

Stanat, Petra; Edele, Aileen

Hypothesenprüfende Forschung

Becker-Mrotzek, Michael [Hrsg.]; Gogolin, Ingrid [Hrsg.]; Roth, Hans-Joachim [Hrsg.]; Stanat, Petra [Hrsg.]: Grundlagen der sprachlichen Bildung. Münster ; New York : Waxmann 2023, S. 261-275. - (Sprachliche Bildung; 10)



Quellenangabe/ Reference:

Stanat, Petra; Edele, Aileen: Hypothesenprüfende Forschung - In: Becker-Mrotzek, Michael [Hrsg.]; Gogolin, Ingrid [Hrsg.]; Roth, Hans-Joachim [Hrsg.]; Stanat, Petra [Hrsg.]: Grundlagen der sprachlichen Bildung. Münster ; New York : Waxmann 2023, S. 261-275 - URN: urn:nbn:de:01111-pedocs-320129 - DOI: 10.25656/01:32012

<https://nbn-resolving.org/urn:nbn:de:01111-pedocs-320129>

<https://doi.org/10.25656/01:32012>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nd/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-Licence: <http://creativecommons.org/licenses/by-nd/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work in the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to alter or transform this work or its contents at all.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Hypothesenprüfende Forschung

Ist Mehrsprachigkeit eine Hürde für Bildungserfolg oder entwickeln Kinder, die mehrsprachig aufwachsen, sogar kognitive Vorteile? Sollten neu zugewanderte Schüler:innen, die „quer“ in das deutsche Schulsystem einsteigen, besser erst einmal separat unterrichtet werden oder sollten sie von Anfang an am Unterricht der Regelklassen teilnehmen? Welche Maßnahmen der Zweitsprachförderung sind effektiv? Sowohl in der Forschung als auch in der Praxis zu sprachlicher Bildung existieren vielfältige Annahmen über Zusammenhänge zwischen Merkmalen, über Unterschiede zwischen Gruppen und über Effekte von Maßnahmen. Diese Annahmen können mehr oder weniger gut theoretisch fundiert und empirisch abgesichert sein. Nicht selten werden im Bildungsbereich, auch in Bezug auf Fragen der sprachlichen Bildung, mit Nachdruck Überzeugungen vertreten, für die es keine gute empirische Rechtfertigung gibt und die sich nicht selten sogar widersprechen.

Daher ist es wichtig, dass Forschung zu sprachlicher Bildung solche Annahmen bzw. Hypothesen überprüft und belastbare Befunde generiert. Wie man bei der Prüfung von Hypothesen vorgehen kann, welche Herausforderungen sich dabei stellen und wie man diesen Herausforderungen begegnen kann, ist Gegenstand des vorliegenden Kapitels (für eine kompakte Darstellung verschiedener Forschungsdesigns der quantitativen Sozialforschung siehe z. B. Stein, 2014). Der Fokus liegt dabei auf der Überprüfung von Kausalhypothesen, die Aussagen über Ursachen und Wirkungen beinhalten (z. B. „sprachsensibler Fachunterricht [Ursache] erhöht den Lernerfolg [Wirkung] von Kindern mit Deutsch als Zweitsprache“). Annahmen über kausale Zusammenhänge sind sowohl aus wissenschaftlicher als auch aus praktischer Sicht besonders relevant, da sie oft Implikationen für die Wahl von Handlungsalternativen haben, z. B. für die Gestaltung sprachlicher Bildungsprozesse. Gleichzeitig ist es besonders anspruchsvoll, Kausalannahmen zu prüfen (Kühnel & Dingelstedt, 2014). In der folgenden Darstellung wird auf verschiedene Beispiele zur Prüfung von Hypothesen aus der Forschung zu sprachlicher Bildung Bezug genommen, um das jeweils beschriebene Vorgehen zu illustrieren. Zwei Beispiele, die zur Illustration der in diesem Kapitel erläuterten Konzepte besonders geeignet sind, werden etwas ausführlicher dargestellt (siehe Kästen) und im Text wiederholt aufgegriffen. Da die Konzepte sukzessive eingeführt werden, empfiehlt es sich, die Darstellung der Beispielstudien zweimal zu lesen – einmal vor und einmal nach Lektüre des Textes.

Beispielstudie 1: Evaluation einer Lehrkräftefortbildung zur Förderung zweisprachlicher Kompetenzen von Schüler:innen (Babinski, Amendum, Knotek, Sánchez & Malone, 2018)

Hintergrund: Obwohl Kinder und Jugendliche, die in ihrer Familie eine andere Sprache sprechen, oft besondere Unterstützung beim Erwerb der Unterrichtssprache benötigen, sind Lehrkräfte oft nicht dafür ausgebildet, Zweitsprachler:innen zu unterrichten. Hierfür sind effektive Fortbildungsangebote wichtig.

Ziel der Studie: Die in den USA durchgeführte Studie untersuchte die Wirksamkeit einer Fortbildung für Lehrkräfteteams, die sich jeweils aus einer Sprachlehrkraft im Bereich *English as a Second Language* (ESL) und aus zwei Fachlehrkräften zusammensetzten. Ziel der Fortbildungsmaßnahme war es, die englischen Sprachkompetenzen von Kindern aus spanischsprachigen Familien zu verbessern. Die Maßnahme umfasste drei Elemente: (1) Systematische Zusammenarbeit von Sprach- und Fachlehrkräften, (2) Vermittlung von Unterrichtsstrategien zur Förderung englischsprachiger Kompetenzen der Schüler:innen und (3) Vermittlung von Ansätzen zur Einbeziehung kultureller Ressourcen der Schüler:innen.

Studiendesign: Die Wirksamkeit der Fortbildung wurde anhand eines experimentellen Designs untersucht. Sprachlehrkräfte wurden gebeten, je zwei Fachlehrkräfte zu gewinnen, um mit ihnen als Team an der Studie teilzunehmen. Aus diesen Lehrkräfteteams wurden Paare gebildet, die sich hinsichtlich der sozioökonomischen Zusammensetzung der Schülerschaft ihrer Schule möglichst ähnlich waren (*Matching*). Anschließend wurde pro Paar jeweils ein Team per Zufall der Experimentalgruppe oder der Kontrollgruppe zugewiesen. Die Lehrkräfteteams in der Experimentalgruppe besuchten im Sommer eine fünftägige Fortbildung (35 Stunden) und nahmen im Verlauf des Schuljahres an vier inhaltlichen Anwendungsmodulen (je 6 Stunden) teil. Darüber hinaus konnten sie Coaching-Angebote wahrnehmen und sich in wöchentlichen Treffen mit ihrem Team abstimmen. Insgesamt nahmen 44 Lehrkräfte und 105 Schüler:innen mit Englisch als Zweitsprache vollständig an der Untersuchung teil. Um die Effekte der Fortbildungsmaßnahme zu prüfen, wurde zum einen der Unterricht der Lehrkräfte im Lauf des Schuljahres dreimal mit Hilfe standardisierter Instrumente beobachtet und zum anderen wurden zu Beginn (Prätest) und zum Ende (Posttest) des Schuljahres die englischen Sprachkompetenzen der Schüler:innen mit Leistungstests erfasst.

Ergebnisse: Die Datenanalysen zeigten, dass die Sprachlehrkräfte der Experimentalgruppe einige der Unterrichtsstrategien, die ihnen in der Fortbildung vermittelt worden waren, häufiger anwendeten als die Sprachlehrkräfte der Kontrollgruppe. Zudem fiel der Lernzuwachs bei den Schüler:innen, die von Lehrkräften der Experimentalgruppe unterrichtet worden waren, bei zwei der sieben getesteten Aspekte der Sprachkompetenz größer aus. Vor allem Schüler:innen mit geringen Ausgangskompetenzen in Englisch profitierten von der Maßnahme.

Einordnung: Die in der Studie geprüften Annahmen konnten weitgehend bestätigt werden. Durch das experimentelle Untersuchungsdesign und durch weitere Maßnahmen zur Kontrolle möglicher Störvariablen (z.B. Kontrolle der Umsetzung der Unterrichtsstrategien durch Beobachtungen, Kontrolle der Zusammensetzung der Schüler:innenschaft in der Experimental- und Kontrollgruppe durch Matching) konnten alternative Erklärungen für die Ergebnisse weitgehend ausgeschlossen werden. Allerdings ist die Stichprobe der Studie recht klein. Dies erhöht die Wahrscheinlichkeit, dass die Experimental- und Kontrollgruppen trotz der Zufallszuweisung doch nicht in allen Merkmalen, die für die Ergebnisse relevant sein könnten, vor Beginn der Studie vergleichbar waren, und es erschwert die Absicherung von kleineren Effekten.

Beispielstudie 2: Effekte von Mehrsprachigkeit auf Kognition (Crivello, Kuzyk, Rodrigues, Friend, Zesiger & Poulin-Dubois, 2016)

Hintergrund: Studien haben immer wieder gezeigt, dass Mehrsprachige im Durchschnitt nicht nur bestimmte sprachliche Vorteile, sondern auch bestimmte kognitive Vorteile aufweisen (Bialystok, 2017). Dies betrifft vor allem kognitive Exekutivfunktionen, die für den Umgang mit Aufmerksamkeitskonflikten wichtig sind¹. Allerdings kann aus dem einfachen Vergleich von Leistungen mehrsprachiger und einsprachiger Personen bei entsprechenden Aufgaben nicht eindeutig auf einen kausalen Effekt der Mehrsprachigkeit geschlossen werden, da auch andere Gruppenunterschiede für die beobachteten kognitiven Vorteile der Mehrsprachigen verantwortlich sein könnten. So wäre es beispielsweise denkbar, dass sich bei mehrsprachigen Personen nicht die Exekutivfunktionen besser entwickeln, sondern dass umgekehrt Personen mit ausgeprägten Exekutivfunktionen eher dazu neigen, ihre Mehrsprachigkeit aufrechtzuerhalten bzw. diese weiterzuentwickeln.

Ziel der Studie und Studiendesign: Um zu prüfen, ob tatsächlich Mehrsprachigkeit für die beobachteten kognitiven Vorteile verantwortlich ist, untersuchte die in Kanada und den USA mit Kleinkindern durchgeführte Studie mit längsschnittlichem Design, welchen Effekt ein zunehmender Grad von Bilingualität auf bestimmte Exekutivfunktionen hat. Dazu wurden die sprachlichen Gewohnheiten und der Wortschatz von 43 monolingualen Kindern und 39 bilingualen Kindern, die in der Familie Englisch und Französisch sprachen, im Alter von 2 Jahren (Messzeitpunkt 1) und 2,5 Jahren (Messzeitpunkt 2) erhoben. Zu Mess-

1 Ein Beispiel für eine Aufgabe mit Aufmerksamkeitskonflikt ist der *Simon Task*, bei dem Personen auf einer Tastatur z.B. immer eine linke Taste drücken sollen, wenn auf dem Bildschirm ein rotes Symbol zu sehen ist, und eine rechte Taste, wenn ein grünes Symbol zu sehen ist, wobei die Symbole an unterschiedlichen Stellen auf dem Bildschirm erscheinen. Generell ist die Reaktion verzögert, wenn das Symbol, bei dem die linke Taste gedrückt werden soll, rechts auf dem Bildschirm erscheint und umgekehrt. Im Durchschnitt ist die Verzögerung bei mehrsprachigen Personen jedoch weniger ausgeprägt als bei einsprachigen Personen.

zeitpunkt 2 wurden zudem mehrere Aufgaben zur Erfassung von Exekutivfunktionen eingesetzt. Die Forschenden prüften die Hypothese, dass Kinder, die im Untersuchungszeitraum den Grad ihrer Bilingualität ausbauen konnten, kognitive Konfliktaufgaben besser lösen als Monolinguale und als Bilinguale mit weniger ausgeprägter Bilingualität. Der Grad der Bilingualität wurde anhand der Zahl der Wörter, die ein Kind in beiden Sprachen beherrscht, operationalisiert. Eine der kognitiven Konfliktaufgaben bestand darin, zunächst mehrfach kleine Klötze in einen kleinen Eimer und große Klötze in einen großen Eimer zu legen (*Pre-switch-Phase*) und anschließend die umgekehrte Zuordnung vorzunehmen (*Post-switch-Phase*). Gezählt wurde die Anzahl richtiger Zuordnungen in der Post-switch-Phase.

Ergebnisse: Übereinstimmend mit der Hypothese konnten die bilingualen Kinder die kognitiven Konfliktaufgaben besser lösen als die monolingualen Kinder; bei anderen kognitiven Fähigkeiten, wie etwa dem Arbeitsgedächtnis, hatten sie hingegen keine Vorteile. In der Gruppe der bilingualen Kinder konnten diejenigen, die ihre Bilingualität im Untersuchungszeitraum ausgebaut hatten, die kognitiven Konfliktaufgaben besser lösen als Kinder, die ihre Bilingualität weniger stark weiterentwickelt hatten.

Einordnung: Studien mit nicht-experimentellem Design können die angenommenen kausalen Effekte weniger eindeutig untermauern als Studien mit experimentellem Design (vgl. Beispielstudie 1). Da es aus offensichtlichen ethischen Gründen nicht möglich ist, die Familiensprache experimentell zu variieren, müssen zur Prüfung von Hypothesen zu Effekten familiärer Mehrsprachigkeit jedoch nicht-experimentelle Untersuchungsdesigns eingesetzt werden. Eine Stärke der beschriebenen Studie besteht in ihrer längsschnittlichen Anlage. Das Design wäre allerdings noch stärker gewesen, wenn die Forschenden die Exekutivfunktionen auch schon zu Messzeitpunkt 1 erhoben und gezeigt hätten, dass sich der Unterschied zwischen der monolingualen und der bilingualen Gruppe über die Zeit vergrößert und dies in besonderem Maße für diejenigen bilingualen Kinder gilt, die ihre Bilingualität ausbauten.

1. Kennzeichen hypothesenprüfender Forschung

Hypothesenprüfende Forschung verfolgt das Ziel, zu überprüfen, ob vorab aufgestellte Annahmen tatsächlich zutreffen. Bei Kausalhypothesen beziehen sich die Annahmen auf die Existenz von Effekten und meistens auch auf deren Richtung (z. B. die Teilnahme der Lehrkräfte an der Fortbildung in Beispielstudie 1 wirkt sich positiv auf die englische Sprachkompetenz ihrer Schüler:innen aus). Es wird in hypothesenprüfenden Studien mindestens eine Alternativhypothese aufgestellt, die aufgrund theoretischer Überlegungen zutreffen sollte (z. B. Mehrsprachigkeit verbessert die Fähigkeit zur Aufmerksamkeitskontrolle), und eine Nullhypothese, die davon ausgeht, dass dies nicht zutrifft (z. B. Mehrsprachigkeit ist irrelevant für die Fähigkeit zur Aufmerksamkeitskontrolle). Nachdem die jeweilige Studie

durchgeführt wurde, wird mithilfe statistischer Analysen entschieden, wie wahrscheinlich es ist, dass das jeweilige Ergebnis zustande gekommen wäre, wenn tatsächlich die Nullhypothese gilt. Ist dies sehr unwahrscheinlich, gilt die Alternativhypothese als bestätigt und die Nullhypothese wird verworfen.

Von hypothesenprüfenden Studien (auch als *explanative* Studien bezeichnet) abzugrenzen sind *explorative* Studien, die einen Sachverhalt erkunden und dazu dienen, Forschungsfragen oder Hypothesen zu generieren, sowie *deskriptive* Studien, die feststellen sollen, wie verbreitet bestimmte Merkmale bzw. Merkmalsausprägungen sind (Döring & Bortz, 2016).

Hypothesen sollten aus Theorien, die für die jeweilige Fragestellung relevant sind, abgeleitet werden, oder sich zumindest auf gesicherte empirische Befunde stützen; sie sollten also nicht „aus dem Bauch heraus“, ohne theoretisch fundierte Begründung, formuliert werden. Hypothesen beziehen sich auf einzelne Merkmale, die unterschiedliche Ausprägungen annehmen können und als „Variablen“² bezeichnet werden. Es geht in hypothesenprüfender Forschung um die Identifikation von Beziehungen dieser Merkmale untereinander, nicht um eine umfassende Beschreibung von Personen- bzw. Personengruppen, von Interaktionsprozessen o. ä. (Döring & Bortz, 2016).

Wichtig ist, dass die Hypothesen empirisch überprüfbar sind und ihr Geltungsbereich abgesteckt wird. Die Hypothesen müssen also so eindeutig und konkret formuliert sein, dass sie sich durch beobachtete Daten bestätigen oder widerlegen lassen, und es muss klar sein, für welche Personen und unter welchen Bedingungen die Annahmen gelten sollten. Hierbei sind zwei Ebenen zu unterscheiden: die konzeptuelle Ebene und die operationale Ebene. Auf der konzeptuellen Ebene wird die jeweilige Hypothese anhand von theoretischen Konzepten formuliert, die sich aus den theoretischen Grundlagen ergeben (z. B. in Beispielstudie 2: Bilingualität und Exekutivfunktionen). In der Regel sind diese Konstrukte nicht direkt messbar. Um die Hypothese prüfen zu können, müssen die Konstrukte in messbare Größen übersetzt werden (z. B. in Beispielstudie 2: Anzahl der Wörter, die ein Kind in beiden Sprachen beherrscht, und Lösung kognitiver Konfliktaufgaben). Diese als *Operationalisierung* bezeichnete Übersetzungsleistung ist für die Aussagekraft der Ergebnisse von zentraler Bedeutung.

Studien, die auf die Prüfung von Kausalhypothesen abzielen, müssen so weit wie möglich sicherstellen, dass sich die Ergebnisse tatsächlich kausal interpretieren lassen und sich die vorliegenden Befunde nicht anders erklären lassen als durch die angenommenen Mechanismen. Nur wenn sich die Veränderung einer *abhängigen Variable* (Wirkung) eindeutig auf den Einfluss einer *unabhängigen Variable* (Ursache) zurückführen lässt, ist eine Studie *intern valide* und kann die aufgestellte Kausalannahme bestätigen. So sollte etwa in der Beispielstudie 1 sichergestellt werden, dass positive Effekte auf die Sprachkompetenz der

2 Bei einer Variable handelt es sich um ein Merkmal, das mindestens zwei einander ausschließende Ausprägungen hat, wie Teilnahme an einer Sprachfördermaßnahme (z. B. Teilnahme vs. keine Teilnahme) oder Familiensprache (z. B. nur Deutsch vs. Deutsch und eine andere Sprache vs. nur eine andere Sprache).

Schüler:innen eindeutig auf die Fortbildungsmaßnahme zurückgehen und nicht etwa darauf, dass Lehrkräfte, die an der Fortbildung teilnehmen, allgemein besonders engagiert sind und besonders guten Unterricht machen. Solche Faktoren, die ebenfalls für die Veränderungen verantwortlich sein könnten, die aber nicht Gegenstand der Untersuchung sind, werden als *Störvariablen* bezeichnet. Wenn Effekte von Störvariablen nicht ausgeschlossen werden können, ist die interne Validität der Untersuchung – also ob die beobachteten Ausprägungen einer abhängigen Variablen auf die unabhängige Variable zurückzuführen sind – eingeschränkt. Entsprechend wird bei der Planung von Studien zur Überprüfung von Kausalhypothesen versucht, das Forschungsdesign so anzulegen, dass der Einfluss potenzieller Störvariablen ausgeschaltet oder zumindest kontrolliert werden kann.

Der Königsweg dazu ist das Experiment mit Zufallszuweisung der Untersuchungsteilnehmenden zu den Ausprägungen der unabhängigen Variablen. Allerdings ist eine solche Zufallszuweisung im Bildungsbereich oft nicht umsetzbar. Im Folgenden wird daher sowohl auf Fragestellungen eingegangen, die sich mit experimentellen Designs untersuchen lassen, als auch auf Fragestellungen, für deren Bearbeitung andere Designs herangezogen werden müssen, wobei der Fokus in diesem Kapitel auf quantitativen Studien liegt.

2. Fragestellungen und Forschungsdesigns

In der Forschung zu sprachlicher Bildung werden etliche Fragestellungen verfolgt, die sich auf theoretische Annahmen beziehen und entsprechend mit hypothesenprüfenden Studien untersucht werden sollten. Viele dieser Annahmen sind kausal. So wird etwa geprüft, ob bestimmte Maßnahmen der Sprachförderung wirksam sind (siehe Beispielstudie 1), oder es wird untersucht, wie sich Mehrsprachigkeit auf bestimmte kognitive Fähigkeiten auswirkt (siehe Beispielstudie 2). Weitere Fragestellungen, die Kausalhypothesen beinhalten, finden sich in verschiedenen anderen Kapiteln dieses Bandes. Drei grundlegende Forschungsdesigns, mit denen sich Kausalhypothesen prinzipiell untersuchen lassen, die sich aber in der internen Validität unterscheiden, werden im Folgenden skizziert (Döring & Bortz, 2016).

2.1 Experimentelle Designs

Zur Prüfung von Kausalhypothesen sind, wie bereits erwähnt, experimentelle Forschungsdesigns am besten geeignet, die hierbei als „Königsweg der Erkenntnis“ (Döring & Bortz, 2016, S. 194) gelten (siehe auch Eifler, 2014). Ziel echter Experimente (Englisch: *experimental studies* bzw. *randomized control trials*) ist es, eindeutige Belege für Effekte einer oder mehrerer unabhängiger Variablen auf eine oder mehrere abhängige Variablen zu generieren. Experimente zeichnen sich durch zwei zentrale Merkmale aus. Das erste Merkmal besteht darin, dass bei ex-

perimentellen Studien die Ausprägungen der unabhängigen Variablen, deren Wirkungen die Studie untersuchen soll, aktiv manipuliert bzw. systematisch variiert werden. Will man also etwa die Wirkungen einer bestimmten Fördermaßnahme überprüfen, wie in Beispielstudie 1, wird eine Experimentalgruppe dieser Maßnahme (auch als *Treatment* bezeichnet) ausgesetzt, eine Kontrollgruppe hingegen nicht. In diesem Fall ist die unabhängige Variable also die Teilnahme an der Maßnahme mit den Ausprägungen „Teilnahme“ versus „keine Teilnahme“. Als abhängige Variablen werden diejenigen Merkmale erfasst, auf die sich das Treatment auswirken soll, in Beispielstudie 1 unter anderem die Sprachkompetenz der Schüler:innen. Beim zweiten zentralen Merkmal echter Experimente handelt es sich um die zufällige (*randomisierte*) Zuweisung der Untersuchungsteilnehmenden zu den Ausprägungen der unabhängigen Variablen, die auch als „Untersuchungsbedingungen“ oder „Versuchsbedingungen“ bezeichnet werden. So wurde in Beispielstudie 1 per Zufall entschieden, welche Lehrkräfteteams die Fortbildung erhielten (Experimentalgruppe) und welche nicht (Kontrollgruppe)³. Um zu vermeiden, dass der Kontrollgruppe ein Treatment vorenthalten wird, das sich als wirksam erweist, erhält sie dieses häufig ebenfalls, allerdings erst nach Abschluss des Experiments (Wartekontrollgruppe). Die jeweiligen abhängigen Variablen werden in der Regel mindestens einmal vor dem Treatment (Prätest) und einmal nach dem Treatment (Posttest) erhoben sowie idealerweise zusätzlich zu noch späteren Zeitpunkten (Follow-up), um zeitverzögerte Effekte zu ermitteln und um Hinweise auf die Nachhaltigkeit der Wirkungen zu gewinnen.

Wenn die untersuchte Maßnahme die erwartete Wirkung hat, sollten sich die Experimental- und die Kontrollgruppe nach dem Treatment in den abhängigen Variablen unterscheiden. Dieser Unterschied kann jedoch nur dann kausal auf das Treatment zurückgeführt werden, wenn die beiden Gruppen vorher ähnlich zusammengesetzt waren – dies soll durch die randomisierte Zuweisung zu den Experimental- und Kontrollbedingungen gewährleistet werden. Bei kleineren Fallzahlen kann es aber auch bei einer Zufallszuweisung passieren, dass sich die Experimental- und Kontrollgruppe in relevanten Merkmalen unterscheiden, etwa wenn der Zufall mehrsprachige Kinder überwiegend der Experimentalgruppe und monolinguale Kinder überwiegend der Kontrollgruppe zuweist. Daher ist es auch bei experimentellen Studien wichtig, dass anhand von Daten einer Prätest-Erhebung geprüft wird, inwieweit die Vergleichbarkeit der Gruppen gewährleistet ist. Unter Umständen kann es zudem sinnvoll sein, die Gruppen im Hinblick auf relevante Merkmale zu parallelisieren, wie in Beispielstudie 1 (*Matching*).

Neben *personenbezogenen Störvariablen* können auch *untersuchungsbedingte Störvariablen* auftreten, etwa wenn die Experimental- und Kontrollgruppen nicht nur in Bezug auf die unabhängige Variable, sondern auch in Bezug auf andere Faktoren unterschiedlich behandelt werden, die sich ebenfalls auf die abhängige

3 In Beispielstudie 1 wurde der randomisierten Zuweisung der Lehrkräfteteams zu den Versuchsbedingungen noch eine Paarbildung vorgeschaltet, um die Vergleichbarkeit der Schüler:innen in der Experimental- und Kontrollgruppe sicherzustellen. Dieser Schritt ist nicht zwangsläufig Teil experimenteller Designs.

Variable auswirken können. In Beispielstudie 1 wäre das etwa dann der Fall, wenn in der Experimentalgruppe viel mit digitalem Unterrichtsmaterial gearbeitet wird, das besonders motivierend ist, in der Kontrollgruppe hingegen nicht. Dies könnte dazu führen, dass die Schüler:innen in der Experimentalgruppe aufmerksamer am Unterricht teilnehmen und sich ihre sprachlichen Kompetenzen aus diesem Grund besonders positiv entwickeln; mit der eigentlich interessierenden Fortbildungsmaßnahme hätte die Veränderung aber nichts zu tun. Ziel hypothesenprüfender Untersuchungen ist es, potenzielle Störvariablen so weit wie möglich auszuschalten oder sie zumindest zu erfassen und in den Analysen zu berücksichtigen.

Die randomisierte Zuweisung zu den Versuchsbedingungen einer experimentellen Studie kann auf verschiedenen Ebenen (z. B. Schüler:innen, Lehrkräfte, Schulklassen) erfolgen, wobei entscheidend ist, wo das Treatment ansetzt. Soll zum Beispiel die Wirkung eines Treatments untersucht werden, das mit Kindern individuell durchzuführen ist, wie etwa eine individuelle Zusatzförderung bei einer Leseschwäche, würden idealerweise die einzelnen Schüler:innen der Stichprobe, die für die Studie ausgewählt wurden, zufällig der Experimental- und der Kontrollgruppe zugewiesen werden. Sollen hingegen Effekte einer Förderung überprüft werden, die in Lerngruppen umzusetzen ist, erfolgt die Zuweisung zu den Untersuchungsbedingungen auf Lerngruppenebene. So ging letztlich auch die Beispielstudie 1 vor, in der die Lehrkräfteteams randomisiert der Experimental- und Kontrollgruppe zugewiesen wurden und mit ihnen die jeweiligen Lerngruppe.

2.2 Quasi-experimentelle Designs

Aufgrund der hohen internen Validität echter Experimente wäre es wünschenswert, ihr Potenzial zur Beantwortung von Fragestellungen zu sprachlicher Bildung verstärkt zu nutzen. Dies ist jedoch oft nicht möglich, so dass häufig auf andere Forschungsdesigns zurückgegriffen werden muss. Wenn etwa die Wirksamkeit von Interventionen bzw. Förderansätzen im Feld untersucht werden soll, aber keine randomisierte Zuweisung zu den Untersuchungsbedingungen möglich ist, muss auf ein quasi-experimentelles Design zurückgegriffen werden. Wie beim Experiment wird dabei eine unabhängige Variable aktiv manipuliert, indem z. B. manche Schüler:innen an einer Intervention teilnehmen (Experimentalgruppe) und andere nicht (Kontrollgruppe). Die Zuweisung zu den Gruppen erfolgt jedoch nicht randomisiert, sondern es werden existierende Gruppen untersucht. In Beispielstudie 1 etwa haben sich die Lehrkräfte auf die Zufallszuweisung zu den Untersuchungsbedingungen eingelassen, was jedoch nicht immer der Fall ist, etwa weil im Schuljahr, in dem die Untersuchung stattfinden soll, nur manche Lehrkräfte die zeitlichen Ressourcen haben, um an einer umfangreichen Fortbildungsmaßnahme teilzunehmen. Zudem können in der Regel nicht die einzelnen Schüler:innen zufällig einer Lehrkraft zugeordnet werden, die an der Maßnah-

me teilnimmt, sondern die gesamte Klasse wird gemeinsam mit ihrer Lehrkraft einer Untersuchungsbedingung zugeordnet. Dadurch unterscheiden sich die Experimental- und Kontrollgruppe häufig systematisch, etwa wenn besonders gut ausgebildete und/oder motivierte Lehrkräfte eher an der Experimentalbedingung teilnehmen. Wäre dies in Beispielstudie 1 der Fall gewesen, könnten Vorteile der Experimentalgruppe, die nach der Fortbildung beobachtet wurden, fälschlich als Effekt dieser Maßnahme gedeutet werden.

Aufgrund solcher personenbezogenen Störvariablen ist die interne Validität von Quasi-Experimenten insgesamt geringer als die echter Experimente und es muss ein höherer Aufwand betrieben werden, um solche Störeinflüsse einzudämmen. Daher ist es bei quasi-experimentellen Studien besonders wichtig, theoriegeleitet bzw. auf Basis früherer Befunde potenzielle Störvariablen zu identifizieren, zu erheben und soweit wie möglich zu kontrollieren. Naheliegende Störvariablen sind in Studien zur Wirksamkeit von Maßnahmen der Sprachförderung unter anderem die Ausbildung der Lehrkräfte – zum Beispiel könnten Lehrkräfte, die eine Ausbildung in „Deutsch als Zweitsprache“ (DaZ) haben, den jeweiligen Ansatz erfolgreicher umsetzen als Lehrkräfte ohne DaZ-Ausbildung –, aber auch die sprachlichen Voraussetzungen der Schüler:innen und ihr außerschulisches Sprachumfeld. Solche Aspekte sollten erfasst und bei der Auswertung der Studienergebnisse berücksichtigt werden. Alternativ können potenzielle Störvariablen konstant gehalten werden, etwa indem man keine bzw. ausschließlich DaZ-Lehrkräfte an der Studie teilnehmen lässt. Eine weitere Möglichkeit, Störvariablen zu kontrollieren, besteht darin, darauf zu achten, dass sie in beiden Bedingungen gleich verteilt sind (z. B. 30 % DaZ-Lehrkräfte). Auch das bereits erwähnte *Matching*, bei dem für jede teilnehmende Person der Experimentalgruppe ein „Zwilling“ für die Kontrollgruppe mit möglichst ähnlichen Ausprägungen auf potenziellen Störvariablen gesucht wird (z. B. hinsichtlich Familiensprache, Bildungsgrad der Eltern, Vorwissen), kann ein sinnvoller Umgang mit Störvariablen sein. Dennoch ist die Unsicherheit von Schlussfolgerungen darüber, was die beobachteten Gruppenunterschiede verursacht hat, bei quasi-experimentellen Untersuchungen größer als bei experimentellen Studien.

2.3 Nicht-experimentelle Designs

Manchmal ist nicht nur eine Randomisierung unmöglich, sondern auch die experimentelle Manipulation der interessierenden unabhängigen Variablen, da die Variablen, deren Wirkungen man bestimmen will, nicht bzw. nicht in ethisch vertretbarer Weise experimentell variiert werden können. Dies betrifft zum Beispiel das Merkmal, ob ein Kind mehrsprachig oder einsprachig aufwächst – darüber entscheiden die Familien bzw. deren Umstände und nicht Leiter:innen von Studien, die sich für die Effekte von Mehrsprachigkeit interessieren. Daher werden zur Überprüfung von Kausalhypothesen zu Fragen sprachlicher Bildung oft auch nicht-experimentelle Designs verwendet, in denen existierende Gruppen bezüg-

lich der interessierenden Merkmale verglichen werden. Es findet also weder eine experimentelle Variation der unabhängigen Variablen noch eine randomisierte Gruppenzuweisung statt. Da sich Kinder aus Familien, in denen mehr als eine Sprache gesprochen wird, im Hinblick auf zahlreiche weitere Merkmale von Kindern aus einsprachigen Familien unterscheiden könnten, lassen sich Gruppenunterschiede in den untersuchten abhängigen Variablen (z. B. Exekutivfunktionen, deutschsprachiger Wortschatz) nicht mit Sicherheit auf den Sprachhintergrund zurückführen. So sind aufgrund der Zuwanderungsgeschichte in Deutschland Mehrsprachigkeit und Bildungshintergrund miteinander konfundiert, da mehrsprachige Kinder öfter in Familien mit geringem Bildungshintergrund aufwachsen als deutschsprachige Kinder, die bislang keine weitere Sprache erworben haben. Findet man also Nachteile für mehrsprachige Kinder im deutschsprachigen Wortschatz, könnte dies auf den Sprachhintergrund zurückzuführen sein oder auch darauf, dass in Familien mit niedrigerem Bildungshintergrund das sprachlich-kognitive Anregungsniveau oft geringer ist. Allgemein ist bei nicht-experimentellen Designs die Gefahr groß, dass sich die untersuchten Gruppen nicht nur hinsichtlich des interessierenden Merkmals (unabhängige Variable), sondern auch hinsichtlich anderer Merkmale (Störvariablen) unterscheiden.

Auch bei nicht-experimentellen Studien ist es daher besonders wichtig, schon bei der Planung genau zu überlegen, welche Merkmale mit der unabhängigen Variable zusammenhängen und sich ebenfalls auf die abhängige Variable auswirken könnten, um den Einfluss dieser Störvariablen zu kontrollieren. Hierzu gibt es verschiedene versuchsplanerische Möglichkeiten, wie etwa das Matching (s. o.), aber auch statistische Verfahren wie das *Propensity Score Matching*, bei dem statistische Zwillinge gebildet werden (Becker, 2011).

Die interne Validität nicht-experimenteller Studien lässt sich insbesondere durch die Umsetzung von Längsschnittdesigns erhöhen, bei denen die jeweiligen Untersuchungseinheiten über einen angemessen langen Zeitraum⁴ wiederholt untersucht werden. Annahmen über kausale Effekte von unabhängigen Variablen auf abhängige Variablen (oder auch Wechselwirkungen zwischen Variablen) lassen sich dann anhand von zeitverzögerten Zusammenhängen prüfen. Vereinfacht ausgedrückt sollten Veränderungen in der unabhängigen Variable zu einem Zeitpunkt t_1 zu Veränderungen in der abhängigen Variable zu späteren Zeitpunkten t_{1+x} führen. Beispielstudie 2 beinhaltet ein solches Vorgehen, da überprüft wurde, ob Bilingualität (unabhängige Variable) mit Veränderungen in den erfassten Exekutivfunktionen (abhängige Variable) zusammenhängt. Eine Schwäche dieser Studie besteht allerdings darin, dass nur die Exekutivfunktionen mehrfach erfasst wurden, nicht aber die Bilingualität. So lässt sich nicht überprüfen, ob die Ergebnisse eindeutig für die angenommene Wirkrichtung sprechen oder ebenso mit der umgekehrten Wirkrichtung vereinbar sind (höher ausgeprägte Exekutivfunktionen fördern die Entwicklung von Bilingualität) bzw. eine Wechselwirkung zwischen den beiden Variablen besteht (höher ausgeprägte Bilingualität fördert

4 Welcher Zeitraum angemessen lang ist, hängt dabei von der jeweiligen Fragestellung ab.

die Entwicklung von Exekutivfunktionen und höher ausgeprägte Exekutivfunktionen fördern die Entwicklung von Bilingualität).

3. Qualitätsmerkmale

Die Belastbarkeit von Ergebnissen hypothesenprüfender Studien (wie allgemein jeder Art von Forschung) hängt von einer ganzen Reihe von Qualitätsmerkmalen ab, von denen einige in den vorherigen Abschnitten bereits angesprochen wurden. Entsprechend ist es wichtig, die einzelnen Schritte von Studien sehr sorgfältig zu planen und durchzuführen. Grundlegend ist dabei, wie schlüssig die Hypothesen von theoretischen Annahmen abgeleitet und die interessierenden Konstrukte definiert werden. Es folgen die ebenfalls grundlegenden Entscheidungen darüber, welches Forschungsdesign zur Prüfung der Hypothese am besten geeignet ist und wie die Konstrukte messbar gemacht werden können, so dass sie das, was gemessen werden soll, möglichst gut abbilden. Zur Sicherung der internen Validität ist es zudem wichtig, potenzielle Störvariablen zu identifizieren, ihren Einfluss wenn möglich auszuschalten oder die Störvariablen zumindest zu erfassen und in den Analysen zu kontrollieren. Eine hohe interne Validität kann insbesondere durch experimentelle Studien erreicht werden, die unter hoch standardisierten Bedingungen durchgeführt werden, etwa wenn eine sprachliche Förderung (z. B. dialogisches Lesen von Kinderbüchern mit Vorschulkindern; Hartung & Ennemoser, 2018) nach einem vorgegebenen Skript durch speziell geschulte Personen in Einzelsitzungen stattfindet. Ein solches Vorgehen kann als erster Schritt im kumulativen Forschungsprozess äußerst sinnvoll sein, um zu prüfen, ob der Förderansatz unter Idealbedingungen wirkt. Zeigt sich dies, ist allerdings weiterhin unklar, inwieweit er auch unter Realbedingungen wirksam ist bzw. auf andere Orte, Zeiten, Operationalisierungen und Personen übertragen werden kann (*externe Validität*). Dies lässt sich mit einem Feldexperiment prüfen, bei dem ähnlich wie in Beispielstudie 1 vorgegangen wird. Bei solchen Studien ist es offensichtlich schwieriger, die Umsetzung des Treatments zu kontrollieren, und daher besonders wichtig, die Implementationsqualität zu erfassen. In Beispielstudie 1 etwa erfolgte dies ansatzweise, indem der Unterricht im Untersuchungszeitraum dreimal beobachtet und mittels standardisierter Verfahren eingeschätzt wurde, ob die Lehrkräfte die vermittelten Unterrichtsstrategien angewendet haben. Die Qualität der Umsetzung wurde jedoch nicht erfasst, was eine Schwäche der Untersuchung darstellt, die auch von den Autor:innen selbst erwähnt wird.

Ein weiteres Qualitätsmerkmal hypothesenprüfender Forschung, das von zentraler Bedeutung ist, betrifft die Auswahl der Stichprobe. Allgemein sollte die Stichprobe die Population, über die Aussagen getroffen werden sollen, möglichst repräsentativ abbilden und ausreichend groß sein, um auch kleinere Effekte identifizieren zu können (*Teststärke*; Cohen, 1988), sofern diese theoretisch erwartet werden. Bei experimentellen Designs muss die Stichprobe nicht im strengen Sinne repräsentativ sein, sie sollte jedoch Personen umfassen, über die eine Aussage

getroffen werden soll. Wird also beispielsweise der Anspruch erhoben, dass ein Treatment sowohl für einsprachige als auch für mehrsprachige Kinder geeignet ist, sollten beide Teilgruppen in der Stichprobe vertreten sein und im Idealfall auch jeweils in ausreichendem Umfang, damit eine Prüfung dieses Generalisierungsanspruchs vorgenommen werden kann.

4. Chancen und Grenzen

Mit hypothesenprüfender Forschung ist die Chance verbunden, theoretisch fundierte Annahmen einem Test zu unterziehen. Dabei ist allerdings zu berücksichtigen, dass jede Studie mit Einschränkungen verbunden ist, auf die in der Regel im Diskussionsteil der Darstellung von Studienergebnissen in wissenschaftlichen Publikationen eingegangen wird. Jede Studie ist mit einer Vielzahl von Entscheidungen verbunden, die man möglicherweise auch anders hätte treffen können und die sich auf die Ergebnisse ausgewirkt haben könnten. Daher ist es wichtig, das Prinzip der Kumulativität von Forschung zu beachten, wonach eine einzelne Studie nicht als eindeutiger „Beweis“ für die Gültigkeit einer Hypothese betrachtet werden kann und es in der Regel mehrerer Studien bedarf, um gut gesicherte Erkenntnisse zu generieren. Wurden ausreichend viele quantitative Studien zu einer Fragestellung durchgeführt, lassen sich deren Ergebnisse statistisch in Metaanalysen zusammenfassen. Metaanalysen können Aussagen darüber treffen, wie groß ein Effekt im Durchschnitt ausfällt, wie stark er über die verschiedenen Studien hinweg streut und welche Faktoren (sog. Moderatoren) damit zusammenhängen, dass die Effekte in den einzelnen Studien größer oder kleiner ausgefallen sind (für einen Überblick zum Verfahren der Metaanalyse siehe z. B. Döring & Bortz, 2016; für Beispiele von Meta-Analysen zu Fragestellungen sprachlicher Bildung unter Bedingungen der Mehrsprachigkeit siehe z. B. Ardasheva, Wang, Adesope & Valentine, 2017 sowie Slavin & Cheung, 2005). Eine Grenze hypothesenprüfender (und allgemein quantitativer) Forschung besteht darin, dass sich die Ergebnisse immer auf Gruppen beziehen. So wurde etwa in Beispielstudie 1 untersucht, inwieweit sich die englische Sprachkompetenz bei Kindern, deren Lehrkräfte an der Fortbildung teilgenommen haben, *im Durchschnitt* positiver entwickelt hat als bei Kindern in der Kontrollgruppe. Dabei wird in der Regel auch die Streuung in den Ergebnissen betrachtet, indem z. B. analysiert wird, wie stark die Ergebnisse der einzelnen Kinder in den Gruppen variieren. In Beispielstudie 1 wurde zudem geprüft, ob Kinder mit unterschiedlichen Ausgangskompetenzen vom Treatment in unterschiedlichem Maße profitiert haben. Über Einzelfälle lassen sich jedoch, anders als bei geeigneten qualitativen Forschungsansätzen (vgl. Kapitel 5.2 in diesem Band), bei quantitativen Studien in der Regel keine genaueren Aussagen treffen.

5. Forschungsdesiderata und Ausblick

Über Grundlagen und wirksame Prinzipien sprachlicher Bildung ist bereits viel bekannt. Insbesondere zu Fragen der Förderung schriftsprachlicher Kompetenzen liegt eine breite empirische Basis vor, die in der Praxis von Bildungseinrichtungen in Deutschland bislang jedoch nur begrenzt aufgegriffen wird. Hier wäre es wichtig, mit Hilfe von Transfer- und Implementationsforschung zu bestimmen, wie eine evidenzbasierte Weiterentwicklung der Praxis gelingen kann (z. B. Schrader, Hasselhorn, Hetfleisch & Goeze, 2020 und Kapitel 5.5 in diesem Band). Gleichzeitig bestehen aber noch viele Forschungslücken, zu denen grundlegende hypothesenprüfende Studien (und auch andere Arten von Untersuchungen) erforderlich sind. Dies betrifft beispielsweise die Förderung mündlicher Sprachkompetenzen, gerade auch bei Kindern und Jugendlichen, die außerhalb der Schule wenig Möglichkeiten haben, ihre Kompetenzen in der Unterrichtssprache weiterzuentwickeln. Die IQB-Bildungstrends etwa haben wiederholt gezeigt, dass die Nachteile von Schüler:innen aus zugewanderten Familien im Bereich Zuhören stärker ausgeprägt sind als beispielsweise im Bereich Orthografie (z. B. Stanat et al., 2022). Dies ist vermutlich darauf zurückzuführen, dass diese beiden Kompetenzbereiche in unterschiedlichem Maße durch außerschulische Lerngelegenheiten beeinflusst werden, die sich wiederum in Abhängigkeit vom Zuwanderungshintergrund (bzw. konkreter: vom Sprachhintergrund) der Schüler:innen unterscheiden. Während alle Kinder orthografische Kompetenzen primär in der Schule erlernen, findet der Erwerb des mündlichen Sprachverstehens auch in anderen Kontexten statt, in denen Schüler:innen aus zugewanderten Familien oft weniger Zugang zur deutschen Sprache haben. Zur Frage, mit welchen Strategien diese Disparitäten effektiv kompensiert werden können, besteht weiterhin erheblicher Forschungsbedarf.

Für die Transfer- und Implementationsforschung wird die Bedeutung einer engen Zusammenarbeit zwischen Forschenden und Partner:innen aus der Praxis betont (z. B. Spiel, 2020 und Kapitel 5.5 diesem Band). Aber auch für hypothesenprüfende Forschung, die im Feld durchgeführt wird, ist eine gute Kooperation wichtig. Pädagogische Einrichtungen und Fachkräfte müssen bereit sein, sich an den Studien zu beteiligen, was oft mit erheblichem Aufwand verbunden ist. Ihre Bereitschaft dazu hängt erfahrungsgemäß unter anderem davon ab, ob die beteiligten Praktiker:innen einen Nutzen für die eigene Arbeit erkennen können und inwieweit sie über die für eine Beteiligung an der Studie erforderlichen zeitlichen Ressourcen verfügen. Selbst bei Studien, die von der Bildungspolitik unterstützt werden, wird nur selten eine Kompensation des zeitlichen Aufwands bereitgestellt, den die Praxispartner:innen investieren müssen.

Allgemein wäre es wichtig, die Kultur der Evidenzbasierung im Bereich der sprachlichen Bildung (wie auch im Bildungsbereich allgemein) weiter auszubauen und Forschung in vielfältiger Weise in Praxis einzubinden (Schrader et al., 2020). Umgekehrt sollte die Praxis verstärkt in die Forschung eingebunden werden, etwa, wenn es um die Identifikation relevanter Forschungsfragen oder Implemen-

tationsbedingungen geht. Die Vorstellung, dass die Forschung etwas herausfindet, was dann anschließend von der Praxis möglichst genau so umgesetzt wird, wie es in den Studien untersucht wurde, wird zunehmend kritisch gesehen (Schrader et al., 2020; Yurkofsky, Peterson, Mehta, Horwitz-Willis & Frumin, 2020). Ein solches Vorgehen kann bei klar umrissenen Maßnahmen (z. B. Programme zu Förderung phonologischer Bewusstheit in der Kita; Schneider, 2019) zielführend sein. Bei Herausforderungen, die komplexere Veränderungen in der Praxis erfordern (z. B. Schulentwicklung zur Verbesserung der sprachlichen Bildung im Fachunterricht), greift dies aber zu kurz. Daher werden inzwischen verschiedene Formen der Zusammenarbeit zwischen Forschung und Praxis umgesetzt, die sich unter anderem darin unterscheiden, welche Rolle die Akteursgruppen jeweils einnehmen (für einen Überblick siehe z. B. Gräsel, 2019; Yurkofsky et al., 2020). Allerdings benötigen auch Ansätze, die auf die Bewältigung komplexer Praxisprobleme einzelner Einrichtungen bzw. Gruppen von Einrichtungen abzielen, zu ihrer Fundierung belastbare Evidenz.

Literatur

- Ardasheva, Y., Wang, Z., Adesope, O. O. & Valentine, J. C. (2017). Exploring effectiveness and moderators of language learning strategy instruction on second language and self-regulated learning outcomes. *Review of Educational Research*, 87, 544–582. <https://doi.org/10.3102/0034654316689135>
- Babinski, L. M., Amendum, S. J., Knotek, S. E., Sánchez, M. & Malone, P. (2018). Improving young English learners' language and literacy skills through teacher professional development: A randomized controlled trial. *American Educational Research Journal*, 55, 117–143. <https://doi.org/10.3102/0002831217732335>
- Becker, M. (2011). Matching-Verfahren und Gruppenvergleiche. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online (EEO), Fachgebiet Methoden der empirischen erziehungswissenschaftlichen Forschung* (S. 1–50). Weinheim: Juventa.
- Bialystok, E. (2017). The bilingual adaptation. How minds accommodate experience. *Psychological Bulletin*, 143(3), 233–262. <https://doi.org/10.1037/bul0000099>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Erlbaum.
- Crivello, C., Kuzyk, O., Rodrigues, M., Friend, M., Zesiger, P. & Poulin-Dubois, D. (2016). The effects of bilingual growth on toddlers' executive function. *Journal of Experimental Child Psychology*, 141, 121–132. <https://doi.org/10.1016/j.jecp.2015.08.004>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Wiesbaden: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Eifler, S. (2014). Experiment. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 195–209). Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-18939-0_11
- Gräsel, C. (2019). Transfer von Forschungsergebnissen in die Praxis. In C. Donie, F. Foerster, M. Obermayr, A. Deckwerth, G. Kammermeyer, G. Lenkse, M. Leuchtner & A. Wildemann (Hrsg.), *Grundschulpädagogik zwischen Wissenschaft und Transfer. Jahrbuch Grundschulforschung* (S. 2–11). Wiesbaden: Springer.

- Hartung, N. & Ennemoser, M. (2018). Ein Förderkonzept im Elementarbereich: Dialogisches Lesen. In C. Titz, S. Weber, A. Ropeter, S. Geyer. & M. Hasselhorn (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln* (Bildung durch Sprache und Schrift, Bd. 2.) (S. 115–127). Stuttgart: Kohlhammer.
- Kühnel, S. & Dingelstedt, A. (2014). Kausalität. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 1017–1028). Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-18939-0_80
- Schneider, W. (2019). Förderung der phonologischen Bewusstheit im Vorschulalter. Bedingungen für den Transfer auf den Schriftspracherwerb. *Die Deutsche Schule*, 111, 344–346. <https://doi.org/10.31244/ddS.2019.03.11>
- Schrader, J., Hasselhorn, M., Hetfleisch, P. & Goeze, A. (2020). Stichwortbeitrag Implementationsforschung: Wie Wissenschaft zu Verbesserungen im Bildungssystem beitragen kann. *Zeitschrift für Erziehungswissenschaft*, 23(2), 9–59. <https://doi.org/10.1007/s11618-020-00927-z>
- Slavin, R. E. & Cheung, A. (2005). A synthesis of research on language of reading instruction for English Language Learners. *Review of Educational Research*, 75, 247–284. <https://doi.org/10.3102/00346543075002247>
- Spiel, C. (2020). Moving beyond the ivory tower – why researchers from the field of education should go ahead. *Zeitschrift für Pädagogische Psychologie*, 34, 1–18. <https://doi.org/10.1024/1010-0652/a000260>
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S. & Henschel, S. (Hrsg.) (2022). *IQB-Bildungstrend 2021: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Münster: Waxmann. <https://doi.org/10.31244/9783830996064>
- Stein, P. (2014). Forschungsdesigns für die quantitative Sozialforschung. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 125–142). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-21308-4_8
- Yurkofsky, M. M., Peterson, A. J., Mehta, J. D., Horwitz-Willis, R. & Frumin, K. M. (2020). Research on continuous improvement: Exploring the complexities of managing educational change. *Review of Research in Education*, 44, 403–433. <https://doi.org/10.3102/0091732X20907363>