









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 nonresponder, 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: 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. 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. Duale Diskrepanz-Methode: Kombination von Endstatus- und Zuwachs-Methoden. Als NR gelten Interventionskinder, die im Lernfortschritt und in der Endleistung diskrepant sind zur Referenzgruppe. |
| Referenzgruppe | Mögliche Referenzgruppen zur Beurteilung der RTI sind: a) Eine normative Referenzgruppe (normative sample) repräsentiert den gesamten Bereich von Fähigkeiten in normierten Tests; b) Eine beschränkt normative Referenzgruppe (limited norm sample) umfasst den Leistungsbereich der Kinder, die an derselben Intervention teilgenommen haben (Interventionsstichprobe). |
| 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.Bo.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 R2) 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 WI-GROT-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 Leseprobleme, 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 Fördereinheiten 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 *Lese-flüssigkeit* (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 T1 (RW AVS): M (SD) | 6.59 (3.82) |
| Leseflüssigkeit T1 (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 Leseflüssigkeit im LDL (T-Wert: 36.14; Walter, 2009). Gemäß den LDL-Werten für den Posttest konnte sich die Stichprobe in der Leseflüssigkeit 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 Leseflüssigkeit 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 r₁₁ = .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)
$$Diff_{krit} = \mathbf{z}_{krit} \times s_e$$

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

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

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: | nbination 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 α 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 α .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, diktierte 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 großenteils 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 | | | | | |
| Normbasiert | | | | | | | |
| 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 | | | | | | |
| Stichprobenbasiert | | | | | | | |
| 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.

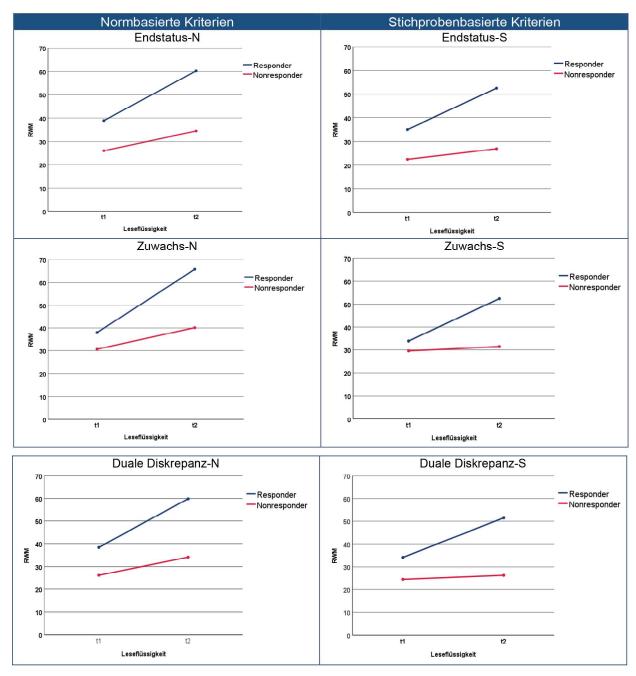


Abb. 1: Prä- und Posttestwerte für Leseflüssigkeit (RWM) der NR- und R-Gruppen bei sechs Kriterien im Vergleich

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

| | NR-Kriterien | | | | | | | | | | | |
|------------------------------------|-------------------|-------|-----------------|-------|-----------------------------|-------|-------------------------|-------|-----------------------|------|-----------------------------------|-------|
| Variablen/ Stichprobengröße (N) | 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) | 0.77 | ₹.001 | 0.93 | ₹.001 | 0.79 | ₹.001 | 0.79 | ₹.001 | 0.95 | .063 | 0.87 | ₹.001 |
| Leseverstehen T1 (130) | 0.78 | ₹.001 | 0.80 | ₹.001 | 0.78 | ₹.001 | 0.79 | .008 | 0.84 | .023 | 0.79 | .021 |
| Lesemotivation T1 (129) | 0.96 | .879 | 1.11 | .691 | 0.89 | .652 | 0.97 | .935 | 0.72 | .277 | 0.78 | .497 |
| RechtschreibenT1 (130) | 0.85 | ₹.001 | 0.83 | ₹.001 | 0.84 | ₹.001 | 0.83 | .001 | 0.87 | .024 | 0.83 | .015 |
| HISEI (128) | 0.97 | .029 | 0.97 | .030 | 0.98 | .033 | 0.99 | .367 | 0.99 | .397 | 0.98 | .305 |
| Trainingsbedingunga (130) | 1.38 | .361 | 2.39 | .023 | 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 | 0.47 | .023 | 0.52 | .041 | 0.46 | .036 |
| 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 *Ns* bei Tempo und Lautstärke wegen fehlender gemeinsamer Lesephasen; *OR* = Odds Ratio; ^a Referenzkategorie = Coach; ^b 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 ^a | .001 | 128 | 103 | .056 | | |
| 6 Synchronität Tempo | .235* | .084 | .331** | .100 | 002 | |

Anmerkung: **p < .o1 (zweiseitig getestet); *p < .o5 (zweiseitig getestet); *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

| | NR-Kriterien | | | | | | | | | | | |
|---------------------------------|-------------------|-------|-----------------|------|-----------------------------|-------|-------------------------|------|-----------------------|------|-----------------------------------|------|
| Variablen (N = 102) | Endstatus Norm | | Zuwachs Norm | | Duale Diskrepanz Norm | | Endstatus Stichprobe | | Zuwachs Stichprobe | | Duale Diskrepanz Stichprobe | |
| | OR | p | OR | р | OR | р | OR | р | OR | р | OR | р |
| Leseflüssigkeit T1 | 0.73 | <.001 | 0.92 | .005 | 0.78 | <.001 | 0.75 | .001 | 0.98 | .532 | 0.89 | .035 |
| 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 | 0.94 | .009 | 0.96 | .027 | 0.95 | .021 | 0.98 | .525 | 0.99 | .528 | 0.99 | .788 |
| Trainingsbedingung ^a | 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 R² 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; ^a 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 Spannbreite überlappt sich großenteils 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 verlaufsdiagnostische 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/eerj.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 [10.9.13]
- Nachtigall, C., & Suhl, U. (2005). Evaluation intraindividueller 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, I., 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, Ř. 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, Á., 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: Verlaufsdiagnostik 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 Bildungskooperation 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