzen, 2019) genügt es jedoch nicht, die Einstellungen zu verändern. Erst die Entwicklung von bewussten *Intentionen*, sich künftig anders (hier: stärker unterrichtsintegriert) verhalten zu wollen, führen zu einer tatsächlichen Verhaltensänderung.

Auf der *Ebene des Schulleitungshandelns* geht es um die Bereitstellung kooperationsförderlicher Rahmenbedingungen. Dazu gehören z. B. strukturell etablierte Zeitgefäße, die für die Entwicklung von Kooperationsformen genutzt werden können. Von verschiedenen Autor:innen wird auch immer wieder die Anzahl der festen Kooperationspartner:innen genannt, die nicht zu groß werden sollte (Widmer-Wolf, 2018). Je größer die Anzahl an Fachpersonen, mit denen zusammengearbeitet werden soll, desto geringer fällt die für jede einzelne Person zur Verfügung stehende Zeit aus und desto eher wird die Kooperation als zusätzliche Belastung empfunden (Arndt & Werning, 2013). In Bezug auf die logopädische Berufsgruppe ziehen Pershey und Rapping (2003) wie auch Baxter et al. (2009) eine weitere Größe heran, nämlich die *Anzahl Schulhäuser*, für die eine Person verantwortlich ist. In ihrer Befragung berichteten die meisten LOG, dass sie zwischen mindestens zwei, häufig sogar zwischen vier Schulhäusern wechseln müssten. Dies beansprucht nicht nur Reise- und damit Arbeitszeit, sondern erfordert auch die Kooperation mit wechselnden Partner:innen.

Es ist deutlich geworden, dass die Zusammenarbeit von RLP, SHP und LOG im Kontext der unterrichtsintegrierten sprachunterstützenden Maßnahmen erwünscht und sinnvoll ist, aber durch die Ausprägung verschiedener Bedingungen, die sich positiv wie negativ auswirken können, deutlich beeinflusst werden kann. Einen besonderen Einfluss nehmen dabei die verschiedenen Rollen ein, die RLP, SHP und LOG im Rahmen der Unterstützung von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten haben. Haben sich die Rollen, Aufgaben und Verantwortlichkeiten der Berufsgruppen seit den letzten Untersuchungen geändert? Zeigt sich dies an einer Steigerung der unterrichtsintegrierten Maßnahmen? Die unterrichtsintegrierte Unterstützung der betroffenen Kinder setzt wie gesagt die multiprofessionelle Zusammenarbeit voraus. Diese wird wiederum von der Häufigkeit des ausgeübten Team-Teachings, den Intentionen, vermehrt unterrichtsintegriert zu arbeiten, wie auch der Anzahl Schulhäuser, an der eine Person arbeitet, beeinflusst, weswegen diese Bedingungen ebenfalls von Interesse sind.

## 2 Fragestellung und Zielsetzung

Die Aufarbeitung des Kenntnisstands zu sprachunterstützenden Maßnahmen (als Oberbegriff für sprachsensibler Unterricht, Sprachförderung, Sprachtherapie, ab jetzt: suM) im Klassenzimmer und der dazu notwendigen multiprofessionellen Kooperation führt zu folgenden Fragestellungen:

- **Forschungsfrage 1:** Wie hängen die unterschiedlichen Rollen von RLP, SHP und LOG mit der Häufigkeit von suM im Klassenzimmer zusammen?
- **Forschungsfrage 2:** Wie beeinflussen die verschiedenen Kooperationsbedingungen die Häufigkeit von suM im Klassenzimmer?
- **Forschungsfrage 3:** Beeinflussen die Kooperationsbedingungen die verschiedenen Berufsgruppen bei der Umsetzung von suM im Klassenzimmer unterschiedlich?

Die geschilderten Zusammenhänge lassen vermuten, dass bei der Häufigkeit der Umsetzung von suM im Klassenzimmer zwischen den verschiedenen Berufsgruppen bzw. deren Rollen (Bedingung auf personaler Ebene) deutliche Unterschiede bestehen. Da der Wirkungsort der RLP in der Regel das Klassenzimmer ist, auch wenn es hier spezifisch um suM (also sprachsensiblen Unterricht) geht, wird diese Berufsgruppe als Referenzgruppe definiert. Vor allem LOG werden ihre suM (Sprachtherapie) bedeutend seltener im Klassenzimmer anbieten als RLP, aber auch weniger als SHP. Letztere werden ihre suM (Sprachförderung) ebenfalls nicht so häufig im Klassenzimmer anbieten wie die RLP, da auch separate Settings zu ihren Arbeitsorten gehören, aber häufiger als LOG, da sie als Speziallehrpersonen das Klassenzimmer als Wirkungsort kennen und nutzen.

Da die unterrichtsintegrierte Unterstützung von Kindern mit Sprach-, Sprech- und Kommunikationsauffälligkeiten stark von der Zusammenarbeit der beteiligten Fachkräfte abhängt, wird die Häufigkeit der suM im Klassenzimmer wahrscheinlich in bedeutender Weise von den Bedingungen auf der institutionellen und der organisationskulturellen Ebene sowie von den Bedingungen auf der Ebene des Schulleitungshandelns beeinflusst. Fachkräfte, die an mehr Schulhäusern (also auch mit mehr Personen) arbeiten, arbeiten seltener in Kooperation und sind daher auch seltener in den Klassenzimmern zu finden als Fachkräfte, die an wenig Schulhäusern arbeiten. Fachkräfte, die häufig in intensiven Zusammenarbeitsformen (Team-Teaching) arbeiten, sind hingegen häufiger in den Klassenzimmern zu finden als solche, die selten im Team-Teaching arbeiten. Und gemäß der Theory of Planned Behavior sollten auch Personen, die starke Intentionen haben, vermehrt unterrichtsintegriert zu arbeiten, häufiger in den Klassenzimmern zu finden sein als solche Personen, die diesbezüglich eher niedrig ausgeprägte Intentionen haben.

Unklar bleibt jedoch, ob und wie sich die Bedingungen in den verschiedenen Berufsgruppen auswirken. Gelten die beschriebenen Einflüsse für alle Fachkräfte gleichermaßen oder wirken sich diese unterschiedlich oder sogar gegensätzlich aus?

Da die Umsetzung von suM im Klassenzimmer nicht in jedem Kontext gleichermaßen sinnvoll ist (Reber, 2012), wird die Häufigkeit entsprechender Maßnahmen anhand verschiedener Kontexte (Schwerpunkte der Sprach-, Sprech- und Kommunikationsauffälligkeiten, Sozialformen des Unterrichts, Fachbereiche, Erwerbsphasen) überprüft.

### 3 Methode

#### 3.1 Untersuchungsdesign

Im August 2021 startete das von der PHBern finanzierte Projekt „SpriCH – Sprachunterstützende Massnahmen in Schweizer Schulen. Studie zur Kooperation multiprofessioneller Teams in integrativen Settings“. Erklärtes Projektziel war es, mithilfe eines Online-Fragebogens RLP, SHP und LOG mit Anstellung an der Regelschule im Kanton Bern zur Häufigkeit der Umsetzung von suM im Klassenzimmer, zur Häufigkeit und Intensität der Zusammenarbeit sowie zu ihren Einstellungen zur multiprofessionellen Kooperation zu befragen. Nachdem der Fragebogen durch mehrere Entwicklungsphasen ging (s. Abschnitt 3.3), konnte die Befragung im März 2022 durchgeführt werden. Die Teilnahme erfolgte anonym, freiwillig und verfolgte ausschliesslich wissenschaftliche Zwecke. Die in diesem Querschnitt erhobenen Daten sollen helfen, die Zusammenarbeitssituation der drei genannten Berufsgruppen im Kontext der unterrichtsintegrierten suM zu beschreiben, bedeutsame Unterschiede zwischen den Berufsgruppen aufzudecken und förderliche wie hinderliche Bedingungen zu identifizieren. Auf dieser Grundlage sollen Maßnahmen entwickelt werden, um die Zusammenarbeit im genannten Kontext zu optimieren (s. Abb. 1).

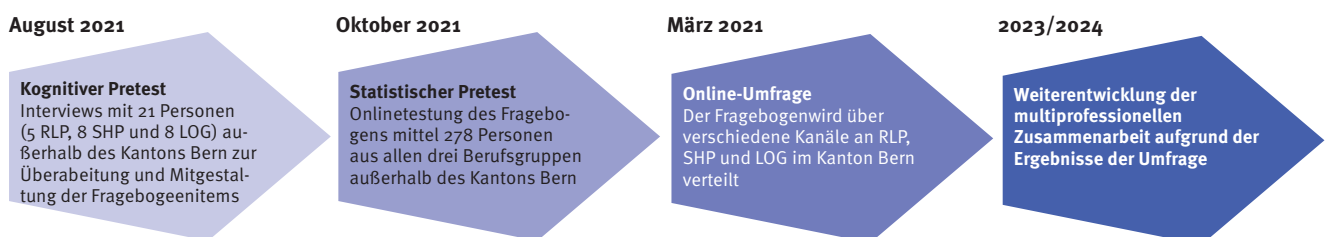


Abb. 1: Darstellung Projektverlauf

#### 3.2 Stichprobenrekrutierung

In einem ersten Schritt wurden die Schulleitungen von sämtlichen Regelschulen im Kanton Bern (N = 389) per Mail zur Teilnahme an der Umfrage eingeladen. Unter diesen gab es solche, die eine Teilnahme aus Mangel an zeitlichen Kapazitäten verweigerten, solche, die es ihren Angestellten freistellten, teilzunehmen, und sogar solche, die die RLP, SHP und LOG ihrer Schule zur Teilnahme verpflichteten. In einem zweiten Schritt, um die Berufsgruppen direkt zu erreichen, wurde die Einladung zur Befragung über die jeweiligen Berufsverbände (Bildung Bern für RLP und SHP und Logopädie Bern für LOG) gestreut. Die Mitglieder von Logopädie Bern aus dem Regelschulbereich (N = 147) wurden dann in einem dritten Schritt noch telefonisch kontaktiert,

um über das persönliche Gespräch noch einmal zur Teilnahme zu motivieren. Die LOG wurden außerdem gebeten, die RLP und SHP in ihrem Schulhaus zur Teilnahme zu bewegen. Insgesamt konnten so Daten von 150 RLP, 97 SHP und 90 LOG ( $N = 337$ ) erhoben werden, welche dann in der Datenanalyse bereinigt und für die deskriptiven und inferenzstatistischen Analysen aufbereitet wurden. Im Verlauf dieses Prozesses mussten die Daten von insgesamt 20 Personen aus dem Datensatz entfernt werden, da die entsprechenden Angaben zum Großteil unvollständig waren oder die Personen nicht zur definierten Zielgruppe gehörten. Die verbliebenen 317 Personen verteilen sich auf die drei Berufsgruppen wie in Tabelle 1 ersichtlich. Während der Rücklauf für die LOG genau ermittelt (60.5%) und damit als gut eingeschätzt werden kann, ist der Rücklauf bei den SHP und den RLP schwierig einzuschätzen, da nicht genau bekannt ist, wie viele der Personen letzten Endes die Einladung erhalten haben. Basierend auf den Statistiken der Bildungs- und Kulturdirektion des Kantons Bern (2022) entspricht der Rücklauf für die SHP 4.0% und für die RLP 1.3%.

Tab. 1: Stichprobenbeschreibung

Merkmale	Stichprobe (N = 317)			Statistischer Vergleich (Effektstärke)
	RLP (n = 142)	SHP (n = 89)	LOG (n = 86)	
<b>Basisvariablen</b>				
Geschlecht weiblich (Anteil in %)	126 (88.7 %)	83 (93.3 %)	84 (97.7 %)	$p = .040$ (schwach)
Alter in Jahren: M (SD)	40.5 (11.3)	42.0 (10.5)	42.0 (11.9)	n.s.
Berufserfahrung in Jahren: M (SD)	17.8 (12.4)	11.3 (9.4)	16.6 (11.2)	$p = .006$ (mittel)
Lektionen pro Woche: M (SD)	20.4 (6.3)	16.3 (5.7)	17.9 (6.3)	$p < .001$ (mittel)
Anzahl Kinder in Verantwortung: M (SD)	33.5 (21.8)	38.1 (31.5)	23.5 (9.5)	n.s.
Davon mit Sprach-, Sprech- und Kommunikations-auffälligkeiten	23.6 %	33.9 %	86.4 %	$p < .001$ (stark)
<b>Prädiktoren</b>				
Anzahl Schulhäuser: M (SD)	1.2 (0.6)	2.2 (1.4)	7.9 (7.2)	$p < .001$ (mittel - stark)
Team-Teaching-Werta: M (SD)	1.6 (1.0)	2.1 (0.8)	0.3 (0.5)	$p < .001$ (schwach - stark)
Intentionswertb: M (SD)	2.0 (0.9)	1.9 (0.9)	0.8 (0.8)	$p < .001$ (stark)

<sup>a</sup> Min.-Max. der Skala: 0-4; <sup>b</sup> Min.-Max. der Skala: 0-3

### 3.3 Fragebogenentwicklung und -beschreibung

Basierend auf Fragebögen von Blechschmidt et al. (2013), Elksnin und Capilouto (1994) und Drossel (2015) wurden verschiedene Items zusammengestellt und adaptiert. U. a. mussten für die genannten Berufsgruppen Parallelversionen des Fragebogens entwickelt werden, damit die Items die jeweiligen Arbeitsschwerpunkte inhaltlich abbilden. Ein Item im Bereich der Intentionen lautet bspw. „Ich würde in Zukunft gerne häufiger in Absprache mit den anderen Fachkräften *sprachsensiblen Unterricht/Sprachförderung/Sprachtherapie durchführen*,“ je nachdem ob RLP, SHP oder LOG den Fragebogen bearbeiten.

Diese Formulierungen wurden im Rahmen kognitiver Interviews (Prüfer & Rexroth, 2005) im August und September 2021 mit insgesamt 21 Personen (5 RLP, 8 SHP und 8 LOG) vorgetestet und laufend überarbeitet. Der Fragebogen wurde daraufhin im Oktober 2021 im Rahmen eines statistischen Pretests von insgesamt 278 Personen (58 RLP, 65 SHP und 155 LOG) aus verschiedenen Kantonen der Deutschschweiz (ausgenommen Bern) beantwortet. Diese Daten wurden in einem nächsten Schritt bereinigt, aufgearbeitet und deskriptiv dargestellt. Aufgrund der statistischen Evaluation des Fragebogens konnten einige der verwendeten Subskalen und Skalennitems als unbrauchbar identifiziert und dementsprechend angepasst oder auch eliminiert werden, was einerseits zur Kürzung der Beantwortungsdauer des Fragebogens auf 20 bis 25 Minuten

führte und andererseits die Ergänzung der Subskala „Schulleitungshandeln“ ermöglichte. Details zur Fragebogenevaluation können der Projekthomepage ([www.phbern.ch/projekt-sprich](http://www.phbern.ch/projekt-sprich)) entnommen werden.

Der Fragebogenteil zur Häufigkeit von suM im Klassenzimmer beinhaltet Items, die nach demselben Muster formuliert sind: „Wie häufig führen Sie Sprachsensiblen Unterricht/Sprachförderung/Sprachtherapie bei Schwierigkeiten im Kontext XY im Klassenzimmer durch?“ Es werden insgesamt 22 verschiedene Kontexte erfragt (Sprach-, Sprech- und Kommunikationsauffälligkeiten: fünf Items; Sozialformen: fünf Items; Fachbereiche: neun Items; Erwerbsphasen: drei Items). Von den neun Items im Bereich der Unterrichtsfächer werden nur die fünf verwendet, die in jeder Schulstufe unterrichtet werden und so von jeder teilnehmenden Fachperson gleichermaßen eingeschätzt werden können.

Die Häufigkeit des Team-Teachings wurde über fünf Items ermittelt, die für die verschiedenen Team-Teaching-Formen (Unterricht und Einzelunterstützung, Unterricht in Stationen, paralleles Unterrichten, Klassen- und Förderunterricht und interaktives Unterrichten) stehen.

Die Antwortskala für diese Items umfasst fünf Stufen: 0 = nie, 1 = 1–2-mal pro Schuljahr, 2 = 1–2-mal pro Monat, 3 = 1–2-mal pro Woche, 4 = fast jeden Tag). Die Skalenstufen wurden möglichst konkret ausformuliert, damit sich die Teilnehmenden bei der Auswahl einer passenden Antwortstufe schneller orientieren und entscheiden konnten (Krosnick & Presser, 2010). Damit wird auch dem von Stommel, Hildebrandt, Senn und Widmer (2014) beschriebenen Problem entgegengewirkt, dass für eine Person „einmal pro Woche“ schon als „häufig“ bewertet wird, von einer anderen Person jedoch als „selten“.

Da für die vorliegende Untersuchung die einzelnen Team-Teaching-Formen nicht von Interesse waren, sondern der Wert lediglich als Prädiktor in die Regressionsgleichung einging, wurden die verschiedenen Werte addiert und gemittelt. Die Werte können analog zu Luder (2021) metrisch interpretiert werden (0 = nie, 1 = selten, 2 = gelegentlich, 3 = oft, 4 = sehr oft).

Die Intentionen wurden ebenfalls über fünf Items erhoben, von denen aber nur eines zum Fokus der vorliegenden Studie passt: „Ich würde in Zukunft gerne öfter Sprachsensiblen Unterricht/Sprachförderung/Sprachtherapie im Klassenzimmer durchführen.“ Die Antwortskala für die Intentionen ist vierstufig (0 = trifft nicht zu, 1 = trifft eher nicht zu, 2 = trifft eher zu, 3 = trifft zu) und kann gemäß Urban und Mayerl (2011) metrisch interpretiert werden.

### 3.4 Statistische Analyse

Die Datenanalyse wurde mit Hilfe der Statistik-Software SPSS 28 durchgeführt. Da die abhängigen Variablen gemäß den Kriterien von Urban und Mayerl (2011, S. 275) als ordinale Daten interpretiert werden müssen, wurden die Zusammenhänge mithilfe von ordinalen Regressionen geprüft.

Als Ergebnis ordinaler Regressionen wird eine Odds Ratio (OR) errechnet. Eine OR zwischen 0 und 1 bedeutet, dass das Erreichen einer höheren Skalenstufe bezüglich der abhängigen Variable unwahrscheinlicher ist als in der Referenzgruppe (hier: RLP), eine OR gleich 1 bedeutet, dass die Wahrscheinlichkeit für einen Stufenanstieg gleich groß wie in der Referenzgruppe ist, und eine OR größer 1 steht dafür, dass die Wahrscheinlichkeit des Stufenanstiegs größer als in der Referenzgruppe ist (Strand, Cadwallader & Firth, 2011). In den grafischen Darstellungen der Ergebnisse werden die OR in Wahrscheinlichkeiten transformiert, um die Zusammenhänge in den Daten nachvollziehbarer zu machen.

Als Gütekriterien für die Regressionsmodelle werden das Nagelkerke- $R^2$  sowie das *proportional-chance-Kriterium* (PCC) herangezogen. Das Nagelkerke- $R^2$  kann gemäß Rohrlack (2009) folgendermaßen interpretiert werden:

- 0.2-0.4 → akzeptabel
- 0.4-0.5 → gut
- > 0.5 → sehr gut

Das PCC beschreibt den Anteil der korrekten Zuteilungen zu den Antwortkategorien, der sich zufällig aus den unterschiedlichen Gruppengrößen ergibt. Der Anteil der korrekten Zuteilungen, der sich aus dem Regressionsmodell ergibt, sollte um mind. 25 % größer als der PCC sein (White, 2013). Erst dann ist die Vorhersagekraft des Modells ausreichend. In den Ergebnistabellen wird von daher nur der um den Faktor 1.25 multiplizierte PCC angegeben.

Die Analyse der abhängigen Variablen zur Beantwortung der Forschungsfragen erfolgt in mehreren Schritten. Zur Beantwortung der ersten Forschungsfrage wird im ersten Schritt lediglich der

Haupteffekt des Faktors „Berufsgruppe“ überprüft (Basismodell). Um Forschungsfrage 2 zu beantworten, erfolgt im zweiten Schritt die Untersuchung der Haupteffekte der Variablen „Anzahl Schulhäuser“, „Team-Teaching-Wert“ und „Intentionswert“, um zu testen, ob weitere bedeutende Zusammenhänge existieren und ob der Effekt der Berufsgruppe neben diesen bestehen bleibt (erweitertes Basismodell). Im dritten Schritt werden zur Beantwortung der Forschungsfrage 3 die Interaktionen der weiteren Faktoren mit dem Faktor „Berufsgruppe“ hinzugenommen, um zu analysieren, ob sich die kooperationsförderlichen Bedingungen in den verschiedenen Berufsgruppen unterschiedlich auswirken (Interaktionsmodell). Im vierten und letzten Schritt werden aus Gründen der Modellsparsamkeit nur noch die signifikanten Interaktionen beibehalten ( finales Modell). Falls es keine signifikanten Interaktionen gibt, entspricht das finale Modell dem erweiterten Basismodell. Diese Modellentwicklung wird aus Platzgründen nicht dargestellt. Die Ergebnisse werden in einer Übersicht der finalen Modelle zusammengefasst und erläutert.

## 4 Ergebnisse

### 4.1 Stichprobencharakterisierung

Es wird deutlich, dass sich die drei Berufsgruppen in verschiedenen Aspekten unterscheiden (Geschlecht, Berufserfahrung in Jahren, Lektionen pro Woche, Anzahl Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten). Hier kommen aller Wahrscheinlichkeit nach grundlegenden Unterschieden zwischen den Berufsgruppen zum Vorschein. Dies fällt vor allem an der Anzahl Kinder mit Sprach-, Sprech- und Kommunikationsauffälligkeiten auf, die dem Berufsbild entsprechend bei den LOG am höchsten ist. Auch die anderen Unterschiede sind wahrscheinlich Ausdruck berufsspezifischer Merkmale.

Die Anzahl Schulhäuser (bzw. Kindergärten), in der eine Fachkraft arbeitet, unterscheidet sich ebenfalls (Kruskal-Wallis:  $H(2) = 172.450$ ,  $p < .001$ ): RLP arbeiten in der Regel nur an einem Schulhaus, SHP an zwei bis drei und LOG an durchschnittlich sieben Schulhäusern. Der Unterschied zwischen RLP und SHP führt zu einem mittleren Effekt; der Unterschied zwischen LOG und den beiden anderen Gruppen zu einem starken Effekt. Die Anzahl Schulhäuser (als kooperationsbedingender Faktor auf der Ebene des Schulleitungshandelns) unterscheidet sich zwischen den drei Berufsgruppen stark, weswegen Auswirkungen auf die Häufigkeit der suM im Klassenzimmer in unterschiedlicher Ausprägung zu erwarten sind.

Dies gilt ebenso für den Team-Teaching-Wert (kooperationsbedingender Faktor auf institutioneller Ebene). Die Unterschiede zwischen den Gruppen sind hochsignifikant (Kruskal-Wallis:  $H(2) = 120.186$ ,  $p < .001$ ). Während sich die Werte der RLP und der SHP nur wenig unterscheiden (schwacher Effekt), sind vor allem die Werte der LOG bedeutend kleiner als die der anderen beiden Gruppen (starker Effekt). LOG arbeiten folglich wesentlich seltener im Team-Teaching, was sich in berufsgruppenspezifischen Effekten zeigen sollte.

Auch die Intentionswerte (kooperationsbedingender Faktor auf organisationskultureller Ebene) fallen signifikant unterschiedlich aus (Kruskal-Wallis:  $H(2) = 84.978$ ,  $p < .001$ ). Während RLP und SHP sich nicht unterscheiden, entsprechen die Unterschiede in den Werten der LOG im Vergleich zu denen der anderen beiden Berufsgruppen starken Effekten und sollten sich gerade in dieser Berufsgruppe bemerkbar machen.

### 4.2 Ergebnisse zu den Schwerpunkten der Sprach-, Sprech- und Kommunikationsauffälligkeiten

Die finalen Modelle bezüglich der abhängigen Variablen im Bereich der Sprach-, Sprech- und Kommunikationsauffälligkeiten werden zusammenfassend in der Tabelle 2 dargestellt. Die Modellgütekriterien bewegen sich alle im guten bis sehr guten Bereich.

**Zur Forschungsfrage 1:** Der Faktor Berufsgruppe führt immer zu einem signifikanten Unterschied. SHP setzen suM im Klassenzimmer (außer im Bereich Grammatik) demnach seltener um als RLP: *Aussprache*: OR = .146; 95 % CI [.044; .485], Wald  $\chi^2(1) = 9.840$ ,  $p = .002$ ; *Wortschatz*: OR = .376; 95 % CI [.209; .675], Wald  $\chi^2(1) = 10.714$ ,  $p = .001$ ; *Sprachverständnis*: OR = .339; 95 % CI [.191; .604], Wald  $\chi^2(1) = 13.537$ ,  $p < .001$ ; *Kommunikation*: OR = .139; 95 % CI [.042; .466], Wald  $\chi^2(1) = 10.229$ ,  $p = .001$ . Für LOG fällt dieser Unterschied noch deutlicher aus, was bedeutet, dass sie für suM selten bis nie im Klassenzimmer sind: *Aussprache*: OR = .018; 95 % CI [.005; .058], Wald  $\chi^2(1) = 44.349$ ,  $p < .001$ ; *Wortschatz*: OR = .008; 95 % CI [.003; .024], Wald  $\chi^2(1) = 79.839$ ,  $p < .001$ ; *Grammatik*: OR = .001; 95 % CI [.000; .006], Wald  $\chi^2(1) = 80.677$ ,  $p < .001$ ; *Sprachver-*

*ständnis*: OR = .015; 95 % CI [.006; .039], Wald  $\chi^2(1) = 74.253$ ,  $p < .001$ ; *Kommunikation*: OR = .012; 95 % CI [.004; .037], Wald  $\chi^2(1) = 59.158$ ,  $p < .001$  (s. Abb. 2).

Tab. 2: Überblick über die finalen Modelle zur Häufigkeit von suM im Bereich „Schwerpunkt der Sprach-, Sprech- und Kommunikationsauffälligkeiten“

Unabhängige Variablen	Odds Ratios				
	Aussprache	Wortschatz	Grammatik	Sprachverständnis	Kommunikation
<b>Haupteffekte</b>					
Beruf (Referenz: RLP)					
SHP	.146**	.376***	.307	.339***	.139**
LOG	.018***	.008***	.001***	.015***	.012***
Anzahl Schulhäuser	.971	1.068*	1.090**	1.060*	1.016
Team-Teaching	1.198	2.328***	1.370*	1.998***	1.218
Intentionen	1.112	1.092	.932	1.055	1.185
<b>Interaktionseffekte</b>					
SHP mit Anzahl Schulhäuser	-	-	-	-	-
LOG mit Anzahl Schulhäuser	-	-	-	-	-
SHP mit Team-Teaching-Wert	1.366	-	1.119	-	1.809*
LOG mit Team-Teaching-Wert	2.886*	-	4.419**	-	3.792**
SHP mit Intentionswert	-	-	-	-	-
LOG mit Intentionswert	-	-	-	-	-
Nagelkerke-R <sup>2</sup>	.499	.642	.619	.576	.532
Korrekte Klassifikation (PCC)	51.7 % (29.7 %)	54.8 % (32.7 %)	50.3 % (29.9 %)	57.2 % (32.3 %)	47.4 % (26.5 %)

Anmerkung: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

**Zur Forschungsfrage 2:** Neben diesen Effekten ist in einigen Fällen noch ein positiver Einfluss der Anzahl Schulhäuser (*Wortschatz*: OR = 1.068; 95 % CI [1.010; 1.131], Wald  $\chi^2(1) = 5.252$ ,  $p = .022$ ; *Sprachverständnis*: OR = 1.060; 95 % CI [1.002; 1.121], Wald  $\chi^2(1) = 4.072$ ,  $p = .044$ ) und des Team-Teaching-Faktors (*Wortschatz*: OR = 2.328; 95 % CI [1.738; 3.114], Wald  $\chi^2(1) = 32.269$ ,  $p < .001$ ; *Sprachverständnis*: OR = 1.998; 95 % CI [1.508; 2.643], Wald  $\chi^2(1) = 23.368$ ,  $p < .001$ ) festzustellen. In beiden Fällen bedeutet das, dass mit einer steigenden Anzahl Schulhäuser oder einem steigenden Team-Teaching-Wert auch die Wahrscheinlichkeit für die Umsetzung von suM im Klassenzimmer steigt. Der Effekt des Team-Teaching-Faktors ist deutlich größer als der der Anzahl Schulhäuser.

**Zur Forschungsfrage 3:** Der Team-Teaching-Wert ist auch der einzige Faktor, der die Berufsgruppen signifikant beeinflusst: Je höher der Team-Teaching-Wert einer Fachperson, desto größer die Wahrscheinlichkeit, dass diese suM im Klassenzimmer durchführt. Dieser Effekt betrifft vor allem LOG (*Aussprache*: OR = 2.886; 95 % CI [1.050; 7.925], Wald  $\chi^2(1) = 4.221$ ,  $p = .040$ ; *Grammatik*: OR = 4.419; 95 % CI [1.487; 13.118], Wald  $\chi^2(1) = 7.157$ ,  $p = .007$ ; *Kommunikation*: OR = 3.792; 95 % CI [1.445; 9.954], Wald  $\chi^2(1) = 7.336$ ,  $p = .007$ ), in einem Fall auch SHP (*Kommunikation*: OR = 1.809; 95 % CI [1.035; 3.161], Wald  $\chi^2(1) = 4.319$ ,  $p = .038$ ) (s. Abb. 3).

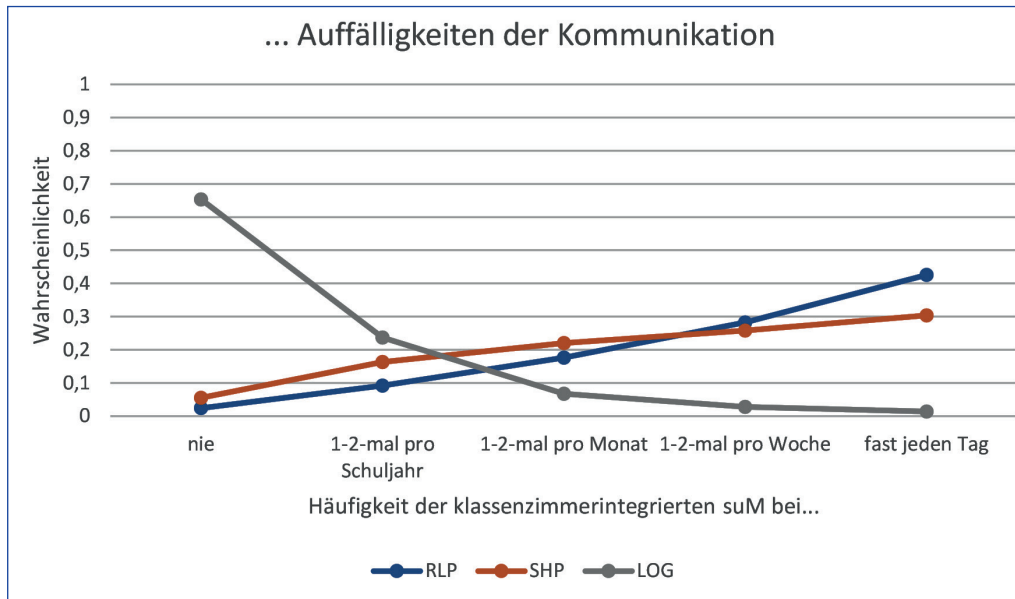


Abb. 2: Modellierter Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer bei Auffälligkeiten der Kommunikation

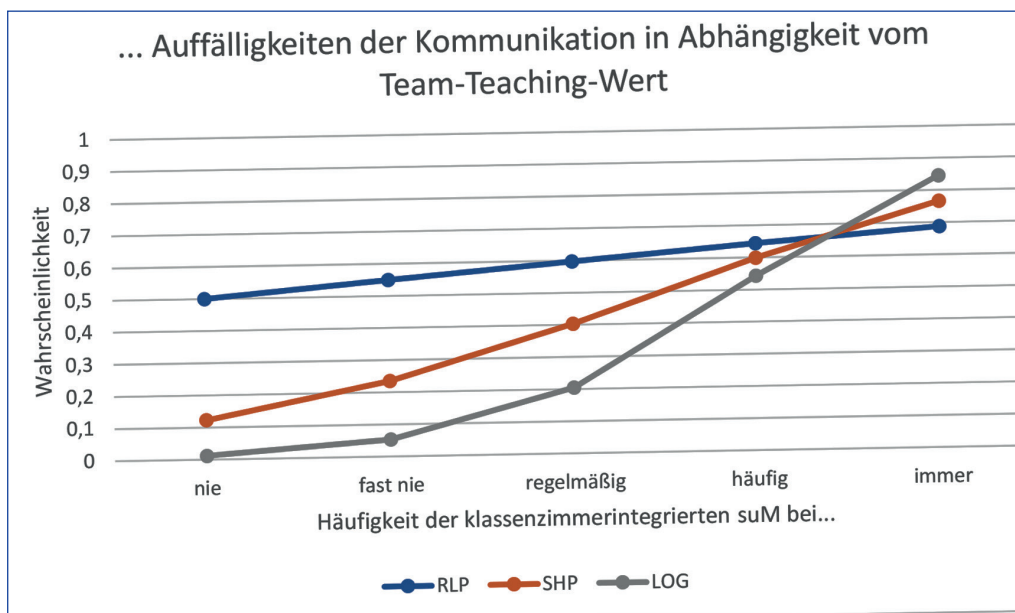


Abb. 3: modellierter Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer bei Auffälligkeiten der Kommunikation in Interaktion mit den Team-Teaching-Werten

### 4.3 Ergebnisse zu den gewählten Sozialformen im Unterricht

Die Übersicht zu den finalen Modellen im Kontext der Sozialformen im Unterricht weist ähnliche Zusammenhänge auf (s. Tab. 3).

**Zur Forschungsfrage 1:** Auch hier wird der Haupteffekt „Berufsgruppe“ in jedem der Modelle signifikant. SHP führen suM im Kontext der Sozialformen deutlich seltener im Klassenzimmer durch als RLP: *Einzelarbeit*: OR = .090; 95 % CI [.020; .418], Wald  $\chi^2(1) = 9.470$ ,  $p = .002$ ; *Partnerarbeit*: OR = .089; 95 % CI [.019; .429], Wald  $\chi^2(1) = 9.102$ ,  $p = .003$ ; *Gruppenarbeit*: OR = .085; 95 % CI [.023; .299], Wald  $\chi^2(1) = 14.750$ ,  $p < .001$ ; *Frontalunterricht*: OR = .012; 95 % CI [.003; .049], Wald  $\chi^2(1) = 37.399$ ,  $p < .001$ ; *Gruppengespräch*: OR = .029; 95 % CI [.137; .419], Wald  $\chi^2(1) = 25.076$ ,  $p < .001$ . Dies gilt gleichermaßen für LOG: *Einzelarbeit*: OR = .005; 95 % CI [.001; .018], Wald  $\chi^2(1) = 56.558$ ,  $p < .001$ ; *Partnerarbeit*: OR = .008; 95 % CI [.002; .031], Wald  $\chi^2(1) = 46.829$ ,  $p < .001$ ; *Gruppenarbeit*: OR = .002; 95 % CI [.001; .008], Wald  $\chi^2(1) = 80.159$ ,  $p < .001$ ; *Frontalunterricht*: OR = .001; 95 % CI [.000; .006], Wald  $\chi^2(1) = 71.302$ ,  $p < .001$ ; *Gruppengespräch*: OR = .014; 95 % CI [.011; .064], Wald  $\chi^2(1) = 63.543$ ,  $p < .001$  (s. Abb. 4).



Tab. 3: Überblick über die finalen Modelle zur Häufigkeit von suM im Bereich „gewählte Sozialform im Unterricht“

Unabhängige Variablen	Odds Ratios				
	Einzelarbeit	Partnerarbeit	Gruppenarbeit	Frontalunterricht	Gruppen-gespräch
<b>Haupteffekte</b>					
Beruf (Referenz: RLP)					
SHP	.090**	.089**	.085***	.012***	.029***
LOG	.005***	.008***	.002***	.001***	.014***
Anzahl Schulhäuser	.566	1.051	1.058	.972	1.151
Team-Teaching	1.709	1.172	1.198	1.084	1.252
Intentionen	.909	1.196	1.017	.784	.862
<b>Interaktionseffekte</b>					
SHP mit Anzahl Schulhäuser	2.006*	-	-	-	-
LOG mit Anzahl Schulhäuser	1.793*	-	-	-	-
SHP mit Team-Teaching-Wert	.875	1.517	2.208**	-	-
LOG mit Team-Teaching-Wert	3.522*	14.732***	6.787***	-	-
SHP mit Intentionenwert	-	1.031	-	2.710**	-
LOG mit Intentionenwert	-	.432*	-	2.912**	-
Nagelkerke-R <sup>2</sup>	.559	.557	.610	.634	.504
Korrekte Klassifikation (PCC)	61.7 % (38.8 %)	60.3 % (36.7 %)	49.5 % (29.1 %)	60.5 % (30.0 %)	51.7 % (28.9 %)

Anmerkung: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

**Zur Fragestellung 2:** Keine der untersuchten Kooperationsbedingungen hat einen signifikanten Einfluss auf die Häufigkeit der suM im Klassenzimmer im Kontext der gewählten Sozialformen im Unterricht.

**Zur Fragestellung 3:** Die Anzahl Schulhäuser beeinflusst die Häufigkeit der suM im Unterricht in beiden Berufsgruppen bei der Sozialform „Einzelarbeit“ (SHP: OR = 2.006; 95 % CI [1.124; 3.582], Wald  $\chi^2(1) = 5.542$ ,  $p = .019$ ; LOG: OR = 1.793; 95 % CI [1.085; 2.962], Wald  $\chi^2(1) = 5.204$ ,  $p = .023$ ). Mit jedem weiteren Schulhaus, für das SHP und LOG zuständig sind, steigt die Wahrscheinlichkeit, dass auch die Sozialform „Einzelarbeit“ für suM im Klassenzimmer genutzt wird.

Der Team-Teaching-Faktor beeinflusst die Berufsgruppe LOG gleich in mehreren Kontexten positiv. Je höher der Team-Teaching-Wert von LOG, desto größer die Wahrscheinlichkeit, dass auch Einzelarbeiten, Partnerarbeiten und Gruppenarbeiten für suM im Klassenzimmer genutzt werden (Einzelarbeiten: OR = 3.522; 95 % CI [1.270; 9.777], Wald  $\chi^2(1) = 5.848$ ,  $p = .016$ ; Partnerarbeiten: OR = 14.732; 95 % CI [4.522; 47.990], Wald  $\chi^2(1) = 19.938$ ,  $p < .001$ ; Gruppenarbeiten: OR = 6.787; 95 % CI [2.354; 19.550], Wald  $\chi^2(1) = 12.569$ ,  $p < .001$ ). Für SHP gilt dies im Fall der Gruppenarbeiten auch (OR = 2.208; 95 % CI [1.245; 3.916], Wald  $\chi^2(1) = 7.344$ ,  $p = .007$ ).

Auch die Intentionen beeinflussen die berufsgruppenspezifischen Häufigkeitswerte. Während im Kontext der „Partnerarbeit“ bei LOG höhere Intentionen für eine Senkung der Wahrscheinlichkeit der Durchführung von suM im Klassenzimmer stehen (LOG: OR = .432; 95 % CI [.198; .944], Wald  $\chi^2(1) = 4.436$ ,  $p = .035$ ), wirkt sich dies im Kontext „Frontalunterricht“ hingegen in beiden Berufsgruppen positiv aus. Mit höheren Intentionen steigt auch die Wahrscheinlichkeit, dass suM auch im Frontalunterricht integriert werden (SHP: OR = 2.710; 95 % CI [1.471; 4.993], Wald  $\chi^2(1) = 10.230$ ,  $p = .001$ ; LOG: OR = 2.912; 95 % CI [1.330; 6.379], Wald  $\chi^2(1) = 7.137$ ,  $p = .008$ ). Mit Blick auf die Abbildung 5 kann dies aber auch anders interpretiert werden. So ist der Anstieg der Häufigkeit von suM im Klassenzimmer zwar messbar, der Effekt besteht

jedoch darin, dass die Häufigkeit, suM im Klassenzimmer durchzuführen, bei RLP mit hohen Intentionen abnimmt: Je eher sich eine RLP vornimmt, unterrichtsintegrierte suM durchzuführen, desto weniger wird sie dies im Frontalunterricht tun. Bei den SHP und LOG hingegen nimmt die Häufigkeit leicht zu (s. Abb. 5).

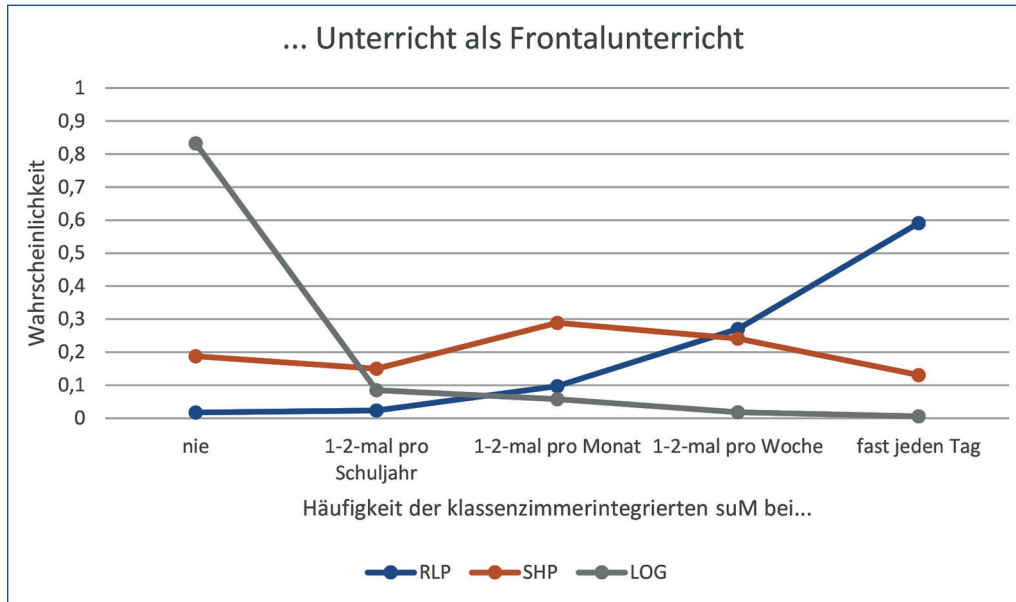


Abb. 4: modellierte Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer bei Unterricht als Frontalunterricht

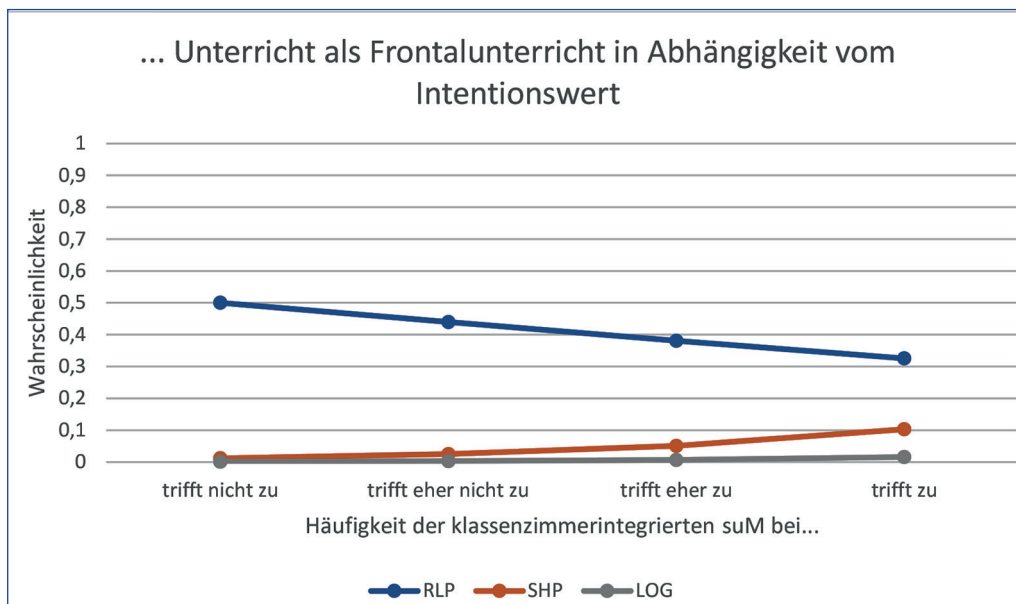


Abb. 5: modellierte Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer bei Unterricht als Frontalunterricht in Interaktion mit den Intentionenwerten

#### 4.4 Ergebnisse zu den Fachbereichen

**Zur Forschungsfrage 1:** Im Bereich der verschiedenen Unterrichtsfächer steht die Zugehörigkeit zur Berufsgruppe LOG immer für eine bedeutend kleineren Häufigkeitswert von suM im Klassenzimmer (*Deutsch*: OR = .003; 95 % CI [.001; .013], Wald  $\chi^2(1) = 55.079$ ,  $p < .001$ ; *Mathematik*: OR = <.001; 95 % CI [.000; .007], Wald  $\chi^2(1) = 30.335$ ,  $p < .001$ ; *Gestalten*: OR = .012; 95 % CI [.002; .070], Wald  $\chi^2(1) = 24.659$ ,  $p < .001$ ; *Musik*: OR = .072; 95 % CI [.006; .871], Wald  $\chi^2(1) = 4.279$ ,  $p = .039$ ; *Bewegung und Sport*: OR = .038; 95 % CI [.008; .176], Wald  $\chi^2(1) = 17.695$ ,  $p < .001$ ). Auf die SHP trifft das nur in den Fächern „Gestalten“ (OR = .301; 95 % CI [.146; .619], Wald  $\chi^2(1) = 10.619$ ,  $p = .001$ ), „Musik“ (OR = .020; 95 % CI [.003; .125], Wald  $\chi^2(1) = 17.555$ ,  $p < .001$ ) und „Bewegung und Sport“ (OR = .156; 95 % CI [.066; .368], Wald  $\chi^2(1) = 18.016$ ,  $p < .001$ ) zu.

Das heißt im Umkehrschluss, dass SHP im Vergleich zu den RLP in den Fächern „Deutsch“ und „Mathematik“ suM im Klassenzimmer gleich häufig durchführen (s. Tab. 4; s. Abb. 6).

Tab. 4: Überblick über die finalen Modelle zur Häufigkeit von suM im Bereich „Fachbereiche“

Unabhängige Variablen	Odds Ratios				
	Deutsch	Mathematik	Gestalten	Musik	Bewegung und Sport
<b>Haupteffekte</b>					
Beruf (Referenz: RLP)					
SHP	.555	.458	.301**	.020***	.156***
LOG	.003***	<.001***	.012***	.072*	.038***
Anzahl Schulhäuser	.966	.944	1.084	.853	.984
Team-Teaching	1.589*	1.255	1.470*	1.376	1.269
Intentionen	.844	.840	1.231	1.093	.998
<b>Interaktionseffekte</b>					
SHP mit Anzahl Schulhäuser	-	-	-	2.784*	-
LOG mit Anzahl Schulhäuser	-	-	-	.889	-
SHP mit Team-Teaching-Wert	1.054	1.391	-	-	-
LOG mit Team-Teaching-Wert	5.419**	7.382*	-	-	-
SHP mit Intentionenwert	-	-	-	-	-
LOG mit Intentionenwert	-	-	-	-	-
Nagelkerke-R <sup>2</sup>	.600	.619	.476	.527	.417
Korrekte Klassifikation (PCC)	61.9 % (48.3 %)	48.8 % (33.9 %)	50.3 % (32.7 %)	55.9 % (38.2 %)	55.6 % (37.4 %)

Anmerkung: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

**Zur Forschungsfrage 2:** Von den weiteren Kooperationsbedingungen wirkt sich lediglich das Team-Teaching im Kontext des Deutschunterrichts (OR = 1.589; 95 % CI [1.037; 2.435], Wald  $\chi^2(1) = 4.520$ ,  $p = .034$ ) und des Unterrichts fürs Gestalten (OR = 1.470; 95 % CI [1.068; 2.022], Wald  $\chi^2(1) = 5.605$ ,  $p = .018$ ) signifikant aus. In beiden Fächern steht ein Anstieg des Team-Teaching-Werts für höhere Häufigkeitswerte von suM im Klassenzimmer. Unabhängig von der Berufsgruppe führen Personen, die im Team-Teaching arbeiten, in den Fächern Deutsch und Gestalten suM mit höherer Wahrscheinlichkeit im Klassenzimmer durch.

**Zur Forschungsfrage 3:** Der Faktor „Anzahl Schulhäuser“ beeinflusst außerdem die Berufsgruppe der SHP im Fach Musik (OR = 2.784; 95 % CI [1.270; 6.104], Wald  $\chi^2(1) = 6.535$ ,  $p = .011$ ). Mit jedem Schulhaus mehr, für das eine SHP verantwortlich ist, steigt die Wahrscheinlichkeit, dass auch der Fachbereich Musik für suM im Klassenzimmer genutzt wird.

Der Team-Teaching-Faktor wirkt sich wiederum auf die Häufigkeit von suM im Klassenzimmer bei den LOG in den Fächern Deutsch- (OR = 5.419; 95 % CI [1.790; 16.428], Wald  $\chi^2(1) = 8.930$ ,  $p = .003$ ) und Mathematik (OR = 7.382; 95 % CI [1.354; 40.246], Wald  $\chi^2(1) = 5.339$ ,  $p = .021$ ) positiv aus. Je höher der Team-Teaching-Wert von LOG, desto größer die Wahrscheinlichkeit, dass suM im Deutsch- oder Matheunterricht oft bis sehr oft durchgeführt werden (s. Abb. 7).

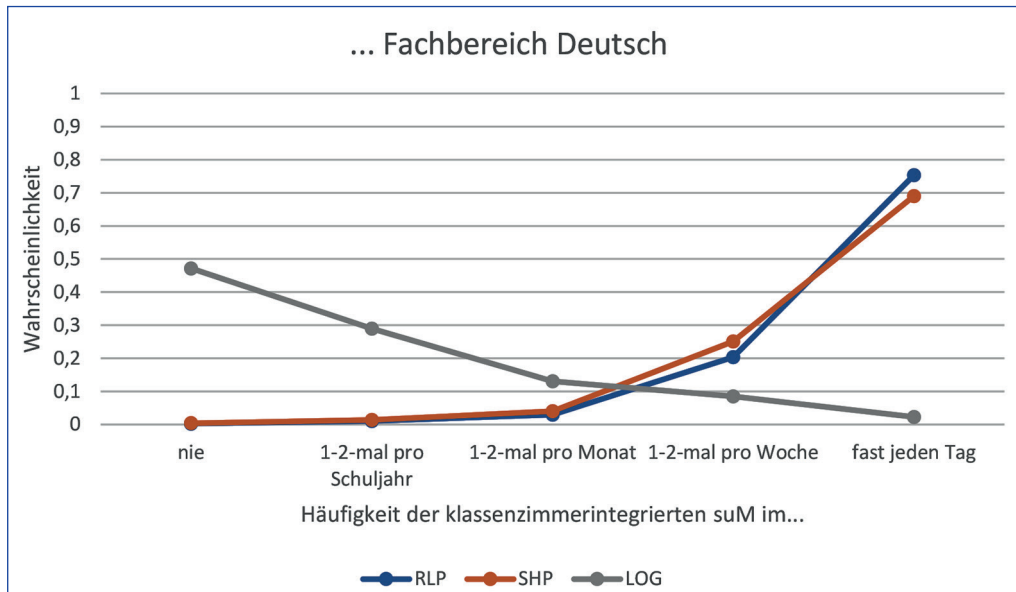


Abb. 6: modellierte Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer im Fachbereich Deutsch

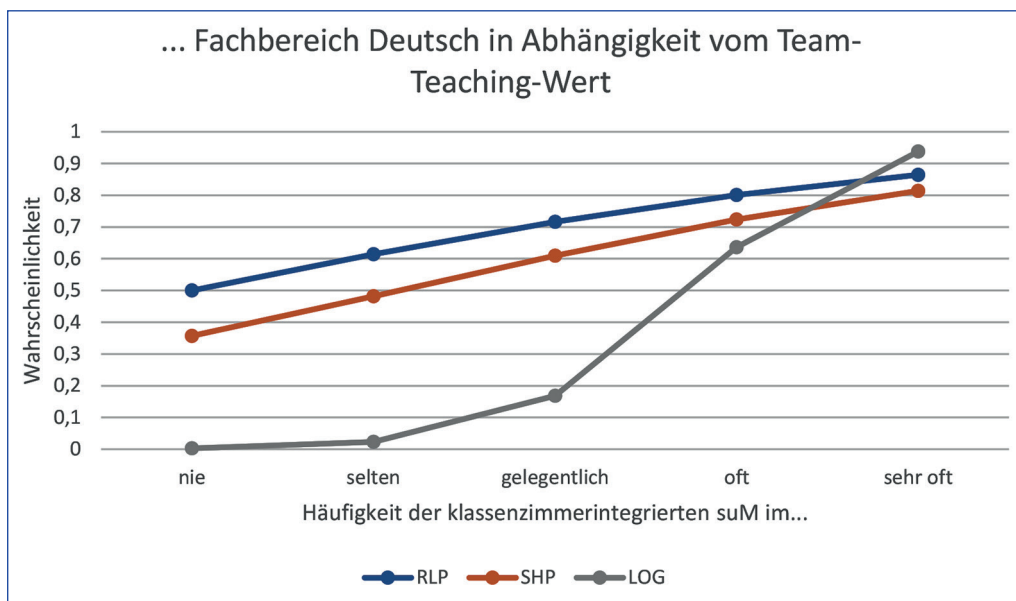


Abb. 7: modellierte Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer im Fachbereich Deutsch in Interaktion mit dem Team-Teaching-Wert

#### 4.5 Ergebnisse zu den Erwerbsphasen

**Zur Forschungsfrage 1:** Bezüglich der verschiedenen Erwerbsphasen gilt, dass die Zugehörigkeit zur Berufsgruppe LOG in allen drei Kontexten für signifikant verminderte Häufigkeitswerte steht (Einführung: OR = .007; 95 % CI [.002; .023], Wald  $\chi^2(1) = 63.316$ ,  $p < .001$ ; Übung: OR = .010; 95 % CI [.004; .026], Wald  $\chi^2(1) = 84.450$ ,  $p < .001$ ; Anwendung: OR = .005; 95 % CI [.002; .016], Wald  $\chi^2(1) = 76.129$ ,  $p < .001$ ). Im Kontext der Phase „Einführung“ gilt dies auch für SHP (OR = .162; 95 % CI [.048; .548], Wald  $\chi^2(1) = 8.580$ ,  $p = .003$ ) (s. Tab. 5). Das heißt im Umkehrschluss, dass SHP, wenn es um die Übung und Anwendung von sprachlichen Kompetenzen geht, statistisch gesehen, diese genauso häufig unterrichtsintegriert durchführen wie RLP (s. Abb. 8).

Tab. 5: Überblick über die finalen Modelle zur Häufigkeit von suM im Bereich „Erwerbsphasen“

Unabhängige Variablen	Odds Ratios		
	Einführung	Übung	Anwendung
<b>Haupteffekte</b>			
Beruf (Referenz: RLP)			
SHP	.162**	.614	.363
LOG	.007***	.010***	.005***
Anzahl Schulhäuser	1.085*	1.063*	1.101**
Team-Teaching (Häufigkeit im Durchschnitt)	1.406*	1.372*	1.363
Intentionen	.925	.983	.901
<b>Interaktionseffekte</b>			
SHP mit Anzahl Schulhäuser	-	-	-
LOG mit Anzahl Schulhäuser	-	-	-
SHP mit Team-Teaching-Wert	1.449	-	1.280
LOG mit Team-Teaching-Wert	4.870**	-	3.947**
SHP mit Intentionswert	-	-	-
LOG mit Intentionswert	-	-	-
Nagelkerke-R <sup>2</sup>	.498	.527	.555
Korrekte Klassifikation (PCC)	52.2 % (30.3 %)	46.4 % (30.2 %)	50.7 % (31.7 %)

Anmerkung: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

**Zur Forschungsfrage 2:** Die Anzahl Schulhäuser wirkt sich bei allen drei Variablen positiv aus (*Einführung*: OR = 1.085; 95 % CI [1.019; 1.156], Wald  $\chi^2(1) = 6.493$ ,  $p = .011$ ; *Übung*: OR = 1.063; 95 % CI [1.006; 1.124], Wald  $\chi^2(1) = 4.678$ ,  $p = .031$ ; *Anwendung*: OR = 1.101; 95 % CI [1.040; 1.165], Wald  $\chi^2(1) = 10.745$ ,  $p = .001$ ), in dem Sinne, dass mit jedem weiteren Schulhaus die Wahrscheinlichkeit für höhere Häufigkeitswerte für suM im Klassenzimmer steigt – auch wenn dieser Anstieg äußerst gering ausfällt.

Der Einfluss des Team-Teachings bezüglich der Variablen „Einführung“ (OR = 1.406; 95 % CI [1.040; 1.904], Wald  $\chi^2(1) = 4.892$ ,  $p = .027$ ) und „Übung“ (OR = 1.372; 95 % CI [1.063; 1.770], Wald  $\chi^2(1) = 5.894$ ,  $p = .015$ ) ist schon etwas deutlicher: Je höher der Team-Teaching-Wert einer Person (egal welche Berufsgruppe), desto größer die Wahrscheinlichkeit, dass suM im Klassenzimmer durchgeführt werden.

**Zur Forschungsfrage 3:** Der Team-Teaching-Wert beeinflusst die Wahrscheinlichkeitswerte der Berufsgruppe LOG im Kontext der „Einführung“ (OR = 4.870; 95 % CI [1.689; 14.041], Wald  $\chi^2(1) = 8.580$ ,  $p = .003$ ) und der „Anwendung“ (OR = 3.947; 95 % CI [1.522; 10.237], Wald  $\chi^2(1) = 7.969$ ,  $p = .005$ ) positiv (s. Abb. 9): Je höher der Team-Teaching-Wert der LOG, desto größer die Wahrscheinlichkeit, dass die Einführung und die Anwendung von sprachlichen Kompetenzen oft bis sehr oft unterrichtsintegriert durchgeführt wird.

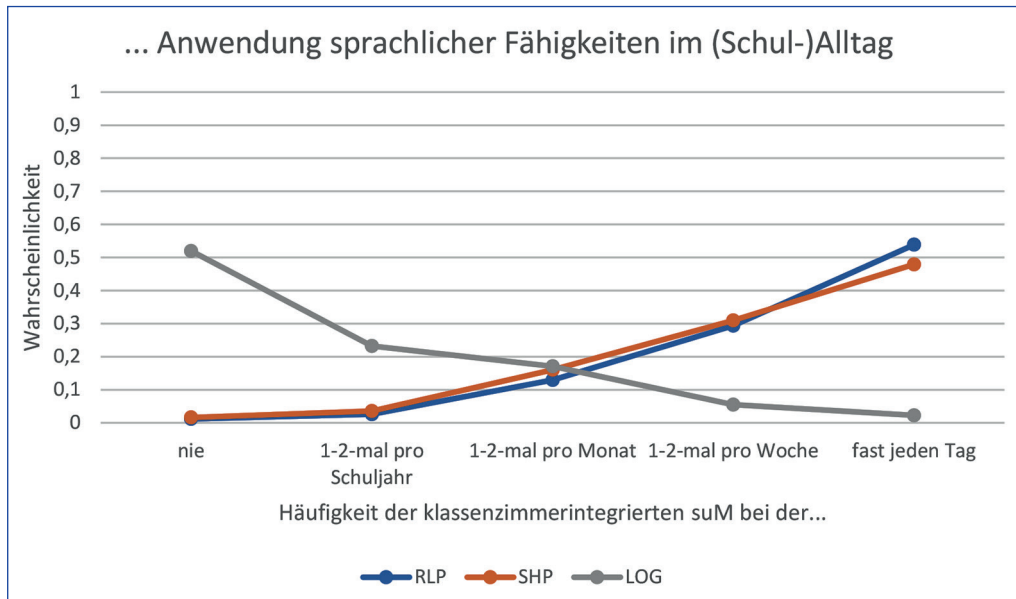


Abb. 8: modellierte Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer bei der Anwendung sprachlicher Fähigkeiten im (Schul-)Alltag

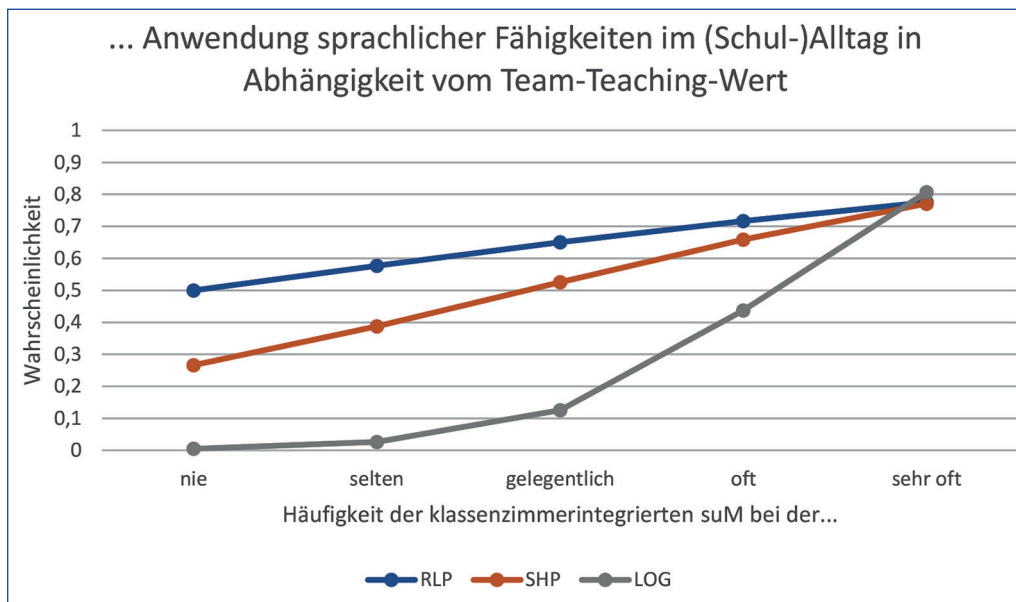


Abb. 9: modellierte Wahrscheinlichkeiten für die Häufigkeit suM im Klassenzimmer bei der Anwendung sprachlicher Fähigkeiten im (Schul-)Alltag in Interaktion mit dem Team-Teaching-Wert

## 5 Diskussion und Schlussfolgerungen

Die Ergebnisse liefern eindeutige Antworten auf die gestellten Fragen. So wird deutlich, dass die Wahrscheinlichkeit für die LOG suM im Klassenzimmer durchzuführen, immer bedeutend kleiner ausfällt als bei der Referenzgruppe der RLP. Das heißt, dass LOG Sprachtherapie unabhängig vom Kontext mit großer Wahrscheinlichkeit selten bis nie in den Klassenzimmern durchführen. Auch die Gruppe der SHP fällt in den meisten Vergleichen mit verringerten Wahrscheinlichkeitswerten auf – lediglich in den Fächern Deutsch und Mathematik sowie zur Übung und Anwendung von neuen Sprachfähigkeiten sind sie (statistisch gesehen) gleichermaßen stark in den Klassenzimmern aktiv wie RLP.

Da die erfolgreiche Umsetzung von suM im Klassenzimmer stark von der Kooperation der beteiligten Personen abhängt, wurden die zentralen Bedingungen der Zusammenarbeit berufsgruppenunabhängig analysiert. So wird der Einfluss der Anzahl Schulhäuser in einigen wenigen Fällen signifikant. Jedes weitere Schulhaus, für das eine Person zuständig ist, steht für eine Stei-

gerung der Wahrscheinlichkeit, dass diese Person häufiger suM im Klassenzimmer durchführt. Dies widerspricht der Erwartung, dass bei einer steigenden Anzahl Schulhäuser die Zusammenarbeit und damit auch die Umsetzung von suM im Klassenzimmer erschwert wird. Im Gegenteil: es scheint eher der Fall zu sein, dass die betroffenen Personen gezwungen sind, zumindest teilweise suM im Klassenzimmer umzusetzen, wenn sie an mehreren Orten arbeiten. Es bleibt aber fraglich, ob auch die Zusammenarbeit dadurch positiv beeinflusst wird oder sogar unter diesem Umstand leidet. Gemäß den Kommentaren der beteiligten RLP, SHP und LOG wird die räumliche Situation eher negativ erlebt (RLP: „So sehe ich die Logopädinnen kaum mehr“, SHP: „Leider sind wir räumlich nicht so nah beieinander“, LOG: „Situation meistens durch die räumliche Trennung erschwert“). Der individuelle Team-Teaching-Wert steht häufig für einen positiven Einfluss auf die Durchführungshäufigkeit von suM im Klassenzimmer. Dabei steht jede Einheit mehr auf der Team-Teaching-Skala für eine deutliche Erhöhung der Wahrscheinlichkeit für die häufigere Durchführung von suM im Klassenzimmer. Der Einfluss der Intentionen wird in keinem der untersuchten Kontexte signifikant.

Die verschiedenen Bedingungen beeinflussen die Berufsgruppen auf unterschiedliche Weise. Im Kontext der Sozialform „Einzelarbeit“ werden beide Berufsgruppen positiv von der Anzahl Schulhäuser beeinflusst – d. h., je mehr Schulhäuser SHP und LOG betreuen, desto wahrscheinlicher ist, dass auch die Sozialform „Einzelarbeit“ häufiger für suM im Klassenzimmer genutzt wird. Im Fachbereich Musik gilt das auch für SHP, aber nicht für LOG.

Der Einfluss der Intentionen, vermehrt unterrichtsintegriert arbeiten zu wollen, auf die Berufsgruppen wird selten deutlich und gestaltet sich unterschiedlich. Während hohe Intentionenwerte bei der Berufsgruppe LOG im Kontext der Sozialform „Partnerarbeit“ zu einer verringerten Wahrscheinlichkeit führt, suM im Klassenzimmer durchzuführen, wirkt sich ihr Einfluss in beiden Berufsgruppen in der Sozialform „Frontalunterricht“ deutlich positiv aus. Je stärker die Intention, vermehrt unterrichtsintegriert zu arbeiten, desto wahrscheinlicher wird es, dass auch der Frontalunterricht häufiger zur Umsetzung suM im Klassenzimmer genutzt wird.

Wiederholt macht sich der Einfluss des Team-Teaching-Werts in den Berufsgruppen bemerkbar – vor allem die LOG werden stark positiv beeinflusst. Je höher der Team-Teaching-Wert einer LOG, desto größer die Wahrscheinlichkeit, dass in vielen der untersuchten Kontexte suM im Klassenzimmer durchgeführt werden. Auch in der Berufsgruppe SHP steht der Team-Teaching-Wert für einen positiven Zusammenhang, wenn auch weniger deutlich und in weniger Kontexten.

An dieser Gegenüberstellung der Ergebnisse werden mehrere zentrale Erkenntnisse deutlich: Auch rund zehn Jahre nach der Untersuchung von Blechschmidt et al. (2013), den optimistischen Erwartungen von Reber (2012) und den progressiven Ideen von Kempe Preti (2010) sind LOG nach wie vor äußerst selten in den Klassenzimmern zu finden, unabhängig von der sprachlichen Auffälligkeit, vom Fach, von der Sozialform im Unterricht und der Erwerbsphase des Kindes – die bevorzugte Arbeitsweise für LOG ist nach wie vor die Einzeltherapie (89.5 % der Angaben) in einem separaten Arbeitszimmer (66.3 % der Angaben). Dies steht auch entgegen der Vorgabe des kantonalen Leitfadens, dass auch Logopädie „in der Regel unterrichtsintegriert“ durchgeführt werden soll (Erziehungsdirektion des Kantons Bern, 2019, S. 26).

Es sind vermutlich eher inhärente Eigenschaften der jeweiligen Berufsgruppen, die zu den dokumentierten Unterschieden in der Häufigkeit unterrichtsintegrierter suM führen. So gehört es zum beruflichen Selbstverständnis der Berufsgruppe LOG, nicht unterrichtsintegriert, sondern separativ zu arbeiten, wie in einigen Kommentaren deutlich wird: „Unterricht ist meiner Meinung nach nicht eine Aufgabe der Logopädin“, „Ich sehe mich als Logopädin eher als Therapeutin und nicht nur als Lehrperson“, „Ich sehe mich eher in einer therapeutischen Rolle als Logopädin und nicht in einer pädagogischen“ und „Ich bin keine Lehrerin und ich unterrichte nicht“. Auch der IBEM-Leitfaden sieht nicht vor, dass LOG im Team-Teaching arbeiten (Erziehungsdirektion des Kantons Bern, 2019, S. 37). Sallat und Siegmüller (2016, S. 249) empfehlen daher, im Kontext der Inklusion den Wechsel „von einer vornehmlich klinisch-einzelfallorientierten Sichtweise hin zu einer inklusiv-unterrichtskompatiblen Expertise“ zu vollziehen. Bei Kempe Preti (2010, 8f.) heißt es passender Weise: „Diesbezüglich gilt es isolierte Positionen zu verlassen, um mit spezifischer Fachkompetenz in Ergänzung zu anderen Professionen einen Beitrag zur integrativen Schulung zu leisten. Die vermehrte Abstimmung von sprachtherapeutischen Interventionen mit curricularen und sozialen Zielen wäre hierbei dienlich“.

An dieser Stelle soll noch einmal betont werden, dass die unterrichtsintegrierte Arbeit die therapeutische Arbeit im Einzelsetting nicht ersetzen, sondern ergänzen soll: „Für den pädagogi-

schen Bereich sind ergänzend zur Therapie Maßnahmen der sprachlichen Prävention und Adaption sowie der Intervention umzusetzen, die [...] unterrichtsintegriert (Schule) implementiert werden sollten“ (Deutsche Gesellschaft für Phoniatrie und Pädaudiologie, 2022, S. 243). Je nach Situation und Erfordernissen ergeben sich unterschiedliche Vor- und Nachteile integrativer und separativer Arbeitsweisen – es ist vielmehr das „Wechselspiel der Settings“ (Braun, 2009, S. 17), das entwicklungsfördernd ist.

Ein solches Wechselspiel der Settings kann im Response-to-Intervention-Modell (RTI) verwirklicht werden. Mahlau (2018) hat die Wirksamkeit eines solchen RTI-Modells mit dem Fokus auf sprachauffällige Kinder nachweisen können. Hartmann et al. (2019) haben skizziert, wie dies im Schweizer Schulsystem aussehen könnte. Dieses bietet grundsätzlich gute Voraussetzungen, um ein solches RTI-Modell zu etablieren. So sind RLP, SHP und LOG gleichermaßen Teil des Schulsystems und haben z.T. sogar ihre Arbeitsorte im selben Schulhaus. Die Wege wären entsprechend kurz. Um intensive Zusammenarbeitsformen installieren zu können, ist u. a. ein Umdenken der beteiligten Fachkräfte notwendig: es müssen neue Aufgaben (u. a. Team-Teaching) im multiprofessionellen Team übernommen werden, die nicht ins klassische Aufgabenprofil gehören (Sallat & Siegmüller, 2016). Aber auch RLP müssen lernen, dass Autonomieparitätsmuster (Luder, 2021) zu überwinden und andere Personen nicht nur in ihren Unterricht hinein-, sondern sie auch entscheidend mitgestalten zu lassen. SHP werden intensiv im Bereich der Kooperation ausgebildet und können hier wertvolle Starthilfe bei der Entwicklung der multiprofessionellen Kooperation leisten.

Diese Entwicklung geschieht immer zum Wohl des Kindes, die Qualität der Maßnahmen muss erhalten oder gar gesteigert werden. Auch das Wohl der beteiligten Fachpersonen muss gefördert werden. Wie aus den Kommentaren herauszulesen war, arbeiten viele der beteiligten Fachpersonen unter hohem Zeitdruck bzw. unter großem zeitlichem Aufwand (die SHP und LOG „haben viel zu wenig Kapazität“ (RLP), „fehlen mir oft die Zeit und die Ressourcen“ (SHP), „Zeit, um in der Klasse zu arbeiten, habe ich nicht“ (LOG)). Und das, obwohl den verschiedenen Fachpersonen gemäß dem kantonalen Leitfaden 232 Stunden der Jahresarbeitszeit bzw. 5.9 Stunden pro Schulwoche für die „Mitarbeit und Zusammenarbeit“ offiziell zur Verfügung stehen (Erziehungsdirektion des Kantons Bern, 2019, S. 29). Gemäß Widmer-Wolf (2018) ist die unterrichtsbezogene Zusammenarbeit eine vielversprechende Lösung für diese Probleme. Neue Modelle der suM in inklusiven Settings können hier auch für Entlastung sorgen und bieten somit eine Chance für alle Beteiligten. Eine entsprechende Entwicklung wäre sehr zu begrüßen.

## 6 Limitationen

Die berichteten Ergebnisse müssen bewusst interpretiert werden. So sind ausschließlich Aussagen über die Quantität der suM möglich, nicht aber über die Qualität der Maßnahmen. Es ist klar, dass sich hinter den Begriffen sprachsensibler Unterricht, Sprachförderung und Sprachtherapie unterschiedlich spezifische Maßnahmen verbergen und die verantwortlichen Fachpersonen unterschiedlich stark im Bereich Sprache spezialisiert sind (Theisel, 2015). Ebenso vorsichtig muss man bei der Übertragung der Ergebnisse auf die vorherrschenden Bedingungen in Deutschland sein, da vor allem Schweizer SHP (als nicht auf Sprache spezialisierte Heilpädagog:innen) und Schweizer LOG (als Teil des pädagogisch-therapeutischen Schulteam) von ihren Pendanten in Deutschland zu unterscheiden sind.

Der Rücklauf der Fragebögen war vor allem bei den RLP und den SHP klein. Es wurde häufig zurückgemeldet, dass keine Kapazitäten für die Teilnahme an Studien vorhanden waren. Häufig wurde die Betreuung von aus der Ukraine geflüchteten Kindern und Jugendlichen bzw. der aktuell vorherrschende Fachkräftemangel in den Schulen als Gründe für die fehlenden Kapazitäten genannt. Dass der Rücklauf bei den LOG eher gut ist, liegt wahrscheinlich daran, dass diese persönlich kontaktiert wurden.

Die Stichprobe ist außerdem – basierend auf den Statistiken der Bildungs- und Kulturdirektion des Kantons Bern (2022) – nicht repräsentativ. So ist die angefallene Stichprobe z. B. weiblicher, arbeitet seltener in Vollzeit und deckt nicht alle Schulstufen gleichermaßen gut ab. Aus der Stichprobenbeschreibung geht außerdem hervor, dass sich die drei Berufsgruppen in vielen Variablen unterscheiden. Theoretisch sollten alle diese Variablen kontrolliert werden, um deren Einfluss ausschließen zu können. Das war hier angesichts der mit jeder weiteren Variable abnehmenden Testpower nicht umsetzbar. Auch wenn die Stichprobe nicht repräsentativ ist, können die gefundenen Unterschiede als grundsätzliche Verschiedenheit zwischen den Berufsgruppen betrachtet



werden, die noch einmal verdeutlichen, dass RLP, SHP und LOG unter sehr verschiedenen Bedingungen arbeiten. Positiv angemerkt werden kann, dass nicht nur motivierte und von der Thematik überzeugte Personen teilgenommen haben, sondern dass durchaus auch kritische Stimmen eingefangen werden konnten.

In den Analysen zu den verschiedenen Fachbereichen sind sehr viele „trifft nicht zu“-Antworten (anzukreuzen, falls das Fach in der jeweiligen Schulstufe nicht unterrichtet wird) zu verzeichnen, selbst bei den Unterrichtsfächern, die in allen Schulstufen unterrichtet werden und somit theoretisch auch von jeder befragten Person eingeschätzt werden könnten. Die Ergebnisse im entsprechenden Bereich müssen also mit Vorsicht betrachtet werden.

Und: Obwohl der Team-Teaching-Faktor immer wieder zu positiven Einflüssen in den Daten geführt hat, muss auch dies mit Vorsicht interpretiert werden: Sind Fachpersonen häufiger in den Klassenzimmern, weil sie in Teams arbeiten oder arbeiten sie häufiger in Teams, weil sie im Klassenzimmer sein wollen? Obwohl nicht geklärt werden kann, was Ursache und was Wirkung ist, scheint offensichtlich, dass multidisziplinäres Team-Teaching professionell entwickelt und gepflegt werden muss, da es sich hier um einen wesentlichen Teil der unterrichtsintegrierten Zusammenarbeit handelt.

## Literatur

- Ajzen, I. (2019). *Behavioral Interventions Based on the Theory of Planned Behavior*. Abgerufen von <http://www.people.umass.edu/aizen/pdf/tpb.intervention.pdf> [30.03.2023].
- Archibald, L. M. D. (2017). SLP-educator classroom collaboration: A review to inform reason-based practice. *Autism & Developmental Language Impairments*, 2, 1-17.
- Arndt, A.-K. & Werning, R. (2013). Unterrichtsbezogene Kooperation von Regelschullehrkräften und Lehrkräften für Sonderpädagogik. Ergebnisse eines qualitativen Forschungsprojektes. In R. Werning & A.-K. Arndt (Hrsg.), *Inklusion: Kooperation und Unterricht entwickeln* (S. 12–40). Bad Heilbrunn: Klinkhardt.
- Bauer, A., Glück, C. W., Jonas, K., Mayer, A., Sallat, S. & Stitzinger, U. (2022). Kernkompetenzen für das Studium der Pädagogik im sonderpädagogischen Förderschwerpunkt Sprache und Kommunikation. *Praxis Sprache*, 67(4), 229–233.
- Baumann, B., Henrich, C. & Studer, M. (2012). *Unterrichtsbezogene Kooperation zwischen Regellehrpersonen und Lehrkräften schulischer Heilpädagogik (IF) und Aspekte guten Unterrichts*, Interkantonale Hochschule für Heilpädagogik. Abgerufen von <https://edudoc.ch/record/104205> [30.03.2023]
- Baxter, S., Brookes, C., Bianchi, K., Rashid, K. & Hay, F. (2009). Speech and language therapists and teachers working together: Exploring the issues. *Child Language Teaching and Therapy*, 25(2), 215–234.
- Beck, A. R. & Dennis, M. (1997). Speech-Language Pathologists' and Teachers' Perceptions of Classroom-Based Interventions. *Language, Speech, and Hearing Services in Schools*, 28(2), 146–153.
- Bildungs- und Kulturdirektion des Kantons Bern (2022). *Lehrpersonen und Kosten der öffentlichen Volksschulen 2021*. Abgerufen von <https://www.bkd.be.ch/de/start/themen/statistik-bildung-und-kultur/daten-publikationen/lehrpersonen-und-finanzen.html> [30.03.2023]
- Blechs Schmidt, A., Besset, S., Läber, J., Reber, K. & Allemann, D. (2013). *Logopädie in Zeiten der Inklusion*. Vortrag an Schweizer Heilpädagogik-Kongress, Bern.
- Braun, W. (2009). Integrierte Sprachförderung: Abgrenzung, Umsetzung, Erfahrungen. *Schweizerische Zeitschrift für Heilpädagogik*, 15(1), 15–23.
- Brimo, D. & Huffman, H. E. (2023). A Survey of Speech-Language Pathologists' and Teachers' Perceptions of Collaborative Service Delivery. *Language, Speech, and Hearing Services in Schools*, 54(3), 873–887.
- Cirrin, F. M., Schooling, T. L., Nelson, N. W., Diehl, S. F., Flynn, P. E., Taskowski, M. et al. (2010). Evidence-Based Systematic Review: Effects of Different Service Delivery Models on Communication Outcomes for Elementary School-Age Children. *Language Speech and Hearing Services in Schools*, 41(3), 233–264.
- Deutsche Gesellschaft für Phoniatrie und Pädaudiologie. (2022). *Therapie von Sprachentwicklungsstörungen. Interdisziplinäre S3-Leitlinie*. Abgerufen von <https://register.awmf.org/de/leitlinien/detail/049-015> [03.04.2023]
- Deutscheschweizer Erziehungsdirektoren-Konferenz. (2016). *Lehrplan 21*. Abgerufen von [https://v-fe.lehrplan.ch/container/V\\_FE\\_Grundlagen.pdf](https://v-fe.lehrplan.ch/container/V_FE_Grundlagen.pdf) [30.03.2023]
- Drossel, K. (2015). *Motivationale Bedingungen von Lehrkooperation. Eine empirische Analyse der Zusammenarbeit im Projekt „Ganz In“*. Münster: Waxmann.
- Drossel, K. & Willems, A. S. (2014). Zum Zusammenhang von Formen der Lehrkooperation des Schulleitungshandelns und des Kooperationsklimas an Ganztagsgymnasien. In R. Strietholt & W. Bos (Hrsg.), *Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen* (S. 129–154). Münster: Waxmann.
- Ehren, B. J. (2000). Maintaining a Therapeutic Focus and Sharing Responsibility for Student Success: Keys to In-Classroom Speech-Language Services. *Language Speech and Hearing Services in Schools*, 31, 219–229.
- Elksnin, L. K. & Capilouto, G. J. (1994). Speech-Language Pathologists' Perceptions of Integrated Service Delivery in School Settings. *Language, speech, and hearing services in schools*, 25, 238–267.
- Erziehungsdirektion des Kantons Bern. (2019). *Integration und besondere Massnahmen in der Volksschule des Kantons Bern (IBEM). Leitfaden zur Umsetzung von Artikel 17 VSG für Lehrpersonen, Schulleitungen und Schulbehörden* (4. Ausgabe). Abgerufen von <https://www.akvb-unterricht.bkd.be.ch/de/start/sonderpaedagogische-massnahmen/einfache-sonderpaedagogische-massnahmen.html> [03.04.2023]
- Glück, C. W., Reber, K., Spreer, M. & Theisel, A. (2013). *Positionspapier. Kinder und Jugendliche mit Förderschwerpunkt Sprache und Kommunikation in inklusiven Bildungskontexten*, Deutsche Gesellschaft für Sprachheilpädagogik. Abgerufen von [https://www.dgs-ev.de/fileadmin/Standpunkte/dgs-Positionen\\_2013\\_Kinder\\_und\\_Jugendliche\\_mit\\_Foerderschwerpunkt\\_Sprache\\_und\\_Kommunikation\\_in\\_inklusive\\_Bildungskontexten.pdf](https://www.dgs-ev.de/fileadmin/Standpunkte/dgs-Positionen_2013_Kinder_und_Jugendliche_mit_Foerderschwerpunkt_Sprache_und_Kommunikation_in_inklusive_Bildungskontexten.pdf) [30.03.2023]
- Gräsel, C., Fußangel, K. & Pröbstel, C. (2006). Lehrkräfte zur Kooperation anregen – eine Aufgabe für Sisyphos? *Zeitschrift für Pädagogik*, 52(2), 205–219.

- Green, L., Chance, P. & Stockholm, M. (2019). Implementation and Perceptions of Classroom-Based Service Delivery: A Survey of Public School Clinicians. *Language, Speech, and Hearing Services in Schools*, 50(4), 656–672.
- Gröhhlich, C., Drossel, K. & Winkelset, D. (2015). Multiprofessionelle Kooperation in Ganztagsgymnasien: Umsetzung und Rahmenbedingungen. In H. Wendt & W. Bos (Hrsg.), *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In* (S. 178–200). Münster: Waxmann.
- Hartas, D. (2004). Teacher and speech-language therapist collaboration: being equal and achieving a common goal? *Child Language Teaching and Therapy*, 20(1), 33–54.
- Hartmann, E., Till, C. & Winkes, J. (2019). Kinder mit Sprachentwicklungsstörungen in der Regelschule. Gelingensbedingungen für erfolgreiche (Sprach-)Förderung und Kooperation zwischen beteiligten Fachpersonen. *Schweizerische Zeitschrift für Heilpädagogik*, 25(2), 46–52.
- Hernandez, S. J. (2013). Collaboration in Special Education: Its History, Evolution and Critical Factors Necessary for Successful Implementation. *US-China Education Review*, 3(6), 480–498.
- Hinz, A. (2004). Vom sonderpädagogischen Verständnis der Integration zum integrationspädagogischen Verständnis der Inklusion!? In I. Schnell (Hrsg.), *Inklusive Pädagogik* (S. 41–74). Bad Heilbrunn, Obb.: Klinkhardt.
- Huber, S. G. & Ahlgrimm, F. (2012). Was Lehrkräfte davon abhält zusammenzuarbeiten – Bedingungen für das Gelingen von Kooperation. In S. G. Huber (Hrsg.), *Jahrbuch Schulleitung 2012. Befunde und Impulse zu den Handlungsfeldern des Schulmanagements* (S. 1–14). Kronach: Carl Link.
- Kauschke, C., Lücke, C., Dohmen, A., Haid, A., Leitinger, C., Männel, C. et al. (2023). Delphi-Studie zur Definition und Terminologie von Sprachentwicklungsstörungen – eine interdisziplinäre Neubestimmung für den deutschsprachigen Raum. *Logos*, 31(1), 2–20.
- Kempe Preti, S. (2010). Logopädisches Angebot an integrativen Schulen. *Schweizerische Zeitschrift für Heilpädagogik*, 16(7-8), 6–12.
- Kolbe, F.-U. & Reh, S. (2008). Kooperation unter Pädagogen. In T. Coelen & H.-U. Otto (Hrsg.), *Grundbegriffe Ganztagsbildung. Das Handbuch* (S. 799–808). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Krosnick, J. A. & Presser, S. (2010). Question and Questionnaire Design. In P. V. Marsden & J. D. Wright (Hrsg.), *Handbook of survey research* (S. 263–312). Bingley, UK: Emerald.
- Luder, R. (2021). *Integrative Förderung in der Schweiz. Eine empirische Studie zur praktischen Umsetzung sonderpädagogischer Unterstützung und Förderung in integrativen Regelklassen in der Schweiz*. Habilitation. Universität Freiburg/CH, Freiburg/CH.
- Lüdtke, U. (2017). Inklusion als multiprofessionelle Aufgabenstellung der ‚Unterrichtsintegrierten Sprachtherapie und Logopädie‘ – Herausforderungen und Ressourcen. In M. Grohnfeldt (Hrsg.), *Kompendium der akademischen Sprachtherapie und Logopädie. Band 2: Interdisziplinäre Grundlagen* (S. 208–222). Stuttgart: Kohlhammer Verlag.
- Mahlau, K. (2016). *Zur Förderung von Kindern mit spezifischen Sprachentwicklungsstörungen nach dem Response-to-Intervention-Ansatz. Kontrollgruppenstudie zur sprachlichen, schulleistungsbezogenen und sozial-emotionalen Entwicklung in unterschiedlichen schulischen Settings*. Frankfurt a.M: Peter Lang GmbH Internationaler Verlag der Wissenschaften.
- Mahlau, K. (2018). *Kinder mit Sprachauffälligkeiten. Förderung in inklusiven Schulklassen*. Stuttgart: Verlag W. Kohlhammer.
- Mahlau, K. & Herse, S. (2017). *Sprechen, Spielen, Spaß – sprachauffällige Kinder in der Grundschule fördern*. München: Ernst Reinhardt Verlag.
- Pershey, M. G. & Rapking, C. I. (2003). A Survey of Collaborative Speech-Language Service Delivery Under Large Case-load Conditions in an Urban School District in the United States. *Journal of Speech-Language Pathology and Audiology*, 27(4), 211–220.
- Pfeiffer, D. L., Pavelko, S. L., Hahs-Vaughn, D. L. & Dudding, C. C. (2019). A National Survey of Speech-Language Pathologists’ Engagement in Interprofessional Collaborative Practice in Schools: Identifying Predictive Factors and Barriers to Implementation. *Language, Speech, and Hearing Services in Schools*, 50(4), 639–655.
- Prüfer, P. & Rexroth, M. (2005). *Kognitive Interviews, Zentrum für Umfragen, Methoden und Analysen -ZUMA-. GESIS-How-to: 15*. Abgerufen von <https://www.ssoar.info/ssoar/handle/document/20147> [30.03.2023]
- Pukas, D. (2008). *Lernmanagement. Einführung in Lern- und Arbeitstechniken* (3., aktualisierte Aufl.). Rinteln: Merkur-Verlag.
- Reber, K. (2012). Sprachheilpädagogik und Sprachtherapie in inklusiven Settings. Perspektiven der Vernetzung zwischen Sprachheilpädagogik, Sprachtherapie und Regelschule. *L.O.G.O.S. Interdisziplinär*, 20(4), 264–275.
- Reber, K. & Schönauer-Schneider, W. (2011). *Bausteine sprachheilpädagogischen Unterrichts* (2. Aufl.). München: Ernst Reinhardt Verlag.
- Reber, K. & Schönauer-Schneider, W. (2017). *Sprachförderung im inklusiven Unterricht. Praxistipps für Lehrkräfte*. München: Ernst Reinhardt Verlag.
- Rohrlack, C. (2009). Logistische und Ordinale Regression. In S. Albers, D. Klapper, U. Konrad, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (3., überarbeitete und erweiterte Auflage, S. 267–282). Wiesbaden: Gabler.
- Sallat, S. & Siegmüller, J. (2016). Interdisziplinäre Kooperation zwischen Institutionen und Professionen. In M. Grohnfeldt (Hrsg.), *Kompendium der akademischen Sprachtherapie und Logopädie. Band 1: Sprachtherapeutische Handlungskompetenzen* (S. 247–265). Stuttgart: W. Kohlhammer Verlag.
- Stommel, S., Hildebrandt, E., Senn, P. & Widmer, R. (2014). Einstellung von Schweizer Lehrkräften zu verschiedenen Formen der Kooperation. *Schulpädagogik heute*, 5(10), 1–15.
- Strand, S., Cadwallader, S. & Firth, D. (2011). *Using Statistical Regression Methods in Education Research*. Abgerufen von <https://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/index.html> [31.03.2023]
- Theisel, A. K. (2015). Unterrichten Sprachheilpädagogen anders? Eine vergleichende Selbsteinschätzung von Förderschullehrkräften „Sprache“ und Regelschullehrkräften zu Qualitätsmerkmalen des Unterrichts. *Empirische Sonderpädagogik*, 7, 320–340.
- Throneburg, R., Calvert, L. K., Sturm, J. J., Paramboulas, A. A. & Paul, P. J. (2000). *A Comparison of Service Delivery Models: Effects on Curricular Vocabulary Skills in the School Setting, Faculty Research and Creative Activity*. Abgerufen von [https://thekeep.eiu.edu/commdis\\_fac/5/](https://thekeep.eiu.edu/commdis_fac/5/)
- Urban, D. & Mayerl, J. (2011). *Regressionsanalyse: Theorie, Technik und Anwendung* (4., überarbeitete und erweiterte Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Vogt, F., Kunz Heim, D., Baez, C., Safi, N. & Zumwald, B. (2022). Kooperation in Klassenteams: Qualität und Wirkungen aus der Sicht von Klassenlehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 40(2), 213–229.
- White, J. L. (2013). Logistic Regression Model Effectiveness: Proportional Chance Criteria and Proportional Reduction in Error. *Journal of Contemporary Research in Education*, 2(1), 4–10.
- Widmer-Wolf, P. (2018). Kooperation in multiprofessionellen Teams an inklusiven Schulen. In T. Sturm & M. Wagner-Willi (Hrsg.), *Handbuch schulische Inklusion* (S. 299–313). Opladen: Verlag Barbara Budrich.

Woerfel, T. & Giesau, M. (2018). *Sprachsensibler Unterricht*, Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. Abgerufen von <https://www.mercator-institut-sprachfoerderung.de/de/themenportal/thema/sprachsensibler-unterricht/> [14.10.2023]

## Zu den Autor:innen

*Christoph Till* ist Bereichsleiter für die Fachwissenschaften und Dozent für Sprachheilpädagogik am Institut für Heilpädagogik der PHBern. Seine Forschungsinteressen liegen im Spracherwerb mehrsprachiger Kinder sowie bei der multiprofessionellen Zusammenarbeit von Regellehrpersonen, schulischen Heilpädagog:innen und Logopäd:innen.

*Jasmin Kolb* ist Hilfsassistentin am Institut für Forschung und Entwicklung an der PHBern und studiert an der Universität Freiburg/CH. Ihre Fachinteressen sind die pädagogische Psychologie sowie die inklusive Schulentwicklung, vor allem in der Zusammenarbeit von Lehrpersonen und anderen Fachkräften.

## Korrespondenzadresse

Dr. Christoph Till  
PHBern, Institut für Heilpädagogik  
Fabrikstrasse 8, CH-3012 Bern  
[christoph.till@phbern.ch](mailto:christoph.till@phbern.ch)



**NEU**

Anita Guggisberg,  
1. Auflage 2023,  
kartoniert: ISBN 978-3-8248-1322-3,  
64 Seiten, EUR 12,50 [D]  
E-Book: ISBN 978-38248-9870-1,  
EUR 8,50 [D]

## Presbyphagie – Schluckbeschwerden im Alter

Ein Ratgeber für Betroffene und Angehörige

Veränderungen des Schluckens und Schluckbeschwerden kommen im Alter häufig vor und werden oft nicht genügend beachtet. Schluckstörungen werden oftmals erst spät erkannt und therapiert. Die Auswirkungen einer Schluckstörung reichen von Gewichtsverlust, Mangelernährung, Flüssigkeitsmangel und Lungenentzündungen bis hin zu sozialem Rückzug und Depression.

Dieser Ratgeber weckt das Bewusstsein für Schluckbeschwerden und Schluckstörungen im Alter. Er informiert über die möglichen lebensbedrohlichen Folgen von Schluckstörungen und gibt alltagstaugliche Tipps, was Betroffene gegen Beschwerden unternehmen können.

In diesem Ratgeber erfahren Sie mehr über:

- das Schlucken beim gesunden Menschen
- die Veränderungen des Schluckens im Alter
- den Unterschied zwischen Schluckstörung und altersbedingten Veränderungen des Schluckaktes
- die Symptome, Ursachen und Folgen einer Schluckstörung
- die diagnostischen Verfahren von Schluckstörungen
- die Therapie von Schluckstörungen
- die Ernährung und Kostform
- die kompensatorischen Maßnahmen und externen Hilfsmittel



Tel.: +49 6126 9320-13 | Fax: +49 6126 9320-50  
bestellung@schulz-kirchner.de | [www.skvshop.de](http://www.skvshop.de)

Lieferung versandkostenfrei innerhalb Deutschlands





„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  
Ostermeierstr. 4  
D-30539 Hannover  
Telefon +49 30 661-6004  
Telefax +49 30 661-6024  
info@dgs-ev.de, www.dgs-ev.de, bestellungen.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 Kontaktdaten 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 Heidelberg <a href="mailto:dgs@theisel.de">dgs@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	Claus Huber, Sabine Krämer, Marc Rauber <a href="mailto:huber@dgs-ev.de">huber@dgs-ev.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	Ellen Bastians <a href="mailto:bastians@dgs-rheinland.de">bastians@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	Anna Sickert <a href="mailto:dgs.sachsen-anhalt@t-online.de">dgs.sachsen-anhalt@t-online.de</a>
Schleswig-Holstein	Regine Vofß-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

11. Jahrgang 2023 | ISSN 2196-6818

#### Herausgeberin

Deutsche Gesellschaft für Sprachheilpädagogik e.V. (dgs)  
Ostermeierstr. 4 | D-30539 Hannover  
Telefon +49 30 661-6004  
Telefax +49 30 661-6024  
[info@dgs-ev.de](mailto:info@dgs-ev.de) | [www.dgs-ev.de](http://www.dgs-ev.de) | [bestellungen.dgs-ev.de](mailto:bestellungen.dgs-ev.de)

#### Redaktion

- Dr. Melanie Eberhardt-Juchem, Bonn
  - Hiltrud von Kannen, Karlstadt
  - Prof. Dr. Susanne van Minnen, Grünberg
  - Irina Ruppert-Guglhör, Rosenheim
  - Prof. Dr. Wilma Schönauer-Schneider, Wetzstetten
- Downloadredaktion: Kerstin Rimpau, München  
Kontakt: [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 eine Wissenschaft interessierte Leserschaft 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 erwoogen 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)  
Ostermeierstr. 4 | D-30539 Hannover

#### Wissenschaftlicher Beirat

Prof. Dr. Margit Berg, Ludwigsburg; Prof. Dr. Anja Blechschmidt, Basel;  
Prof. Dr. Wolfgang Braun, Zürich; Prof. Dr. Solveig Chilla, Flensburg;  
Dr. Jürgen Cholewa, Heidelberg; Prof. Dr. Kirsten Diehl, 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; Prof. Dr. phil. Vanessa Hoffmann, Hamburg;  
V. Prof. Dr. Kristina Jonas, Köln; Prof. Dr. Tanja Jungmann, Oldenburg;  
Prof. Dr. Simone Kannengieser, Basel; Prof. Dr. Ulrich von Knebel, Hamburg;  
Prof. Dr. Anette Kracht, Landau; Jun. Prof. Dr. Ulla Licandro, Oldenburg;  
Prof. Dr. Carina Lüke, Würzburg; Hannah Manowita, Gießen;  
Prof. Dr. Kathrin Mahlau, Greifswald; Dr. Dana-Kristin Gaigulo, München;  
Prof. Dr. Andreas Mayer, München; Prof. Dr. Christiane Miosga, Hannover;  
Prof. Dr. Sandra Neumann, Erfurt; Dr. Antje Orgassa, Nijmegen;  
Prof. Dr. Claudia Osburg, Hamburg; Dr. Stephanie Riehemann, Köln;  
Prof. Dr. Stephan Sallat, Halle/Saale; Dr. Christoph Schiefele, Ludwigsburg;  
Marc Schmidt, Strassen; Prof. Dr. Christof Schreiber, Gießen;  
Prof. Dr. Markus Spreer, Leipzig; Prof. Dr. Anja Starke, Bremen;  
Dr. Ulrich Stitzinger, Hannover; PD Dr. Katja Subellok, Dortmund;  
PD Dr. Anja Theisel, Heidelberg; Prof. Dr. Tanja Ulrich, Duisburg/Essen;  
Dr. Julia Winkes, Fribourg

#### Datenbanken

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

#### Erscheinungsweise

2 Ausgaben 2022: 15. Mai, 15. November

#### Satz und Gestaltung

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

Vertretungsberechtigte Geschäftsführer:

Dr. Ullrich Schulz-Kirchner, Martina Schulz-Kirchner