



PROSAIOS – Konzeptualisierung eines Bewertungsinstruments für Wirksamkeits- und Wirkungsstudien in der schulischen Sprachförderung*

PROSAIOS – Conceptualisation of an Assessment Tool for Efficacy and Effectiveness Studies for Language Support Methods in Schools

Lea Wiehe^{1**}, Romy Räling^{2,3**} und Maja Stegenwallner-Schütz^{2,4}

1 Universität Potsdam, 2 Humboldt-Universität zu Berlin, 3 Freie Universität Berlin, 4 Universität Koblenz,

** gemeinsame Erstautor:innenschaft

Zusammenfassung

Hintergrund: Pädagogische Entscheidungen von Förderkräften über Interventionsmethoden in Lehr-Lernsettings bei Schüler:innen mit sprachlich-kommunikativem Unterstützungsbedarf sollten auf einer *evidenzbasierten Praxis* (EbP) basieren. Dabei werden externe, interne sowie soziale Evidenz berücksichtigt. Um zu entscheiden, ob ein Wirkvorteil einer Fördermaßnahme auch unter praxisnahen Bedingungen zu erwarten ist, ist die differenzierte Bewertung von Interventionsstudien hinsichtlich der Evaluationsdimensionen *Wirksamkeit* und *Wirkung* erforderlich.

Ziele: In der Synthese empirischer Evidenz stellt sich die Frage, wie Interventionsstudiendesigns von Sprachfördermaßnahmen den Evaluationsdimensionen zugeordnet werden können, um gültige Schlussfolgerungen bezüglich des Geltungsbereichs und der Anwendbarkeit der Erkenntnisse zu ermöglichen. Es wird ein Instrument konzipiert, das die Einordnung von Evaluationsergebnissen hinsichtlich der Evidenzdimensionen ermöglicht.

Methode: Durch eine systematische Literaturrecherche wurden fünf Wirksamkeits-Wirkungsindikatoren sowie verschiedene Bewertungsstrategien identifiziert.

Ergebnisse: Fünf Indikatoren zur Wirksamkeits-Wirkungsdifferenzierung werden für die schulische Sprachförderung adaptiert und in PROSAIOS (als Akronym für die identifizierten Indikatoren: **P**rovider, **S**ample, **I**ntervention, **O**utcome und **S**etting) mit einem spektralen Verständnis zusammengefasst. Die Anwendung wird anhand einer Beispielstudie verdeutlicht.

Diskussion: PROSAIOS ermöglicht die Differenzierung hinsichtlich der Wirksamkeits- und Wirkungsdimension von Interventionsstudien zu Sprachfördermaßnahmen im schulischen Kontext. Es ermöglicht dadurch die Analyse der Generalisierbarkeit von positiv evaluierten Sprachfördermaßnahmen in pädagogischen Kontexten.

Schlüsselwörter

Evidenzbasierte Pädagogik; Evidenzbasierte Praxis; EbP; Evaluation; externe Evidenz; Wirksamkeit; Wirkung

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

Abstract

Background: Decisions by teachers about intervention methods for students with speech and language support needs should be based on evidence-based practice (EBP). EBP takes into account external, internal and social evidence. Distinguishing intervention studies with regard to the evaluation dimensions of efficacy and effectiveness is crucial for evaluating whether an intervention will be effective under real-life practice.

Objectives: In the synthesis of empirical evidence, the question arises as to how study designs of language interventions can be assigned to the evaluation dimensions. An instrument is designed that enables the classification of intervention studies with regard to the evidence dimensions.

Methods: A systematic literature review identified five indicators for efficacy and effectiveness as well as different assessment strategies.

Results: Five indicators for differentiating between effectiveness and efficacy are adapted for school-based language interventions and summarized in PROSAIOS (acronym for the identified indicators: provider, sample, intervention, outcome and setting) with a spectral understanding. The application is illustrated using an example study.

Discussion: PROSAIOS enables differentiation between the efficacy and effectiveness dimension of studies on language interventions. It thus enables the analysis of the generalizability of positively evaluated language interventions in pedagogical contexts.

Keywords

Evidence-based pedagogy; evidence-based practice; EBP; evaluation; external evidence; effectiveness; efficacy

1 Einleitung

Pädagogische Entscheidungen zur Gestaltung, Planung und Organisation von Lehr-Lernprozessen erfordern eine kritische Auseinandersetzung mit wissenschaftlichen Erkenntnissen sowie den Transfer in die pädagogische Praxis (Kultusministerkonferenz [KMK], 2004). Dies gilt ebenso für Lehrpersonen, die sich mit der schulischen Förderung sprachlicher und kommunikativer Fähigkeiten befassen (Blumenthal & Mahlau, 2015; Rauch & Hartig, 2018). Die evidenzbasierte Pädagogik (EbP) – ein Paradigma für die pädagogische Entscheidungsfindung – stützt sich u. a. auf die sogenannte externe Evidenz. Diese umfasst empirische Studien, die nach wissenschaftlichen Standards durchgeführt und ausgewertet wurden. Die externe Evidenz ergänzt die interne Evidenz, also die Expertise der Fachkräfte, und die soziale Evidenz, also die Perspektive der Schüler:innen. Dieser Artikel konzentriert sich auf die Dimensionen externer Evidenz von schulischen Sprachfördermaßnahmen, insbesondere auf die Konzepte Wirksamkeit und Wirkung. Die Unterscheidung dieser beiden Evidenzdimensionen spielt sowohl für die Einordnung wissenschaftlicher Studien als auch für den Transfer ihrer Ergebnisse in die pädagogische Sprachförderpraxis eine zentrale Rolle, da sie eine differenzierte Bewertung des Geltungsbereichs der Erkenntnisse ermöglicht. Darüber hinaus erfordert die Auseinandersetzung mit dem Thema der Evidenzbasierung aus der Forschungsperspektive eine klare Trennung dieser beiden Evidenzdimensionen. Im Vorfeld eines systematischen Reviews zur Untersuchung spezifischer Sprachfördermethoden im Unterricht (vgl. Wiehe, Råling & Stegenwallner-Schütz, 2024) zeigt sich, dass bisher veröffentlichte Instrumente aus dem medizinischen Kontext zur Unterscheidung der Evidenzdimensionen *Wirksamkeit* und *Wirkung* nicht uneingeschränkt auf Studien zu Sprachfördermaßnahmen übertragbar sind. Aus diesem Grund wurde das Bewertungsinstrument PROSAIOS konzipiert (Akronym für die Bewertungsindikatoren: Provider, Sample, Intervention, Outcome und Setting). PROSAIOS ermöglicht es, *Wirksamkeit* und *Wirkung* differenziert zu erfassen und trägt somit zur Bestimmung der Gültigkeit der wissenschaftlichen Erkenntnisse und deren Transfermöglichkeiten in die pädagogische Sprachförderpraxis bei. Im Folgenden werden nun die Konzepte der evidenzbasierten Praxis genauer erläutert und auf die schulische Sprachförderpraxis übertragen, sowie die Konzeptualisierung des PROSAIOS-Instrumentes vorgestellt und an einem Beispiel erläutert.

2 Evidenzbasierte Pädagogik und die externe Evidenz

Evidenzbasierte Pädagogik

Die evidenzbasierte Pädagogik überführt ein ursprünglich für die Medizin entwickeltes Paradigma für Handlungsentscheidungen in die Pädagogik (Blumenthal & Mahlau, 2015; Davies, 1999). Dieses Handlungsparadigma beruht auf der Prämisse, dass eine qualitativ hochwertige medizinisch-therapeutische Versorgung durch die systematische Orientierung an wissenschaftlichen Belegen erreicht werden kann (Sackett et al., 1996) und bietet somit auch einen Orientierungsrahmen für pädagogische Entscheidungen (Egert et al., 2022). Die Notwendigkeit einer evidenzbasierten Praxis (EbP) in der Pädagogik wird mit der Absicht begründet, pädagogische Handlungsentscheidungen von unreflektierten oder tradierten Handlungspraktiken oder individuellen Überzeugungen zu lösen (Kuhr & Kulawiak, 2018). Da jedoch die Komplexität pädagogischer Handlungen eine ausschließlich empirie-basierte Entscheidungsfindung nicht zulässt, wird im Sinne einer EbP nach Dollaghan (2007) eine triangulative Entscheidungsfindung in Lehr-Lernsettings präferiert. Diese berücksichtigt die folgenden drei Informationsquellen: die interne, die soziale und die externe Evidenz (siehe Abb. 1). Unter der internen Evidenz werden die individuellen Erkenntnisse und Erfahrungen von Fachkräften aus der pädagogischen Praxis zusammengefasst (Blumenthal & Mahlau, 2015; Kuhr & Kulawiak, 2018). Im Fall der sprachförderpädagogischen Betrachtung entspräche dies der Sprachförderexpertise der jeweiligen Fachkraft. Auf dieser Ebene können auch individuelle Wirksamkeitsüberzeugungen eine Rolle spielen. Die soziale Evidenz, die als zweite Informationsquelle der EbP herangezogen werden sollte, berücksichtigt die Präferenzen der Schüler:innen (Blumenthal & Mahlau, 2015; Kuhr & Kulawiak, 2018). Im Rahmen der Sprachförderpädagogik können die individuellen Bedürfnisse der Schüler:innen einbezogen werden und Komorbiditäten, potenzieller Leidensdruck oder das Selbstkonzept Berücksichtigung finden. Im Rahmen der dritten Informationsquelle der EbP, der externen Evidenz, werden wissenschaftliche Nachweise über die Wirksamkeit oder Wirkung von pädagogischen Maßnahmen berücksichtigt (Blumenthal & Mahlau, 2015; Kuhr & Kulawiak, 2018). Im Fall der schulischen Sprachfördermaßnahmen sind dies die Ergebnisse von Evaluationsstudien zu einzelnen Fördermethoden, -konzepten oder -programmen und ihre Übertragbarkeit auf die schulische Sprachförderpraxis, beispielsweise, ob sie sich auf die unterrichtsintegrierte Förderung übertragen lassen.

Problemstellung in der EbP

Ein wesentliches Hindernis für die Umsetzung evidenzbasierter Pädagogik liegt zunächst in der geringen Anzahl empirischer Studien, die externe Evidenz für Sprachfördermaßnahmen im schulischen Lehr-Lern-Setting liefern. Insbesondere fehlen Studien, die über schriftsprachliche Fähigkeiten hinausgehen und beispielsweise auch lautsprachliche Fähigkeiten einbeziehen (z. B. Smith-Lock et al., 2013). Auch gestaltet sich der Transfer wissenschaftlicher Evaluationsergebnisse in die Praxis schwierig (Souvignier & Hasselhorn, 2023). Hinzu kommt, dass sich für Sprachfördermaßnahmen in der schulischen Anwendung kaum synthetisierte Befunde in Form von systematischen Überblicksarbeiten über Interventionsstudienbefunde finden (z. B. in Form von Systematic Reviews, vergleichenden Meta-Analysen). Systematische Zusammenfassungen bieten u. a. den Vorteil, dass Verzerrungen reduziert werden und eine höhere statistische Sicherheit der Befunde besteht. Fachkräfte können durch sie einen schnelleren Überblick über Forschungsstände bekommen, wobei Einzelevidenzen nicht eigenständig verglichen sowie Schlussfolgerungen getroffen werden müssen. Zusätzlich wird die Diskussion über die EbP durch konzeptionelle Unklarheiten in der Differenzierung externer Evidenzdimensionen erschwert, insbesondere hinsichtlich der Abgrenzung von Wirksamkeit und Wirkung sowie der Triangulierung der drei Informationsquellen. Genau diese definitorische Verschwommenheit trägt u. a. zum erschwerten Transfer von externer Evidenzen in die Praxis bei. In einem ersten Schritt sollen daher nachfolgend die Evidenzdimensionen definitorisch vorgestellt werden.

Externe Evidenz und ihre Evidenzdimensionen.

Streiner (2002) beschreibt zwei Dimensionen von externer Evidenz (Wirksamkeit und Wirkung), wobei diese beiden Dimensionen durch eine dritte Dimension, die Effizienz, ergänzt werden können (vgl. Beushausen & Grötzbach, 2018). Alle drei Evidenzdimensionen weisen eine hohe Nützlichkeit auf, erfüllen jedoch unterschiedliche Funktionen:

- (1) *Wirksamkeitsstudien* (engl. *efficacy studies* oder *explanatory trials*) liefern empirische Belege für die Wirksamkeit spezifischer, trennbarer Einflussgrößen unter idealisierten, klinisch kontrollierten experimentellen Bedingungen (Eisele-Metzger & Breuer, 2021). Wirksamkeitsstudien zeichnen sich daher durch eine hohe interne Validität aus und lassen in Abhängigkeit der Studiendurchführung, der Auswertung und der Berichterstattung Aussagen über die Vertrauenswahrscheinlichkeit der Studieneffekte zu. Die Untersuchung einer Fördermethode unter Idealbedingungen maximiert die Wahrscheinlichkeit einen Wirkvorteil zu beobachten, wenn dieser existiert (Singal et al., 2014). In der Konsequenz ist jedoch die Praxisnähe von Wirksamkeitsstudien reduziert.
- (2) *Wirkungsstudien* (engl. *effectiveness studies* oder *pragmatic trials*) finden unter alltäglichen Anwendungsbedingungen statt (Eisele-Metzger & Breuer, 2021) und weisen durch ihre geringere Standardisierung der Studienbedingungen eine höhere Praxisnähe als Wirksamkeitsstudien auf. Wirkungsstudien zeichnen sich durch eine hohe externe Validität aus, die eine bessere Generalisierbarkeit der Ergebnisse auf die gängige Behandlungs- oder Förderpraxis erlaubt. Wirkungsstudien sollten auf Wirksamkeitsstudien folgen: Nach dem Nachweis eines potenziellen Fördereffekts unter Idealbedingungen sollte also ein Effekt auch unter Realbedingungen überprüft werden (Streiner, 2002).
- (3) Bei *Effizienzstudien* (engl. *efficiency studies*) liegt der Fokus auf einer Kosten-Nutzen-Abwägung (Beushausen & Grötzbach, 2018). Hier werden Behandlungs- oder Förderfortschritte in den ökonomischen Kontext, wie investierte Zeit oder Personalressourcen, gestellt. Da in diesem Artikel der Fokus auf die Differenzierung von Wirksamkeits- und Wirkungsstudien gelegt wird, wird auf Effizienzstudien in diesem Artikel nicht weiter eingegangen.

Unterscheidung von Wirksamkeits- und Wirkungsstudien.

Für Fachkräfte in der schulischen (Sprachförder-)Pädagogik, Administrator:innen und Bildungspolitiker:innen ist die Unterscheidung zwischen Wirksamkeit und Wirkung von Fördermethoden besonders relevant. Ihre Schlussfolgerungen zur externen Evidenz sollten berücksichtigen, ob ein Wirkvorteil möglicherweise ausschließlich unter Idealbedingungen oder auch unter praxisnahen Bedingungen erbracht wurde (Flay et al., 2005). Neben der Relevanz für Fachkräfte in der schulischen Sprachförderpraxis ist auch für Forscher:innen eine klare Differenzierung zwischen *Wirksamkeit* und *Wirkung* entscheidend. Studien sollten mit einem Bewusstsein für die Art und den Zweck der jeweiligen Evaluationsdimension geplant und durchgeführt werden, um dies in der Ergebnisinterpretation und Wissenschaftskommunikation angemessen darzulegen. Wie Streiner (2002) betont, kann keine einzelne Studie beide Evaluationsdimensionen und die damit zusammenhängenden Ansprüche erfüllen. Daher müssen Forscher:innen bewusst entscheiden, an welcher Stelle des Spektrums zwischen Wirksamkeit und Wirkung ihre Untersuchung verortet wird. Aktuell werden die Begriffe Wirksamkeit und Wirkung in veröffentlichten Evaluationsstudien häufig unscharf oder synonym verwendet (Lawrence et al., 2023). Dies erschwert eine präzise Differenzierung der Evidenzdimensionen externer Evidenz und damit die Übertragung von Studienergebnissen in die Praxis.

Eine Unterscheidung der Evidenzdimensionen erfolgt u. a. über die Betrachtung der Studienziele und der Fragestellung, des Studiensettings und/oder des Studiendesigns (Porzolt et al., 2015), d. h. auf Basis der Kriterien externer Validität. Einige Autor:innen nehmen an, dass die Unterscheidung zwischen Wirksamkeits- und Wirkungsstudien spektral ist (Gartlehner et al., 2006). Im Rahmen dieser Annahme wird davon ausgegangen, dass die Unterscheidung in binäre Kategorien angesichts der Vielschichtigkeit der Studiendesigns nicht zielführend ist. Stattdessen werden Studien auf einem Wirksamkeit-Wirkungsspektrum positioniert (Singal et al., 2014; Wieland et al., 2017) und mithilfe verschiedener Indikatoren bestimmt (Abb. 1).

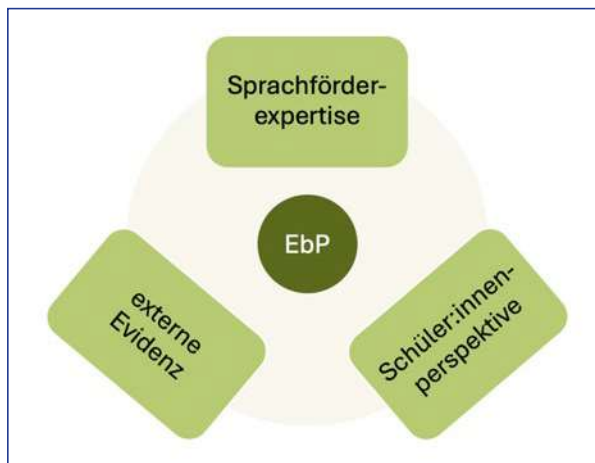


Abb. 1: Die drei Informationsquellen der EbP in der schulischen Sprachförderpraxis (in Anlehnung an Beushausen & Grötzbach, 2018, S. 19; Blumenthal & Mahlau, 2015, S. 409).

Nachdem nun die Begriffe *EvP*, *externe Evidenz* und spezifisch die Evidenzdimensionen *Wirksamkeit* und *Wirkung* definiert wurden und die Relevanz der Differenzierung der beiden Dimensionen dargelegt werden konnte, wird nachfolgend die Konzeptualisierung für ein Bewertungsinstrument für Sprachfördermaßnahmen vorgestellt.

3 Zielsetzung und Fragestellung

Dieser Beitrag verfolgt zum einen das Ziel, über die Einordnung externer Evidenz zu informieren und die Notwendigkeit und Funktion der Evidenzdimensionen *Wirksamkeit* und *Wirkung* sowie deren Differenzierung aufzuzeigen. Zum anderen wird das Ziel verfolgt, ein Bewertungsinstrument zur Unterscheidung der Evidenzdimensionen *Wirksamkeit* und *Wirkung* für Studien zu Sprachfördermaßnahmen zu konzipieren. In der Synthese empirischer Evidenz stellt sich die Frage, wie Studiendesigns zu Sprachförderinterventionen den Evidenzdimensionen bestmöglich zugeordnet werden können, um gültige Schlussfolgerungen bezüglich des Grades an Evidenzbasierung zu ermöglichen.

Verfügbare Instrumente und Entscheidungshilfen zur Unterscheidung der Evidenzdimensionen stammen überwiegend aus dem medizinischen Kontext und sind daher nicht vollständig auf pädagogische Studien übertragbar. Die Spezifika von (sprachförder-) pädagogischen Kontexten an Schulen finden bisher keine Berücksichtigung. Zum Zwecke der Konzeptualisierung eines Bewertungsinstruments für Studien im Sprachförderkontext an Schulen werden auf Basis des aktuellen Forschungsstandes etablierte Instrumente systematisch analysiert, die bereits Indikatoren zur Unterscheidung von Evidenzdimensionen identifiziert haben. Auf dieser Grundlage erfolgt eine gezielte Adaption der Indikatoren, um sie für den spezifischen Kontext der Sprachförderung zu operationalisieren.

Das Instrument soll für die folgenden Zwecke konzipiert werden:

1. Für die Forschung soll das Instrument dazu dienen, in Synthesen, wie systematischen Reviews, Zuordnungen von Studien zu den Evaluationsdimensionen *Wirksamkeit* und *Wirkung* vorzunehmen (vgl. Wiehe, Råling & Stegenwallner-Schütz, 2024) und somit Aussagen darüber treffen zu können, ob Sprachfördermethoden sowohl unter Idealbedingungen als auch Realbedingungen überprüft wurden.
2. Gleichfalls kann das Instrument Forscher:innen als eine Planungshilfe für die Konzeptionierung von Studiendesigns für schulische Sprachförderevaluationen dienen (in Anlehnung an PRECIS(-2); Loudon et al., 2015; Thorpe et al., 2009)
3. Das Instrument soll zudem Fachkräfte in der schulischen Praxis unterstützen, das Ausmaß der Generalisierbarkeit von Studienergebnissen anhand der Einordnung in die Evidenzdimensionen der externen Evidenzen zu erfassen.

4 Methodik

Um das Ziel zu verfolgen, ein Bewertungsinstrument (PROSAIOS) für die Differenzierung von Wirksamkeits- und Wirkungsstudien in pädagogischen Settings zu konzipieren, wurden zunächst wissenschaftliche Veröffentlichungen, in deren Rahmen eine Differenzierung vorgenommen wird, sowie etablierte Instrumente oder Indikatorenkataloge mit einer systematischen Literaturrecherche identifiziert. Dabei wurden Veröffentlichungen aus den medizinischen, pädagogischen und multidisziplinären Bereichen herangezogen, um den aktuellen Forschungsstand zu erfassen. Auf Grundlage der identifizierten Veröffentlichungen wird anschließend ein bereichsspezifisches Instrument für die Unterscheidung von Wirksamkeits- und Wirkungsstudien für den (sprachförder-)pädagogischen Bereich konzipiert.

4.1 Identifizierte Veröffentlichungen

Im Rahmen einer systematischen Literaturrecherche konnten insgesamt neun Publikationen (Instrumente, Indikatorenkataloge bzw. empirische Studien zu Indikatorenanwendungen), die den Anspruch haben eine Differenzierung der Evidenzdimensionen *Wirksamkeit* und *Wirkung* vorzunehmen, identifiziert werden (siehe Tab. 1; Alphs & Bossie, 2016; Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Loudon et al., 2015; Singal et al., 2014; Thorpe et al., 2009; Wieland et al., 2017). Mit einer systematischen Literaturrecherche in englischer Sprache in *Web of Science* (Suchstring: (ALL=(efficacy) OR ALL=(pragmatic) OR ALL=(pragmatic trail)) AND (ALL=(efficacy) OR ALL=(explanatory trail) OR ALL=(explanatory)) AND (ALL=(standards) OR ALL=(difference) OR ALL=(evaluation) OR ALL=(criteria) OR ALL=(framework) OR ALL=(tool) OR ALL=(rating) OR ALL=(review))) wurden sieben Publikationen gefunden (Alphs & Bossie, 2016; Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Singal et al., 2014; Thorpe et al., 2009; Wieland et al., 2017). Eine Publikation wurde durch eine ergänzende Handsuche identifiziert (Loudon et al., 2015).

Die fachlichen Kontexte der neun Studien sind divers (siehe Tab. 1). Eine der neun Veröffentlichungen entstand im pädagogischen (Lake et al., 2020), sechs Veröffentlichungen im medizinischen (Gartlehner et al., 2006; Loudon et al., 2015; Singal et al., 2014; Thorpe et al., 2009; Wieland et al., 2017) und zwei weitere Veröffentlichungen in einem multidisziplinären Kontext (Flay et al., 2005; Gottfredson et al., 2015).

Neben den Fachdisziplinen sind auch die Zwecke, zu denen die Indikatoren zur Wirksamkeits-Wirkungs-Differenzierung aufgestellt wurden, divers (siehe Tab. 1). Drei Veröffentlichungen sind konkrete Instrumententwicklungen (Alphs & Bossie, 2016; Gartlehner et al., 2006; Wieland et al., 2017), vier Veröffentlichungen Indikatorenkataloge (Flay et al., 2005; Gottfredson et al., 2015; Loudon et al., 2015; Singal et al., 2014; Thorpe et al., 2009), die als Standards verstanden werden können, sowie eine Anwendungsstudie (Lake et al., 2020).

Wird der Anwendungszeitpunkt betrachtet, kann festgestellt werden, dass sieben der neun Veröffentlichungen für die nachträgliche Einordnung von Studien bzgl. der Differenzierung von Wirksamkeit und Wirkung intendiert sind (Alphs & Bossie, 2016; Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Singal et al., 2014; Wieland et al., 2017) und zwei für die Studienplanung (PRECIS(-2): Loudon et al., 2015; Thorpe et al., 2009). Die nachträgliche Kategorisierung von Studien ist somit mithilfe von PRECIS(-2) nicht intendiert. Das Instrument wird dennoch häufiger im Kontext der Differenzierung des Wirksamkeit-Wirkungsspektrums diskutiert (Zwarenstein et al., 2017) und wurde daher in die hier vorliegende Literaturanalyse inkludiert.

Die Veröffentlichungen beziehen sich teilweise inhaltlich aufeinander (siehe Tab. 1). Zum einen finden sich konkrete Weiterentwicklungen von Instrumenten oder Standards (PRECIS-Familie: Alphs & Bossie, 2016; Loudon et al., 2015; Thorpe et al., 2009; Standards der Society of Prevention Research: Flay et al., 2005; Gottfredson et al., 2015), zum anderen referieren einige Autor:innen auf Veröffentlichungen anderer und begründen somit die Auswahl bestimmter Indikatoren (siehe Tab. 1).

Es zeigt sich, dass die Kontexte, in denen die neun identifizierten Studien entstanden sind, und auch die Zwecke, zu denen die Indikatoren zur Unterscheidung von Wirksamkeits- und Wirkungsstudien beschrieben werden, divers sind.

Tab. 1: Identifizierte Veröffentlichungen und ihre Charakteristika mit Indikatoren zur Differenzierung von Wirksamkeits- und Wirkungsstudien

Indikator	Gartlehner et al. (2006)	Flay et al. (2005)	Gottfredson et al. (2015)	Singal et al. (2014)	Lake et al. (2020)	Thorpe et al. (2009)	Loudon et al. (2015)	Alphs & Bossie (2016)	Wieland et al. (2017)
		Standards der Society of Prevention Research				PRECIS	PRECIS-2	ASPECT-R	Prä-Delphi
Zweck	Instrumententwicklung für nachträgliche Studien-einordnung	Indikatorenkatalog zur Orientierung für Praxis, Administration und Politik		Indikatorenkatalog	Anwendungsstudie	Indikatorenkatalog für Studien-designentwicklung		Instrumententwicklung für nachträgliche Studien-einordnung	Instrumententwicklung für nachträgliche Studien-einordnung in systematischen Reviews
Bezugs-wissenschaft	Medizin	multidisziplinär		Medizin	Pädagogik	Medizin	Medizin	Medizin	Medizin
Verweise auf inhaltliche Bezüge	-	-	Flay et al. (2005)	Flay et al. (2006); Gartlehner et al. (2006)	Gartlehner et al. (2006); Singal et al. (2014)	-	Thorpe et al. (2009)	Thorpe et al. (2009)	Gartlehner et al. (2006); Loudon et al. (2015); Thorpe et al. (2009)
Ratingsystem	dichotom	dichotom		dichotom	dichotom	dichotom	5-stufige Likert-Skala	6-stufige Likert-Skala	5-stufige Likert-Skala
Indikatoren-umfang (k)	7	23	25	5	2	10	9	6	17
									4

Tab. 2: Vorkommen von Indikatoren zur Differenzierung von Wirksamkeits- und Wirkungsstudien in den identifizierten Veröffentlichungen (ausführliche Itembeschreibung, siehe Appendix)

Indikator	Gartlehner et al. (2006)	Flay et al. (2005)	Gottfredson et al. (2015) Standards der Society of Prevention Research	Singal et al. (2014)	Lake et al. (2020)	Thorpe et al. (2009)	Loudon et al. (2015)	Alphs & Bossie (2016)	Wieland et al. (2017)		Vorkommenshäufigkeit von Indikator in Veröffentlichungen
						PRECIS	PRECIS-2	ASPECT-R	Prä-Delphi	Post-Delphi	
Provider (Spezialisierung Förderkraft)				+	+	+	+	+	+		6 von 9
Sample (Population)	+	+		+		+	+	+	+	+	8 von 9
Intervention (Stringenz des Förderprotokolls)				+		+	+	+	+	+	6 von 9
Outcome Ergebnismessung)	+	+	+			+	+	+	+		6 von 9
Setting	+			+	+		+	+	+	+	7 von 9
Follow-up		+	+			+	+	+	+		6 von 9
Adhärenz und Intention-to-treat-Analyse (ITT)	+	+	+			+	+	+	+		6 von 9
Verblindung		+	+						+		3 von 9
Rekrutierung							+		+		2 von 9
Studiendauer	+								+		2 von 9
Bewertung von unerwünschten Ereignissen	+		+								2 von 9
Stichprobengröße	+										1 von 9

Anmerkung: + = Indikator kommt als Item in der Publikation vor

4.2 Indikatorenauswahl im Rahmen der identifizierten Veröffentlichungen

Die neun identifizierten Publikationen variieren in Bezug auf die Indikatorenauswahl, auf deren Basis zwischen Wirkungs- und Wirksamkeitsstudien unterschieden wird (Alphs & Bossie, 2016; Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Singal et al., 2014; Wieland et al., 2017; siehe Tab. 2).

Im Instrument von Gartlehner und Kolleg:innen (2006) finden sich sieben Indikatoren (Stichprobe, Striktheit der Auswahlkriterien, Outcome, Studiendauer, Bewertung von unerwünschten Ereignissen, Angemessenheit der Stichprobengröße, Intention-to-treat-Analyse) zur Unterscheidung von Wirksamkeits- und Wirkungsstudien. Flay und Kolleg:innen (2005) definieren 23 Indikatoren für Wirksamkeitsstudien (Spezifität des Wirksamkeitsstatements ($k = 1$ Standard), Studienbeschreibung und Ergebnisse ($k = 6$ Standards), Kontroll- bzw. Vergleichskonditionen ($k = 5$ Standards), Generalisierbarkeit ($k = 1$ Standard), Präzision des Ergebnisses ($k = 10$ Standards)). Laut der Autor:innen sind Wirkungsstudien darüber zu definieren, dass sie die Indikatoren von Wirksamkeitsstudien erfüllen und zusätzlich zwölf weiteren Indikatoren entsprechen (Erfüllung der Wirksamkeitsindikatoren ($k = 1$ Standard), Programmbeschreibung und Ergebnisse ($k = 6$ Standards), Klarheit des Kausalschlusses ($k = 1$ Standard), Generalisierbarkeit der Ergebnisse ($k = 2$ Standards), Präzision des Ergebnisses ($k = 2$ Standards)). Diese Indikatoren wurden durch Gottfredson und Kolleg:innen (2015) erneuert und überarbeitet. So wird in der Überarbeitung bspw. auf die Stichprobengrößenbeschreibung verzichtet, aber dafür bspw. die Bewertung von unerwünschten Ergebnissen ergänzt (siehe Tab. 2).

Singal und Kolleg:innen (2014) postulieren fünf Indikatoren für die Zuordnung von Studien zur Wirksamkeits- oder Wirkungsdimension (Studiendesign, Heterogenität der Stichprobe, Maß der Interventionsstandardisierung, Umgang mit fehlenden Daten, Ergebnisbericht) und lehnen sich, laut eigener Aussage, dabei an Flay und Kolleg:innen (2006) und Gartlehner und Kolleg:innen (2016) an. Wieland und Kolleg:innen (2017) bestimmen die Wirksamkeit-Wirkungsindikatoren im Rahmen des sogenannten Rating of Included Trials on the Efficacy Effectiveness Spectrum (kurz RITES) mithilfe eines Delphi-Verfahrens. Es wurden insgesamt 17 potentielle Indikatoren (Proband:innen ($k = 2$), Setting und Intervention ($k = 5$), weitere Aspekte ($k = 10$)) in das Expert:innenvoting inkludiert. Diese 17 Indikatoren wurden nach dem Voting auf vier Indikatoren reduziert (Proband:innen, Setting, Flexibilität der Intervention, klinische Relevanz und Vergleich). Lake und Kolleg:innen (2020) definieren in Anlehnung an Gartlehner und Kolleg:innen (2006) sowie Singal und Kolleg:innen (2014) zwei Items zur Differenzierung: das Interventionssetting sowie die Art der Förderkraft.

Im Rahmen des Instruments *PRECIS* von Thorpe und Kolleg:innen (2009) werden zehn Items zur Konstruktion von Wirksamkeits- und Wirkungsstudien vorgeschlagen (Proband:innen: Exklusions- und Inklusionskriterien; Intervention und Flexibilität: Flexibilität der Experimentalintervention, Art der Kontrollintervention, Expertise der Förderkraft in der Experimental- und Kontrollintervention; Follow-up und Ergebnisse: Follow-Up-Intensität, primäre Studienergebnisse; Compliance/Adhärenz: Proband:innencompliance, Förderkraftadhärenz; Analyse: Analyseart der primären Studienergebnisse). In der Weiterentwicklung des Instruments, dem *PRECIS-2*, finden sich neun Items (Loudon et al., 2015). Änderungen im Vergleich mit dem *PRECIS* sind folgende: Der Indikator *Proband:innen* wurde in zwei neue Indikatoren *Setting* und *Exklusions- und Inklusionskriterien* aufgeteilt sowie die Indikatoren *Rekrutierung* und *Organisation* ergänzt. Die Expertise der Förderkraft in der Experimental- und Kontrollintervention sowie Flexibilität der Kontrollintervention wurden als Indikatoren entfernt. Alphs und Bossie (2016) haben in Anlehnung an *PRECIS* das post-hoc Bewertungstool *ASPECT-R* entwickelt. In *ASPECT-R* werden sechs der ursprünglichen *PRECIS*-Indikatoren, welche die Autor:innen für besonders relevant für die Unterscheidung der Wirksamkeits- und Wirkungsdimensionen einschätzen (Proband:innen; Intervention und Flexibilität; Expertise der Förderkraft; Follow-Up-Intensität und -dauer; primäre Studienergebnisse; Compliance/Adhärenz), inkludiert.

Zwölf verschiedene Indikatoren lassen sich übereinstimmend in den neun Veröffentlichungen zur Unterscheidung der Evidenzdimensionen Wirkung- und Wirksamkeit mit unterschiedlicher Vorkommenshäufigkeit identifizieren. In sechs bis acht der neun Veröffentlichungen, und somit besonders frequent, kamen die folgenden sieben Indikatoren vor: die Beschreibung vom Provider (insb. der Spezialisierung der Förderkraft), dem Sample (insb. der Heterogenität der Proband:innengruppe), der Intervention (insb. der Stringenz des Förderprotokolls), dem Outcome (insb. die Art und Weise der Ergebnismaße), dem Setting (insb. die Ähnlichkeit zur Praxis), Follow-Ups und Adverse Events. Seltener – zwischen ein und dreimal in den neun Veröffentli-

chungen – kamen folgende Indikatoren vor: die Rekrutierung, die Studiendauer, der Stichprobenumfang, die Verblindung und Intention-to-treat-Analyse (kurz ITT) (siehe Tab. 2). Die genauen Items in den Veröffentlichungen sind im Appendix gelistet.

4.3 Bewertungsstrategien im Rahmen der identifizierten Veröffentlichungen

In Bezug auf die Bewertungsart im Rahmen der identifizierten Veröffentlichungen kann festgestellt werden, dass bei sechs der neun Studien Indikatoren aufgestellt wurden, die dichotom bewertet werden: Das Kriterium wird somit erfüllt oder nicht erfüllt (Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Singal et al., 2014; Thorpe et al., 2009; Wieland et al., 2017; siehe Tab. 1). Eine Besonderheit bei der Bewertung von Wirkungsstudien bei Flay und Kolleg:innen (2005) sowie der Überarbeitung durch Gottfredson und Kolleg:innen (2015) besteht darin, dass nicht nur eine dichotome Entscheidung über die Erfüllung von Indikatoren getroffen werden muss, sondern, dass Wirkungsstudien zusätzlich zu den Wirksamkeitsindikatoren weitere Indikatoren erfüllen müssen (siehe Kapitel 3.2). Im Gegensatz zum Ansatz von Flay und Kolleg:innen (2005) bzw. Gottfredson und Kolleg:innen (2015) wählen die Autor:innen von vier weiteren Publikationen (Gartlehner et al., 2006; Lake et al., 2020; Thorpe et al., 2009; Singal et al., 2014) ein gemeinsames Indikatorenset für Wirkungs- und Wirksamkeitsstudien. In diesem Ansatz erfolgt die Kategorisierung in Wirksamkeits- und Wirkungsstudien über die (Nicht-)Erfüllung bestimmter Indikatoren.

Im Gegensatz zu den dichotomen Ratingskalen wird in Alphas und Bossie (2016), Loudon und Kolleg:innen (2015) sowie Wieland und Kolleg:innen (2017) eine spektrale Interpretation des Wirksamkeit-Wirkungsspektrums über eine 5-stufige Likert-Skala vorgenommen, wodurch ein differenziertes Verständnis von Wirksamkeit und Wirkung umgesetzt wird.

Die identifizierten Indikatoren (siehe Kap. 4.2) und Rating-Optionen (siehe Kap. 4.3) auf Grundlage der systematischen Literaturrecherche bilden die Grundlage für die Konzeptualisierung des Instruments PROSAIOS für die Bewertung von Studien in schulischen Sprachförderkontexten.

5 Ergebnisse: Konzeptualisierung des Bewertungsinstruments PROSAIOS und die Anwendung an einer Beispielstudie

Nachfolgend wird zunächst die Konzeptualisierung des Bewertungsinstruments *PROSAIOS* (Akronym für *Provider, Sample, Intervention, Outcome, Setting*) mit den gewählten Indikatoren und den Definitionen der Skalenpole der Ratingskala vorgestellt. Hierbei handelt es sich zunächst um die Vorstellung des Pilotkonzepts. Eine Validierung ist vorgesehen, jedoch bisher noch nicht erfolgt. Im Anschluss an die Vorstellung der Indikatoren in PROSAIOS wird eine ausgewählte Interventionsstudie exemplarisch mit dem Instrument bewertet, um die Anwendung und das Reasoning der Bewertungen zu verdeutlichen.

5.1 Das Instrument *PROSAIOS*

Basierend auf den Ergebnissen der oben vorgestellten Itemkataloge bzw. Instrumente zur Einordnung von Evaluationsstudien hinsichtlich der Evidenzdimensionen *Wirksamkeit* und *Wirkung* (siehe Kapitel 4), wurden folgende fünf Indikatoren zur Unterscheidung von Wirkungs- und Wirksamkeitsstudien identifiziert: Provider (Förderkraft), Sample (Stichprobe), Intervention, Outcome und Setting. Die Auswahl der Indikatoren basiert zunächst auf einer hohen übereinstimmenden Nennung der Indikatoren – mindestens in sechs der neuen Veröffentlichungen in der o. g. systematischen Literaturrecherche (siehe Tab. 2) – als valides, differenzierendes Kriterium. Darüber hinaus zeichnen sich die Indikatoren als besonders relevant für die sprachförderpädagogische Forschung aus. Um die Komplexität gering und die Handhabbarkeit des Instrumentes zu gewährleisten, erfolgte eine Beschränkung auf fünf Indikatoren. Trotz einer ebenfalls hohen Vorkommenshäufigkeit der beiden Indikatoren *Adhärenz* und *Follow-Up* in den neun identifizierten Veröffentlichungen (siehe Tab. 2) werden diese nicht in das PROSAIOS-Instrument integriert. Grund hierfür ist, dass Studienberichte von Interventionsstudien zu Sprachfördermaßnahmen bisher seltener Informationen zur Beurteilung dieser beiden Kriterien aufführen, so dass eine Bewertung nicht durchgängig möglich wäre. Hinzu kommt, dass in der Delphi-Studie von Wieland und Kolleg:innen (2017) zudem beide Kriterien von Fachkräften als nicht relevant für die Differenzierung von Wirksamkeits- und Wirkungsstudien bewertet wurden und daher auch nicht in der finalen Version des *RITES-Instruments* inkludiert wurden.

Die fünf ausgewählten Indikatoren *Provider* (*Spezialisierung der Förderkraft*), *Sample* (*Population*), *Intervention* (*Stringenz des Förderprotokolls*), *Outcome* (*Ergebnisse*) und *Setting* weisen die höchste Anschlussfähigkeit an die bestehende Studienlage (siehe Tab. 2) auf. Die jeweiligen Indikatorenkonzepte wurden übernommen, jedoch in der Definition für den (Sprach-)Förderkontext adaptiert und umformuliert, um den Spezifika (sprachförder-)pädagogischer Forschung gerecht zu werden (siehe Abb. 2). Für die Operationalisierung der Indikatoren werden Leitfragen sowie eine Festlegung der Skalenpole formuliert (siehe Abb. 2). Im Folgenden werden die Indikatoren einzeln vorgestellt.

Provider

Der Indikator Provider (vgl. Förderkraft) wird mithilfe von zwei Leitfragen operationalisiert, um den Grad der Spezialisierung umfassend abbilden zu können. Wirksamkeitsstudien zeichnen sich v. a. durch eine hohe Spezialisierung der Fachkräfte und eine hohe Expertise in Bezug auf die anzuwendende Interventionsmethode aus. Diese Expertise ließe sich in der ersten Leitfrage über den Grad der Fortbildung und des Schulungsumfanges zur Interventionsmethode messen (siehe Abb. 2). Keine oder eine geringere Erfahrung mit den anzuwendenden Interventionsmethoden und -techniken sowie ein geringerer Grad an Fortbildung mit der jeweiligen zu evaluierenden Interventionsmethode entspräche eher der Wirkungsdimension. Ein hohes Maß an Fortbildung und eine große Erfahrung im Umgang mit der Methode würde hingegen stärker für die Wirksamkeitsdimension sprechen.

Um die Ausprägung der Spezialisierung der Förderkraft auch über die Profession noch weiter differenzieren zu können, wurde dieser Indikator um eine weitere Operationalisierungsfrage zum Professionsverständnis ergänzt (siehe Abb. 2). In der schulischen Sprachförderung sind verschiedene hochgradig qualifizierte Professionen mit unterschiedlichen Bezugswissenschaften (vgl. Lüttke, 2010; Mußmann, 2020) tätig. Zum einen arbeiten Personen in der Förderung, die eine therapeutisch-klinische Spezialisierung besitzen und somit auch in der schulischen Förderung einem Heilauftrag folgen, wie beispielsweise Sprachtherapeut:innen oder Sprachheilpädagog:innen. Medizin, Therapie- oder Gesundheitswissenschaften sind die primären Bezugswissenschaften dieser Professionen (vgl. Lüttke & Stitzinger, 2015; Maihack, 2014). Der Grad der Vertrautheit, eine medizinisch definierte Symptomatik oder Beeinträchtigung zu behandeln, ist für diese Profession als hoch einzuschätzen. Zum anderen sind in der schulischen Sprachförderung auch zu einem überwiegenden Teil Personen tätig, die eine hohe pädagogische Qualifikation, beispielsweise als allgemeine Fachlehrkraft oder als Sonderpädagog:in, mitbringen. Die Ausrichtung des Förderziels stellt die Bildungsteilhabe von betroffenen Personen in den Vordergrund. Die primären Bezugswissenschaften dieser Professionen sind die Erziehungswissenschaften und die Fachdidaktiken. Das pädagogische Handeln im Bereich der schulischen Sprachförderung orientiert sich dabei primär an Bildungszielen und ist ausgerichtet auf die Optimierung der Teilhabe an Lehr-Lernprozessen. Der Grad der Vertrautheit, in diesem Sinne Sprachfördermaßnahmen für den Abbau und die Vermeidung von sprachbasierten Lernbarrieren einzusetzen (Spreer, 2020), ist für diese Profession als hoch einzuschätzen. Die Einordnung in die Evidenzdimensionen Wirksamkeit und Wirkung für die jeweilige Profession stellt sich hier als eine besondere Herausforderung dar, da keine Wertung eines Professionsverständnisses erfolgen soll. Wirksamkeitsstudien zeichnen sich vor allem dadurch aus, dass auf individueller Ebene unter Anwendung einer spezifischen Intervention die Überwindung einer Beeinträchtigung ermöglicht wird (vgl. Heilauftrag). Diese Ausrichtung geht in der Regel einher mit einem hohen Grad der therapeutischen Ausrichtung der Förderkraft in der Intervention. Wirkungsstudien zur unterrichtsimmanenten Sprachförderung an allgemeinen Schulen verfolgen primär einen Bildungsauftrag und erfordern eine hohe pädagogische Expertise für die individuelle Förderung von Schüler:innen in einer sprachlich heterogenen Lerngruppe.

Da der Indikator Förderkraft mittels zweier Leitfragen operationalisiert wurde und somit zwei Bewertungsergebnisse möglich sind, wird aus den beiden Bewertungen der Mittelwert gebildet, damit der Indikator Förderkraft einwertig in die Gesamtbewertung eingeht.

Sample

Der Indikator Sample (vgl. Stichprobe, Population) wird über den Grad der Homogenität der Stichprobe definiert. Wirksamkeitsstudien zeichnen sich durch eine besonders hohe Homogenität der Stichproben aus, welche über die Definition strenger Inklusions- und Exklusionskriterien erreicht wird. Ist die Stichprobe heterogen, so dass sie der Bezugsgruppe in der Praxis stark äh-

nelt, und gelten eher weniger strenge Inklusions- und Exklusionskriterien, wird dieser Indikator eher zugunsten der Wirkungsdimension entschieden.

Intervention

Mit dem Indikator *Intervention* wird der Standardisierungsgrad der Intervention erhoben. Wirksamkeitsstudien sind durch geringe Freiheitsgrade der Förderkraft, die Intervention durchzuführen, charakterisiert. Liegt ein strenges Studienprotokoll vor, welches eine geringe Flexibilität bzgl. der Interventionsanwendung erlaubt, wird eine Zuordnung zur Wirksamkeitsdimension vorgenommen. Ist das Studienprotokoll weniger strikt und der Förderkraft wird eine hohe Flexibilität in der Interventionsanwendung gewährt, so kann sich für die Wirkungsdimension entschieden werden.

Outcome

Mit dem Indikator *Outcome* (vgl. Ergebnisse) erfolgt eine Einschätzung der Form der Überprüfung von Interventionseffekten. Dies erfolgt beispielsweise über die Einordnung des Direktheitsgrades der Interventionseffektmaße (Gartlehner et al., 2006). Werden zur Überprüfung der Interventionseffekte vorrangig Maße angelegt, die einen direkten, spezifischen Übungseffekt anhand der Messung einer sprachlichen Teilfähigkeit oder Veränderung der Symptomatik überprüfen, spricht dies eher für die Wirksamkeitsdimension. Werden die Ergebnisse einer Intervention hingegen eher über indirekte Maße erhoben – wird beispielsweise ein Ersatzmarker für den Fördererfolg gemessen oder ein schulisch bedeutsamer Marker (z. B. Verbesserung der schulischen Leistung oder der Lebensqualität im Anschluss an die Förderung von grammatischen Fähigkeiten) – so würde man dies eher der Wirkungsdimension zuordnen.

Setting

Die Leitfrage zum *Setting* bezieht sich auf den Grad der Abweichung des jeweiligen Studiensettings von den schulischen Alltagsbedingungen der Sprachförderung an den jeweiligen Einrichtungen. Dabei erfolgt die Einschätzung unabhängig von der jeweiligen Schulart, d. h. unabhängig davon, ob es sich um ein Förderzentrum oder eine allgemeine Schule handelt. Wirkungsstudien werden dadurch charakterisiert, dass sie einen hohen Bezug zur Realwelt aufweisen. Dies entspricht im schulischen Sprachförderkontext dem gewöhnlichen Schulalltag. Wirksamkeitsstudien zeichnen sich durch ein Studiensetting unter Idealbedingungen aus, in dem möglichst viele Parameter kontrolliert werden und unübliche Ressourcen zur Verfügung stehen. Im Sprachförderkontext entspräche dies Studien, die bspw. in Lernlaboren oder universitären, internen Praktika durchgeführt werden, bei denen die Kontrolle ablenkender Störreize stärker gegeben ist.

Indikator	Leitfrage	Wirkung				Wirksamkeit		keine Angabe
		max. Wirkung	sehr niedrig	niedrig	hoch	sehr hoch	max. Wirksamkeit	
Provider Förderkraft	Wie hoch ist der Umfang der Schulung für die Interventionsmethode durch die Anwender:innen?	geringes Maß an Erfahrung mit und Fortbildung zu der Interventionsmethode					hohes Maß an Erfahrung mit und Fortbildung zu der Interventionsmethode	
	Wie hoch ist der Grad der therapeutischen Ausrichtung der Förderkraft in der Förderung?	Bildungsauftrag					Heilauftrag	
Sample Stichprobe	Wie hoch ist der Grad der Homogenität in der Gruppe der Teilnehmenden?	gering durch Abbildung der typischen Heterogenität der einzelnen Lerngruppen					hoch durch strenge Exklusionskriterien, (Komorbiditäten ausgeschlossen, Selektionsbias)	
Intervention	Wie hoch ist der Grad der Standardisierung bei der Förderung?	freieres Studienprotokoll hohe Flexibilität					striktes, standardisiertes Studienprotokoll geringe Flexibilität	
Outcome	Wie hoch ist der Grad der Direktheit der Maße der Interventionseffekte?	indirekte Maße					direkte Maße	
Setting	Wie hoch ist der Grad der Abweichung des Studiensettings vom üblichen Schulalltag?	reale Welt Schulalltag					Ideal-Bedingung akademisch	

Abb. 2: Das Instrument PROSAIOS mit den Indikatoren, Leitfragen und der Ratingskala

Ratingskala

Die Indikatoren werden auf einer 4-stufigen Ratingskala (sehr niedrig = 1, niedrig = 2, hoch = 3, sehr hoch = 4) mit der zusätzlichen Option „Keine Angabe“ bewertet. Hierbei werden vier Stufen für das Rating angeboten, um die Interpretierbarkeit einer Tendenz in Richtung einer Evaluationsdimension zu erzielen. Die Ratingskala soll bei der Einordnung der Wirksamkeits-Wirkungsdimensionen der einzelnen Indikatoren helfen, sodass auch Abstufungen in der Ausprägung beider Dimensionen berücksichtigt werden können. Für die Pilotierung des Instrumentes werden zunächst nicht alle Skalenstufen, sondern lediglich die Skalenpole, formuliert, um der Menge an Möglichkeiten, die in Studien zu finden sind, noch Interpretationsraum zu lassen. Dies wird für die anstehende Validierung vorgesehen.

Über die Ratingwerte der fünf Indikatoren kann im Anschluss ein Mittelwert gebildet werden. Jeder Indikator geht dabei mit gleichem Anteil in den Durchschnittswert ein. Fehlende Angaben werden nicht in den Mittelwert inkludiert. Eine Interventionsstudie mit einem hohen Durchschnittswert ($MW > 2,5$ bis max. $MW = 4$) wäre somit eher der Evidenzdimension der Wirksamkeit zuzuordnen, wohingegen ein geringer Durchschnittswert (min. $MW = 1$ bis $MW = 2,5$) eher auf die Charakteristika einer Wirkungsstudie hinwiese. Die Validität dieses Cut-Off-Werts ist noch nicht überprüft, aber für einen späteren Zeitpunkt intendiert.

Um die Anwendung des Instrumentes zu verdeutlichen, soll nun im nachfolgenden Kapitel 5.2 eine Studie exemplarisch mit PROSAIOS bewertet werden.

5.2 Bewertungsbeispiel

Das Instrument PROSAIOS und dessen Indikatoren werden im Folgenden an einem Beispiel illustriert. Dazu wird eine Interventionsstudie zur Förderung der Grammatikproduktion bei Kindern mit Sprachentwicklungsstörung im schulischen Setting (Smith-Lock, et al. 2013) vorgestellt und bewertet. Smith-Lock und Kolleg:innen (2013) nutzen ein Pre-Post-Testdesign mit multipler Baseline und vergleichen eine Experimental- ($n = 22$) mit einer Kontrollgruppe ($n = 18$). Die Studie fand in einer australischen Sprachförderschule statt, die sich durch reguläre, kleine Lerngruppen (max. 12 Schüler:innen, Lehrkraft, Lehrassistent:in) und additive Sprachförderung in Kleingruppen in Zusammenarbeit mit Sprachtherapeut:innen und Sonderpädagog:innen auszeichnet (S. 268). Die Intervention fand im Rahmen des regulären Schulalltags einmal wöchentlich über einen Zeitraum von acht Wochen statt. Die Experimentalgruppe erhielt in der Sprachförderstunde eine Förderung mit dem Interventionsziel der Produktion morphologischer Markierungen (Possessivmarkierung, Personalform (3. Person Singular), Markierung der Vergangenheitsform (Präteritum)). Dabei kamen folgende Methoden zum Einsatz: die metasprachliche Einführung der sprachlichen Strukturen (sog. *Direct Instructions*), spezifische Strukturierungen des Inputs (sog. *Focused Stimulation*), Modellierungen (sog. *Recasting*) und Nachsprechprompts (sog. *Imitation*) (S. 271 f.). Die Kontrollgruppe erhielt im selben Umfang wie die Interventionsgruppe eine reguläre Sprachförderstunde mit dem Ziel der Sprachverständnisförderung (Äußerungen mit räumlichen Präpositionen). Alle Schüler:innen wurden zu Beginn der Intervention hinsichtlich ihrer sprachlichen Fähigkeiten und in Bezug auf die Testitems in einem grammatischen Elizitationstest mit geübten und ungeübten Items getestet. Die Überprüfung erfolgte zweimal vor Studienbeginn und einmal nach Durchführung der Studie. Im Folgenden wird die Studie in Bezug auf die o.g. Indikatoren mit PROSAIOS eingeschätzt.

Provider

Die Lehrkräfte und Lehrassistenten der Interventionsgruppe erhielten einen dreistündigen Workshop zum Interventionsdesign, zur Abfolge und den Bestandteilen der Interventionsmaßnahme einschließlich der Sprachfördertechniken. Eine Auseinandersetzung mit der Methodik erfolgte über Übungen im Rollenspiel und entsprechendes Feedback durch das Studienpersonal. Es fanden wöchentliche Treffen mit dem Team der schuleigenen Sprachtherapeut:innen zu den Förderzielen und den geplanten Aktivitäten statt. Der Auftakt der wöchentlichen Interventionsseinheiten erfolgte über eine metasprachliche Einführung der Zielstrukturen (via direct instruction) durch das Sprachtherapeut:innen-Team und hatte somit auch Modellcharakter für die beteiligten Lehrkräfte und -assistenten. Darüber hinaus gab es ein genaues Manual zum Vorgehen. Der Umfang der Schulung für die Interventionsmethode kann somit als hoch eingestuft werden (siehe Abb. 3).

Die vorwiegende Umsetzung der Interventionsmethode oblag Lehrkräften und Lehrassistenten, deren Handlungsentscheidungen primär ein Bildungs- als ein Heilauftrag zugeschrieben

wird. Die schulischen Sprachtherapeut:innen wurden unterstützend in die Intervention einbezogen. Daher wird der Grad der therapeutischen Ausrichtung der Förderkraft in der Förderung als niedrig eingeschätzt (siehe Abb. 3).

Sample

Die an der Studie teilnehmenden Schüler:innen erfüllten strenge Inklusionskriterien. Es lag für alle teilnehmenden Schüler:innen eine mittels psychometrischer Testverfahren ermittelte klinische Diagnose einer Sprachentwicklungsstörung vor, wobei das kognitive Potential sowie das Hörvermögen innerhalb des Normbereichs liegen musste. Moderate Beeinträchtigungen rezeptiver und produktiver Sprachfähigkeiten waren ein relevantes Inklusionskriterium, wohingegen weitere klinische Diagnosen und Komorbiditäten als Exklusionskriterium galten. Alle Kinder besuchten die erste Klasse. Somit liegt ein hoher Grad an Homogenität unter den Studienteilnehmenden vor (siehe Abb. 3).

Intervention

Die angewandte Intervention folgte einem detaillierten Förderprotokoll, in dem Förderziele, Förderitems und -techniken vorgegeben waren. Der Ablauf jeder Förderstunde wurde in detaillierten Skripten vorgegeben. Auch wurden sämtliche Fördermaterialien für die einzelnen Förderaktivitäten zur Verfügung gestellt. Zudem erfolgte für jede Äußerung der Schüler:innen ein festgelegtes Scoring, um individuelle Fortschritte im Verlauf der Interventionsphase zu messen und zu erfassen, ob alle Schüler:innen die a priori festgelegte Anzahl an Lerngelegenheiten pro Förderaktivität erhielten. Verschiedene Formen von Modellierungen der kindlichen Äußerungen wurden mit den Lehrkräften eingeübt, die finale Umsetzung erfolgte jedoch spontan und ohne schriftliche Vorgabe. Der Grad der Standardisierung der Intervention ist somit als hoch einzuschätzen (siehe Abb. 3).

Outcome

In Smith-Lock und Kolleg:innen (2013) wurden Interventionseffekte mittels eines selbsterstellten Grammatik-Elizitationstests gemessen, wodurch psychometrische Verfahren zur Überprüfung der Interventionseffekte nicht zum Einsatz kamen. Der Elizitationstest enthielt 30 geübte und ungeübte Items und sollte die Entwicklung in der zielsprachlich korrekten Produktion morphologischer Strukturen erfassen, indem die Testergebnisse nach der Interventionsphase mit zwei Testzeitpunkten vor der Intervention verglichen wurden. Die Überprüfung der Übungseffekte erfolgte sowohl im Einzel- als auch im Gruppenvergleich. Somit werden in der Studie Maße verwendet, die direkt sprachliche Fähigkeiten überprüfen. Der Grad der Direktheit wird somit als hoch eingestuft (siehe Abb. 3).

Setting

Die Studie fand in einer regulären australischen Förderschule (Language Development Center) statt. Die Intervention war Teil des regulären Unterrichtsgeschehens und fand regulär im Rahmen des Sprachförderkonzeptes der Schule statt. Die Förderintensität entsprach dem gängigen Umfang. Die einzige Abweichung vom alltäglichen Geschehen stellten zusätzliche Supervisionen der Lehrkräfte durch das Studienteam dar, durch die jede Lehrkraft einmalig ein Feedback zur Interventionstreue erhielt. Daher wird von einem niedrigen Abweichungsgrad vom schulischen Alltag im Rahmen des Studiensettings ausgegangen (siehe Abb. 3).

Finale Einordnung

Für die abschließende Einordnung der Studie von Smith-Lock und Kolleg:innen (2013) hinsichtlich ihrer Evidenzdimension wurden Punktwerte vergeben, die der Zuteilung zur Ratingskala entsprechen. Eine Übersicht über die Bewertung ist der Abb. 3 zu entnehmen. Für die vorliegende Studie ergäbe sich laut PROSAIOS-Instrument ein Mittelwert von 2,7, der auf dem Spektrum zwischen Wirkungs- und Wirkungsstudie eher einer Wirksamkeitsstudie zuzuordnen wäre. Die Studienautor:innen ordnen die Studie als Wirkungsstudie ein (Smith-Lock et al., 2013) und begründen die Einordnung mit der Nähe des Studiensettings zur gängigen Sprachförderpraxis an der ausgewählten Schule.

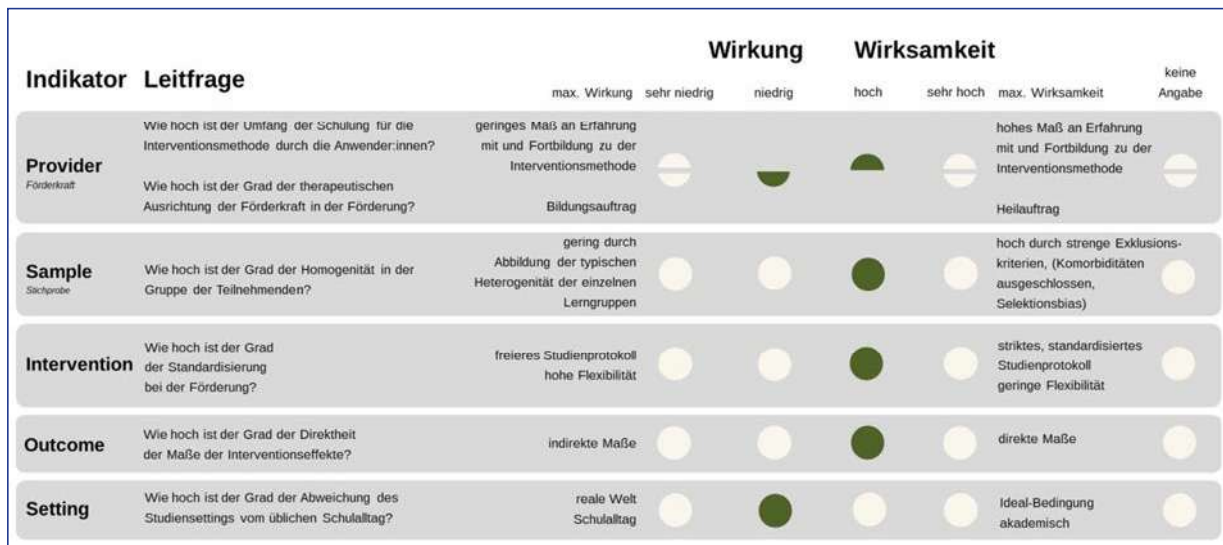


Abb. 3: Die Bewertung der Studie von Smith-Lock und Kolleg:innen (2013) mit PROSAIOS

6 Diskussion

Die aktuellen wissenschaftlichen und bildungspolitischen Diskussionen zeigen, dass Evaluationsstudien, die den Wirkvorteil von pädagogischen Maßnahmen und Interventionen messen, immer stärker an Bedeutung gewinnen. Auch in Bezug auf die schulische Sprachförderung sind Fachkräfte zunehmend gefordert, evidenzbasierte Maßnahmen und Methoden der Sprachförderung in die pädagogische Praxis zu übertragen. Im Kontext der triangulierenden EBP entspricht dies der Nutzung externer Evidenz als Informationsquelle. Veröffentlichte Evaluationsstudien zu Sprachfördermaßnahmen legen jedoch nahe, dass die Differenzierung der Evidenzdimensionen – insbesondere Wirksamkeit und Wirkung – im pädagogischen Kontext bisher unzureichend berücksichtigt wird. Dies hat zur Folge, dass auch die systematische Reihenfolge bei der Prüfung von Wirkvorteilen zu Sprachfördermaßnahmen, d. h. (1) Prüfung der Wirksamkeit, (2) Prüfung der Wirkung, (3) Prüfung der Effizienz, ebenfalls kaum umgesetzt wird. Sinnvolle Transferentscheidungen für die Praxis können jedoch vor allem aus Wirkungsstudien abgeleitet werden. Diese sollten jedoch erst erfolgen, wenn Wirksamkeitsstudien belegen, dass eine Sprachfördermaßnahme unter Idealbedingungen einen Wirkvorteil erzielt. Die unzureichende Berücksichtigung dieser Abfolge kann dazu führen, dass Sprachfördermaßnahmen voreilig übertragen werden, ohne dass deren tatsächliche Wirkung in der schulischen Sprachförderpraxis ausreichend geprüft wurde. Das Instrument *PROSAIOS* wurde daher mit dem Ziel konzipiert, veröffentlichte Evaluationsstudien zur schulischen Sprachförderung präzise in Wirksamkeits- und Wirkungsstudien einteilen zu können.

Um die Evidenzlage anhand veröffentlichter Evaluationsstudien analysieren zu können, wurde *PROSAIOS* als Entscheidungshilfe entwickelt. Ähnlich wie bei Wieland und Kolleg:innen (2017), Alphs und Bossie (2016) sowie Loudon und Kolleg:innen (2015) baut *PROSAIOS* auf einem spektralen Verständnis der Wirksamkeits-Wirkungsdimensionen auf. Dieses Verständnis spiegelt sich auch in der Bewertungsmöglichkeit anhand der 4-stufigen Ratingskala wider. Darüber hinaus war es erforderlich, die identifizierten Indikatoren für die Unterscheidung zwischen Wirksamkeits- und Wirkungsstudien an die Gegebenheiten der schulischen Sprachförderung anzupassen. Grundsätzlich wurden alle Indikatoren vom medizinischen, gesundheitswissenschaftlichen oder pädagogischen Anwendungsbereich auf die schulische Sprachförderung übertragen.

Ein Anpassungsbedarf zeigte sich besonders am Indikator *Provider*. Das *PROSAIOS*-Instrument bewertet diesen Indikator anhand von zwei Fragestellungen. Neben dem Grad der Erfahrung und der Fortbildung zu einer Interventionsmethode, wird auch der Grad der therapeutischen Ausrichtung berücksichtigt. Dies ermöglicht es, die unterschiedlichen primären Bezugswissenschaften der verschiedenen Professionen einzubeziehen, die im internationalen schulischen Kontext mit der Aufgabe betraut sind, Sprachfördermaßnahmen anzubieten. Im medizinischen oder gesundheitswissenschaftlichen Kontext steht die Spezialisierung der Fachkräfte in der Art der Gesundheitsversorgung im Vordergrund. Die Spezialisierung ist jedoch in

der pädagogischen Praxis leicht misszuverstehen, da die an der schulischen Sprachförderung beteiligten Personen je nach Bildungssystem oder Bildungseinrichtung ihre Qualifikationen in unterschiedlichen Professionen erwerben können. Somit ist die Unterscheidung bezüglich der therapeutischen Ausrichtung des professionellen Handelns zielführender, um Wirksamkeitsstudien von Wirkungsstudien zu unterscheiden.

In Bezug auf die vier anderen Indikatoren *Sample*, *Intervention*, *Outcome* und *Setting* erfolgte die Anpassung an die Gegebenheiten der schulischen Sprachförderung zu einem geringeren Maße im Vergleich zu den primär medizinisch-gesundheitswissenschaftlichen Einordnungen der Evidenzdimensionen. Hier wurden vor allem Formulierungen auf den Kontext der Sprachförderung angepasst.

Mithilfe von PROSAIOS kann eine Einordnung auf dem Wirksamkeit-Wirkungsspektrum erfolgen, basierend auf einer 4-stufigen Ratingskala. Dabei sind lediglich die Skalenpole definiert, während eine detaillierte Beschreibung der Skalenpunkte zwischen den Polen noch aussteht. Das Vorgehen in PROSAIOS entspricht dem im überwiegenden Teil der Referenzliteratur verwendeten Ansatz, bei dem Bewertungen auf einer Ratingskala vorgenommen werden. Auch hier wird überwiegend nur eine Beschreibung der Skalenpole genutzt (Loudon et al., 2015; Thorpe et al., 2009; Wieland et al., 2017). Ausschließlich Alphs und Bossie (2016) verwenden eine Skala, deren sämtliche Skalenpunkte detailliert beschrieben sind, was zu einer höheren Eindeutigkeit der Skalenbewertung führt.

Eine ausschließliche Beschreibung der Skalenpole ermöglicht hingegen eine größere Offenheit und Flexibilität des Instruments, wodurch es auf viele verschiedene Studienformate angewendet werden kann. Der Verzicht auf die Beschreibung der Skalenpunkte kann jedoch auch zu einer gewissen Uneindeutigkeit führen, die insbesondere unerfahrenere Bewerter:innen verunsichern könnte. Die Herausforderung bei vorgegebenen Beschreibungen aller Skalenpunkte besteht darin, eine Formulierung zu finden, die für möglichst viele Studienformate und -kontexte gleichermaßen gelten kann. Aufgrund der genannten Vor- und Nachteile ist eine Überprüfung, ob die Beschreibung aller Skalenpunkte im Gegensatz zur reinen Beschreibung der Skalenpole Vorteile bietet, im Rahmen einer Validierungsstudie intendiert.

Zur Einordnung auf dem Wirksamkeit-Wirkungsspektrum mithilfe der Ratingskala muss außerdem angeführt werden, dass die Kategorisierung anhand des Mittelwertes ebenfalls bisher noch nicht validiert ist, wobei auch diese noch folgen soll. Unabhängig davon kann jedoch diskutiert werden, ob eine Kategorisierung mit dem vorgeschlagenen Schwellenwert von 2,5 theoretisch valide ist. Ein solcher an einem Gesamtmittelwert ausgerichteter Schwellenwert führt zu einem ausgewogenen Wirksamkeit-Wirkungsspektrum, für das angenommen wird, dass die Verteilung von Wirksamkeits- und Wirkungsstudien symmetrisch ist. Dieser Ansatz vermeidet auch eine Hierarchisierung der Evidenzdimensionen. Ein anderes Vorgehen zur Berechnung eines Gesamtmittelwertes findet sich bspw. in der Familie der PRECIS-Studien (Alphs & Bossie, 2016; Loudon et al., 2015; Thorpe et al., 2009), bei denen die Einschätzung durch grafische Darstellungen, wie sogenannte Radardiagramme (auch Netz- oder Spinnennetzdiagramm genannt), vorgenommen wird.

Auch setzt die derzeitige Berechnung der Indikatorenbewertungen zu einem Mittelwert voraus, dass alle Indikatoren gleichwertig zur Kategorisierung auf dem Wirksamkeit-Wirkungsspektrum beitragen. Es ist jedoch auch möglich, dass einzelne Indikatoren unterschiedlich stark zur Differenzierung beitragen. Die Gleichgewichtung aller Indikatoren in PROSAIOS entspricht weitestgehend der Referenzliteratur (Alphs & Bossie, 2016; Loudon et al., 2015; Thorpe et al., 2009; Wieland et al., 2017). Die Validierungsstudie von Gartlehner und Kolleg:innen (2006) vertritt ein dichotomes Verständnis der Wirksamkeit und Wirkung und basiert auf der Kategorisierung der Indikatoren als (Nicht-)Zutreffen. Ihre Ergebnisse zeigen, dass ein ausgeglichenes Verhältnis von Sensitivität und Spezifität erreicht wird, wenn sechs statt sieben der ursprünglich intendierten Indikatoren zutreffen. Ein solch gewichtetes Verhältnis könnte auch auf PROSAIOS anwendbar sein und sollte im Rahmen einer Validierungsstudie geprüft werden.

Bei der Zusammenstellung der Indikatoren ist ein potentieller Bias zu berücksichtigen, da die identifizierten Referenzliteraturquellen aufeinander aufbauen (Alphs & Bossie, 2016; Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Loudon et al., 2015; Singal et al., 2014; Thorpe et al., 2009; Wieland et al., 2017; siehe Tab. 1). Da die Indikatorenauswahl auf Grundlage einer primären Häufigkeitsanalyse dieser Referenzliteratur erfolgte, könnte hier eine Verzerrung entstanden sein. Allerdings bedeutet dies auch, dass die aktuelle Studienlage berücksichtigt wurde. Insofern weist PROSAIOS eine hohe Anschlussfähigkeit an bestehende Instrumente auf.

PROSAIOS ist u. a. für die nachträgliche Bewertung bereits veröffentlichter Studien hinsichtlich ihrer Evidenzdimension konzipiert. In dieser Hinsicht verfolgt PROSAIOS ein ähnliches Ziel wie der überwiegende Teil der identifizierten Referenzliteratur (Alphs & Bossie, 2016; Flay et al., 2005; Gartlehner et al., 2006; Gottfredson et al., 2015; Lake et al., 2020; Singal et al., 2014; Wieland et al., 2017). Lediglich die PRECIS-Versionen dienen der prospektiven Studienkategorisierung (Loudon et al., 2015; Thorpe et al., 2009). PROSAIOS kann, wie in Kapitel 3 bei der Zielbeschreibung von PROSAIOS bereits beschrieben, zukünftig auch für die prospektive Kategorisierung genutzt werden.

PROSAIOS wurde mit dem Ziel entwickelt, mit einer begrenzten Anzahl an Indikatoren, die standardmäßig in Publikationen berichtet werden, Aussagen über die Evidenzdimension treffen zu können. Informationen zu den Indikatoren *Follow-Up* und *Adhärenz* werden häufig in empirischen Studien zu Sprachfördermaßnahmen nicht berichtet und wurden daher nicht in das Instrument aufgenommen. Sollte sich dieser Studienberichtsstandard ändern, könnte diese Entscheidung hinsichtlich der Konstruktion von PROSAIOS revidiert werden. Ein weiterer Grund für die Auswahl spezifischer Indikatoren war der Anspruch, PROSAIOS möglichst zeiteffizient und eindeutig anwenden zu können. In der anstehenden Validierung und Evaluation des Bewertungsinstrumentes wird zudem die Anwendbarkeit für verschiedene Nutzer:innengruppen (z. B. Administration, Forscher:innen und Fachkräften in der schulischen Praxis) überprüft, um zu untersuchen, ob das Instrument für die verschiedenen Nutzer:innengruppen geeignet ist.

Die beispielhafte Bewertung der Studie von Smith-Lock und Kolleg:innen (2013) mit PROSAIOS verdeutlicht, wie die verschiedenen Indikatoren innerhalb von PROSAIOS bewertet werden können. Die Wahl dieser Studie ist besonders interessant, da die Autor:innen selbst die Zuordnung der Studie zu den Evidenzdimension *Wirksamkeit* oder *Wirkung* diskutieren. Smith-Lock und Kolleg:innen (2013) kommen dabei zu dem Schluss, dass es sich bei der Studie um eine Wirkungsstudie handelt und begründen dies mit der Nähe des Studiensettings zur gängigen schulischen Sprachförderpraxis. Die mit PROSAIOS vorgenommene Bewertung tendiert jedoch eher zur Kategorisierung als Wirksamkeitsstudie. Das Fördersetting wird ebenfalls, wie bei Smith-Lock und Kolleg:innen (2013), als tendenziell nah am Schulalltag bewertet, ebenso wie der Grad der therapeutischen Ausrichtung (Teilfrage beim Indikator *Provider*). Weitere Indikatoren, wie *Outcome*, *Sample*, *Intervention* und *Provider* (insbesondere zur Erfahrung und Fortbildung der Fachkräfte mit der spezifischen Interventionsmaßnahme), werden als *hoch* bewertet und implizieren dadurch eher eine Tendenz zur Wirksamkeitsstudie. Dieses Beispiel verdeutlicht die Relevanz der verschiedenen Wirksamkeits- und Wirkungsindikatoren für die Bewertung der Evidenzdimensionen in der Sprachförderpraxis.

Die Auseinandersetzung mit der aktuellen Literatur zeigt eine fehlende Unterscheidung zwischen Wirksamkeits- und Wirkungsstudien bzw. eine Kritik an der vermeintlich mangelnden Übertragbarkeit evidenzbasierter Praxis in pädagogische Kontexte. Dazu ist es aus unserer Sicht zwingend erforderlich zwischen den Perspektiven der Forscher:innen und der Fachkräften in der schulischen Praxis zu unterscheiden:

Aus Forschungsperspektive ist ein hoher Grad an *Treatment Fidelity* (vgl. Interventionstreue) erwünscht, d. h., es besteht ein großes Interesse, dass die Anwender:innen Interventionen möglichst auf die gleiche Art und Weise durchführen, um Störeffekte zu vermeiden. Ein zu hoher Grad an Kontrolle entspricht jedoch nicht den Realbedingungen und beschränkt daher die Aussagekraft über die Wirkung im alltäglichen schulischen Kontext.

Aus Anwender:innenperspektive birgt sowohl eine Überbetonung von Wirksamkeitsstudien als auch Wirkungsstudien Risiken. Im Falle von Wirksamkeitsstudien bleibt die Frage der Generalisierbarkeit in alltäglichen Handlungskontexten offen. Im Falle von Wirkungsstudien ist häufig eine unzureichende Kontrolle von möglichen Einflussvariablen gegeben, so dass ein Risiko besteht, dass Interventionen keine Effekte zeigen, obwohl sie möglicherweise für einen Teil der Teilnehmenden wirksam sein könnten. Die Interpretation empirischer Daten zur Evaluation von Sprachfördermaßnahmen erfolgt daher stets in einem Spannungsfeld zwischen kontrollierten Idealbedingungen (Wirksamkeitsstudien) und Realbedingungen (Wirkungsstudien).

Zukünftig ist geplant, das Bewertungsinstrument PROSAIOS umfassend zu evaluieren und zu validieren, um eine auf Gütekriterien geprüfte Anwendung zu sichern. Darüber hinaus ist es für zukünftige Studien, die Wirkvorteile von schulischen Sprachfördermaßnahmen und Interventionen untersuchen, erforderlich, eindeutig anzugeben, welche Evidenzdimension – Wirksamkeit oder Wirkung – mit der Studie geprüft wird. Einige der eingangs vorgestellten Indikatorenkataloge (Loudon et al., 2015; Thorpe et al., 2009) bieten eine Orientierung für die Studienplanung und

können somit Forscher:innen bei dieser Entscheidung unterstützen. Zukünftig soll PROSAIOS diesen Bewertungsprozess sowohl prospektiv als auch retrospektiv erleichtern.

Literatur

- Alphs, L. D., & Bossie, C. A. (2016). ASPECT-R-A Tool to Rate the Pragmatic and Explanatory Characteristics of a Clinical Trial Design. *Innovations in Clinical Neuroscience*, 13(1–2), 15–26.
- Kultusministerkonferenz [KMK]. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften* (i. d. F. vom 16.05.2019). Abgerufen von https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf [23.03.2025]
- Blumenthal, Y., & Mahlau, K. (2015). Effektiv fördern – Wie wähle ich aus? Ein Plädoyer für die Evidenzbasierte Praxis in der schulischen Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 66, 408–421.
- Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies*, 47(2), 108–121. doi: 10.1111/1467-8527.00106
- Dollaghan, C. A. (2007). *Evidencebased Practice in Communication Disorders*. Baltimore: Paul H. Brookes.
- Egert, F., Galuschka, K., Groth, K., Hasselhorn, M., & Sachse, S. (2020). Evidenzbasierung vorschulischer sprachlicher Bildung und Förderung: Was man darunter versteht und bisher darüber weiß. In K. Blatter, K. Groth, & M. Hasselhorn (Hrsg.), *Evidenzbasierte Überprüfung von Sprachförderkonzepten im Elementarbereich* (S. 4–27). Wiesbaden: Springer VS.
- Eisele-Metzger, A., & Breuer, C. (2021). *Cochrane Glossar*. Abgerufen von <https://www.cochrane.de/cochrane-glossar#e> [16.12.2024]
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175. doi: 10.1007/s11121-005-5553-y
- Gartlehner, G., Hansen, R. A., Nissman, D., Lohr, K. N., & Carey, T. S. (2006). A simple and valid tool distinguished efficacy from effectiveness studies. *Journal of Clinical Epidemiology*, 59(10), 1040–1048. doi: 10.1016/j.jclinepi.2006.01.011
- Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation. *Prevention Science*, 16(7), 893–926. doi: 10.1007/s11121-015-0555-x
- Lawrence, R. E., Bernstein, A., Jaffe, C., & Goldberg, T. E. (2023). In clinical trials, efficacy vs. effectiveness language is confusing. *Journal of Clinical Epidemiology*, 159, 345–347. doi: 10.1016/J.JCLINEPI.2023.05.022
- Lake, J. K., Tablon Modica, P., Chan, V., & Weiss, J. A. (2020). Considering efficacy and effectiveness trials of cognitive behavioral therapy among youth with autism: A systematic review. *Autism*, 24(7), 1590–1606. doi: 10.1177/1362361320918754
- Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K. E., & Zwarenstein, M. (2015). The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*, 350, h2147–h2147. doi: 10.1136/bmj.h2147
- Lüdtke, U. (2010). Relationale Didaktik in Sprachpädagogik und Sprach-Therapie: Historische Einbettung und aktuelle Forschung. *MitSprache: Fachzeitschrift Für Sprachheilpädagogik*, 2010(1), 21–46.
- Lüdtke, U., & Stitzinger, U. (2015). *Pädagogik bei Beeinträchtigungen der Sprache*. München, Basel: UTB. doi: 10.36198/9783838585994
- Kuhr, L., & Kulawiak, P. R. (2018). Wissen, was wirkt: Modelle evidenzbasierter Pädagogik. *Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF)*, 7, 1–8. doi: 10.5771/1613-0707-2016-6-13
- Maihack, V. (2014). Sprachtherapie im Kontext der sprachtherapeutischen Berufe. In M. Grohnfeldt (Hrsg.), *Grundwissen der Sprachheilpädagogik und Sprachtherapie* (S. 29–39). Stuttgart: Kohlhammer.
- Mußmann, J. (2020). *Inklusive Sprachförderung in der Grundschule* (2. Aufl.). München: UTB. doi: 10.36198/9783838555119
- Porzolt, F., Galito, N., Toledo-Arruda, A., Thomaz, T., Moraes, C., Guerra, T., Leão, M., Migowski, A., da Silva, A. R. A., & Weiß, C. (2015). Efficacy and effectiveness trials have different goals, use different tools, and generate different messages. *Pragmatic and Observational Research*, 47. doi: 10.2147/por.s89946
- Rauch, D., & Hartig, J. (2018). Warum sollten Sprachförderkonzepte überprüft werden? Evidenzbasierung als Prinzip der Qualitätssicherung. In M. Becker-Mrotzek, H.-J. Roth, M. Hasselhorn, & P. Stanat (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung umsetzen und überprüfen* (1. Aufl., S. 65–77). Stuttgart: Kohlhammer.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, 312(7023), 71–72. doi: 10.1136/bmj.312.7023.71
- Singal, A. G., Higgins, P. D. R., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, 5. doi: 10.1038/ctg.2013.13
- Smith-Lock, K. M., Leita, S., Lambert, L., & Nickels, L. (2013). Effective intervention for expressive grammar in children with specific language impairment. *International Journal of Language and Communication Disorders*, 48(3), 265–282. doi: 10.1111/1460-6984.12003
- Souvignier, E., & Hasselhorn, M. (2023). Transfer und Implementation. In M. Becker-Mrotzek, I. Gogolin, H.-J. Roth, & P. Stanat (Hrsg.), *Grundlagen der sprachlichen Bildung* (Vol. 10, Sprachliche Bildung, S. 291–298). Münster, New York: Waxmann. doi: 10.25656/01:32014
- Spreer, M. (2020). Sprachsensibler Fachunterricht plus X: Notwendige Spezifizierungen der Sprachsensibilität zur unterrichtlichen Unterstützung von Schülerinnen und Schülern mit dem Förderschwerpunkt Sprache. *Sprachförderung und Sprachtherapie*, 1, 11–15.
- Streiner, D. L. (2002). The 2 “Es” of Research: Efficacy and Effectiveness Trials. *The Canadian Journal of Psychiatry*, 47(6), 552–556. doi: 10.1177/070674370204700607
- Thorpe, K. E., Zwarenstein, M., Oxman, A. D., Treweek, S., Furlong, C. D., Altman, D. G., Tunis, S., Bergel, E., Harvey, I., Magid, D. J., & Chalkidou, K. (2009). A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology*, 62(5), 464–475. doi: 10.1016/j.jclinepi.2008.12.011

- Wieland, L. S., Berman, B. M., Altman, D. G., Barth, J., Bouter, L. M., D'Adamo, C. R., Linde, K., Moher, D., Mullins, C. D., Treweek, S., Zwarenstein, M., & Witt, C. (2017). Rating of Included Trials on the Efficacy–Effectiveness Spectrum: development of a new tool for systematic reviews. *Journal of Clinical Epidemiology*, 84, 95–104. doi:10.1016/j.jclinepi.2017.01.010
- Wiehe, L.*, Råling, R.* & Stegenwallner-Schütz, M. (2024) A systematic review of effectiveness and efficacy of language-specific intervention methods for learners with heterogeneous language abilities in school-based settings. *PROSPERO* 2024, CRD42024569673. Abgerufen von https://www.crd.york.ac.uk/prospERO/display_record.php?ID=CRD42024569673 [16.12.2024]* gemeinsame Erstautor*innenschaft
- Zwarenstein M. (2017). 'Pragmatic' and 'explanatory' attitudes to randomised trials. *Journal of the Royal Society of Medicine*, 110(5), 208–218. doi: 10.1177/0141076817706303

Zu den Autor:innen

Lea Wiehe studierte Patholinguistik (B.Sc.) und Rehabilitationspädagogik (M.A.) und promoviert derzeit in den Rehabilitationswissenschaften an der Humboldt-Universität zu Berlin zu dem Thema Blickbewegungen beim Lesen bei Kindern mit und ohne Lesebeeinträchtigung. Ihre Forschungsschwerpunkte liegen im Bereich der erwerbsbedingten Beeinträchtigungen der Schriftsprache sowie der evidenzbasierten Förderung von Kindern und Jugendlichen mit (schrift-)sprachlichen Beeinträchtigungen in pädagogischen Settings. Derzeit arbeitet sie an der Universität Potsdam in der Grundschulpädagogik Deutsch.

Romy Råling ist Diplom-Patholinguistin und hat am Linguistik Department der Universität Potsdam im Fach Kognitionswissenschaft zum Thema Erwerbsalter (age of acquisition) and Typikalität (semantic typicality) promoviert. Derzeit ist sie Gastdozentin zur Vertretung der Professur im Arbeitsbereich Sonderpädagogik mit dem Schwerpunkt „Entwicklung des Lernens“ im Fachbereich Erziehungswissenschaft und Psychologie an der Freien Universität Berlin. Ihre Forschungsschwerpunkte liegen in den Bereichen der Diagnostik und Therapie von Sprachentwicklungsstörungen sowie der Evidenzbasierung von Interventionsmethoden.

Maja Stegenwallner-Schütz hat an der Universität Potsdam Patholinguistik (B.Sc.) und Experimental Clinical Linguistics (IECL) studiert und im Fach Kognitionswissenschaft zum Thema Development of Syntactic and Pragmatic Aspects of Language in Children with Developmental Disorders promoviert. Seit dem Wintersemester 2022/23 ist sie Gastprofessorin für Pädagogik bei Beeinträchtigungen der Sprache und Kommunikation an der Humboldt-Universität zu Berlin und ab dem SoSe 2025 Juniorprofessorin für Förderpädagogik/Sprache an der Universität Koblenz. Ihre Forschungsschwerpunkte liegen in den Bereichen Sprachentwicklung, -diagnostik und -förderung im schulischen Kontext.

Korrespondenzadressen

Lea Wiehe (lea.wiehe@uni-potsdam.de)
Universität Potsdam
Department Grundschulpädagogik
Grundschulpädagogik/Deutsch
Universitätscampus II Golm
Haus 16
Karl-Liebknecht-Str. 24-25
14476 Potsdam OT Golm

Romy Råling (romy.raeling@hu-berlin.de)
Humboldt-Universität zu Berlin
Kultur-, Sozial- und Bildungswissenschaftliche Fakultät
Institut für Rehabilitationswissenschaften
Pädagogik bei Beeinträchtigungen der Sprache und Kommunikation
Unter den Linden 6
10099 Berlin

Maja Stegenwallner-Schütz (stegenwa@uni-koblenz.de)
Universität Koblenz
Institut für Förderpädagogik
Universitätsstraße 1
D-56070 Koblenz

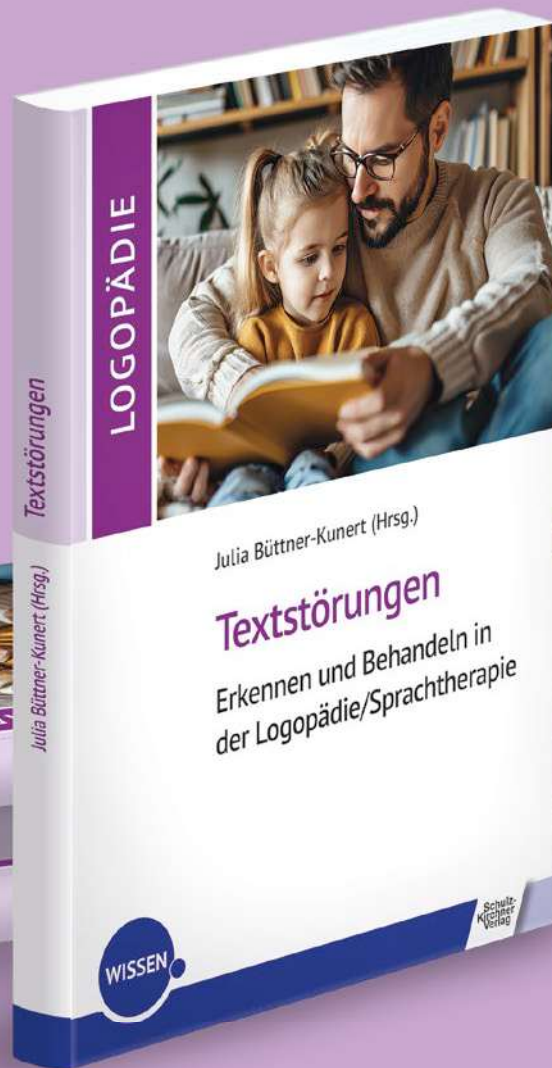
Indikator	Gartlehner et al. (2006)	Flay et al. (2005) Standards der Society of Prevention Research	Gottfredson et al. (2015)	Singal et al. (2014)	Lake et al. (2020)	Thorpe et al. (2009) PRECIS	Loudon et al. (2015) PRECIS-2	Alphas & Bossie (2016) ASPECT-R	Wieland et al. (2017)		Vorkommenshäufigkeit von Indikator in Veröffentlichungen
									Prä-Delphi	Post-Delphi	
Provider (Spezialisierung Förderkraft)				Item "Providers" (Hochgradig spezialisiert und ausgebildet)	Item "2" (Therapeut:innen abhängig vom Forscher:innenteam)	Item "Experimental intervention practitioner expertise" (Erfahrene, ausgewählte Behandler:innen)	Item "Organisation" (Hochgradig spezialisierte Fachkraft im Vergleich zu den üblichen Bedingungen)	Item "Medical Practice Setting/ Practitioner Expertise - Experimental" (Hohe Erfahrung durch Fachkräfte)	Item „Setting and Interventions: Organization“ (Hoher Spezialisierungsgrad, Erhöhung des Personalbestands und der Ressourcen)		6 von 9
						Item "Experimental intervention practitioner expertise" (Erfahrene, ausgewählte Behandler:innen)		Item "Medical Practice Setting/ Practitioner Expertise - Comparison" (Hohe Erfahrung durch Fachkräfte)			
Sample (Population)	Item "2" (Strenge Auswahlkriterien)	Item "Defined Sample" (Stichproben- und Rekrutierungsbericht)		Item „Study population“ (Homogene Teilnehmendenzusammensetzung durch Ausschlusskriterien)		Item "Participants: Participant eligibility criteria" (Strenge Auswahlkriterien)	Item "Eligibility" (Wenig Ähnlichkeit zur üblichen Praxis)	Item "Participant eligibility criteria" (Strenge Auswahlkriterien)	Item „Participants: Eligibility“ (Ausschlusskriterien mittels Tests, unüblich für gängige Praxis)	Item „Participants characteristics“ (Homogene Teilnehmendenzusammensetzung durch Ausschlusskriterien und Auswahlprozess, unüblich für gängige Praxis)	8 von 9
Intervention (Stringenz des Förderprotokolls)				Item „Intervention“ (Strenge Anweisungen, Standardisierung, hohe Kontrolle parallel verlaufender Behandlungen)		Item „Interventions and expertise: Experimental intervention Flexibility“ (Unflexibles und strenges Behandlungsprotokoll)	Item "Flexibility (delivery)" (Strenge Behandlungsprotokoll)	Item „Intervention Flexibility - Experimental“ (Unflexibles und strenges Behandlungsprotokoll)	Item „Setting and Interventions: Flexibility (delivery)" (Unflexibles, strenges Behandlungsprotokoll)	Item „Flexibility of intervention(s)" (Strenge Behandlungsprotokoll, hohe Kontrolle von parallel verlaufenden Behandlungen)	6 von 9
						Item "Interventions and expertise: Comparison intervention" (Eingeschränkte Flexibilität)		„Intervention Flexibility - Comparison“ (Eingeschränkte Flexibilität)	Item „Setting and Interventions: Flexibility of the comparison intervention“ (Begrenzte Flexibilität)	Item "Flexibility (adherence)" (Ausschlussmöglichkeit und Maßnahme zur Erhöhung der Compliance)	

Indikator	Gartlehner et al. (2006)	Flay et al. (2005) Standards der Society of Prevention Research	Gottfredson et al. (2015)	Singal et al. (2014)	Lake et al. (2020)	Thorpe et al. (2009) PRECIS	Loudon et al. (2015) PRECIS-2	Alphs & Bossie (2016) ASPECT-R	Wieland et al. (2017)		Vorkommenshäufigkeit von Indikator in Veröffentlichungen
									Prä-Delphi	Post-Delphi	
Outcome (Ergebnismessung)	Item „3“ (Objektive und subjektive Maße des Behandlungsergebnisses)	Item „Outcomes—What is Measured“ (Messung von Ergebnissen der öffentlichen Gesundheit oder des Verhaltens)	Item „Measures and their properties“ (Messung von Ergebnissen der öffentlichen Gesundheit; psychometrisch solide Messungen oder Informationsquellen)			Item „Primary trial outcome“ ((In) direktes Maß für ummittelbares Behandlungsergebnis, für Messung ggf. Training notwendig)	Item „Primary outcome“ (Geringer Grad der Relevanz für Teilnehmende)	Item „Primary Trial Outcomes“ (Direkte und sofortige Interventionsergebnisse, für Messung ggf. Training notwendig, Ergebnisart ggf. ungewöhnlich in Praxis)	Item „Other design aspects: Primary outcome“ (Behandlungsergebnis nicht verfügbar in üblicher Praxis oder Messung zu einem ungewöhnlich frühen Zeitpunkt)		7 von 9
		Item „Outcomes—Measurement Properties“ (psychometrisch solide Messung)									
Setting	Item „1“ (Geringe Ähnlichkeit zu üblichen Bedingungen)		Item „Question“ (Wirksamkeit der Intervention unter Idealbedingungen)	Item „Setting“ (Ressourcenintensives Setting mit Idealbedingungen)	Item „1“ (Geringe Ähnlichkeit zu üblichen Routinen der klinischen Praxis)		Item „Setting“ (Geringe Ähnlichkeit zu üblichen Bedingungen)	Item „Medical Practice Setting/ Practitioner Expertise - Experimental“ (Hohe Erfahrung in Setting)	Item „Setting and Interventions: Setting“ (Spezialisierteres Setting als übliches Setting)	Item „Clinical relevance of experimental and comparison intervention(s)“ (Maßnahmen und -dauer untypisch für übliche Gesundheitsversorgung)	7 von 9
Follow-up		Item „Outcomes“ (Mindestens ein Follow-Up)	Item „Measures and Their Properties“ (Mindestens ein Follow-Up)			Item „follow-up and outcomes“ ((In) direktes Maß für ummittelbares Behandlungsergebnis, für Messung ggf. Training notwendig)	Item „Follow-up“ (Umfangreiche Follow-Ups)	Item „Follow-Up Intensity / Duration“ (Häufigkeit und längere Konsultationen als in üblicher Praxis)	Item „Other design aspects: Follow-up“ (Häufigere oder längere Konsultationen als in üblicher Praxis)		6 von 9

Indikator	Gartlehner et al. (2006)	Flay et al. (2005) Standards der Society of Prevention Research	Gottfredson et al. (2015)	Singal et al. (2014)	Lake et al. (2020)	Thorpe et al. (2009) PRECIS	Loudon et al. (2015) PRECIS-2	Alphs & Bossie (2016) ASPECT-R	Wieland et al. (2017) Prä-Delphi Post-Delphi	Vorkommenshäufigkeit von Indikator in Veröffentlichungen
Adhärenz und Intention-to-treat-Analyse (ITT)	Item "7" (ITT)	Item „Measures“ (Compliance-, Adhärenz-, Akzeptanzmessung)	Item „Measures and their properties“ (Compliance-, Adhärenz-, Akzeptanzmessung)			Item "Participant compliance with "prescribed" intervention" (Engmaschiges Monitoring der Interventions-treue) Item "Practitioner adherence to study protocol" (Engmaschiges Monitoring der Interventions-treue) Item "Analysis of primary Outcome" (ITT)	Item "Flexibility (adherence)" (Ge-ringe Flexibilität im Monitoring und strenge Einhaltung der Interventions-treue) Item "Primary ana-lysis" (ITT)	Item "Participant compliance" (Engmaschiges Monitoring der Interventions-treue)	Item "Other design aspects: Practitioner adherence to study protocol" (Engmaschiges Monitoring der Interventions-treue)	6 von 9
Verblindung		Item "Clarity of Causal Inference: Assignment" (Randomisierung wenn möglich)	Item "Clarity of Causal Inference: Assignment" (Kontrollkondition, Gruppenzuweisung muss Bias reduzieren)						Item „Other design aspects: Blinding of participants and personnel to the intervention“ (Verblindung von Teilnehmenden oder Untersucher:innen)	3 von 9
Rekrutierung							Item „Recruitment“ (hoher Aufwand für Rekrutierung)		Item "Participants: Recruitment" (Gezielte Maßnahmen zur Rekrutierung)	2 von 9
Studiendauer	Item "4" (Kürzere Behandlungsdauer)								Item „Other design aspects: Study duration“ (kürzere Behandlungsdauer als in gängiger Praxis)	2 von 9
Bewertung von unerwünschten Ereignissen	Item "5" (Erhebung von unerwünschten Ereignissen)		Item „efficacy claims“ (Bericht aller Ergebnisse: negative, positive und nicht-signifikanter) Item „Measures and their properties“ (Messung iatrogener Effekte)							2 von 9



NEU



Julia Büttner-Kunert (Hrsg.),
1. Auflage 2025,
kartoniert: ISBN 978-3-8248-1351-3, 208 Seiten,
E-Book: ISBN 978-3-8248-9897-8, PDF,
EUR 32,00 [D]

Textstörungen

Erkennen und Behandeln in der Logopädie/Sprachtherapie

Geschriebene oder gehörte Texte sind in unserem Alltag allgegenwärtig. Ohne den kompetenten Umgang mit sprachlichen Makrostrukturen sind Kinder und Erwachsene mit Sprachstörungen in jeder Lebensphase in ihrer Teilhabe beeinträchtigt. Bislang fehlt ein Buch für die Logopädie/Sprachtherapie, das gezielt die Textebene behandelt und anwendungsbezogene Handlungsempfehlungen für die Diagnostik und Therapie sowohl für Kinder als auch für Erwachsene bietet.

Der vorliegende Sammelband schließt diese Lücke: Er enthält die Grundlagen der Textlinguistik für die Logopädie/Sprachtherapie und gibt einen Überblick über Methoden und Verfahren für die Diagnostik auf Textebene.

In sechs Beiträgen behandeln ausgewiesene Expert:innen die Anwendung von Diagnostikverfahren und stellen geeignete Therapieansätze bei unterschiedlichen Zielgruppen, untergliedert nach Kindes-/Jugendalter und Erwachsenenalter, vor. Die Therapieansätze werden zusätzlich anhand von Praxisbeispielen konkretisiert und veranschaulicht.

WISSEN



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

Lieferung versandkostenfrei innerhalb Deutschlands



Schulz-
Kirchner
Verlag