

Stichwortbeitrag Implementationsforschung: Wie Wissenschaft zu Verbesserungen im Bildungssystem beitragen kann

Josef Schrader · Marcus Hasselhorn · Petra Hetfleisch · Annika Goeze

Online publiziert: 10. März 2020
© Der/die Autor(en) 2020

Zusammenfassung Mit dem Programm evidenzbasierter Bildungsreform geht die Erwartung einher, auf der Grundlage gesicherten empirischen Wissens die pädagogische Praxis ebenso wie ihre politische Gestaltung zu verbessern. Empirische Studien bestätigen allerdings ebenso wie Erfahrungsberichte aus dem Feld eine fortdauernde Kluft zwischen Forschung, Politik und Praxis. Daher findet Implementationsforschung wachsende Aufmerksamkeit, ablesbar an einer wachsenden Zahl empirischer Studien und einer breiten konzeptuellen und methodologischen Debatte. Der Stichwortbeitrag sichtet die konzeptuelle Diskussion zur Implementationsforschung im Bildungsbereich und darüber hinaus. Zudem wird analysiert, was Implementations-

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11618-020-00927-z>) enthalten.

Prof. Dr. J. Schrader (✉)

Deutsches Institut für Erwachsenenbildung (DIE) – Leibniz-Zentrum für Lebenslanges Lernen e. V.,
Heinemannstraße 12–14, 53175 Bonn, Deutschland
E-Mail: schrader@die-bonn.de

Abteilung Erwachsenenbildung / Weiterbildung, Eberhard Karls Universität Tübingen, Münzgasse
11, 72070 Tübingen, Deutschland

Prof. Dr. M. Hasselhorn

Abteilung Bildung und Entwicklung, Deutsches Institut für Internationale Pädagogische
Forschung (DIPF) – Leibniz-Institut für Bildungsforschung und Bildungsinformation, Rostocker
Str. 6, 60323 Frankfurt a. M., Deutschland
E-Mail: hasselhorn@dipf.de

Dr. P. Hetfleisch · Prof. Dr. A. Goeze

Philosophische Fakultät, Professur Erwachsenenbildung und Weiterbildung, Technische Universität
Chemnitz, Reichenhainer Str. 41, 09126 Chemnitz, Deutschland

Dr. P. Hetfleisch

E-Mail: petra.hetfleisch@phil.tu-chemnitz.de

Prof. Dr. A. Goeze

E-Mail: annika.goeze@phil.tu-chemnitz.de

studien im Feld des organisierten Lehrens und Lernens über die Gelingensbedingungen erfolgreicher Implementationen von evidenten Interventionen aussagen. Der Forschungsüberblick zeigt, dass eine wachsende Zahl von Studien mit wissenschaftlich fundierten Interventionen zur Verringerung der beklagten Kluft beitragen konnten. Zugleich fehlt es noch an theoretisch fundierten Einsichten zu den institutionellen, organisationalen und personalen Bedingungen wirksamer Implementationsstudien.

Schlüsselwörter Implementationsforschung · evidenzbasierte Bildungsreform · Empirische Bildungsforschung · Literaturreview

Implementation research: How science can contribute to improvements in the education system

Abstract The program of evidence-based educational reform is accompanied by the expectation to improve pedagogical practice as well as its political formation based on proven empirical knowledge. However, empirical research as well as field reports confirm a continuing gap between research, politics and practice. Therefore, implementation research is gaining attention, as shown by a growing number of empirical studies and a broad conceptual and methodological debate. This paper reviews the conceptual discussion on implementation research in the education sector and beyond. It also analyzes what implementation studies in the field of organized learning and instruction say about the success conditions for productive implementations of evident interventions. The research overview shows that an increasing number of studies with scientifically sound interventions have contributed to reducing the criticized gap. At the same time, there is still a lack of theoretically sound insights into the institutional, organizational and personal conditions for effective implementation studies.

Keywords Empirical educational research · Evidence based policy and practice in education · Implementation research · Review of literature

1 Die Bedeutung von Wissenschaft für Verbesserungen im Bildungsbereich – Erwartungen und Erfahrungen

Die Frage, *warum* und *wozu* sich Wissenschaft mit pädagogischer Praxis beschäftigen solle, wird in der pädagogischen Tradition gewöhnlich wie folgt beantwortet: Es gehe um eine Theorie nicht nur *von*, sondern auch *für* Praxis. Ein solches Verständnis eint ganz unterschiedliche Richtungen und „Schulen“, seit sich die Pädagogik als eine eigenständige Form der wissenschaftlichen Reflexion von Phänomenen der Erziehung und Bildung in Distanz zur Praxis etablierte. Kant hielt 1803 eine Vorlesung zur Einführung in die Pädagogik und stellte sie als eine Theorie für einen „zukünftig möglich bessern Zustande des menschlichen Geschlechts“ (Kant 1803, S. 17) vor. Schleiermacher (1983) betrachtete die Pädagogik noch als „eine rein mit der Ethik zusammenhängende, aus ihr abgeleitete angewandte Wissenschaft“, die der älteren Generation aufzeige, wie sie die Heranwachsenden auf die „größte-

ren Lebensgemeinschaften“ vorbereiten könne. Dilthey (1986) ging gegen Ende des 19. Jahrhunderts mit dem Entwurf einer „positiven Geisteswissenschaft“ (Herrmann 1971, S. 214) über diese Theorietradition hinaus, indem er auch die Entwicklung von Regeln und Techniken für die Lösung praktischer Probleme einforderte. Dazu sei es notwendig, die gesetzmäßig verlaufende psychische Entwicklung mit den Methoden der Naturwissenschaften zu *erklären*, zugleich aber die historisch konkreten, ganzheitlichen Lebenswirklichkeiten und die dort wirkenden Bedingungen und Einflüsse zu *verstehen*, gestützt auf hermeneutische Verfahren, typisierende Beschreibungen und Vergleiche.

Auch als sich in Deutschland die Pädagogik von ihren geisteswissenschaftlichen Traditionen löste (vgl. Dahmer und Klafki 1968), sich zur Erziehungswissenschaft entwickelte und empirischen Methoden öffnete, wurde der überlieferte Anspruch aufrechterhalten. Für Roth (1967, S. 119) bestand die Aufgabe empirischer Forschung darin, „die angeblichen Fakten, das scheinbar unabänderlich Gegebene, unter der produktiven Fragestellung, die die pädagogische Idee entwickelt, auf die noch verborgenen pädagogischen Möglichkeiten hin herauszufordern“. Und als schließlich die empirische Bildungsforschung mit den international-vergleichenden Large Scale Assessments zu Beginn dieses Jahrhunderts die Aufmerksamkeit nicht nur der Forschung, sondern auch von Politik und Öffentlichkeit auf sich zog, tat sie dies mit dem Anspruch, nicht nur besseres, weil empirisch geprüftes, sondern auch nützlicheres, weil auf übergreifende Probleme bezogenes Wissen für Praxis und – historisch erstmals betont – für Politik anzubieten (vgl. Schrader 2014). So kam es zu einer in der deutschen Bildungsgeschichte einmaligen Institutionalisierung der wissenschaftlichen Politik- und Praxisbegleitung, die den gelegentlichen „gelehrten Rat“, wie er seit der ersten „empirischen Wendung“ bevorzugt wurde, um kontinuierliches Monitoring erweiterte und auf Dauer stellte (vgl. Tenorth 2014).

Begleitet wird der weithin akzeptierte Anspruch einer praxis- und politikrelevanten Wissenschaft allerdings von einer andauernden Klage über die „Kluft“ zwischen Forschung, Praxis und Politik. Untersuchungen der Wissenschafts-, der Wissensverwendungs-, der Steuerungs- und der empirischen Bildungsforschung verweisen auf unterschiedlich gelagerte Probleme und Ursachen: So wird Forschungswissen nicht genutzt, weil das Wissenschaftssystem Wissenstransfer nicht belohnt (vgl. Thomsen und Anger 2018), die Außenkommunikation eher Reputation als Aufklärung priorisiert (vgl. Weingart 2017), die Evidenz wissenschaftlichen Wissens oftmals als zu schwach bewertet wird, um es der Praxis zuzumuten (vgl. Mosteller und Boruch 2002), Akteure der pädagogischen Praxis nicht hinreichend vorbereitet sind, um z. B. als Schulleitungen eine datenbasierte Personal- und Organisationsentwicklung umzusetzen (vgl. Böse et al. 2018), die Interessen handelnder Akteure einflussreicher als die Befunde der Forschung sind, so dass z. B. die Einführung der Ganztagschule am Widerstand der Lehrerverbände scheitert (vgl. Heinemann 2017), oder weil Bildungspolitik wissenschaftliche Evidenz zwar als Ressource für Legitimation, nicht jedoch als Anregung für Reformen nutzt (vgl. Köller 2017 zum Themenfeld G8/G9). Die keineswegs vollständige Auflistung zeigt, dass ein Austausch wissenschaftlichen Wissens mit Politik und Praxis zwar zunehmend stattfindet, aber keineswegs immer mit Erfolg. Der Steuerbarkeit des Bildungssystems und der Bildungspolitik mit den

Mitteln der Wissenschaft sind offenkundig Grenzen gesetzt (vgl. Baumert 2016, S. 243).

Eine vergleichende Forschung zu den Wirkungen wissenschaftlicher Transferangebote und ihrer Nutzung mit Rücksicht auf unterschiedliche wissenschaftliche Richtungen, unterschiedliche Wissensformen und Forschungstypen liegt bislang nicht vor. Aber auch ohne solche Untersuchungen haben die Befunde das Potenzial, um wissenschaftliche Selbstverständnisse zu irritieren. So gehört nicht nur die Klage über eine andauernde Kluft zwischen Wissenschaft, Politik und Praxis zu den Topoi jener Wissenschaften, die sich mit Erziehungs- und Bildungsprozessen beschäftigen, sondern auch das Versprechen, diese Kluft zu überwinden (vgl. Century und Cassata 2016, S. 169 f.).

Auf die Diskrepanz zwischen Anspruch und Wirklichkeit wurde unterschiedlich reagiert. Die geisteswissenschaftliche Pädagogik etablierte mit dem Theorie-Praxis-Diskurs eine gleichsam metatheoretische Reflexionsform. Noch in der zweiten Hälfte des 20. Jahrhunderts schien es, als könne die Einheit des Nachdenkens über Erziehung und Bildung in „Theorie“ und „Praxis“ gesichert werden: Die Praxis könne von der Wissenschaft Aufklärung, Orientierung und Kritik erwarten und die Wissenschaft aus der Praxis Erfahrung, Anregungen und Korrektur ihrer Theoriebildung, und zwar dann, wenn bestimmte Prinzipien des Denkens beachtet würden. Dazu zählten die Orientierung an der Bildsamkeit der Lernenden, die Aufforderung zu ihrer Selbsttätigkeit, die Überführung gesellschaftlicher Determinanten in pädagogische sowie die Konzentration auf die gemeinsame Aufgabe der Höherentwicklung der Menschheit (vgl. Benner 1980, S. 494). Die Suche nach einer Einheit von Theorie und Praxis endete allerdings mit der Einsicht, dass bereits die Annahmen *einer* „Theorie“ und *einer* „Praxis“ nicht haltbar sind: Mit der Öffnung der Pädagogik für die Sozialwissenschaften ging zum einen eine wissenschaftstheoretische Neuorientierung einher, die zeigte, dass die Erziehungswissenschaft nicht nur theoretisches, sondern auch empirisches Wissen erarbeitet, dazu normatives (als Philosophie der Erziehung) und praktisches (als Unterrichtskunde) (vgl. Brezinka 1975), ohne dass eine dieser Formen des Wissens die jeweils anderen „regieren“ könnte. Zum anderen hat sich die Expansion der Erziehungswissenschaften an Hochschulen seit den 1960er Jahren nicht als *Ausdifferenzierung* (so Benner 1980, S. 485) der *einen* Pädagogik vollzogen, sondern als *Erweiterung* der Schulpädagogik um weitere Pädagogiken, die mit der Institutionalisierung des lebenslangen Lernens als Berufs-, Sozial- und Erwachsenenpädagogik in die Hochschulen drängten.

Die empirische Bildungsforschung dagegen hat die Kluft zwischen Forschung sowie Politik und Praxis als ein forschungsmethodisches Problem interpretiert. Um zu verhindern, dass evidenzbasierte Bildungsreform eine bloße Fiktion (vgl. Spiel 2009) bleibe, werden seit einigen Jahren, ergänzend vor allem zu Large Scale Assessments, vermehrt Interventions- und Implementationsstudien gefordert (vgl. Heid 2011, S. 504). Implementationsstudien wurden gar zu einer Bringschuld der Wissenschaft erklärt (vgl. Hasselhorn 2009). Sie haben in einem weiter gefassten Verständnis zum Ziel, (pädagogische) Interventionen im Feld zu erproben und Bedingungen, Prozesse bzw. Wirkungen ihrer Implementierung zu untersuchen. Bei einem enger gefassten Verständnis müssen diese Interventionen vorab bereits bzgl. ihrer Wirkungsannahmen wissenschaftlich geprüft sein. Implementationforschung macht

die Vermittlung bzw. die Übernahme von wissenschaftlich begründetem Wissen durch Politik und Praxis selbst zum Gegenstand von Forschung (vgl. Prenzel 2005). Grundlegend ist dabei die Annahme, dass Implementation immer mit Transformation einhergeht und daher mit Selektionen, Fehl- und Umdeutungen, Neu-Kontexturierungen, mithin mit „Implementationsbrüchen“ zu rechnen ist (van Ackeren et al. 2011, S. 174). Trotz teils vehementer Plädoyers werden Implementationsstudien in Überblicksdarstellungen allerdings immer noch als Desiderat bewertet. Diese Einschätzung findet sich für die Unterrichtsforschung bereits bei Gräsel und Parchmann (2004, S. 197). Hasselhorn und Hasselhorn (2017, S. 51) merken in einem jüngeren Beitrag gar an, dass „eine eigenständige Implementationsforschung im Bereich der empirischen Bildungsforschung [...] sich bisher [...] nicht etabliert“ habe (so auch Tabak et al. 2012, S. 338).

Für die Suche danach, *wie* Wissenschaft auch für Bildungspraxis und -politik nützlich sein kann, drängen sich daher zwei Fragen auf, die den Fokus des Stichwortbeitrages bilden: Die erste Frage richtet sich auf die konzeptuelle und methodologische Diskussion zur Implementationsforschung, die den Anspruch erhebt, Erkenntnis- und Veränderungsinteressen zugleich zu verfolgen, indem sie mit den Mitteln der Forschung die Kluft zwischen Forschung und Praxis reduzieren will. Auf der Grundlage eines weiten Verständnisses von Implementationsforschung geben wir dazu im Abschn. 2 einen narrativen Literaturüberblick. Die zweite Frage richtet sich auf einen ausgewählten Schlüsselaspekt der Implementationsforschung, nämlich die Einflussfaktoren auf Implementationen. Im Abschn. 3 wird hierzu ein Überblick über empirische Befunde zu hemmenden und förderlichen Einflussfaktoren auf Implementationsprozesse und -ergebnisse gegeben. Dabei orientieren wir uns an einem engen Verständnis von Implementationsforschung und beschränken uns auf solche Studien, in denen Interventionen zur Verbesserung von Lehr-Lernprozessen in organisierten Bildungskontexten umgesetzt werden, für deren Wirkungen wissenschaftliche Evidenz vorliegt. Hierfür wurde ein gesonderter, systematischer Literaturüberblick vorliegender empirischer Befunde erarbeitet, der sich – anders als der breit angelegte, narrative Überblick in Abschn. 2 – ausschließlich auf einen Literaturcorpus zu Einflussfaktoren beschränkt, verbunden mit transparenten Kriterien für Ein- oder Ausschluss. Betrachtet man beide Fragen in ihrem Zusammenhang, so sucht der Beitrag auszuloten, inwieweit Implementationsforschung nicht nur Potenzial für eine wissenschaftlich fundierte Verbesserung von Politik und Praxis besitzt, sondern darüber hinaus die Möglichkeit bietet, im Sinne Kants (1992, S. 45 f.) Theorien über die Bedingungen der Möglichkeit der Verbesserung von Politik und Praxis mit den Mitteln der Wissenschaft in empirischen Studien zu entwickeln und zu prüfen (vgl. auch Heid 2011, S. 504).

2 Theoretische und methodische Herausforderungen der Implementationsforschung – Ein narrativer Literaturüberblick

Recherchiert man Beiträge zur Implementationsforschung in einem weiten Sinn und bezieht die internationale Forschung benachbarter Disziplinen mit ein, stößt man auf eine überaus lebendige, seit fast vier Jahrzehnten geführte Diskussion über Stra-

tegien zur Unterstützung der Implementation wissenschaftlicher Erkenntnisse durch Politik und Praxis, die seit mehr als einem Jahrzehnt auch in der Bildungsforschung aufgegriffen wird. In diesem Umfeld hat sich eine Implementation-Science-Bewegung (vgl. Fixsen et al. 2005) etabliert, die ihren Ausdruck in wissenschaftlichen Vereinigungen¹, einschlägigen Publikationsorganen (z. B. die Zeitschriften „Implementation Science“ und „Translational Science“) und gelegentlich auch in Studiengängen findet (vgl. Ullrich et al. 2017). Beklagt wurde – und wird nach wie vor – ein Mangel an Evidenz, wie Implementation praktisch gelingen kann (z. B. Goldman et al. 2001, S. 1593; Spiel et al. 2018, S. 339). Im Folgenden zeichnen wir die konzeptuelle und methodologische Debatte zu der Frage nach, *wie* Implementationsforschung Erkenntnis- und Veränderungsinteressen zugleich bedienen möchte, welche Annahmen dabei geteilt werden und wo offene Fragen fortbestehen. Als Schwerpunkte der Diskussion wurden die Themen „Selbstverständnis der Implementationsforschung“, „Stellenwert von Theorien, Modellen und Handlungsrahmen“, „Untersuchungsdesigns“, „Anforderungen an statistische Analysen sowie Datenqualität“ sowie „ethische Fragen“ ausgewählt. Entsprechend wurden systematische Literaturrecherchen zu den Suchbegriffen „*theor* of implement*/*, implementation theor*/*, framework of implement*/*, implementation framework*/*, model of implement*/*, implementation model*/*, implementation (research) AND ethi*/**“ in den Datenbanken Fachportal Pädagogik, ERIC, PSYINDEX und PsycINFO durchgeführt. Beiträge aus den Gesundheits- und den Sozialwissenschaften wurden nicht ausgeschlossen. Schließlich wurde die identifizierte Literatur gemäß den Schwerpunkten kategorisiert und nach den Prinzipien eines narrativen Literaturüberblicks (vgl. Robinson und Lowe 2015) aufbereitet.

2.1 Begriffliche Einordnung

Vielfältige Forschungsansätze beschäftigen sich mit der Vermittlung und Rezeption wissenschaftlichen Wissens in Politik und Praxis. Dazu gehören neben der *Implementationsforschung* die *Diffusionsforschung*, die *Evaluationsforschung*, die *Transfer- und Scaling-up-* sowie die *Disseminationsforschung*. Die *Diffusionsforschung*² legt ihren Fokus auf die Beobachtung der gleichsam „natürlichen“, nicht von mandatierten Akteuren des Bildungssystems vorangetriebenen Verbreitung von Innovationen in der pädagogischen Fläche. In gewisser Weise leistet die Bildungsberichterstattung einen elementaren Beitrag zur Diffusionsforschung, wenn sie z. B., wie aktuell, die Verbreitung digitaler Medien in Lehr-Lernkontexten der Erwachsenen- und Weiterbildung im Zeitverlauf dokumentiert (vgl. Schmid et al. 2017). Die

¹ So. z. B. die „Division 16 Working Group on Translating Science to Practice“ (Forman et al. 2013) innerhalb der American Psychological Association, die disziplinübergreifende Global Implementation Initiative, die disziplinübergreifende European Implementation Collaborative, das (disziplinübergreifende) National Implementation Research Network oder etwa das an der University of North Carolina at Chapel Hill gegründete National Implementation Research Network (NIRN), das sich mit Informationen, Tools und Schulungsangeboten an Wissenschaftler und PraktikerInnen wendet.

² „Diffusion is the passive, untargeted, unplanned, and uncontrolled spread of new interventions. Diffusion is part of the diffusion-dissemination-implementation continuum, and it is the least focused and intense approach.“ (Rabin und Brownson 2012, S. 25)

*Evaluationsforschung*³ konzentriert sich auf die *Bewertung* eines Programms oder einer Reform zur Verbesserung pädagogischer Praxis (vgl. Kuper 2005). Die *Transfer- und Scaling-up-Forschung*⁴ untersucht, wie durch Top-Down-, Bottom-Up- oder hybride Strategien (vgl. Schneider und McDonald 2007b; Gräsel 2010) Interventionen für eine (flächendeckende) Nutzung bereitgestellt werden können; im Falle der Scaling-up-Forschung insbesondere auch für erweiterte Nutzerzahlen, -gruppen und -settings. Zur *Disseminationsforschung*⁵ lassen sich Aktivitäten zählen, in denen spezifische Akteure versuchen, forschungsbasiertes Wissen in der Gesellschaft zu verbreiten und dies zu erforschen, in denen also die Kommunikation evidenten Wissens im Vordergrund steht, das i. d. R. zielgruppen- und kontextspezifisch aufbereitet wird. Im Bildungsbereich leisten dies Institutionen wie das What-Works-Clearing-House, das Clearing-House Unterricht an der Technischen Universität München oder das Netzwerk „Evidence Informed Policy and Practice in Education in Europe“ (EIPPEE), die Forschungswissen in systematischen Reviews bündeln. Die Disseminationsforschung hat mit der Transferforschung gemein, dass beide vor allem das Handeln der Anbieterseite untersuchen, da eine „Übergabe“ im Fokus steht (vgl. Graham et al. 2006, S. 16), weniger die Seite der Nutzenden, die in der Implementationsforschung (vgl. Kremer 2003, S. 71) wie in der Scaling-up-Forschung besondere Beachtung finden. Gegenüber den skizzierten Forschungslinien widmet sich die *Implementationsforschung* dem Prozess der Umsetzung einer Intervention in einem bestimmten Setting, mit dem Ziel, die institutionellen, organisationalen und personalen Bedingungen zu identifizieren, unter denen diese erfolgreich implementiert werden kann. So betonen Damschroder et al. (2013, S. 1 – Hervorh. i. Orig.): „Implementation research focuses on understanding how programs are implemented, translated, replicated, and disseminated in ‘real-world’ settings. It expands the focus of traditional research from discovering *what* works to also discovering *how* the implementation works in *specific contexts*“. Von *Implementationsforschung im weiteren Sinne* kann dann gesprochen werden, wenn zu Fragen der (Wirksamkeit von) Implementation pädagogischer Reformen und Interventionen geforscht wird. Dabei sind zwar meist zumindest Teile der Reformen wissenschaftlich legitimiert, aber nicht zwingend bzgl. ihrer Wirkungsannahmen auf Evidenz geprüft. Von *Implementationsforschung im engeren Sinne* sprechen wir hingegen dann, wenn die Inhalte der zu implementierenden Reformmaßnahmen nicht nur theoretisch fundiert sind, sondern bereits Evidenz für ihre Wirksamkeit vorliegt.

Die vorgenommenen begrifflichen Abgrenzungen sind nicht trennscharf und unumstritten, werden aber weithin geteilt. Die Implementationsforschung widmet sich anders als die Diffussionsforschung gezielten Interventionen in die pädagogische

³ Evaluation wird hier verstanden als die *wissenschaftlich fundierte, kriteriengeleitete Bewertung* von Voraussetzungen, Prozessen und/oder Ergebnissen des Real-Werdens wissenschaftlicher Erkenntnisse in Politik bzw. Praxis unter Berücksichtigung jeweils relevanter Akteursgruppen (vgl. Kuper 2005).

⁴ „Scale-up is the enactment of interventions whose efficacy has already been established in new contexts with the goal of producing similarly positive impacts in larger, frequently more diverse populations“ (Schneider und McDonald 2007a, S. 4). Scaling-up-Forschung „examines factors that influence the effectiveness of interventions as they are brought to scale across settings“ (Lynch et al. 2012, S. 306).

⁵ „Dissemination is an active approach of spreading evidence-based interventions to the target audience via determined channels using planned strategies“ (Rabin und Brownson 2012, S. 26).

Praxis unter aktiver Beteiligung der Wissenschaft. Anders als die Disseminationsforschung legt die Implementationsforschung den Fokus nicht vornehmlich auf die Kommunikation wissenschaftlichen Wissens, sondern auf seine Umsetzung unter Alltagsbedingungen. Gegenüber einer Interventionsforschung, die laborähnliche Bedingungen im Feld herstellt, um die gezielte Veränderung abhängiger Variablen durch die Variation unabhängiger (personaler oder institutioneller) Variablen kausal interpretieren zu können (vgl. Hascher und Schmitz 2010, S. 7f.), ist die Implementationsforschung vor allem an den spezifischen (Gelingens-)Bedingungen für die erfolgreiche Einführung von Reformen und Interventionen in der Praxis interessiert (zu ihrer Integration s. Spiel et al. 2018).

2.2 Modelle und Handlungsrahmen in der Implementationsforschung

Einen zweiten Schwerpunkt der Diskussion bildet die Entwicklung von Modellen oder Handlungsrahmen, die Implementationsprozesse beschreiben, erklären bzw. anleiten können. Überblicksdarstellungen enthalten diverse Modelle, die

- a) auf Einflussfaktoren und/oder Prozesse von Implementationen bezogen sind (vgl. Nilsen 2015),
- b) die Relevanz institutioneller Rahmenbedingungen und/oder die Interessen handelnder Akteure behandeln,
- c) die unterschiedlichen Wissensformen in Forschung, Politik und Praxis herausstellen oder
- d) die Herausforderungen in der Koordination der Handlungen in sozialen Prozessen betonen (vgl. Eldredge et al. 2016).

Mit Einflussfaktoren des Implementationserfolgs beschäftigt sich etwa der Consolidated Framework for Implementation Research (CFIR; Damschroder et al. 2009), der hier beispielhaft erwähnt wird. Der CFIR unterscheidet insgesamt fünf Bereiche potenzieller Gelingensbedingungen der Implementation einer Intervention:

1. *Charakteristika einer Intervention;*
2. *Charakteristika der äußeren Rahmenbedingungen* (die Bedarfe und Ressourcen der Adressaten, der „Kosmopolitanismus“ der Organisation und die erwartbaren Incentives);
3. Merkmale der *inneren Rahmenbedingungen* wie z. B. sozialstrukturelle Merkmale der implementierenden Organisation oder ihres Implementationsklimas;
4. Merkmale der *implementierenden Individuen*; und schließlich
5. *Merkmale des Implementationsprozesses.*

Reviews über weitere Rahmenmodelle bieten Meyers et al. (2012), Tabak et al. (2012) sowie Nilsen und Bernhardsson (2019).

Die Diskussion um solche i. d. R. induktiv entwickelten Modelle hat manche ihrer systematischen Grenzen verdeutlicht. So wird „Kontext“ zumeist unterspezifiziert und mit „Setting“ gleichgesetzt (wie im CFIR) und auf all das bezogen, was nicht zur Intervention selbst gehört, ohne aber z. B. an etablierte Unterscheidungen von Mikro-, Meso- und Makroebenen anzuschließen (vgl. Nilsen und Bernhardsson 2019, S. 14, S. 18f.). Zudem werden fördernde und hemmende Einflussfaktoren oft

nicht danach unterschieden, ob sie aus der jeweils gewählten theoretischen Perspektive als „variable Größen“ und prinzipiell veränderbar gelten (wie z. B. das Handeln der Implementierenden und der Nutzenden) oder aber als „Konstanten“ (wie etwa die „Kultur“ einer Organisation oder die institutionelle Struktur eines Bildungsbereichs), die sich oft eher langsam wandeln als dass sie gezielt verändert würden. Schließlich fehlt es weitgehend an einer Spezifizierung der Wirkmechanismen, mit denen die jeweiligen Faktoren Implementationsprozesse und -ergebnisse beeinflussen (vgl. Nilsen und Bernhardsson 2019, S. 2).

In einer Befragung von über 200 Implementationsforschenden aus insgesamt 12 Staaten durch Birken et al. (2017) zeigte sich, dass Rahmenmodelle vor allem aufgegriffen wurden, um Implementationsprozesse planen, Einflussfaktoren und Determinanten identifizieren und die Datenerhebung anleiten zu können. Nur 15 % der Befragten gaben an, theoretische Annahmen über Implementationsprozesse und ihre Wirkungen prüfen zu wollen (vgl. auch Munter et al. 2016). Insgesamt scheinen bisherige Rahmenmodelle der Implementationsforschung stärker durch den Wunsch zur Verbesserung von Politik und Praxis motiviert zu sein als durch einen Anspruch auf Theorieentwicklung oder das Absichern von Erklärungsansätzen, die über die ‚lokalen‘ Theorien hinausgehen, die auf das Treatment bezogen sind. Insofern sind bisher vorliegende Modelle oftmals nicht von Frameworks als bloßen Handlungsanleitungen zu unterscheiden (vgl. Albers et al. 2017).

2.3 Untersuchungsdesigns für Implementationsstudien

Die Diskussion um angemessene Untersuchungsdesigns von Implementationsstudien im Bildungsbereich hat wichtige Impulse aus der medizinischen Forschung erhalten, in der miteinander verschränkte Interventions- und Implementationsstudien eine lange Tradition haben. Typisch sind hier sogenannte „Efficacy“-Studien, in denen die Wirksamkeit einer nachweislich wirksamen Intervention unter *optimalen* (Feld-)Bedingungen erprobt wird. In „Effectiveness“-Studien geht es um die Frage, ob die in Interventionsstudien unter laborähnlichen Bedingungen nachweisbaren Effekte auch unter Alltagsbedingungen erhalten bleiben, also z. B. dann, wenn nicht Forschende, sondern Praktiker die Intervention selbst durchführen, ggfls. auch für neue Adressatengruppen. Effectiveness-Studien versuchen zu klären, wer von der Intervention profitiert und wie lange nach der Intervention diese Wirkungen in welchen Alltags-Kontexten anhalten (vgl. Flay 1986, S. 455). Auch Fragen nach den möglichen ‚Nebenwirkungen‘ (vgl. Gräsel und Parchmann 2004, S. 205) und personalen und kontextuellen ‚Gelingensbedingungen‘ werden bisweilen adressiert. Schließlich wird auch die Frage behandelt, ob die Wirkungen auch vor dem Hintergrund variabler Implementationsstrategien erhalten bleiben bzw. wie lange und wie weitreichend diese sind (vgl. Landsverk et al. 2012, S. 226f.).

In solchen gestuften Forschungsstrategien haben randomisierte Feldstudien im Kontrollgruppendesign (Randomized Controlled Field Trials/RCFTs) einen hohen Stellenwert. Connolly et al. (2018) berichten für die Jahre 1980–2016 von mehr als 1000 RCTs im Feld der Bildung, mit steigender Tendenz, besonders häufig in den USA, zahlreich auch in Großbritannien, deutlich seltener in Kontinentaleuropa. Selbstverständlich sind RCFT’s nicht die einzige Option für das Design von Imple-

mentationsstudien, insbesondere wenn es um die Wahrnehmungs-, Rezeptions- und Nutzungslogiken, also um die „Köpfe“ implementierender Akteure geht. Und umgekehrt sind nicht alle RCFTs Implementationsstudien im hier explizierten engeren Verständnis, da sie gelegentlich auch, wie in der Pharmaindustrie, aus vorwiegend kommerziellen Interessen für Interventionen durchgeführt werden, bei denen die Zulassung von Therapien für zahlungskräftige Kunden im Vordergrund steht, nicht ihre breite Implementierung für alle (Rothwell et al. 2016). Das besondere Potenzial von RCFTs für Implementationsstudien besteht jedoch in der experimentellen Prüfung theoretisch begründeter Annahmen unter Feldbedingungen (vgl. Connolly et al. 2018).

Gegen RCFTs wurde eingewandt, dass die Theorieprüfung zumeist auf Treatment und Outcome beschränkt sei und keine Erkenntnisse zum Implementationsprozess generiere. Diese Kritik führte zu einem Plädoyer für Mixed-Methods-Designs auch in der Implementationsforschung (vgl. Green et al. 2015). Über Mixed-Method-Designs wird in der empirischen Bildungs- und Sozialforschung seit einigen Jahrzehnten intensiv diskutiert. Eine entfaltete methodologische Debatte (vgl. Creswell und Creswell 2018) hat Leistungen und Grenzen einer sequenziellen oder simultanen Verknüpfung quantitativer und qualitativer Erhebungs- und Auswertungsverfahren in explorativen oder explanativen Studien aufgezeigt. In der Implementationsforschung eröffnen Mixed-Methods-Designs u. a. eine „dichte Beschreibung“ der Implementationsprozesse und der Interessen der beteiligten Akteure, gestützt auf semistrukturierte Interviews, standardisierte Fragebögen, Fokusgruppen, Dokumentenanalysen oder (nicht-)teilnehmende Beobachtung. Eine solche Forschungsstrategie bietet Optionen für die Steigerung der Adaptivität von Interventionen, für die Optimierung von Sampling-Strategien, für eine verbesserte Beurteilung der ökologischen Validität von Interventionen, für die Analyse der Interaktion von Akteuren und damit für die Prüfung von Kausalannahmen zu Wirkungen und Nebenwirkungen von Interventionen im Mehrebenensystem der Bildung (vgl. Palinkas 2014).

Ein weiterer Einwand betrifft die Verallgemeinerbarkeit der Befunde von RCFTs. In RCFTs erfolgt eine zufällige Zuweisung von Personen zu Untersuchungsbedingungen. Dies ließe sich auf die institutionellen und organisationalen Rahmenbedingungen erweitern, unter denen eine Implementation erfolgen soll. Auch auf dieser Ebene könnten Randomisierungen erfolgen. Allerdings wird Connolly et al. (2018, S. 289) zufolge die Verallgemeinerbarkeit von Befunden auf andere Kontexte bisher kaum diskutiert. Allenfalls differentielle Analysen für Subgruppen ließen sich als ein erster Schritt in diese Richtung finden.

Schließlich wird die Bedeutung von Umsetzungstreue (Treatment-Fidelity), die für RCFTs zur Sicherung interner Validität und zur Replizierbarkeit der Befunde unverzichtbar ist, in der Implementationsforschung inzwischen differenzierter bewertet. Umsetzungstreue bildete als Qualitätsmerkmal einer Intervention innerhalb der Implementationsforschung längere Zeit einen eigenen Forschungsschwerpunkt (z. B. Stains und Vickrey 2017). Zur Messung der Umsetzungstreue finden sich oftmals Kriterien wie die Befolgung von Vorgaben (adherence), die Dosierung (dose), die Qualität der Übermittlung (quality of delivery), die Ansprechbarkeit bzw. das Engagement der Beteiligten (responsiveness) oder die Differenzierung des eingeführten Programms (program differentiation) (vgl. Allen et al. 2012, S. 282, S. 286).

Inzwischen aber wird die Beachtung von Anforderungen externer Validität stärker betont, die zumeist die Adaptivität eines Treatments erfordert (vgl. Southam-Gerow und McLeod 2013).

2.4 Herausforderungen an Datenqualität und statistische Analysen

Unabhängig davon, ob Implementationsstudien auf Theorieprüfung zielen, ist die psychometrische Qualität der erfassten Variablen für die Belastbarkeit der Befunde von fundamentaler Bedeutung. Dies gilt zumal dann, wenn Variablen im Prozess der Implementation und in Abstimmung mit den Akteuren im Feld neu eingeführt und entsprechende Daten erhoben werden (vgl. Lyon et al. 2017). In solchen Fällen werden ideographische (d. h. sich aus den konkreten Fällen ergebende und ggf. die Interessen der Stakeholder berücksichtigende; vgl. Lewis et al. 2015) und nomothetische (sich aus der Prüfabsicht theoretischer Annahmen ergebende) Datenerhebungen (vgl. Lyon et al. 2017) miteinander verknüpft. Die Frage der Datenqualität wird forciert dadurch, dass eine ökonomische Erfassung von Variablen angesichts begrenzter Ressourcen bei Organisationen, Akteuren und Adressaten oft unabweisbar ist. Allerdings ist die Diskussion um das Measurement in der Implementationsforschung noch wenig entwickelt (vgl. Proctor und Brownson 2012, S. 261).

Einen weiteren Diskussionschwerpunkt bilden Fragen der statistischen Analyse von Daten aus Implementationsstudien, insbesondere dann, wenn sie als Mixed-Methods-Studien durchgeführt werden. Für RCFTs ergibt sich eine erste Herausforderung bereits aus der Entscheidung, ob die randomisierte Zuweisung der Untersuchungseinheiten zu Treatmentbedingungen auf der Ebene von Individuen oder von Clustern durchgeführt wird (vgl. Connolly et al. 2018, S. 285). Der Mehrwert von RCFTs wurde häufig in einem günstigen Verhältnis von Stichprobenumfang und statistischer Power gesehen (Slavin 2002, S. 17). Wenn aber mit Cluster-Randomisierungen gearbeitet wird, weisen die Daten regelmäßig eine Mehrebenenstruktur auf, die in statistischen Analysen angemessen berücksichtigt werden muss. Und in Mixed-Methods-Designs sind die erfassten Einflussfaktoren und Outcome-Variablen in aller Regel auf unterschiedlichen Handlungsebenen (z. B. Schulleitung und Kollegium, Lehrerinnen und Schüler) angesiedelt. Sie erfordern also eine *integrierte* und nicht nur *additive* Analyse quantitativer und ggf. auch qualitativer Daten.

2.5 Ethische Prinzipien und moralisches Handeln in Implementationsstudien

Ethische Standards empirischer Forschung wurden insbesondere in der experimentellen Forschung entwickelt und zielen auf den Schutz von Persönlichkeitsrechten sowie auf das informierte Einverständnis von Studienteilnehmenden mit der Erhebung und der Verwendung von Daten. Solche Standards gelten selbstverständlich auch in Implementationsstudien. Da Implementationsstudien aber nicht allein der Forschung dienen, sondern immer auch auf die Veränderung sozialer Praxen gerichtet sind, stellen sich weitergehende ethische Fragen. Die Diskussion zur Implementationsforschung hat gezeigt, dass eine einfache Extrapolation von Standards z. B. der (klinischen) Forschung in das Praxisfeld ebensowenig möglich ist wie umge-

kehrt eine bloße Übertragung berufsethischer Standards in das Forschungsfeld (vgl. Gopichandran et al. 2016).

Das betrifft zunächst etablierte Standards wie das informierte Einverständnis der untersuchten Personen, das bei Cluster-Randomisierungen, wie sie in Implementationsstudien verbreitet sind, schwierig einzuholen ist (vgl. Gopichandran et al. 2016, S. 7). Zudem werden Implementationsstudien häufig partizipativ gestaltet, mit Spielraum für Adaptivität. Damit geht einher, dass ggf. nicht alle ethisch relevanten Fragen vorab benannt und entschieden werden können, sondern iterativ bearbeitet werden müssen, z. B. im Blick auf die Frage, ob geplante Randomisierungen angesichts der Bedarfe spezifischer Adressatengruppen weiterhin gerechtfertigt werden können (vgl. Goodyear-Smith et al. 2015, S. 2, S. 5).

In Implementationsstudien übernehmen Forschende nicht nur ethische Verantwortung gegenüber den teilnehmenden Individuen, sondern auch gegenüber sozialen Gruppen, Organisationen und deren Umfeld. Mit diesen Studien geht der Anspruch einher, dass sich nicht nur die handelnden, sondern auch die verantwortlichen (politischen) Akteure zumindest zur Unterstützung der Implementation verpflichten (vgl. Spiel et al. 2018). Dies macht es erforderlich, in ethische Verfahrensweisen nicht nur jene einzubeziehen, von denen Daten erhoben werden, sondern auch weitere „Stakeholder“ (vgl. Douglas et al. 2015, S. 1830f.; Gopichandran et al. 2016, S. 5), insbesondere wenn es sich um langjährige Prozesse handelt, in denen sich möglicherweise der rechtliche Status und die Entscheidungskompetenz der Untersuchten ändern. Schließlich werden in Implementationsstudien inzwischen häufig digitale Medien genutzt (vgl. Waycott et al. 2015). Wenn anfallende Daten für personalisierte Interventionen genutzt werden, stellt sich u. a. die Frage, wann und unter welchen Bedingungen die Erforschten Zugang zu „ihren“ Daten erhalten, um autonom über die Fortsetzung ihrer Teilnahme entscheiden zu können (vgl. Skär und Söderberg 2018, S. 1044, S. 1046). Eine besondere ethische Herausforderung ergibt sich ferner aus der Tatsache, dass Implementationsstudien häufig Befunde aus einem kulturellen Kontext in andere Kontexte übertragen. Offen ist dann, ob forschungs- und berufsethische Standards in beiden Kulturen geteilt werden. Hier werden vorbereitende oder begleitende ethnographische Studien empfohlen, die die Perspektiven aller Beteiligten einbeziehen und die Risiken kultureller Kolonisation begrenzen (vgl. Sutoris 2018).

In der einschlägigen Literatur liegen bislang nur wenige Studien vor, in denen das moralische Handeln der Forschenden in Implementationsprozessen erfasst wird. Bei den wenigen Ausnahmen wird deutlich, dass Forschende und politisch Handelnde ethische und evidenzorientierte Erwägungen z. B. für die Beteiligung von Personen mit spezifischen Risiken unterschiedlich gewichten (vgl. Beima-Sofie et al. 2019). Kontroverse „Fälle“ könnten helfen, Forschende auf ethisch reflektierte und moralisch vertretbare Entscheidungen in Implementationsstudien vorzubereiten (vgl. Weidema et al. 2012).

3 Einflussfaktoren auf die Implementation evidenzbasierter Lehr-Lernangebote – Ein systematischer Literaturüberblick zu einem Schlüsselaspekt der Implementationsforschung

Während im vorigen Abschnitt Schwerpunkte der konzeptuellen und methodologischen Diskussion zur Implementationsforschung nachgezeichnet wurden, greifen wir im Folgenden einen Schlüsselaspekt der empirischen Implementationsforschung mit der Frage auf: „Welche Faktoren bedingen, dass die Implementation einer evidenzbasierten Intervention in reale pädagogische Handlungsfelder (nicht) gelingt“, so dass es ggf. zu „Implementationsbrüchen“ (van Ackeren et al. 2011, S. 174) kommt?

Zur Frage der Einflussfaktoren auf Implementationen in Bildungskontexten findet man bereits Einzelstudien (z. B. Eickelmann 2011; Long et al. 2016; Hettfleisch et al. 2017), konzeptuelle Arbeiten (z. B. Beelmann und Karing 2014; Hasselhorn et al. 2014) sowie einige Reviews (z. B. Durlak und DuPre 2008; Baeten und Simons 2014). Hier ergänzen wir diese Literatur um einen systematischen Überblick empirischer Befunde zu der spezifischen Forschungsfrage, welche Faktoren in pädagogischen Handlungsfeldern *organisierten* Lehrens und Lernens die (auch) *mikrodidaktische Implementation evidenzbasierter Lehr-Lernangebote* durch *pädagogische Professionelle* (z. B. Lehrende) beeinflussen. Dabei berücksichtigen wir anders als vorherige Ansätze erstens *alle* Bildungsbereiche des organisierten Lehrens und Lernens – von der Frühpädagogik bis hin zur Erwachsenen- und Weiterbildung. Zweitens verzichten wir darauf, die Recherche auf spezifische Implementationsgegenstände (wie sonst häufig, z. B. Nathan et al. 2018) oder bestimmte Publikationsorgane einzugrenzen. Drittens setzen wir das Recherche-Enddatum auf den 31. Juli 2019 (vgl. Abb. 1), *ohne* zeitliche Limitation für die Vergangenheit. Neben diesen gezielt aufgehobenen Begrenzungen wurden jedoch auch Schwerpunkte im Sinne der Forschungsfrage gesetzt: So fanden nur empirische Studien Berücksichtigung, die Einflussfaktoren zum *primären* Untersuchungsgegenstand hatten. Zudem wurden vor dem Hintergrund der oben differenzierten Forschungslinien nicht explizit nach Transfer-, Diffusions-, Disseminations- oder Evaluationsforschung gesucht, sondern Einflussfaktoren auf *Implementationen* fokussiert, die Lehr-Lernangebote zum Gegenstand haben. Da uns die Bedingungen der Möglichkeit *wissenschaftlich fundierter* Verbesserungen interessieren, wurde nur *empirische* Implementationsforschung im *engeren* Verständnis (s. oben) einbezogen. Zudem entschieden wir uns, nur Lehr-Lernangebote zu berücksichtigen, die von *professionellen PädagogInnen* – nicht z. B. von Eltern oder Peers – implementiert werden, um im Ergebnis nicht personenbezogene Einflussfaktoren von Akteuren ausweisen zu müssen, die nicht typisch für das professionelle Bildungssystem sind. Diese Suchstrategie führt dazu, dass die Befunde selbstredend nur einen spezifischen Ausschnitt aktueller Erkenntnisse der Implementationsforschung zu Einflussfaktoren darstellen und keinen Anspruch auf Vollständigkeit erheben; zumal aus forschungspragmatischen Gründen alle Implementationsforschung aus der Suche ausgeschlossen werden musste, die ohne den Terminus „implement*“ in den relevanten Suchabschnitten auskommt (z. B. nur von „Transfer“ spricht, damit aber de facto Implementation meinen könnte). Die erfassten Studien werden umfassend in Tab. 2 des Anhangs beschrieben. Als

Interpretationsrahmen für die Ergebnisse wurde die Grundheuristik des Angebots-Nutzungs-Wirkungs-Modells (ANW-Modell) für organisierte Lehr-Lernsituationen in Anlehnung an Helmke (2015) herangezogen.

3.1 Methodisches Vorgehen

Ende März 2018 fand ein erster Durchgang der systematischen Literaturrecherche statt, auf die Anfang August 2019 ein Update für den Erscheinungszeitraum April 2018 bis 31. Juli 2019 erfolgte. Tab. 1 enthält einen Überblick der in fünf Datenbanken zu beiden Recherchezeitpunkten jeweils identisch umgesetzten Suchprozeduren.

Die sich an die Datenbankrecherche anschließenden Arbeitsschritte sind in Abb. 1 illustriert. Umgesetzt wurden diese von der Person aus der Autorengruppe, die auch die Datenbankrecherchen durchgeführt hatte.⁶

Aufgrund der in Abb. 1 (Schritt 3a & 4b) festgelegten Einschlusskriterien konnten von den dublettenbereinigten 17.715 Datenbankesamttreffern nur 33 in die Auswertung eingehen, die zwischen 2003 und 2019 in referierten Zeitschriften publiziert wurden. Denn im Schritt 3a fielen knapp 90% der Datenbanktreffer allein deshalb heraus, weil sie sich auf Handlungskontexte wie z. B. Gesundheitswesen,

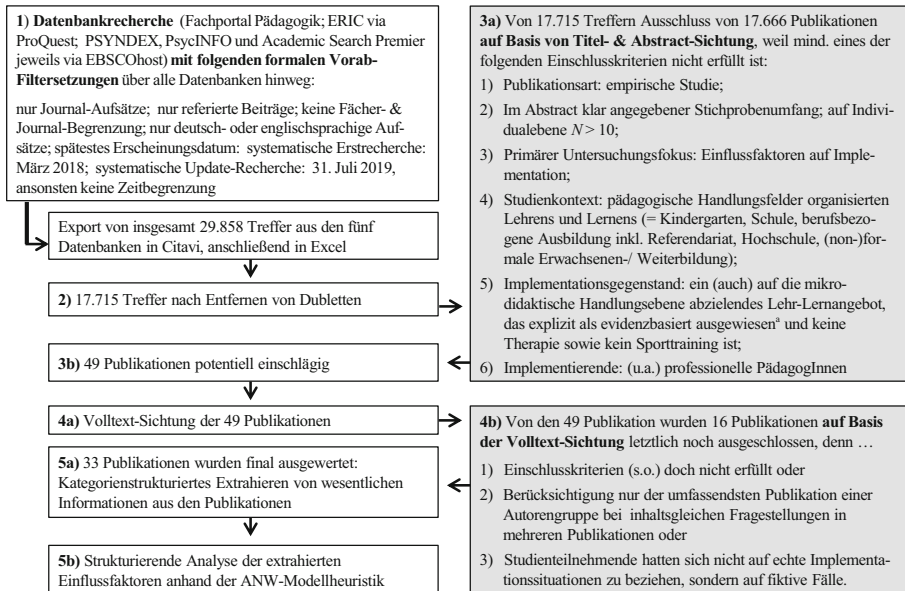
Tab. 1 Übersicht Recherchestrategie der Datenbankrecherche

Datenbank	Kombinationen von jeweils zwei eingesetzten Suchbegriffen und Suchfeldern pro Suchdurchlauf			
ERIC ^a	<i>implement*</i>	AND	<i>determin*</i>	... als zweiter Suchbegriff jeweils im Suchfeld
Academic Search Premier ^a	(als erster Suchbegriff jeweils im Suchfeld „Titel“ bzw. „Schlüsselwort“)	AND	<i>barrier</i>	„Abstract“
PsycINFO ^a		AND	<i>influen*</i>	
PSYINDEX ^a		AND	<i>factor</i>	
Fachportal Pädagogik ^b	<i>implement*</i>	AND	<i>condition</i>	
	(als erster Suchbegriff jeweils im Suchfeld „Titel“ bzw. „Schlüsselwort“)	AND	<i>Hindernisse</i>	... jeweils als zweiter Suchbegriff mal im Suchfeld
		AND	<i>Einfluss</i>	„Schlüsselwort“, mal „Titel“, mal „Freitext“
		AND	<i>Bedingung</i>	
		AND	<i>Faktor</i>	
		AND	<i>determin*</i>	
		AND	<i>Herausforderung</i>	
		AND	<i>barrier*</i>	

^aDie Recherche in der Datenbank ERIC fand via ProQuest statt, die in ASP, PsycINFO und PSYINDEX jeweils via EBSCOhost. Jeweiliger Suchmodus in EBSCOhost: „Find all my search terms“

^bFachportal Pädagogik ermöglichte zum ersten systematischen Recherchedurchgang Ende März 2018 noch keine Stichwortsuche über „Abstract“ als Suchfeld. Daher wurden hier ersatzweise andere Suchfelder für den jeweils zweiten Suchbegriff herangezogen. Um bei der späteren systematischen Update-Recherche im August 2019 nicht von dieser systematischen Recherchestrategie grundlegend abzuweichen, wurde an dieser Suchstrategie festgehalten, wenngleich im August 2019 eine Stichwortsuche über „Abstract“ als Suchfeld mittlerweile möglich gewesen wäre

⁶ Der letzte Schritt 5b umfasste zum einen die strukturierende Aufbereitung und Analyse der Einflussfaktoren entlang des zugrunde gelegten ANW-Modells; zum anderen bedeutete er ein inhaltliches Bündeln von Einzelfaktoren innerhalb der jeweiligen ANW-Modellelemente (s. hierzu Tab. 3 und 4 im Online-Anhang). Schritt 5b wurde von einer zweiten Person aus der Autorengruppe anteilig überprüft. Sehr wenige Abweichungen wurden konsensuell aufgelöst.



^a Den AutorInnen ist bewusst, dass eine fehlende Explizierung von Evidenz im Titel und Abstract einer Studie kein eindeutiges Indiz dafür ist, dass der Implementationsgegenstand nicht doch wissenschaftlich geprüft wirksam ist. Aus forschungspragmatischen Gründen konnte jedoch im Erstzugriff nicht im Volltext aller 17.715 Publikationen danach gesucht werden.

Abb. 1 Flussdiagramm zur Illustration des Gesamtvorgehens zum systematischen Literaturüberblick

Sozial- und Familienarbeit, Wirtschaft oder Kriminologie beziehen – nicht aber auf Kontexte organisierten Lehrens und Lernens. Von den verbleibenden rund 10% an Publikationen wurden rund 25% ausgeschlossen, weil sie keine empirischen Studien, sondern konzeptuelle Beiträge, Erfahrungsberichte, Überblicksbeiträge oder Sonstiges umfassten. Ausgeschlossen wurden Publikationen auch, wenn der Stichprobenumfang auf Individualenebene bei $N \leq 10$ lag oder nicht präzisiert war, sie Einflussfaktoren auf Implementation nur am Rande behandelten, der fokussierte Implementationsgegenstand nicht (auch) auf die Mikroebene organisierten Lehrens und Lernens abhob und kein Lehr-Lernangebot umfasste, Implementierende keine professionellen PädagogInnen waren und/oder der Implementationsgegenstand nicht eindeutig als wissenschaftlich geprüft deklariert wurde. Ausgeschlossen wurden Publikationen aus forschungspragmatischen Gründen, wenn im Abstract unklare oder keine Angaben zu in Abb. 1 (Schritt 3a) aufgelisteten Einschlusskriterien erkennbar waren.

3.2 Beschreibung und Deutung des inhaltsanalytisch ausgewerteten Studiencorpus

Die final identifizierten 33 Publikationen beziehen sich auf 30 unterschiedliche empirische Studien (vgl. Übersicht in Tab. 2 im Anhang). Von diesen wurden 25 im pädagogischen Handlungsfeld Schule (davon zwei in der Lehrerbildung) und 5 im

Kindergarten durchgeführt. Insbesondere die angelsächsischen Ganztags schulangebote (19 Studien durchgeführt in den USA, zudem insgesamt 7 aus Australien, Großbritannien etc.) erlauben umfangreichere pädagogische Interventionen anhand größerer, stabiler Personengruppen (20 Studien mit mehr als 150 Teilnehmenden auf Individualebene – B, C, E, F, G, H, I, J, K, L, M, N, O, U, W, X, Z, a, b/c, f).

Insgesamt wurden in 11 Publikationen (B, M, O, Q, R, S, V, W, Z, e, g) Daten von Akteuren nicht nur einer pädagogischen Handlungsebene erhoben. Das bedeutet jedoch nicht, dass diese Daten immer auch mehr Ebenenanalytisch ausgewertet wurden. Die Mehrebenenstruktur (d.h. mindestens zwei Handlungsebenen⁷) findet in 11 Publikationen in quantitativen Analysen Berücksichtigung (A, C, F, O, P, R, T, Z, b, c, f). Dabei kann mit Blick auf den Stichprobenumfang Studie N Einflussfaktoren in einer Mehrebenenstruktur analysieren (– mehr als 30 Teilnehmende auf der 2. Ebene ($N=69$ LehrerInnen) und mehr als 10 SchülerInnen je LehrerIn auf 1. Ebene (insg. $N=1721$ SchülerInnen)). Insgesamt werden in 27 Publikationen statistische Auswertungen berichtet. Überwiegend werden Zusammenhänge zwischen jeweils nur zwei Merkmalen untersucht bzw. in separaten Analysen geprüft, kaum jedoch medierende oder moderierende Zusammenhänge zwischen mehreren Variablen (Ausnahmen z.B. in den Studien Q, R, Z & f). In 10 Publikationen werden (auch) qualitativ inhaltsanalytische Auswertungen realisiert (s. Tab. 2 im Anhang).

Bei den Implementationsgegenständen sind genuin lehr-lernbezogene Interventionen im Sinne pädagogischer (Fach-)Inhalte in der deutlichen Minderzahl: nur 6 Studien beschäftigen sich mit Mathematik, Leseverstehen, Classroom Management etc.: D, H, I, L, X, d. Blickt man auf die Länder und die Geldgeber der Studien, spiegeln möglicherweise die Implementationsgegenstände überwiegend problematische ‚soziale Epidemien‘, die vor Ort als besonders veränderungsbedürftig gelten: In den Studien, die in den USA (19 Studien) und Canada (1 Studie), Australien (1 Studie), Südafrika und Uganda (3 Studien) sowie den Bahamas (1 Studie) durchgeführt wurden, dominieren gesundheitsbezogene Interventionen (AIDS, ungewollte Schwangerschaft, Drogen, Übergewicht/Bewegungsmangel), in Skandinavien (3 Studien) z.B. Anti-Bullying Programme. Die Interventionen sind teils eingebunden in staatlich vorgegebene Curricula (z.B.: E) oder explizit staatlich gefördert und eingefordert (z.B.: F, G, K, U, b, c). Sie werden fast alle von Drittmittelgebern finanziert, die sich der Bekämpfung dieser gesellschaftlichen Nöte verschrieben haben (vgl. Tab. 2 im Anhang). Im Mittelpunkt vorliegender Implementationsstudien stehen also überwiegend Lehr-Lernangebote, die Gesundheit und prosoziales Verhalten durch gezielte Interventionen evidenzbasiert fördern und deren Einsatz als politisch gewollte, weil staatlich finanzierte Reaktion auf gesellschaftlich unerwünschte Deviationen verstanden werden können, nicht jedoch Lehr-Lernangebote, die einen Verbesserungsimpetus für Kernaufgaben des Bildungssystems beinhalten. Insgesamt scheint das forschersiche Interesse an Einflussfaktoren auf Implementation im engen Sinne (s. oben) jedoch in den letzten Jahren deutlich anzusteigen. So

⁷ Differenziert wird hier explizit zwischen *Handlungsebenen* (Mikro-, Meso- & Makroebene) und *Auswertungsebenen* (z.B. Klasse/Lehrer (= Gruppenebene) und Schüler (= Individualebene), die jedoch derselben Handlungsebene angehören).

befinden sich im Corpus doppelt so viele Publikationen aus den Jahren 2013 bis 2019 (22 Publikationen) gegenüber dem gesamten Zeitraum davor (11 Publikationen).

Unter den hier berücksichtigten Studien finden sich 8 RCFTs, deren Datenerhebung auf unterschiedlichen Methoden basiert (C, F, G, M, N, R/b/c, T, a); daneben finden sich mixed methods designs allerdings auch in Studien, die keine RCFT-Designs beinhalten (B, J, L, U, W, d, f). Methodisch besonders komplexe und in der Durchführung aufwändige Forschungsdesigns, mit deren Hilfe hinderliche und förderliche Einflussfaktoren auf Implementationsprozesse und -ergebnisse in (large scale) RCFT (mit mixed methods design) auf mehreren Ebenen analysiert werden können, scheinen vor allem dort zu entstehen, wo der Leidens- bzw. Rechtfertigungsdruck von Praxis bzw. Politik besonders hoch ist. Die Reduktion der Wissenschafts-Praxis-Kluft durch eine Suche nach *evidenzbasierten* Abhilfe-Möglichkeiten in Form von hilfreichen Veränderungswissen ist jedenfalls häufig in den Einleitungskapiteln der entsprechenden Studien zu finden (z. B.: E, G, K, M, P, Q, R/b/c, S, W, a, e). Die hier untersuchten Implementationsstudien scheinen also nicht vornehmlich theorieprüfend oder -generierend und damit wissenschaftsimmanent motiviert zu sein. Vielmehr scheint die (Handlungsnotwendigkeit der) Praxis der Wissenschaft vorauszuweichen und diese zu begründen.

Für diese Deutung spricht auch, dass die meisten identifizierten Studien einen sich erst allmählich festigenden Forschungsdiskurs widerspiegeln, keine ausgereifte „Implementation Science“. Mehrere Beobachtungen begründen diese Einschätzung: So wird der bisher bereits erarbeitete ‚genuine‘ Theorie- und Modellstand der „Implementation Science“ (s. oben, Abschn. 2) wenig rezipiert: nur 12 von 33 Publikationen ziehen zum Verorten der untersuchten Einflussfaktoren in Implementationskontexten explizit Modelle heran (B, J, M, N, O, P, Q, R, S, T, a, f). Selbst wenn eine Bezugnahme erfolgt, wird meist nur auf Modelle rekurriert, die den Einfluss der Lernenden auf den Implementationserfolg nicht oder nur am Rande in den Blick nehmen (Ausnahmen: J, N). Die oben angesprochenen Aspekte der Ethikdiskussion, die für die Implementationsforschung spezifisch sind, werden nur in einem Ausnahmefall (e) aufgegriffen. Schließlich verweist eine zusätzlich durchgeführte Recherche der Forschungsprofile aller (Co-)AutorInnen der identifizierten 33 Publikationen eher auf Fachexpertise zu den Implementationsgegenständen denn auf Expertise zum Thema „Implementationsforschung“.

3.3 Befunde: Einflussfaktoren auf Implementation evidenzbasierter Lehr-Lernangebote

Zur Verortung der förderlichen und hemmenden Einflussfaktoren wurde die Grundheuristik des Angebots-Nutzungs-Wirkungs-Modells herangezogen. Mit diesem Modell werden pädagogische Implementationssituationen – analog zu organisierten Lehr-Lernsituationen – als mehrfache, je eigenlogische Co-(Re-)Konstruktions- und Nutzungsprozesse verschiedener Akteure, z. B. Lehrende sowie Lernende, konzeptualisiert (vgl. zur Übertragung der ANW-Grundheuristik auf Implementationssituationen Hetfleisch 2018). Durch diese Reflexionsfolie finden nicht nur potentielle Einflussfaktoren unterschiedlicher Handlungsebenen (Mikro-, Meso- & Makroebene) Berücksichtigung, sondern es werden auch potentielle Einflussfaktoren

analysierbar, die in anderen implementationsbezogenen Rahmenmodellen nicht (in dieser Ausdifferenzierung) modelliert sind – etwa die Lernenden als ‚Endnutzer‘ evidenzbasierter Lehr-Lernangebote. Deren Wahrnehmung und Rezeption sowie deren Nutzungsweisen der zu implementierenden Lernangebote vor dem Hintergrund (1) ihrer Lernvoraussetzungen, (2) ihrer Sozialisations- und Lebenskontexte sowie (3) implementations-, d. h. lehr-lernsituationsspezifischer Kontextbedingungen sind regelmäßig nicht *Kernbestandteil* von Implementationsmodellen⁸.

Insgesamt werden 194 förderliche bzw. hemmende Faktoren in den 33 Publikationen berichtet (s. Tab. 3 im Online-Anhang). Inhaltlich gebündelt resultierten 86 verschiedene Einflussfaktoren, die in den Abb. 2 und 3 entlang der Heuristik des Angebots-Nutzungs-Wirkungs-(ANW)-Modells verortet sind.

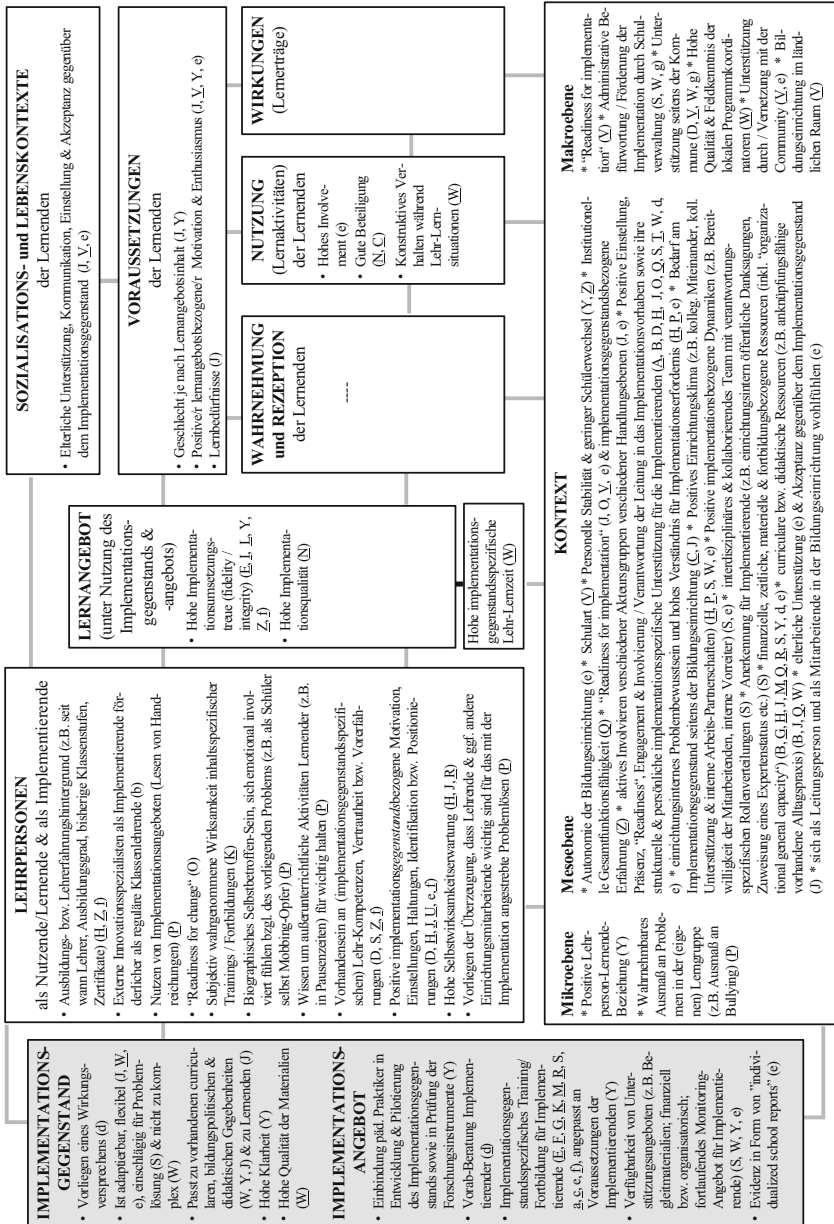
Betrachtet man die in den beiden Abbildungen zusammengefassten Ergebnisse zu den berichteten förderlichen (Abb. 2) und hemmenden Einflussfaktoren (Abb. 3)⁹, so fallen relative Lücken bei den Feldern „Voraussetzungen“, „Wahrnehmung und Rezeption“ sowie „Nutzung“ der Lernenden ins Auge. Dies markiert eine auffällige Diskrepanz zwischen der hohen Wirkungserwartung, die mit dem Motiv zusammenhängt, einer bestimmten (problembehafteten) Klientel zu neuen Verhaltensmöglichkeiten verhelfen zu wollen, und der tatsächlich geringen Empirie zur Untersuchung von Einflussfaktoren bei dieser Klientel. Die Rolle der ‚Endnutzer‘ von Implementationsangeboten ist nicht häufig Gegenstand der bisherigen Implementationsforschung (Ausnahmen: C, J, N, V, W, Y, e, f, g sowie die Korrelationsstudie D¹⁰). Besonders intensiv beforscht und empirisch besonders häufig als förderliche Einflussfaktoren berichtet werden Kontextfaktoren wie die ideelle Unterstützung der Implementierenden durch die Organisationsleitung (12 mal berichteter Faktor), die Unterstützung durch finanziell-materielle Ressourcen (11 mal berichtet) sowie ein positives Einrichtungsklima (5 mal berichtet), aber auch Trainings für Implementierende (15 mal berichtet, vgl. neben „Kontext“ zudem Feld „Implementationsangebot“ in Abb. 2).

Die untersuchten Implementationsstudien scheinen eher aus einer „top down“-Perspektive die chronologisch plausiblen „Nadelöhre“ zu fokussieren, ohne sie auch theoretisch zu reflektieren. Der Blick auf die Adressaten als Träger von Einflussfaktoren der Implementation scheint ebenfalls noch kaum theoretisch fundiert oder

⁸ So modelliert bspw. der CFIR ‚Endnutzer‘ in Bezug auf ihre Voraussetzungen und Ressourcen als Bestandteil (nur) „äußerer Rahmenbedingungen“, weist ihnen darin eine eher passive Einflussrolle zu, anstelle sie ebenfalls auch als aktive „individuals involved with the intervention and/or implementation process“ (Damschroder et al. 2009, S. 5/15) zu modellieren, wie es hingegen die ANW-Modellierung vornimmt.

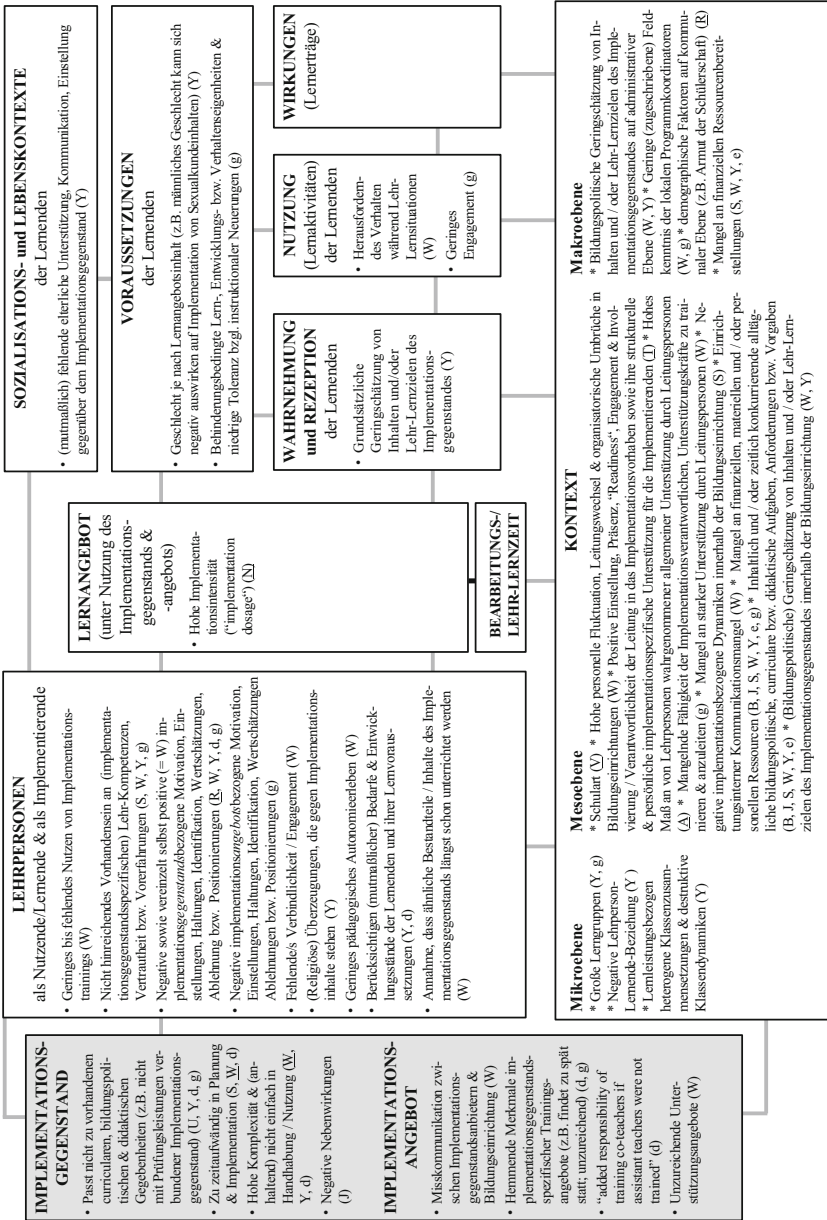
⁹ Eine detaillierte Auflistung der in den 33 Publikationen ermittelten Einflussfaktoren und ihrer jeweiligen „abhängigen Variablen“ findet sich im Online-Anhang in den Tab. 3 und 4. In den Abb. 2 und 3 sind die jeweils adressierten abhängigen Variablen der Übersicht halber nicht mit aufgeführt.

¹⁰ Bei der Studie D (ebenso wie bei Studie O) handelt es sich um eine reine Korrelationsstudie. Daher sind keine Kausalitäten überprüft worden. Dennoch sind in Studie D berichtete Faktoren im vorliegenden Beitrag i. S. von Einflussfaktoren (s. Abb. 2 und 3 sowie Online-Anhang-Tab. 3 & 4) und mutmaßliche abhängige Variablen (s. Online-Anhang-Tab. 3 & 4) aufgenommen worden. Denn manche auf systematische Zusammenhänge hin geprüfte Variablen lassen aus theoretischer Sicht durchaus spezifische Chronologie-, d. h. Wirkrichtungsannahmen plausibel erscheinen (z. B. „Vorwissen der Lehrperson“ beeinflusst „Implementationsprozess“ (Studie D); „Readiness for change/implementation“ beeinflusst „Implementation“ (Studie O)).



in Klammer gesetzt sind die Verweise auf die Publikationen, in denen der jeweilige Faktor als förderlicher Einfluss ausübend berichtet wurde (zur Buchstaben-Publikationszuordnung siehe Tab. 2 bzw. Literaturberichts-separates Literaturverzeichnis), Publikationsverweise zu Befinden, die aus Signifikanztests stammen, sind unterstrichen.

Abb. 2 ANW-Modellheuristik zur Verortung der als förderlich berichteten Einflussfaktoren



In Klammer gesetzt sind die Verweise auf die Publikationen, in denen der jeweilige Faktor als hemmendes Einflusssfaktor berichtet wurde (zur Buchstaben-Publikationsanordnung siehe Tab. 2 bzw. Literaturüberblick-separates Literaturverzeichnis). Publikationsverweise zu Befunden, die aus Signifikanzttests stammen, sind unterstrichen.

Abb. 3 ANW-Modellheuristik zur Verortung der als hemmend berichteten Einflussfaktoren

empirisch belegt zu sein. Dies mag einerseits überraschen, zumal die Lernenden mit ihren Lernvoraussetzungen, Rezeptions- und Nutzungslogiken den Kern bzw. Ausgangspunkt jeglicher Erziehungs- und Bildungswirklichkeit bilden, auf die jede professionelle pädagogische Intervention ausgerichtet sein müsste. Andererseits sind die Implementationsgegenstände evidenzbasiert, haben sich also bereits unter kontrollierten Bedingungen bei anderen Lernenden als wirksam erwiesen. Aus forschungsökonomischen bzw. -pragmatischen Gründen könnte daher bewusst darauf verzichtet worden sein. Gegen diese Annahme des bewussten Aussparens spricht jedoch, dass die Rolle des ‚Endnutzers‘ auch in den entsprechenden theorieinspirierten Modellen bisher kaum reflektiert wird (s. oben). Kontextmerkmale der Mikro- und der Makroebene – relativ betrachtet zur Mesoebene – wurden bisher empirisch ebenfalls eher wenig ausgeleuchtet (vgl. Abb. 2 und 3, Feld „Kontext“) oder theoretisch durchdrungen.

Ferner ist bemerkenswert, dass es sich sowohl bei den in den Studien als förderlich als auch als hemmend berichteten Einflussfaktoren zumeist um „variable Größen“ handelt, die grundsätzlich (pädagogisch) beeinflusst werden können, nicht um eher schwer veränderbare „Konstanten“. Betrachtet man die Ergebnisse zu hemmenden Einflussfaktoren (Abb. 3), so fällt auf, dass – relativ zu den als förderlich berichteten Faktoren – deutlich weniger Studien dazu überhaupt Befunde liefern (vor allem S, W, Y, d, g). Zweitens wird deutlich, dass der Grad an Fidelity ein Faktor ist, dessen Wirkung uneindeutig ist: Innerhalb des vorliegenden Studiencorpus wird in insgesamt 8 Publikationen der Einfluss der Umsetzungstreue auf Lernerträge untersucht. Ein hohes Maß davon erweist sich mehrheitlich als förderlich (E, I, L, M, Y, Z, f), aber nur bis zu einem gewissen Grad. Denn zu viel Umsetzungstreue wird vereinzelt zum hemmenden Faktor (*N*), weil dann Adaptation an die Verhältnisse vor Ort nicht mehr gelingt. Zudem zeigt der Einflussfaktor wider Erwarten in Teilen doch keinen Effekt auf spezifische Lehr-Lernziele (s. Tab. 4 im Online-Anhang: I, N, T, Z). Ähnlich verhält es sich mit einer umsetzungstreuen „Dosierung“: Sie erscheint förderlich (G, a, f) solange die Zeitinvestition nicht zum hemmenden Faktor wird (*N*).

Aussagekräftigere Befunde zum Umgang mit dem Spannungsfeld von interner und externer Validität, von „Wahrheit“ und „Nützlichkeit“, von Fidelity und Adaptivität, der „Dosierung“ etc. liegen bislang nicht vor. Es bleibt häufig bei allgemeinen Empfehlungen, eine Balance zu finden. Nur in Ausnahmefällen wird genau diese empirisch offene, theorie- und zugleich praxisrelevante Frage des günstigsten Ausmaßes an Adaptionsspielräumen, Dosierungen etc. zum Gegenstand (quasi-)experimenteller Implementationsforschung erhoben (vgl. hierfür L).

4 Perspektiven der Implementationsforschung im Bildungsbereich

Einleitend haben wir die Frage aufgeworfen, wie jene Wissenschaften, die sich mit Prozessen der Bildung und Erziehung beschäftigen, auf die Bedrängnis reagieren, als Wissenschaften *von* Praxis empirisch fortwährend über die Grenzen ihrer Relevanz *für* Praxis belehrt zu werden. Die empirische Bildungsforschung, die im Fokus dieses Beitrages steht, begreift die Überwindung dieser Kluft als eine methodische

Herausforderung. Von der Erweiterung ihres Repertoires um Implementationsstudien erhofft sie sich theoretisch informierte und empirisch gesättigte Einsichten in die Bedingungen der Möglichkeit wissenschaftlich fundierter Verbesserungen in der Bildungspraxis.

Das Ziel des Stichwortbeitrages bestand darin, die konzeptuelle Diskussion ebenso wie den empirischen Stand der Implementationsforschung systematisch zu sichten, um für die Bildungsforschung Erträge und offene Fragen zu identifizieren. Konzeptuelle Übereinkünfte, aber auch Differenzen zeigten sich sowohl in begrifflichen Abgrenzungen zur Diffusions-, Evaluations-, Transfer-, Scaling-Up- und Disseminationsforschung als auch in der Erarbeitung von Rahmenmodellen, die Implementationsstudien anzuleiten und Einflussfaktoren zu systematisieren erlauben. Dies gilt auch für die Verständigung auf angemessene Untersuchungsdesigns, in der die besonderen Potenziale von RCFTs anerkannt, aber zugleich auch die Notwendigkeit von Mixed-Methods-Designs gesehen wird (vgl. Slavin 2002), um theoretische Annahmen zu Einflussfaktoren auf Implementationsprozesse und -ergebnisse prüfen zu können. Schließlich wurden besondere ethische Herausforderungen in Implementationsstudien identifiziert, wenn auch noch nicht codifiziert.

Diese konzeptuellen Fortschritte haben dazu beigetragen, dass die Zahl an Implementationsstudien im Bildungsbereich in den vergangenen Jahren erkennbar gestiegen ist. Dies zeigte sich auch an einem Kernbereich der Bildungsforschung, der Implementation evidenter mikrodidaktischer Interventionen durch pädagogische Praktiker, die hier fokussiert wurden. Die systematische Recherche dazu deutet an, dass die gestiegene Zahl an Implementationsstudien oft veränderten Erwartungen und Förderpraxen der Politik folgt. Deutlich wurden zudem kulturelle Differenzen zwischen Staaten sowie institutionelle Differenzen zwischen Bildungsbereichen. Für den Bereich der Schule ist auffällig, dass unter den final berücksichtigten Publikationen mehr Studien an ihren Erziehungs- als an ihren Unterrichtsauftrag anschließen, insofern Interventionen zur sozialen und gesundheitlichen Prävention und Förderung häufiger erforscht werden als solche zur Optimierung curricular gerahmter Lehr-Lernprozesse.

Die vorgenannten Befunde beziehen sich jedoch – dies muss einschränkend gesagt werden – auf einen spezifischen Aspekt der Implementationsforschung, nämlich auf die Einflussfaktoren (nicht) gelingender Implementationen. Dabei führt schon allein die Fokussierung auf ausschließlich pädagogische Handlungsfelder im Rahmen des systematischen Forschungsüberblicks in Abschn. 3 zu einem Ausschluss von fast 16.000 der rund 17.700 zuerst identifizierten Publikationen. Das Vorgehen zielte darauf ab zu eruieren, welche Faktoren empirischen Studien zufolge in pädagogischen Handlungsfeldern organisierten Lehrens und Lernens die vornehmlich mikrodidaktische Implementation evidenzbasierter Lehr-Lernangebote durch pädagogische Professionelle (z. B. Lehrende) beeinflussen. Diese Schwerpunktsetzung führte dazu, dass die 33 inhaltlich ausgewerteten Publikationen und die im Abschn. 3.3 thematisierten Befunde nur einen Ausschnitt aktueller Erkenntnisse der Implementationsforschung darstellen (können) (vgl. zur Ausschlußbegründung einer Vielzahl an Implementationsstudien Abschn. 3.1).

Neben unabweisbaren Fortschritten in der gesamten Implementationsforschung zeigen sich aber auch Grenzen und offene Fragen, und zwar in theoretischer, me-

thodischer und datenanalytischer Hinsicht. Dies wird in der Zusammenschau der Ergebnisse nicht nur des systematischen (vgl. Abschn. 3), sondern vor allem des narrativen (vgl. Abschn. 2) Forschungsüberblicks deutlich. Viele Implementationsstudien fokussieren das (evidenzbasierte) Treatment. Personale, organisationale und institutionelle Einflussfaktoren, die den Implementationsprozess und auch das Implementationsergebnis beeinflussen, werden zwar sehr häufig empirisch erfasst, aber kaum theoretisch reflektiert, z. B. im Blick auf ihre moderierenden oder mediierenden Einflüsse oder im Blick auf ihr Zusammenspiel in einem ebenenübergreifenden Wirkgefüge. Die hier erfassten Studien haben den Kontakt zu den zahlreich vorhandenen Rahmenmodellen der Implementationsforschung selbst noch kaum hergestellt; und diese Rahmenmodelle folgen eher einer Pragmatik in der Entwicklung von Untersuchungsplänen als explizit theoretischen Annahmen. Dies gilt sowohl für Merkmale der so genannten „Endnutzer“, die eine Implementation beeinflussen (vgl. Hetfleisch 2018), als auch für den „Kontext“ (vgl. Damschroder et al. 2009). Zudem haben wir aufgrund des hier zugrunde gelegten Reviewverständnisses die Frage ausgeklammert, inwieweit die unterschiedenen Einflussfaktoren und die von ihnen erwarteten Wirkungen theoretisch angemessen konzipiert wurden oder ob die empirischen Analyseverfahren diesen Annahmen angemessen ausgewählt und durchgeführt wurden. Diese theoretischen Desiderate können auch die methodischen Weiterentwicklungen (vgl. Abschn. 2.3) der vergangenen Jahre nicht beheben. So werden RCFTs auch heute noch häufig als „single-point-in-time interventions“ (vgl. Hedges 2018) durchgeführt. Weiterentwicklungen zielen auf die Etablierung adaptiver oder sequentieller Designs, in denen nach der besten Kombination der Komponenten einer Intervention gesucht wird (vgl. Collins et al. 2007). Auch optimierte RCFTs sind allerdings dem Risiko ausgesetzt, dass aus einer experimentellen Studie ein „natürliches Experiment“ wird, z. B. wenn aufgrund veränderter politischer oder institutioneller Rahmenbedingungen die Implementation angepasst werden soll oder muss – mit Folgen auch für die Prüfung theoretischer Annahmen und die Anwendbarkeit datenanalytischer Verfahren.

Herausforderungen für Datenanalysen – quantitativ wie qualitativ – ergeben sich darüber hinaus aus dem Sachverhalt, dass Implementationsstudien in der Regel eine Mehrebenenstruktur aufweisen, sei es, dass Cluster-Randomisierungen vorliegen, sei es, dass Daten von Akteuren unterschiedlicher Handlungsebenen anfallen. Eine solche Mehrebenenstruktur stellt die Analyse der Daten vor nicht-triviale Herausforderungen, die oft nicht angemessen berücksichtigt werden.

Nimmt man die Antworten auf die beiden eingangs formulierten Leitfragen zusammen, so erarbeitet die Implementationsforschung im Bildungsbereich derzeit einen wachsenden Corpus an „Fallbeispielen“, der den research-to-practice-gap zu reduzieren vermag. Theoretisch verallgemeinerbare Einsichten in die Bedingungen der Möglichkeit, mit den Mitteln der Wissenschaft zu Verbesserungen in Politik und Praxis beizutragen, liegen allerdings noch nicht vor.

Auf diesem Weg voranzukommen, ist voraussetzungsreich. Die erste Voraussetzung besteht darin anzuerkennen, dass die Kluft zwischen Wissenschaft und Praxis für funktional differenzierte Gesellschaften (vgl. Luhmann 2002) konstitutiv ist. Eine „Einheit“ von Theorie und Praxis kann es nicht geben, und man sollte sie aus demokratietheoretischen Gründen auch nicht wünschen, da Wissenschaft, Praxis und Po-

litik ihre je eigene Verantwortung haben, die sie weder abtreten noch stellvertretend übernehmen können. Metatheoretische Reflexionen, wie sie die Allgemeine Pädagogik gepflegt hat und wie sie mit systemtheoretischen Anleihen gelegentlich auch in der empirischen Bildungsforschung anklingen (vgl. Baumert 2016), klären den Blick für das grundsätzlich Mögliche, aber sie leiten nicht dazu an, aus theoretisch begründeten Erfahrungen zu lernen, unter welchen Bedingungen die Überwindung der Kluft wahrscheinlicher wird.

Um ihr theoretisches Potenzial zu erweitern, könnte sich die erziehungs- und bildungswissenschaftliche Implementationsforschung zukünftig stärker als bislang benachbarten Forschungsfeldern öffnen. So fokussiert beispielsweise die Forschung zur Educational Governance die Handlungsstrukturen und die Handlungskoordination im Mehrebenensystem der Bildung (Langer und Brüsemeister 2019). Analysiert wird das Zusammenwirken staatlicher, zivilgesellschaftlicher und marktlicher Akteure in der Steuerung des Bildungssystems auf der Basis institutionalistischer und systemtheoretischer Annahmen (für die Schule vgl. z.B. Fend 2008; für die Erwachsenenbildung Schrader 2011). Üblich ist es u.a., hierarchische Anweisungen, Mehrheitsentscheidungen, Verhandlungen oder wechselseitige Beobachtungen als *Formen* der Handlungskoordination zu unterscheiden; als ihre *Medien* Macht oder Geld, Normen und Gesetze oder (professionell und organisationskulturell) geteilte Werte. Bereits solche einfachen Unterscheidungen könnten helfen, die Wirkmechanismen, die dem vielfach hervorgehobenen Einflussfaktor „Organisationskultur“ zugeschrieben werden, theoretisch zu präzisieren und empirisch zu prüfen.

Unterspezifiziert ist auch, was in der Implementationsforschung oft im Sinne einer Sammelkategorie als „Kontext“ bezeichnet wird, insbesondere dann, wenn es um Einflussfaktoren auf der Makroebene geht. Implementationsstudien werden zwar oft im Rahmen politischer Reformen durchgeführt, aber die bildungswissenschaftlich akzentuierte Implementationsforschung hat den Kontakt zu einer politikwissenschaftlichen Policy Implementation Research noch nicht hergestellt. Das mag u.a. daran liegen, dass sich die politikwissenschaftliche Forschung in Deutschland lange Zeit kaum für Fragen der Bildungspolitik interessiert hat (Schmid und Schuhen 2018, S. 517; vgl. auch Busemeyer 2015; Nikolai und Rothe 2016). Im angelsächsischen Sprachraum stellt sich die Situation jedoch anders dar. Hier hat sich seit langem eine politik- und sozialwissenschaftlich fundierte und empirisch vielfältige Forschung etabliert, die sich mit den Formen (polity), den Inhalten (policy) sowie den Prozessen (politics) von Bildungspolitik und ihren Effekten im Bildungssystem beschäftigt (vgl. Odden 1991; Honig 2006). Die empirischen Befunde verweisen u.a. auf die Bedeutung des Implementationsprozesses für die Wirkung von Reformen (vgl. Young und Lewis 2015, S. 4). Jüngere Beiträge werben für einen intensiveren Austausch zwischen einer (bildungswissenschaftlichen) Implementationsforschung und der Policy Implementation Research (vgl. Spiel et al. 2018). Auch McDonnell und Weatherford (2016) werben dafür, „das Politische“ in der Implementationsforschung stärker anzuerkennen. Ohne einen politischen Willen zur Erforschung der Implementation einer Reformidee wird es kaum möglich sein, hochwertige und aussagefähige Implementationsstudien umzusetzen.

Methodische und datenanalytische Herausforderungen ergeben sich aus der Tatsache, dass in Implementationsstudien nicht nur *geplante* Eingriffe in die Praxis

stattfinden, sondern dass sich immer wieder, etwa bei der Adaption von Implementationsgegenständen und -angeboten, „natürliche Experimente“ einstellen. Um die damit verbundenen Herausforderungen angemessen zu bewältigen, können die bildungsökonomische und die bildungssoziologische Forschung helfen, die statistische Verfahren für die Identifizierung kausaler Zusammenhänge für nicht-experimentelle und querschnittliche Daten entwickelt haben. Die Bildungsökonomie hat solche Verfahren zuerst bei der Evaluation arbeitsmarktpolitischer Programme erprobt (vgl. grundlegend Heckman et al. 1999; Blundell und Costa-Dias 2009) und nutzt sie inzwischen auch für die Evaluation bildungspolitischer Reformen (vgl. Thomsen und Anger 2018). Und immer dort, wo die Adaption von Implementationen an alltägliche Bedingungen die theoretisch begründete oder die „statistische“ Prüfung kausaler Annahmen ausschließt, könnte die Implementationsforschung qualitative und quantitative Daten vergleichend und typisierend analysieren, entweder, um das Besondere des Einzelfalls im Sinne Diltheys besser zu „verstehen“, oder aber, um den Fallvergleich als eine Ressource für Theoriebildung zu nutzen, wie es in der Governance-Forschung praktiziert wird (vgl. Maag Merki 2014).

Herausforderungen stellen sich aber auch für die Förderer und Partner von Wissenschaft. Zwar wird allenthalben ein verbesserter Wissenschafts-Praxis-Transfer gefordert (so im Rahmenprogramm Empirische Bildungsforschung, BMBF 2017); gleichzeitig aber wurden noch keine Qualitätskriterien entwickelt, die das Transfer- bzw. Implementationspotenzial von Studien zu beurteilen erlauben, z. B. anhand von Untersuchungsdesigns oder Implementationsstrategien. Die Umsetzung gestufter Forschungsdesigns, die Implementationsstudien einschließen, erfordern aber nicht nur eine angemessene Förderung, sondern auch den Aufbau von Expertise in größeren Forschungseinheiten innerhalb und außerhalb von Universitäten. Idealer Weise verfügen diese Einheiten auch über institutionelle Strukturen für eine verlässliche und qualitätsgesicherte, im Idealfall flächendeckende Vermittlung wissenschaftlichen Wissens an Politik und Praxis, mit dem Einzelprojekte in der Regel überfordert sind. Nötig ist zudem eine veränderte Kultur in der Zusammenarbeit von Forschung und Praxis. Dazu können Forschungs-Praxis-Partnerschaften beitragen (vgl. Hartmann und Decristan 2018), die problemorientiert arbeiten und die institutionelles und personales Vertrauen fördern, das für kontrollierte Interventionen angesichts des Risikos, zu scheitern, unverzichtbar ist. Leitend für diese Partnerschaften sollte die erziehungswissenschaftlich begründete Frage sein, unter welchen Bedingungen die Verbesserung von Bildung und Erziehung möglich werden (vgl. Tenorth 2003).

Hedges (2018) hat betont, dass die empirische Bildungsforschung Antworten finden müsse auf die Replikationskrise, die die Medizin und die Psychologie bereits erschüttert habe, auf die Frage nach der Verallgemeinerbarkeit ihrer Befunde und auf die Herausforderung, Forschungsdesigns an die Komplexität der Bildungswirklichkeit zu adaptieren. Er empfiehlt dazu u. a. eine standardisierte Protokollierung der Durchführung von Untersuchungen (S. 9), eine größere Aufmerksamkeit für die Heterogenität von Treatment-Effekten, so dass Inferenz-Populationen präziser beschrieben werden können (S. 14), und schließlich „much greater attention to the problem of construct validity of cause in complex interventions“ (S. 15). Zur Bewältigung aller drei Herausforderungen können Implementationsstudien beitragen. Wenn Wissenschaft theoretisch und methodisch offen ist auch für die Kontingenzen

der Erziehungs- und Bildungswirklichkeit und nicht nur für das, was theoretisch vorab erwartet werden kann, wenn sie sich für die fördernden und hemmenden Einflussfaktoren, für Implementationslogiken und -brüche interessiert und damit für die Bedingungen der Erziehungswirklichkeit (Dilthey 1986), kann sie lernen, was sie selbst anders und besser machen kann, um die Wahrscheinlichkeit wissenschaftlich fundierter Verbesserungen von Bildung in und durch Politik und Praxis zu erhöhen.

Danksagung Die AutorInnen danken den anonymen GutachterInnen für wertvolle Überarbeitungshinweise zum Manuskript. Alle möglicherweise verbliebenen Fehler liegen selbstverständlich in der Verantwortung der AutorInnen.

Funding Open Access funding provided by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Anhang

Literaturverzeichnis der im Abschn. 3 systematisch ausgewerteten Studien – mit laufender Nummerierungszuordnung

- (A) Ahtola, A., Haataja, A., Kärnä, A., Poskiparta, E., & Salmivalli, C. (2013). Implementation of anti-bullying lessons in primary classrooms: How important is head teacher support? *Educational Research*, 55(4), 376–392.
- (B) Bonde, A. H., Stjernqvist, N. W., Sabinsky, M. S., & Maindal, H. T. (2018). Process evaluation of implementation fidelity in a Danish health-promoting school intervention. *BMC Public Health*, 148:1407. (10 Seiten)
- (C) Borman, G. D., Grigg, J., Rozek, C. S., Hanselman, P., & Dewey, N. A. (2018). Self-affirmation effects are produced by school context, student engagement with the intervention, and time: Lessons from a district-wide implementation. *Psychological Science*, 29(11), 1773–1784.
- (D) Burchard, M., Fisler, J., & Dormer, J. (2017). Interactions between teachers' attribution for student learning and implementation of evidence-based practices. *AASA Journal of Scholarship and Practice*, 14(3), 28–44.
- (E) Caldwell, L. L., Younker, A. S., Wegner, L., Patrick, M. E., Vergnani, T., Smith, E. A., & Flisher, A. J. (2008). Understanding leisure-related program effects by using process data in the HealthWise South Africa Project. *Journal of Park and Recreation Administration*, 26(2), 146–162.

- (F) Conroy, M. A., Sutherland, K. S., Algina, J., Ladwig, C., Werch, B., Martinez, J., Jessee, G., & Gyure, M. (2019). Outcomes of the BEST in CLASS Intervention on teachers' use of effective practices, self-efficacy, and classroom quality. *School Psychology Review, 48*(1), 31–45.
- (G) Drake, P. M., Firpo-Triplett, R., Glassman, J. R., Ong, S. L., & Unti, L. (2015). A randomized-controlled trial of the effects of online training on implementation fidelity. *American Journal of Sexuality Education, 10*(4), 351–376.
- (H) Foley, L. S. (2011). Exploring K-3 teachers' implementation of Comprehension Strategy Instruction (CSI) using Expectancy-Value Theory. *Literacy Research and Instruction, 50*, 195–215.
- (I) Guo, Y., Dynia, J. M., Logan, J. A. R., Justice, L. M., Breit-Smith, A., & Kaderavek, J. N. (2016). Fidelity of implementation for an early-literacy intervention: Dimensionality and contribution to children's intervention outcomes. *Early Childhood Research Quarterly, 37*, 165–174.
- (J) Halliday, A. J., Kern, M. L., Garrett, D. K., & Turnbull, D. A. (2019). Understanding factors affecting positive education in practice: An Australian case study. *Contemporary School Psychology*. <https://doi.org/10.1007/s40688-019-00229-0> (18 Seiten)
- (K) Hanley, S., Ringwalt, C., Vincus, A. A., Ennett, S. T., Bowling, J. M., Haws, S. W., & Rohrbach, L. A. (2009). Implementing evidence-based substance use prevention curricula with fidelity: The role of teacher training. *Journal of Drug Education, 39*(1), 39–58.
- (L) Hetfleisch, P., Goeze, A., & Schrader, J. (2014). Implementation eines wissenschaftlich erprobten, didaktischen Konzepts: Der Einfluss pädagogischer Autonomie auf die Wirksamkeit in der Praxis. *Zeitschrift für Erziehungswissenschaft, 17*(2), 297–322.
- (M) Huang, K.-Y., Nakigudde, J., Rhule, D., Gumikiriza-Onoria, J. L., Abura, G., Kolawole, B., Ndyabangi, S., Kim, S., Seidman, E., Ogedegbe, G., & Brotman, L. M. (2017). Transportability of an evidence-based early childhood intervention in a low-income African country: Results of a cluster randomized controlled study. *Prevention Science, 18*(8), 964–975.
- (N) Humphrey, N., Barlow, A., & Lendrum, A. (2018). Quality matters: Implementation moderates student outcomes in the PATHS curriculum. *Prevention Science, 19*(2), 197–208.
- (O) Hustus, C. L., & Owens, J. S. (2018). Assessing readiness for change among school professionals and its relationship with adoption and reported implementation of mental health initiatives. *Child & Youth Care Forum, 47*, 829–844.
- (P) Kallestad, J. H., & Olweus, D. (2003). Predicting teachers' and schools' implementation of the Olweus Bullying Prevention Program: A multilevel study. *Prevention & Treatment, 6*(1), Article 21.
- (Q) Lau, E. Y., Saunders, R. P., Beets, M. W., Cai, B., & Pate, R. R. (2017). Factors influencing implementation of a preschool-based physical activity intervention. *Health Education Research, 32*(1), 69–80.
- (R) Little, M. A., Sussman, S., Sun, P., & Rohrbach, L. A. (2013). The effects of implementation fidelity in the Towards No Drug Abuse dissemination trial. *Health Education, 113*(4), 281–296.

- (S) Locke, J., Wolk, C. B., Harker, C., Olsen, A., Shingledecker, T., Barg, F., Mandell, D., & Beidas, R. (2017). Pebbles, rock, and boulders: The implementation of a school-based social engagement intervention for children with autism. *Autism, 21*(8), 985–994.
- (T) Locke, J., Shih, W., Kang-Yi, C. D., Caramanico, J., Shingledecker, T., Gibson, J., Frederick, L., & Mandell, D. S. (2019). The impact of implementation support on the use of a social engagement intervention for children with autism in public schools. *Autism, 23*(4), 834–845.
- (U) Locke, J., Lawson, G. M., Beidas, R. S., Aarons, G. A., Xie, M., Lyon, A. R., Stahmer, A., Seidman, M., Frederick, L., Oh, C., Spaulding, C., Dorsey, S., & Mandell, D. S. (2019). Individual and organizational factors that affect implementation of evidence-based practices for children with autism in public schools: A cross-sectional observational study. *Implementation Science, 14*:29. (9 Seiten)
- (V) McDaniel, S., Kim, S., Kwon, D., & Choi, Y.-J. (2018). Stakeholder perceptions of contextual factors related to PBIS implementation in high need schools. *Journal of Children and Poverty, 24*(2), 109–122.
- (W) Mihalic, S. F., Fagan, A. A., & Argamaso, S. (2008). Implementing the LifeSkills Training drug prevention program: Factors related to implementation fidelity. *Implementation Science, 3*:5. (16 Seiten)
- (X) Morrier, M. J., Hess, K. L., & Heflin, L. J. (2011). Teacher training for implementation of teaching strategies for students with Autism Spectrum Disorders. *Teacher Education and Special Education, 34*(2), 119–132.
- (Y) Mukoma, W., Flisher, A. J., Ahmed, N., Jansen, S., Mathews, C., Klepp, K.-I., & Schaalma, H. (2009). Process evaluation of a school-based HIV/AIDS intervention in South Africa. *Scandinavian Journal of Public Health, 37*(S2), 37–47.
- (Z) Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes: State-wide scale-up of school-wide positive behavior intervention and supports. *Journal of Behavioral Health Services & Research, 39*(4), 417–433.
- (a) Pas, E. T., Bradshaw, C. P., Becker, K. D., Domitrovich, C., Berg, J., Musci, R., & Ialongo, N. S. (2015). Identifying patterns of coaching to support the implementation of the Good Behavior Game: The role of teacher characteristics. *School Mental Health: A Multidisciplinary Research and Practice Journal, 7*(1), 61–73.
- (b) Rohrbach, L. A., Dent, C. W., Skara, S., Sun, P., & Sussman, S. (2007). Fidelity of implementation in project Towards No Drug Abuse (TND): A comparison of classroom teachers and program specialists. *Prevention Science, 8*(2), 125–132.
- (c) Rohrbach, L. A., Gunning, M., Sun, P., & Sussman, S. (2010). The project Towards No Drug Abuse (TND) dissemination trial: Implementation fidelity and immediate outcomes. *Prevention Science, 11*(1), 77–88.
- (d) Shernoff, E. S., & Kratochwill, T. R. (2007). Transporting an evidence-based classroom management program for preschoolers with disruptive behavior problems to a school: An analysis of implementation, outcomes, and contextual variables. *School Psychology Quarterly, 22*(3), 449–472.

- (e) Storey, K. E., Montemurro, G., Flynn, J., Schwartz, M., Wright, E., Osler, J., Veugelers, P. J., & Roberts, E. (2016). Essential conditions for the implementation of comprehensive school health to achieve changes in school culture and improvements in health behaviours of students. *BMC Public Health*, 16:1133. (11 Seiten)
- (f) Wang, B., Stanton, B., Deveaux, L., Lunn, S., Rolle, G., Adderley, R., Poirier, M., Koci, V., Marshall, S., & Gomez, P. (2017). Multi-year school-based implementation and student outcomes of an evidence-based risk reduction intervention. *Implementation Science*, 12:16. (13 Seiten)
- (g) Wilson, K. P., & Landa, R. J. (2019). Barriers to educator implementation of a classroom-based intervention for preschoolers with autism spectrum disorder. *Frontiers in Education*, 4, Article 27. (10 Seiten)

Tab. 2 Übersicht Charakteristika final in den systematischen Literaturüberblick inkludierter empirischer Studien ($N = 30$ Studien, repräsentiert in $N = 33$ Publikationen^{a)})

Publikation & Land Studien-durchführung	Implementationsgegenstandsbereich	Bildungsbereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinsturmentarten	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(A) Ahola et al. (2013) Finnland	KiVa anti-bullying programme	Schule	Large-scale Evaluationsstudie	$N = 93$ LehrerInnen aus 27 Grundschulen	Lehrer-Fragebogen; Lehrer-Unterrichtselbstbeobachtungshefte ** Deskriptive Statistik; Mehrbenennanalytische Regressionsmodelle (Analyseebenen: Schule, Lehrer)	Interne Konsistenz der beiden Fragebogenskalen: Cron α -Werte von 0,80 und 0,90	-	-	Drittmittelgeber für Entwicklung von KiVa programme; Finnish Ministry of Education
(B) Bonde et al. (2018) Dänemark	„We Act together for health“ – mit IVAC-educational methodology (→ Förderung Ernährungs- & Bewegungsverhalten, wellbeing & social capital)	Schule	Quasi-experimentelle Interventionsstudie mit Kontrollgruppe	$N = 289$ SchülerInnen; $N = 22$ LehrerInnen; $n = 4$ Schulleitungen; $N = 52$ Eltern aus $N = 8$ Schulen	Feldbeobachtungen; teilnehmende Unterrichts-fremdbeobachtungen; Lehrer-, Eltern- Fragebögen; Lehrer-, Schüler-, Eltern-Interviews; Schüler-Fokusgruppen ** zusammenfassende & inhaltstrukturierende Inhaltsanalyse mit anschließender cross-cutting Analyse; Datentriangulation durch Informationskombination & -abgleich aus verschiedenen Datensorten	-	MRC process evaluation framework (Moore et al. 2015); The Multilevel Implementation Quality Framework (Domitrovich et al. 2008)	-	Aspekte der Rekrutierung Studienteilnehmender; Umsetzung ethischer Standards fürs Datenmanagement; Drittmittelgeber: Technical University of Denmark; Steno Diabetes Center Copenhagen

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-tellnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(C) Borman et al. (2018) USA	Self-affirmation intervention by Cohen et al. (2006; 2009)	Schule	Randomisierte Interventionsstudie mit Kontrollgruppe (RCT)	N = 1048 Schü-lerinnen aus 11 Schulen	Nutzung distriktrweiter Schulstatistiken; Report Cards der Schüler; schriftliche Schülerarbeiten ** Deskriptive Statistik; Inhaltsanalyse; multilevel growth models	Kodierung aller schriftlichen Schülerarbeiten durch 2 trainierte, unabhängige Kodierer – Klärung von Unstimmigkeiten durch 3. Kodierer; Interrater-Reliabilität: Cohens Kappat: 0,75	–	–	Aspekte der Rekrutierung Studien-tellnehmender; Doppelblindversuch
(D) Burchard et al. (2017) USA	Evidenzbasierte Fremdsprachen-, Les-, Schreib- & Mathematik- unterrichtsbezogene Lehrstrategien	Schule	Fragebogenstudie	N = 78 Lehr-erinnen	Lehrer-Fragebogen ** Deskriptive Statistik; Korrelationsanalysen	–	–	–	Aspekte der Rekrutierung Studien-tellnehmender
(E) Caldwell et al. (2008) Südafrika	Health Wise and the Motivation Lesson (→ Prävention riskanten Freizeitverhaltens vor allem ,sexuell riskanten' Verhaltens & Drogen)	Schule	Interventionsstudie im Längsschnitt- & Kontrollgruppendesign mit mehreren Kohorten	N = 2193 Schü-lerinnen N = 41 Lehr-erinnen aus 9 Schulen	Lehrer-Unterrichtselbstbeobachtung; Schüler- & Lehrer-Fragebogen ** Deskriptive Statistik & visuelle Prüfung der Prozessdaten wegen geringer Lehrer-Teilnahme; t-Tests (Analyseebene: Schule); Verallgemeinerte lineare Modellanalysen (GLM) mit Messwiederholung (Analyseebene: Schule)	Interne Konsistenzen zu den Motivationskalen: 1. Welle Cron α : 0,58–0,78 2. Welle Cron α : 0,65–0,79 3. Welle Cron α : 0,64–0,82 4. Welle Cron α : 0,67–0,83	–	–	Finanz- & verantwortungsbezogene Aspekte zur Begründung von Wirksamkeitsstudien im Feld; Drittmittelgeber: National Institute of Drug Abuse

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(F) Comroy et al. (2019) USA	BEST in CLASS (→ effective instructional practices addressing problem behaviors)	Kinder-garten	Geklusterte multisite randomisierte Pre-Post-Interventionsstudie mit Kontrollgruppe (RCT)	N = 186 early childhood teachers N = 465 Kinder n = 26 Coaches	Nicht-teilnehmende strukturierte & teils videographie-rende Unterrichts-fremdbeobachtung; Lehrer-Fragebögen; Coaches-Fragebögen ** Deskriptive Statistik; Chi ² -Tests mit Schulebene als Cluster; two-level mit full information maximum likelihood estimation und fixed effects (Analyseebenen: Lehrer, Schule); Single-level full information maximum likelihood estimation; Benjamini-Hochberg false discovery rate (BH-FDR) procedure; two-level logistische Regressionsmodellierungen	Zweifaches Daten-Abspeichern durch zwei Forschende mit Prüfung auf Fehlerfreiheit; Interne Konsistenzen der Beobachtungs-skalen; Cron α -Werte von 0,85–0,89; Unterrichts/fremdbeobachtung (auch der Coaching-gebete für Lehrende) durch mehrfach erneut trainierte & re-zertifizierte, unabhängig voneinander je 23 % der Unterrichtsbeobachtungen doppelt beobachtende Personen – Inter-rater Übereinstimmung (point-by-point-Methode); 0,90–0,93 & ICC-Werte von 0,44–0,75; Interne Konsistenzen der Fragebogenskalen; Cron α -Werte von 0,81–0,90	–	Practice-based coaching (Snyder et al. 2015)	Aspekte der Rekrutierung Studien-teilnehmender; Drittmittel/geber: U.S. Department of Education, National Institutes of Health/National Center for Advancing Translational Sciences Clinical and Translational Science Award

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(G) Drake et al. (2015) USA	Reducing the Risk: Building Skills to Prevent Pregnancy STD & HIV (RTR)	Schule	Randomisierte Pre-Post-Interventionsstudie mit Kontrollgruppe (RCT)	N = 162 classroom-based educators & community-based educators	Educator-Implementationsprotokolle; Educator-Web-Logfiles; Educator-Pre-Post-Fragebögen; Nichtteilnehmende Unterrichts-fremdbeobachtungen mit Audioaufzeichnung; Post-Interviews *** Deskriptive Statistik; Lineare Regressionsmodellierungen	Unterrichts/renzbeobachtung zur Validierung der Educator-Implementationsprotokolleinträge, je zwei Beobachter mit anschl. konsensuellem Diskurs über die Beobachtungsstratgs zur Konsistenzverbesserung	-	-	Separate Adressdatenspeicherung; Wartelisten-kontrollgruppe; Aspekte der Rekrutierung Studien-einnehmender; Drittmittelgeber: National Institute on Mental Health
(H) Foley (2011) USA	Evidence-based comprehension strategy instruction (CSI → Leseverstehen)	Schule	Explanatory design	N = 197 K-Lehrerinnen aus 40 Schulklassen	Lehrer-Fragebogen ** Deskriptive Statistik; Multiple Regressionsanalysen	Interne Konsistenz des Gesamtinstrumentariums: Cron $\alpha = 0,90$; interne Konsistenzen der verschiedenen Skalen von Cron $\alpha = 0,56-0,80$; Konfirmatorische Faktorenanalysen zur Prüfung der Konstruktvalidität	-	Expectancy-Value Theory (Bandura & Locke 2003; Vroom 1964) Heuristik einfacher Lehrerrenamtionen zur Prüfung (Abrami et al. 2004)	Aspekte der Rekrutierung Studienteilnehmender

Tab. 2 (Fortsetzung)

Publikation & Land Studien- durchführung	Implementations- gegenstand	Bildungs- bereich	Studien- design	Studien- teilnehmende	Datenerhebung & Datenanalyse	Qualitätsicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(I) Guo et al. (2016) USA	Early literacy intervention	Kinder- garten	Experimentelle Pre-Post- Interventi- onsstudie	N= 83 Lehre- rInnen; N= 291 Kindergar- tenkinder	Lehrer-Fragebogen; Lehrer- Implementationsprotokolle; Nicht-teilnehmende Fremd- beobachtung videographierter Unterrichtssituationen; Kin- dergartenkinder-Leistungstest ** Deskriptive Statistik; Multi- level mediation structural equation model (zwei Ana- lyseebenen: Klasse/Lehrer, Schüler)	Doppelkodierung durch trainierte Kodierer von 10 % der Unter- richtsvideos, Interrater- Reliabilität: ICC= 0,95; Interrater-Reliabilität Leistungstest (point-by- point-Methode)=0,94; Interne Konsistenz Leistungstest: Cron α-Werte von 0,75–0,93; Interrater-Reliabilität: 0,99; Außenkriterium zur Validitätseinschät- zung: Korrelationen zu ähnlichen Assessments zwischen 0,61 und 0,71	–	Theoretisches Konzept von <i>implementation-fidelity</i> als ein- oder drei- dimensionales Konstrukt (mit adherence & do- sage, participant responsiveness und program differentiation)	–
(J) Halliday et al. (2019) Australien	Positive Psy- chology In- terventions: MoodGYM, Three Good Things, Gratitude Letter, Meaning Through Photo- graphy, Counting Kindness & Best Possible Selves (→ zur Förderung von Wohlbefin- den, Resilienz & Reduktion depressiver Symptome)	Schule	Quasi- experi- mentelle Interventi- onsstudie mit Warte- liste	N= 143 SchülerInnen N= 8 pastoral care teachers	Halbstrukturierte Lehrer- & Schüler-Fokusgruppen; halbstrukturiertes Schu- ler-Interview; Schüler-, Eltern-, Lehrer-Fragebogen; Dokumentenanalyse ** Deskriptive Statistik; „fol- lowing a thread“-Methode zur Datentriangulation; inhaltsana- lytische Auswertung der Daten auf induktiver Basis, inkl. Mustererkennungsmethode bei qual. Daten nach Braun & Clarke (2006)	Gegenprüfung des no- tizenbasierten Berichts zur Schüler-Fokusgruppe durch zweiten Lehrer; Besprechung inhaltsana- lytischer Auswertungs- ergebnisse innerhalb des Forscherteams	Eigen-Kombinati- on aus Ecological framework for understanding durch zweiten Lehrer; Mentation (Durlak und DuPe 2008), The Multilevel Implementa- tion Quality Framework (Do- mitrovich et al. 2008) & Concep- tual Framework for Implementa- tion (Fixsen et al. 2005)	–	Aspekte der Rekrutierung Studienteil- nehmender; Wartelistenkon- trollgruppe; Keine wortwörtliche Transkription der Schüler- Fokusgruppe

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementationsgegenstand	Bildungsbereich	Studien-design	Studien-tellnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(K) Hanley et al. (2009) USA	Evidence-based school-based Substance Use Prevention Programs/Curricula	Schule	Teilstudie einer Evaluationsstudie im Längsschnittdesign	N= 276 für Drogenprävention hauptverantwortliche, trainierte LehrerInnen	Fragebogen ** Deskriptive Statistik; Kumulative logistische Regressionsanalysen mit Berücksichtigung der geschichteten Stichprobe	-	-	-	Drittmittelgeber: National Institute of Drug Abuse
(L) Heftleisch, Goetz, & Schrader (2014) Deutschland	„Lehrende lernen mit Videofällen“ – computer-gestütztes Videofallarbeitskonzept	Referendarat	Quasi-experimentelle Pre-Post-Interventionsstudie im Feld	N= 261 ReferendarInnen; N= 19 LehrerInnen aus 6 Seminarstandorten	Freie schriftliche Fallanalysen der Referendare; Fragebögen & Wissenstest Referendare; Nicht-teilnehmende, strukturierte Seminararbeitsprotokollierung für Manipulation-Check ** Qualitative & quantifizierende Inhaltsanalyse auf Basis eines deduktiv gebildeten Kategoriensystems; Deskriptive Statistik; ANOVAs; ANCOVAs mit geplanten paarweisen Kontrasten	Performanztest: Domänenspezifische Kompetenzanwendungssituation via freier schriftlicher Fallanalysen; Vergleichbarkeit Pre- & Posttest-Videofälle auf Basis unabhängiger Fachdidaktik-Experteneinschätzungen; Kodierung der Fallanalysen durch 2 trainierte, unabhängige Kodierer – Interrater-Reliabilität: Cohens Kappa = $K > 0,80$; Seminararbeitsprotokollierung	-	Professionstheoretische Ansätze: Pädagogische Autonomie (Tenorth 2003; van Buer 1990)	Drittmittelgeber: Deutsche Forschungsgemeinschaft

Tab. 2 (Fortsetzung)

Publikation & Land Studien- durchführung	Implementations- gegenstand	Bildungs- bereich	Studien- design	Studien- teilnehmende	Datenerhebung & Datenanalyse	Qualitätsicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(M) Huang et al. (2017) Uganda	Lehrer-Förderungskonzept des „Parent Corps“; Schülerprogramm des „ParentCorps“	Lehrerfortbildung; Schule	Gechlusterte randomisierte Pre-Post-Interventionsstudie mit Kontrollgruppe (RCT)	N= 79 Lehrerinnen; N= 10 SchullehrerInnen aus 10 Schulenten; N= 154 Eltern	Lehrer- & Schulleiter-Fragebögen; Eltern-Fragebogen zur Erfassung von Schüler-Lernerträgen; Lehrer-Implementationsprotokolle; Lehrer-Pre-Post-Wissenstests; Nicht-teilnehmende Unterrichts-fremdbeobachtung ** Deskriptive Statistik; Linear Mixed Effect Models (Analyseebene: Schule)	Unterrichts/fremdbeobachtung durch 2 voneinander unabhängige Beobachtende; Interrater-Übereinstimmung zwischen ICC=0,77–0,79; Pre-Post-Wissenstests; Interne Konsistenzen verschiedener Fragebogenskalen: Cron α -Werte von 0,59–0,85	Consolidated Framework for Implementation Research (Damschroder et al. 2009); Teacher Training Implementation Model (TTIM) von Reinke und Kollegen (2013)	Theory of Change	Wartelistenkontrollgruppe; Aspekte der Rekrutierung Studienleitnehmender; Vorliegen eines Ethik-Votums; Drittmittelgeber: National Institutes of Health
(N) Humphrey, Barlow, & Lendrum (2018) United Kingdom	PATHS (= Promoting Alternative Thinking Strategies) Curriculum	Schule	Major cluster RCT im longitudinal, natural variation design (→ Teilstudie: Daten aus dem ersten Jahr dieser Studie)	N= 69 Lehrerinnen; N= 1721 SchülerInnen aus 23 Grundschulen	Nicht-teilnehmende Unterrichts-fremdbeobachtung; Schüler-Selbstberichte; Lehrer-Fragebögen ** Deskriptive Statistik; Hierarchische lineare modeling (zwei Analyseebenen: Klasse/Lehrer, Schüler)	Unterrichts/fremdbeobachtung; Videobasiertes Training der 3 Beobachtenden; Interrater-Reliabilität: ICC=0,91; Interne Konsistenzen verschiedener Fragebogenskalen: Cron α -Werte von 0,86–0,92 und Test-Retest r von 0,80–0,92; Konfirmatorische Faktorenanalysen zur Prüfung der Konstruktvalidität; hohe konvergente, diskriminante und prä-diktive empirisch belegte Validität von Skalen	Integrated theoretical model of program implementation von Berkel und Kollegen (2011)	Theoretisches Implementationskonzept, das <i>implementation fidelity</i> (= <i>what is implemented?</i>) und <i>implementation quality</i> (= <i>how well is it implemented?</i>) als zwei eigenständige Teildimensionen von <i>Implementation</i> modelliert (Berkel et al. 2011)	Aspekte der Rekrutierung Studienleitnehmender; Vorliegen eines Ethik-Votums; Drittmittelgeber: National Institute for Health Research

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/Framework zum Vororten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(O) Hustus et al. (2018) USA	Daily Report Card (→ Intervention bei lernabriglichem Verhalten) & Behavior/color Wheel (→ Instrument zur Dokumentation regelverletzenden SchülerInnen-Verhaltens)	Schule	Fragebogen-studie	n = 194 LehrerInnen; n = 47 school mental health staff n = 13 building administrators (= (assistent) principals & deans of students) von 11 Schulen 2er Distrikte	Fragebögen; Records des assoziierten Projekts ** Deskriptive Statistik; Korrelationsanalysen (zwei Analyseebenen: Individual-ebene und auf Schulebene aggregierte Daten)	Interne Konsistenz der Fragebogenskaten: Cron α -Werte von 0,84–0,91	Scale-up model (Adelman & Taylor 1997)	Readiness for change (Armenakis et al. 1993)	Aspekte der Rekrutierung Studienteilnehmender
(P) Kallestad & Olweus (2003) Norwegen	Olweus Bullying Prevention Program	Schule	Quasi-experimentelles Studiendesign (→ Teilstudie einer large-scale Längsschnittstudie)	N = 89 LehrerInnen aus 37 Schulen	Lehrer-Fragebögen (teils Pre-Post-Erhebung) ** Deskriptive Statistik; Mehrbenennanalytische Regressionsanalysen (Analyseebenen: Schule, Klasse/Lehrer)	Interne Konsistenzen verschiedener Fragebogenskaten: Cron α -Werte von 0,68–0,80	Implementation model von Fullan (1982); Implementation model von Green & Kreuter (1999)	Selbstwirksamkeitserwartung (Bandura 1986) bezogen auf Implementationshandeln; Professionstheoretische Ansätze: Pädagogische Autonomie (Hargreaves 1992)	Drittmittelgeber: Norwegian Research Council; Rådet for psykisk helse; Ministry of Children & Family Affairs; Norwegian Public Health Association

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementationsgegenstand	Bildungsbereich	Studien-design	Studien-tellnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsverfahren	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(Q) Lau et al. (2017) USA	Move Outside – Physical Activity Intervention	Kinder-garten	Prospective observational study design mit Querschnittsauswertung (→ Teil-daten aus einem drei-jährigen RCT)	N= 8 Kindergarten-kindern; N= 36 Kindergarten-LehrerInnen; N= 504–720 Kinder	Lehrer-Fragebögen; Nicht-teilnehmende Unterrichts-fremdbeobachtung ** Deskriptive Statistik; Bayesian Pädanalysen (Analyseseinheit: Kindergartenklassen)	Unterrichts/fremdbeobachtung; Interrater-Reliabilitätsbestimmung zwischen 2 Beobachtern auf Basis von 10% der Beobachtungen: >0,80	Ecological framework for understanding effective implementation (Durlak und DuPre 2008)	–	Aspekte der Rekrutierung Studententeilnehmer: Drittmittelgeber: Eunice Kennedy Shriver National Institute of Child Health & Human Development; National Institute of General Medical Sciences, both at the National Institutes of Health
(R) Little et al. (2013) USA	Towards No Drug Abuse (TND)-Program	Schule	Randomisierte Pre-Post (follow-up)-Interventionsstudie mit Kontrollgruppe (RCT)	N = 50 LehrerInnen; N = 41 Schullad-ministratorenInnen; aus 43 Schulen	Schuladministratoren-Fragebögen; Lehrer-Pre-Post-Fragebögen; Schüler-Pre-Post-follow-up-Fragebögen; Nicht-teilnehmende Unterrichts/fremdbeobachtung; ** Deskriptive Statistik; Verallgemeinerte gemischte lineare Modellanalysen (Analyseebenen: Schulschritte, Schulen, Lehrer); Mediationsanalysen	Unterrichts/fremdbeobachtung seitens trainierter Beobachter, die um die verschiedenen Interventionsbedingungen wussten; interne Konsistenz der Beobachtungsgesamtskala für fidelity: Cron $\alpha = 0,96$; Interne Konsistenzen der verschiedenen Skalen zwischen Cron $\alpha = 0,56$ und 0,88	Ecological framework for understanding effective implementation (Durlak und DuPre 2008); The Multilevel Implementation Quality Framework von Domitrovich und Kollegen (2008)	–	Drittmittelgeber: National Institute on Drug Abuse

Tab. 2 (Fortsetzung)

Publikation & Land Studien- durchführung	Implementations- gegenstand	Bildungs- bereich	Studien- design	Studien- teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(S) Locke et al. (2017) USA	„Remaking Recess“; A school-based social engagement intervention for children with autism	Schule	Rekonstruktiv angelegte Interviewstudie (→ Teilstudie eines RCT)	n = 15 Schül- (kon)rektorInnen; n = 10 Lehr-erInnen; n = 14 Klas- senlehrer- assistentIn- nen, -hilfen, BeraterIn	Halbstrukturierte Einzelin- terviews ** Qualitative Inhaltsanalyse auf Basis eines deduktiv und induktiv gebildeten Kategoriensystems	Kodierung aller Tran- skripte durch 2 un- abhängige Kodierer; Interrater-Relia- bilitätsanalyse auf Basis von 20 % zu- fällig ausgewählten Transkripten; $k \geq 0,85$	The Multilevel Implementation Quality Framework von Domitrovich und Kollegen (2008)	-	Aspekte der Rekrutierung Studienteil- nehmender; Verantwor- tungsbezogene Aspekte zur Begründung von Wirksam- keitsstudien im Feld
(D) Locke et al. (2019) USA	„Remaking Recess“; A school-based social engagement intervention for children with autism	Schule	Randomi- sierte Pre-Post- follow-up- gruppen- interven- tionsstudie (RCT)	N = 31 Schü- lerInnen mit Autismus N = 11 Leh- rerInnen N = 17 Class- room assi- stants, 1:1- Assistenten & noontime aides aus 12 Schulen	Schüler-Fragebögen; Nicht- teilnehmende Unterrichts- fremdbeobachtungen; Lehrer-Unterrichtselbstbe- obachtungsbögen; Soziale Netzwerkanalyse ** Deskriptive Statistik; se- parate multilevel models (Analyseebenen: multiple Zeitpunkte, genestet auf Schülerebene, genestet in Schulen); Gemischte lineare Regressionsmodellierungen (Analyseebene: Schule); ordinal logistic mixed models oder logistische mixed models	Blind-Unter- richts/zenzbeob- achtungen durch trainierte, unabhän- gig Beobachtende, Rating-Übereinstim- mung Beobachtung von (a) fidelity: 0,85 & (b) Schülerverhalten: 0,82 mit Kappa-Wert von 0,82; Power-Analy- se: Teststärke von 0,80 zur Aufdeckung kleiner Unterschiedseffekte in der AV „soziale Netzwerkinklusion von Schülern“; Prü- fung der statistischen Voraussetzungen für Regressionsanalysen	Conceptual model of im- plementation research (Proce- ssor et al. 2009)	-	Aspekte der Rekrutierung Studienteil- nehmender; Drittmittelgeber: Autism Science Foundation & FARFund Early Career Award

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(U) Locke et al. (2019) USA	Discrete trial training, Pivotal response training & Visual Interventions bei SchülerInnen mit Autismus)	Schule	Cross-sectional observational study	n = 67 LehrerInnen n = 85 Classroom assistants & aides aus 67 Klassen von 52 Schulen	Lehrer-(Selbstreflexions-)Beurteilung & Ratings; Lehrer-Fragebögen ** Deskriptive Statistik; Korrelationsanalysen; lineare Regressionsmodellierungen	Interne Konsistenzen der Fragebogenskalen: Cron α-Werte von 0,66–0,98	-	-	Aspekte der Rekrutierung Studien-teilnehmender; Drittmittel/geber: US National Institute of Mental Health
(V) McDaniel et al. (2018) USA	Positive Behavior Intervention and Supports (PBIS)	Schule	Fragebogen-studie	n = 121 LehrerInnen/ Schulumitglieder (z. B. Beratende, Sozialarbeit-erInnen); n = 40 Schulad-ministrato-rInnen	Lehrer- & Administratoren-Fragebögen ** Deskriptive Statistik; ANOVAs (Prüfung von Unterschiedseffekten in Einschätzung von Einflussfaktoren zw. Akteursgruppen & Kontextbe-dingungen – z. B. städtische/ ländliche/gemischte Schulen)	Prüfung der Items bzgl. Wording, augenscheinlich inhaltlicher Trennschärfe & Notwendigkeit der einzelnen Items nach internen Prüfdurchgängen zudem durch ein Team externer Wissenschaftler; Interne Konsistenz des Fragebo-gens: Cron α-Wert von 0,82	-	-	Aspekte der Rekrutierung Stu-dienteilnehmender

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementationsgegenstand	Bildungsbereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(W) Mihalic, Fagan, & Argamaso (2008) USA	LifeSkills Training (LST) drug prevention program	Schule	Prozessevaluationsstudie	N= 432 Schulen (Daten von LehrerInnen, LST-ProgrammkordinatorInnen; instructors; Blueprint Mitarbeitern) aus 105 sites	KoordinatorIn-, Trainer- & Lehrer-Fragebögen (teils zu mehreren Messzeitpunkten) & Lehrer-Berichte; Nicht-teilnehmende Unterrichts/fremdbeobachtungen; Interviews mit „key participants“ **	Interne Konsistenzen verschiedener Skalen; Cron α -Werte von 0.49–0.87; Unterrichts/renzbeobachtung durch Schullehrer zur Vermeidung von Beeinflussungen; stete Supervision der Beobachtenden & zusätzliche Beobachtungen jeweils durch Blueprint-Mitarbeitende, um Beobachtungen lokaler Beobachter zu validieren; Beobachtungsübereinstimmungen: 89,7 %	–	–	Aspekte der Rekrutierung Studienteilnehmer; Drittmitteler: Office of Juvenile Justice and Delinquency Prevention
(X) Morrier, Hess, & Hefflin (2011) USA	Evidence-based teaching strategies advocated for children with Autism Spectrum Disorders	Schule	Evaluationsstudie	N= 185 LehrerInnen von auch autistischen SchülerInnen	Lehrer-Fragebogen ** Deskriptive Statistik; <i>t</i> -Tests für unabhängige Stichproben; Multiple Regressionsanalysen; ANOVAs	Fragebogentwürfe wurden 4 Experten für Autismus & sonderpädagogisches Handeln vorgelegt, um Konstruktvalidität zu prüfen; Pilotierung des Fragebogens mit anschließender Überarbeitung	–	–	Aspekte der Rekrutierung der Studienteilnehmenden

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(Y) Mukoma et al. (2009) Südafrika	HIV/AIDS-Intervention	Schule	Prozessevaluationsstudie mit zwei Messzeitpunkten (→ Teilstudie eines RCT)	N= 25 Lehrerinnen; N= 122 Schülerinnen aus 12 Schulen	Lehrer-Implementationsprotokolle; Nicht-teilnehmende Unterrichtsformbeobachtungen; Halbstrukturierte Lehrer- (Einzel- oder Gruppen-)Interviews; Schüler-Fokusgruppen **	Unterrichts/renzi/überbeobachtung; Alle Transkripte & Beobachtungsprotokolle wurden Zeile für Zeile seitens der Erstautorin kodiert; zur Erhöhung der Reliabilität: mind. zweifache Kodierung und Wieder-Kodierung aller Transkripte innerhalb von drei Wochen (→ Keine konkreten Kennwertangaben)	–	–	Aspekte der Rekrutierung der Studien-teilnehmenden; Vorliegen eines Ethik-Votums; Drittmittelgeber: EU-Projekt-förderung EC-INCO
(Z) Pas & Bradshaw (2012) USA	SW-PBIS (School-wide Positive Behavioral Interventions and Supports)	Schule	Pre-Post-Evaluationsstudie	N= 421 Schulen (Daten von PBIS intervention coaches; externe EvaluatorenInnen; PBIS-Team-Mitglieder; SchülerInnen) aus 24 Schuldis-trikten	Fragebogen an externe Evaluatoren & PBIS-Interventionsberater; Pre-Post-Schüler-Leistungstests ** Deskriptive Statistik; Strukturegleichungsmodellierung; Zur Berücksichtigung der Mehrebenenstruktur (Schuldis-trikte, Schulen); Huber-White corrections zur Anpassung der Standardfehler	Interne Konsistenzen der verschiedenen Skalen: Cron α -Werte zwischen 0,65 bis 0,91; Verwiesen wird auf andere Publikationen, in denen Test-Retest-Reliabilitäts- und Validitäts-schätzungen zu den Instrumenten angegeben sind	–	–	Drittmittelgeber: National Institute of Mental Health, Centers for Disease Control and Prevention; Institute of Education Sciences

Tab. 2 (Fortsetzung)

Publikation & Land Studierendurchführung	Implementationsgegenstand	Bildungsbereich	Studiendesign	Studienteilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(a) Pas et al. (2015) USA	PAX Good Behavior Game	Schule	Randomisierte Pre-Post-Interventionsstudie mit Kontrollgruppe (RCT) pro Kohorte	N = 210 LehrerInnen 16 Schulen, verteilt über 3 Kohorten je einjähriger Interventionen/Implementationen	Lehrer-(Pre-Post-)Fragebögen; Lehrer-Implementationsprotokolle; Nichtteilnehmende Unterrichts-fremdbeobachtung ** Deskriptive Statistik; Growth Mixture Modelle; Multinominale logistische Regressionsanalysen	Unterrichts/fremdbeobachtung mit Doppelkodierung zu 15 % der Lehrenden der ersten Erhebungsphase & darauf basierenden Interrater-Reliabilitätsanalysen: >0,80; Interne Konsistenzen der verschiedenen (Beobachtungs-)Skalen zwischen Cron $\alpha = 0,64-0,94$	The Multilevel Implementation Quality Framework von Domitrovich und Kollegen (2008); Interactive Systems Framework von Wandersman und Kollegen (2008)	-	Aspekte der Rekrutierung Studienteilnehmender; Vorliegen eines Ethik-Votums; Drittmittelgeber: Institute of Sciences
(b) Rohrbach et al. (2007) USA	Towards No Drug Abuse (TND)-Program	Schule	Randomisierte Pre-Post-Interventionsstudie (RCT)	n = 13 LehrerInnen; N = 2331 SchilernerInnen aus 18 Schulen, aus 9 Schulsdiaktriken; n = 5 externe Programm-spezialistInnen	Schüler-Pre-Post-Fragebögen; Nichtteilnehmende Unterrichts/fremdbeobachtungen ** Deskriptive Statistik; Mehrbenenanalytische gemischte lineare Regressionsmodellierungen (Zwei Analyseebenen: Schulen, Klasse/Lehrer)	Unterrichts/fremdbeobachtung durch vorab trainierte Beobachter, dabei 77 % aller Unterrichtseinheiten durch je 2 Beobachter beobachtet, Interrater-Reliabilität: von 0,60-0,89; Interne Konsistenz der Beobachtungsskalen: Cron α -Werte von 0,66-0,76; Interne Konsistenzen der Fragebogenskalen: Cron α -Werte von 0,73-0,94	-	-	Aspekte der Rekrutierung der Studienteilnehmenden; Drittmittelgeber: National Institute on Drug Abuse

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementationsgegenstand	Bildungsbereich	Studien-design	Studien-tellnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(c) Rohrbach et al. (2010) USA	Towards No Drug Abuse (TND)-Program	Schule	Randomisierte Pre-Post-Interventionsstudie mit Kontrollgruppe (RCT)	N = 54 Lehrern; N = 2983 Schülerinnen aus 65 Schulen, aus 14 Schulschuldistrikt	Schüler-Pre-Post-Fragebögen; Nicht-teilnehmende Unterrichts-fremdbeobachtungen** Deskriptive Statistik; F-Tests & Chi ² -Tests; Verallgemeinerte gemischt-lineare Regressionsmodelle (Analysesebene: Schule bzw. Klasse/Lehrer)	Unterrichtsfremdbeobachtung seitens trainierter Beobachter, die um die verschiedenen Interventionsbedingungen wussten; interne Konsistenz der Beobachtungsgesamtskala für fidelity: Cron $\alpha = 0,96$	–	–	Aspekte der Rekrutierung der Studienteilnehmenden; Drittmittelgeber; National Institute on Drug Abuse
(d) Shernoff & Kratochwill (2007) USA	Evidence-based Classroom management strategies & „The Incredible Years Classroom Management Program for teacher“	Kinder-garten/pre-school	Randomisiertes multiple probe design mit mehreren Messzeitpunkten	N = 8 Kindergarten-Lehrerinnen N = 13 Kinder-garten/pre-schools	Halbstrukturierte Einzelinterviews; Lehrer-Fragebögen (vor, während & nach der Implementierung); Lehrer-Beobachtungsbögen zur Erfassung von Schüler-Lernfortschritten zu mehreren Messzeitpunkten; „Time logs“ von Lehrern & Beratern während des Lehrentrainings & der Implementierung; Fremdbeobachtungen von Schülern**	Interne Konsistenzen der Fragebogenskalen: Cron α -Werte von 0,52–0,97; Schüler-Beobachtungen wurden von trainierten Beobachtern durchgeführt, die keine Kenntnis hatten von Bedingungen, Zielen & Hypothesen der Studie; Interobserver-Übereinstimmung auf Basis von 15 % der Beobachtungseinheiten und anhand der point-by-point-Methode: zwischen 86 bis 100 %	–	Behavioral Consultation Framework von Kratochwill und Kollegen (1999)	Aspekte der Rekrutierung Studienteilnehmender

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-teilnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungsinstrumentarien	Modell/Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(e) Storey et al. (2016) Canada	Comprehensive School Health (CSH)	Schule	Sekundärana-lyse der Daten aus fünf Interview-studien mit jeweils community-based participatory research designs	n = 45 LehrerInnen; n = 46 SchullehrInnen; n = 34 school health facilitator	Halbstrukturierte Einzelinterviews mit Schulleitungen & school health facilitators; Fokusgruppen mit Lehrern ** Induktive qualitative Inhaltsanalyse nach Miles und Huberman (1994)	Durchführung der inhaltsanalytischen Auswertung und interpretativen Analyse von 3 Autoren im regelmäßigen Diskurs; Peer-Debriefing mit 5 Autoren zur Sicherung (a) ihrer Responsivität & ihres Reflektiert-Seins & (b) der Genauigkeit durch Prüfen evtl. Verzerrungen; Member checking mit 3 zentralen Projekt-Stakeholdern zum Verfeinern der Ergebnisse & Erhöhung der Vertrauens der Daten	-	-	Aspekte der Rekrutierung Studienleitender; Vorliegen eines Ethik-Votums bei allen Primärstudien; Drittmittelgeber: MSI Foundation of Alberta und Collaborative Research; Innovation Opportunities Team Grant from Alberta Innovates – Health Solutions
(f) Wang et al. (2017) Bahamas	Focus on Youth in The Caribbean (= evidence-based HIV prevention program/sexual risk-reduction intervention)	Schule	Evaluationsstudie im Längsschnittdesign	N = 341 LehrerInnen; N = 4411 SchülerInnen aus 80 Schulen	Lehrer-Pre-Post-Fragebögen; Lehrer-Implementationsprotokoll; Schüler-Fragebögen & Wissenstest ** Deskriptive Statistik; Clusteranalysen; Korrelationsanalysen mit „Klasse“ als Analyseebene (Pearson & Spearman); ANOVA mit Tukey HSD-Tests; Mixed-effects Modellanalysen (Analyseebene/n: Klasse und/oder Schule); Strukturgleichungsmodellierungen (Analyseebenen: Klasse/Lehrer und Schule)	Interne Konsistenzen der Skalen zu den Schüler-Lerntragsvariablen zwischen Cron $\alpha = 0,81-0,85$	The Exploration, Preparation, Implementation and Sustainment (EPIS) framework (Aarons et al. 2011)	-	Aspekte der Rekrutierung Studienleitender; Vorliegen eines Ethik-Votums; Drittmittelgeber: National Institute of Child Health and Human Development

Tab. 2 (Fortsetzung)

Publikation & Land Studien-durchführung	Implementations-gegenstand	Bildungs-bereich	Studien-design	Studien-tellnehmende	Datenerhebung & Datenanalyse	Qualitätssicherung Erhebungs- bzw. Auswertungs-instrumentarien	Modell/ Framework zum Verorten der Einflussfaktoren	Theorien/ Konzepte zum Modellieren der Einflussfaktoren	Ethische Aspekte explizit benannt/ ggf. Geldgeber
(g) Wilson & Landa (2019) USA	Early Achievements intervention (→ Förderung sozialer und kommunikativer Fähigkeiten autistischer Vorschulkinder) & allgemein evidenzbasierte Interventionen	Kinder-garten- pre- school	Qualitative Studie als Teilstu-die einer größer angeleg-ten sich iterativ vollze-henden Entwick-lungs- & Evaluati-onsstudie	n = 4 Educa-tors und n = 4 in-structional Assistants aus ins-gesamt 4 Schulen; n = 3 schul-übergreifen-de (special education) & Administra-tors	Halbstrukturierte mündliche & schriftliche Feedbacks ** Qualitative, zweischrittige Inhaltsanalyse auf Basis induktiven, an der Grounded Theory orientierten Vorgehens	Vollständige Doppel-kodierung durch 2 trainierte, unabhängige Kodierer; Kriterien für Kodierung & finale Berücksichtigung von Aussagen: Jeweilige Barriere wurde (a) mehr-mals und (b) von beiden Akteursgruppen genannt sowie (c) von beiden Kodieren kodiert	–	–	Aspekte der Rekrutierung Studienteilneh-mender; Drittmittelgeber: Institute of Educational Sciences

^aPublikationen desselben Studien-/Projektkontextes sind jeweils einheitlich in kursiv (R, b, c) oder fett (S, T) dargestellt

Literatur

- van Ackeren, I., Zlatkin-Troitschanskaia, O., Binnewies, C., Clausen, M., Dormann, C., Preisendörfer, P., Rosenbusch, C., & Schmidt, U. (2011). Evidenzbasierte Schulentwicklung: Ein Forschungsüberblick aus interdisziplinärer Perspektive. *Die Deutsche Schule*, *103*(2), 170–184.
- Albers, B., Mildon, R., Lyon, A. R., & Shlonsky, A. (2017). Implementation frameworks in child, youth and family services—results from a scoping review. *Children and Youth Services Review*, *81*, 101–116.
- Allen, J. D., Linnan, L. A., & Emmons, K. M. (2012). Fidelity and its relationship to implementation effectiveness, adaptation, and dissemination. In R. C. Brownson, G. A. Colditz & E. K. Proctor (Hrsg.), *Dissemination and implementation research in health: Translating science to practice* (S. 281–304). New York: Routledge.
- Baeten, M., & Simons, M. (2014). Student teachers' team teaching: models, effects, and conditions for implementation. *Teaching & Teacher Education*, *41*, 92–110.
- Baumert, J. (2016). Leistungen, Leistungsfähigkeit und Leistungsgrenzen der empirischen Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, *19*(1), 215–253.
- Beelmann, A., & Karing, C. (2014). Implementationsfaktoren und -prozesse in der Präventionsforschung: Strategien, Probleme, Ergebnisse, Perspektiven. *Psychologische Rundschau*, *65*(3), 129–139.
- Beima-Sofie, K. M., Trinidad, S. B., Ngur, K., Heffron, R., Baeten, J. M., Sohn-Stewart, G. C., & Kelley, M. (2019). Lessons from PrEP: A qualitative study investigating how clinical and policy experts weigh ethics and evidence when evaluating preventive medications for use in pregnant and breastfeeding women. *AIDS and Behavior*, *23*(7), 1858–1870.
- Benner, D. (1980). Das Theorie-Praxis-Problem in der Erziehungswissenschaft und die Frage nach Prinzipien pädagogischen Denkens und Handelns. *Zeitschrift für Pädagogik*, *26*(4), 485–497.
- Birken, S. A., Powell, B. J., Shea, C. M., Haines, E. R., Kirk, M. A., Leeman, J., Rohweder, C., Damschroder, L., & Presseau, J. (2017). Criteria for selecting implementation science theories and frameworks: Results from an international survey. *Implementation Science*, *12*, 124.
- Blundell, R., & Costa-Dias, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, *44*(3), 565–640.
- BMBF (2017). Bundesministerium für Bildung und Forschung (2017). *Rahmenprogramm Empirische Bildungsforschung des Bundesministeriums für Bildung und Forschung*. Bonn: BMBF. [http://www.empirische-bildungsforschung-bmbf.de/media/content/Rahmenprogramm%20empirische%20Bildungsforschung%20\(BITV\).pdf](http://www.empirische-bildungsforschung-bmbf.de/media/content/Rahmenprogramm%20empirische%20Bildungsforschung%20(BITV).pdf). Zugegriffen: 30. Aug. 2017.
- Böse, S., Neumann, M., Becker, M., Maaz, K., & Baumert, J. (2018). Kooperationsbereit oder Innovationsgegner? Schulleiterprofile im Kontext der Implementation von Schulreformen. *Zeitschrift für Erziehungswissenschaft*, *21*(6), 1157–1186.
- Brezinka, W. (1975). *Von der Pädagogik zur Erziehungswissenschaft. Eine Einführung in die Metatheorie der Erziehung*. Weinheim: Beltz.
- Busemeyer, M. R. (2015). Bildungspolitik. In G. Wenzelburger & R. Zohlhöfer (Hrsg.), *Handbuch Policy-Forschung* (S. 615–640). Wiesbaden: Springer.
- Century, J., & Cassata, A. (2016). Implementation research: finding common ground on what, how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215.
- Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. *American Journal of Preventative Medicine*, *32*(5), 112–118.
- Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016. *Educational Research*, *60*(3), 276–291.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed-methods approaches* (5. Aufl.). Los Angeles: SAGE.
- Dahmer, I., & Klafki, W. (1968). *Geisteswissenschaftliche Pädagogik am Ausgang ihrer Epoche – Erich Weniger*. Weinheim: Beltz.
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, *4*, 50.
- Damschroder, L. J., Peikes, D., & Petersen, D. (2013). Using implementation research to guide adaptation, implementation, and dissemination of patient-centered medical home models. Rockville, MD: Agency for Healthcare Research and Quality. <https://pcmh.ahrq.gov/sites/default/files/attachments/using-implementation-research-brief.pdf>. Zugegriffen: 23. Nov. 2019.

- Dilthey, W. (1986). *Pädagogik. Geschichte und Grundlinien des Systems* (4. Aufl., Gesammelte Schriften, Bd. 9). Göttingen: Vandenhoeck & Ruprecht.
- Douglas, N.F., Campbell, W.N., & Hinckley, J.J. (2015). Implementation science. Buzzword or game changer? *Journal of Speech, Language, and Hearing Research*, 58(6), 1827–1836.
- Durlak, J.A., & DuPre, E.P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350.
- Eickelmann, B. (2011). Supportive and hindering factors to a sustainable implementation of ICT in schools. *Journal for Educational Research Online*, 3(1), 75–103.
- Eldredge, L.K.B., Markham, C.M., Ruitter, R.A.C., Fernández, M.E., Kok, G., & Parcel, G.S. (2016). *Planning health promotion programs: an intervention mapping approach* (4. Aufl.). San Francisco: Jossey-Bass Inc.
- Fend, H. (2008). *Neue Theorie der Schule: Eine Einführung in das Verstehen von Bildungssystemen* (2. Aufl.). Wiesbaden: Springer VS.
- Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., & Wallace, F. (2005). *Implementation research: a synthesis of the literature*. Tampa: University of South Florida.
- Flay, B.R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15(5), 451–474.
- Forman, S.G., Shapiro, E.S., Coddling, R.S., Gonzales, J.E., Reddy, L.A., Rosenfield, S.A., Sanetti, L.M.H., & Stoiber, K.C. (2013). Implementation science and school psychology. *School Psychology Quarterly*, 28(2), 77–100.
- Goldman, H.H., Ganju, V., Drake, R.E., Gorman, P., Hogan, M., Hyde, P.S., & Morgan, O. (2001). Policy implications for implementing evidence-based practices. *Psychiatric Services*, 52(12), 1591–1597.
- Goodyear-Smith, F., Jackson, C., & Greenhalgh, T. (2015). Co-design and implementation research. Challenges and solutions for ethics committees. *BMC Medical Ethics*, 16, 78.
- Gopichandran, V., Luyckx, V.A., Biller-Andorno, N., Fairchild, A., Singh, J., Tran, N., Saxena, A., Lanois, P., Reis, A., Maher, D., & Vahedi, M. (2016). Developing the ethics of implementation research in health. *Implementation Science*, 11, 161.
- Graham, I.D., Logan, J., Harrison, M.B., Straus, S.E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in knowledge translation: time for a map? *The Journal of Continuing Education in the Health Profession*, 26(1), 13–24.
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13(1), 7–20.
- Gräsel, C., & Parchmann, I. (2004). Implementationsforschung – oder: der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft*, 32(3), 196–214.
- Green, C.A., Duan, N., Gibbons, R.D., Hoagwood, K.E., Palinkas, L.A., & Wisdom, J.P. (2015). Approaches to mixed methods dissemination and implementation research: methods, strengths, caveats, and opportunities. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 508–523.
- Hartmann, U., & Decristan, J. (2018). Brokering activities and learning mechanisms at the boundary of educational research and school practice. *Teaching and Teacher Education*, 74, 114–124.
- Hascher, T., & Schmitz, B. (Hrsg.). (2010). *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen*. Weinheim: Juventa.
- Hasselhorn, M. (2009). Psychologische Erkenntnisse müssen in die Gesellschaft getragen werden: Von der Bringschuld der Psychologie. *Psychologische Rundschau*, 60, 245–246.
- Hasselhorn, M., & Hasselhorn, J. (2017). Quantitative methodische Zugänge der empirischen Bildungsforschung. In M. Heinrich, C. Kölzer & L. Streblow (Hrsg.), *Forschungspraxen der Bildungsforschung: Zugänge und Methoden von Wissenschaftlerinnen und Wissenschaftlern* (S. 39–53). Münster: Waxmann.
- Hasselhorn, M., Köller, O., Maaz, K., & Zimmer, K. (2014). Implementation wirksamer Handlungskonzepte im Bildungsbereich als Forschungsaufgabe. *Psychologische Rundschau*, 65(3), 140–149.
- Heckman, J., LaLonde, R., & Smith, J. (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter & D. Card (Hrsg.), *Handbook of labor economics* (S. 1865–2097). Amsterdam: Elsevier.
- Hedges, L.V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, 11(1), 1–21.
- Heid, H. (2011). Über Bedingungen der Anwendung erziehungswissenschaftlichen Wissens. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen der Empirischen Bildungsforschung* (S. 490–508). Wiesbaden: VS.

- Heinemann, U. (2017). *Bewegter Stillstand. Die paradoxe Geschichte der Schule nach PISA*. Weinheim: Beltz Juventa.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze-Velber: Kallmeyer.
- Herrmann, U. (1971). *Die Pädagogik Wilhelm Diltheys. Ihr wissenschaftstheoretischer Ansatz in Diltheys Theorie der Geisteswissenschaften*. Göttingen: Vandenhoeck & Ruprecht.
- Hetfleisch, P. (2018). *Wie kann Forschung wirksam werden? Gelingensbedingungen der Implementation evidenzbasierter Erkenntnisse in pädagogische Praxis – Ein Beitrag zur Verzahnung pädagogischer Interventions- und Implementationsforschung*. Dissertation. Tübingen: Universität Tübingen. <https://doi.org/10.15496/publikation-24441>.
- Hetfleisch, P., Goeze, A., & Schrader, J. (2017). Wie PraktikerInnen wissenschaftliche Befunde verwenden. Selektions- und Rezeptionsprozesse bei der Implementation eines evidenzbasierten Trainingskonzepts. *Zeitschrift für Pädagogik*, 63(2), 182–205.
- Honig, M.I. (Hrsg.). (2006). *New directions in education policy implementation. Confronting complexity*. Albany: State University of New York Press.
- Kant, I. (1803). Über Pädagogik. Königsberg: Nicolovius. http://www.deutschestextarchiv.de/book/show/kant_paedagogik_1803. Zugegriffen: 23. Nov. 2019
- Kant, I. (1992). *Kritik der reinen Vernunft* (Hrsg. von W. Weischedel). Frankfurt a. M.: Suhrkamp.
- Köller, O. (2017). Verkürzung der Gymnasialzeit in Deutschland: Folgen der G8-Reform in den Ländern der Bundesrepublik Deutschland. Essen: Stiftung Mercator GmbH. https://www.stiftung-mercator.de/media/downloads/3_Publikationen/Neu_Verkuerzung_der_Gymnasialzeit_in_Deutschland.pdf. Zugegriffen: 23. Nov. 2019.
- Kremer, H.-H. (2003). *Implementation didaktischer Theorie – Innovationen gestalten. Annäherungen an eine theoretische Grundlegung im Kontext der Einführung lernfeldstrukturierender Curricula*. Paderborn: Eusl.
- Kuper, H. (2005). *Evaluation im Bildungssystem*. Stuttgart: Kohlhammer.
- Landsverk, J., Brown, C.H., Chamberlain, P., Palinkas, L., Ogihara, M., Czaja, S., Goldhaber-Fiebert, J.D., Rolls Reutz, J.A., & Horwitz, S.M. (2012). Design and analysis in dissemination and implementation research. In I.R.C. Brownson, G.A. Colditz & E.K. Proctor (Hrsg.), *Dissemination and implementation research in health. Translating science to practice* (S. 225–260). New York: Oxford University Press.
- Langer, R., & Brüsemeister, T. (Hrsg.). (2019). *Handbuch Educational Governance Theorien*. Wiesbaden: Springer VS.
- Lewis, C. C., Stanick, C.F., Martinez, R. G., Weiner, B.J., Kim, M., Barwick, M., & Comtois, K. A. (2015). The Society for Implementation Research Collaboration Instrument Review Project: A methodology to promote rigorous evaluation. *Implementation Science*, 10, 2.
- Long, A. C.J., Hagermoser Sanetti, L. M., Collier-Meek, M. A., Gallucci, J., Altschaefl, M., & Kratochwill, T.R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of School Psychology*, 55, 1–26.
- Luhmann, N. (2002). *Das Erziehungssystem der Gesellschaft*. Frankfurt a. M.: Suhrkamp.
- Lynch, S.J., Pyke, C., & Grafton, B.H. (2012). A retrospective view of a study of middle school science curriculum materials: Implementation, scale-up, and sustainability in a changing policy environment. *Journal of Research in Science Teaching*, 49(3), 305–332.
- Lyon, A.R., Connors, E., Jensen-Doss, A., Landes, S.J., Lewis, C.C., McLeod, B.D., Rutt, C., Stanick, C., & Weiner, B.J. (2017). Intentional research design in implementation science: Implications for the use of nomothetic and idiographic assessment. *Transl Behav Med.*, 7, 567–580.
- Maag Merki, K. (2014). Das quasi-experimentelle Design in der Educational Governance-Forschung? In K. Maag Merki, R. Langer & H. Altrichter (Hrsg.), *Educational Governance als Forschungsperspektive: Strategien. Methoden. Ansätze* (Bd. 2, S. 51–83). Wiesbaden: Springer VS.
- McDonnell, L.M., & Weatherford, M.S. (2016). Recognizing the political in implementation research. *Educational Researcher*, 45(4), 233–242.
- Meyers, D.C., Durlak, J.A., & Wandersman, A. (2012). The Quality Implementation Framework. A synthesis of critical steps in the implementation process. *American Journal of Community Psychology*, 50(3–4), 462–480.
- Mosteller, F., & Boruch, R. (Hrsg.). (2002). *Evidence matters: randomized trials in education research*. Washington, DC: The Brookings Institute.
- Munter, C., Cobb, P., & Shekell, C. (2016). The role of program theory in evaluation research: a consideration of the what works clearinghouse standards in the case of mathematics education. *American Journal of Evaluation*, 37(1), 7–26.

- Nathan, N., Elton, B., Babic, M., McCarthy, N., Sutherland, R., Presseau, J., Seward, K., Hodder, R., Booth, D., Yoong, S.L., & Wolfenden, L. (2018). Barriers and facilitators to the implementation of physical activity policies in schools: a systematic review. *Preventive Medicine, 107*, 45–53.
- Nikolai, R., & Rothe, K. (2016). Bildung in der Vergleichenden Politikwissenschaft. In H.-J. Lauth, M. Kneuer, G. Pickel & G. Erdmann (Hrsg.), *Handbuch Vergleichende Politikwissenschaft* (S. 733–745). Wiesbaden: Springer.
- Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science, 10*, 53.
- Nilsen, P., & Bernhardsson, S. (2019). Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Services Research, 19*, 189.
- Odden, A.R. (Hrsg.). (1991). *Education policy implementation*. Albany: State University of New York Press.
- Palinkas, L.A. (2014). Qualitative and mixed methods in mental health services and implementation research. *Journal of Clinical Child & Adolescent Psychology, 43*(6), 851–861.
- Prenzel, M. (2005). Zur Situation der Empirischen Bildungsforschung. In H. Mandl & B. Kopp (Hrsg.), *Impulse für die Bildungsforschung. Stand und Perspektiven. Dokumentation eines Expertengespräch* (S. 7–21). Berlin: Akademie Verlag.
- Proctor, E.K., & Brownson, R.C. (2012). Measurement issues in dissemination and implementation research. In R.C. Brownson, G.A. Colditz & E.K. Proctor (Hrsg.), *Dissemination and implementation research in health. Translating science to practice* (S. 261–279). New York: Oxford University Press.
- Rabin, B.A., & Brownson, R.C. (2012). Developing the terminology for dissemination and implementation research. In R.C. Brownson, G.A. Colditz & E.K. Proctor (Hrsg.), *Dissemination and implementation research in health. Translating science to practice* (S. 23–51). New York: Oxford University Press.
- Robinson, P., & Lowe, J. (2015). Literature reviews vs systematic reviews. *Australian and New Zealand Journal of Public Health, 39*(2), 103.
- Roth, H. (1967). Die realistische Wendung in der pädagogischen Forschung. In H. Thiersch & H. Türken (Hrsg.), *Heinrich Roth: Erziehungswissenschaft, Erziehungsfeld und Lehrerbildung. Gesammelte Abhandlungen 1957–1967* (S. 113–126). Hannover: Schroedel.
- Rothwell, L.E., Greene, J.A., Podolsky, S.H., & Jones, D.S. (2016). Assessing the Gold Standard—lessons from the history of RCTs. *New England Journal of Medicine, 374*(22), 2175–2181.
- Schleiermacher, F. (1983). *Pädagogische Schriften* (Bd. 1). Frankfurt a. M.: Ullstein.
- Schmid, J., & Schuhen, M. (2018). Bildungspolitik. In K. Mause, C. Müller & K. Schubert (Hrsg.), *Politik und Wirtschaft: Ein integratives Kompendium* (S. 497–521). Wiesbaden: Springer.
- Schmid, U., Goertz, L., & Behrens, J. (2017). *Monitoring Digitale Bildung. Die Weiterbildung im digitalen Zeitalter*. Gütersloh: Bertelsmann Stiftung.
- Schneider, B., & McDonald, S.-K. (2007a). Introduction. In B. Schneider & S.-K. McDonald (Hrsg.), *Scale-up in education: Ideas in principle* (S. 1–15). Lanham: Rowman & Littlefield.
- Schneider, B., & McDonald, S.-K. (Hrsg.). (2007b). *Ideas in principle* (Scale-up in education, Bd. 1). Lanham: Rowman & Littlefield.
- Schrader, J. (2011). *Struktur und Wandel der Weiterbildung*. Bielefeld: Bertelsmann.
- Schrader, J. (2014). Analyse und Förderung effektiver Lehr-Lernprozesse unter dem Anspruch evidenzbasierter Bildungsreform. *Zeitschrift für Erziehungswissenschaft, 17*(2), 193–223.
- Skär, L., & Söderberg, S. (2018). The importance of ethical aspects when implementing eHealth services in healthcare: A discussion paper. *Journal of Advanced Nursing, 74*(5), 1043–1050.
- Slavin, R.E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.
- Southam-Gerow, M.A., & McLeod, B.D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: introduction to special issue. *Clinical Psychology: Science and Practice, 20*(1), 1–13.
- Spiel, C. (2009). Evidenzbasierte Bildungspolitik und Bildungspraxis – eine Fiktion? Problemaufriss, Thesen, Anregungen. *Psychologische Rundschau, 60*(4), 255–256.
- Spiel, C., Schober, B., & Strohmeier, D. (2018). Implementing intervention research into public policy—the “I²-Approach”. *Prevention Science, 19*(3), 337–346.
- Stains, M., & Vickrey, T. (2017). Fidelity of implementation: an overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE—Life Sciences Education, 16*(1), 1–11.

- Sutoris, P. (2018). Ethically scaling up interventions in educational development: a case for collaborative multi-sited ethnographic research. *Comparative Education*, 54(3), 390–410.
- Tabak, R. G., Khoong, E. C., Chambers, D. A., & Brownson, R. (2012). Bridging research and practice: models for dissemination and implementation research. *American Journal of Preventive Medicine*, 43(3), 337–350.
- Tenorth, H.-E. (2003). „Wie ist Bildung möglich?“. Einige Antworten – und die Perspektive der Erziehungswissenschaft. *Zeitschrift für Pädagogik*, 49(3), 422–430.
- Tenorth, H.-E. (2014). Politikberatung und Wandel der Expertenrolle oder: Die Expertise der Erziehungswissenschaft. In R. Fatke & J. Oelkers (Hrsg.), *Das Selbstverständnis der Erziehungswissenschaft: Geschichte und Gegenwart* (Zeitschrift für Pädagogik: Beiheft 60, S. 139–171). Weinheim: Beltz Juventa.
- Thomsen, S. L., & Anger, S. (2018). Die Notwendigkeit ökonomischer Politikberatung für eine evidenzbasierte Bildungspolitik: Verkürzung und Verlängerung der Schulzeit am Gymnasium. *Perspektiven der Wirtschaftspolitik*, 19(3), 167–184.
- Ullrich, C., Mahler, C., Forstner, J., Szecsenyi, J., & Wensing, M. (2017). Teaching implementation science in a new master of science program in Germany. A survey of stakeholder expectations. *Implementation Science*, 12, 55.
- Waycott, J., Morgans, A., Pedell, S., Ozanne, E., Vetere, F., Kulik, L., & Davis, H. (2015). Ethics in evaluating a sociotechnical intervention with socially isolated older adults. *Qualitative Health Research*, 25(11), 1518–1528.
- Weidema, F., Molewijk, B., Widdershoven, G., & Abma, T. A. (2012). Enacting ethics: Bottom-up involvement in implementing moral case deliberation. *Health Care Analysis*, 20(1), 1–19.
- Weingart, P. (2017). Wissenschaftskommunikation unter digitalen Bedingungen. Funktionen, Akteure und Probleme des Vertrauens. In P. Weingart, H. Wormer, A. Wenninger & R. F. Hüttel (Hrsg.), *Perspektiven der Wissenschaftskommunikation im digitalen Zeitalter* (S. 31–59). Weilerswist: Velbrück Wissenschaft.
- Young, T., & Lewis, W. D. (2015). Educational policy implementation revisited. *Educational Policy*, 29(1), 3–17.