

Henschel, Sofie

Evaluation von Konzepten der Sprachbildung und -förderung

Becker-Mrotzek, Michael [Hrsg.]; Gogolin, Ingrid [Hrsg.]; Roth, Hans-Joachim [Hrsg.]; Stanat, Petra [Hrsg.]: Grundlagen der sprachlichen Bildung. Münster ; New York : Waxmann 2023, S. 277-290. - (Sprachliche Bildung; 10)



Quellenangabe/ Reference:

Henschel, Sofie: Evaluation von Konzepten der Sprachbildung und -förderung - In: Becker-Mrotzek, Michael [Hrsg.]; Gogolin, Ingrid [Hrsg.]; Roth, Hans-Joachim [Hrsg.]; Stanat, Petra [Hrsg.]: Grundlagen der sprachlichen Bildung. Münster ; New York : Waxmann 2023, S. 277-290 - URN: urn:nbn:de:0111-pedocs-320131 - DOI: 10.25656/01:32013

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-320131>

<https://doi.org/10.25656/01:32013>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

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

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

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

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



Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Evaluation von Konzepten der Sprachbildung und -förderung

1. Einleitung

Bei einer Evaluation handelt es sich ganz allgemein um die systematische Bewertung einer Sache, z. B. eines Gegenstandes, Produkts oder Prozesses (Scriven, 1972). Im Gegensatz zu alltäglichen Wertungen und Einschätzungen (z. B. „Das Buch ist gut.“), zielen wissenschaftliche Evaluationen darauf ab,

- Qualität, Wirksamkeit und Nützlichkeit von unterschiedlichen Maßnahmen (z. B. zur Förderung des bildungssprachlichen Wortschatzes)
- mithilfe sozialwissenschaftlicher Verfahren wie Befragungen, Beobachtungen oder Tests
- auf der Grundlage empirisch gewonnener qualitativer oder quantitativer Daten
- anhand von unterschiedlichen Evaluationskriterien (z. B. Akzeptanz, Praktikabilität, Wirksamkeit, Effizienz oder Nachhaltigkeit)

zu bewerten (Bortz & Döring, 2016).

In den USA wurden erste Evaluationen im Bildungsbereich bereits in den 1930er bzw. 1940er Jahren im Rahmen des Progressive Education Movement durchgeführt (Watras, 2006). Auch in Deutschland gewannen Evaluationen (bildungs-) politischer Reformmaßnahmen (u. a. Schulversuche mit integrierten Gesamtschulen) in den 60er und 70er Jahren kurzzeitig an Bedeutung (Stockmann & Meyer, 2010). Doch erst seit der so genannten „empirischen Wende“ um die Jahrtausendwende wird die Leistungsfähigkeit des deutschen Bildungssystems im nationalen und internationalen Vergleich regelmäßig evaluiert (KMK, 2015).

Im Bereich der Sprachbildung und -förderung haben sich (wissenschaftliche) Evaluationen erst in den letzten Jahren zunehmend etabliert. Vorangetrieben wurde diese Entwicklung u. a. durch die von Bund und Ländern seit Beginn der 2000er Jahre angestoßenen Maßnahmen der durchgängigen Sprachbildung und -förderung, zu denen beispielsweise die Programme „Förderung von Kindern und Jugendlichen mit Migrationshintergrund“ (FörMig), „Sprachliche Bildung und Mehrsprachigkeit“ sowie „Bildung durch Sprache und Schrift“ (BiSS) zählen. In diesen bundesweiten Programmen, die zumeist sowohl den Elementar- als auch den Schulbereich einbeziehen, wurden und werden sowohl aus der Praxis heraus (bottom-up) als auch durch Wissenschaftler:innen entwickelte oder die Bildungs-

administration vorgegebene (top-down) Maßnahmen der Sprachbildung und -förderung in Praxisverbünden (z.B. mehrere Schulen einer Stadt oder Region) umgesetzt. Begleitende wissenschaftliche Evaluationen, die etwa die Qualität des zugrundeliegenden Konzepts, die Praktikabilität und Akzeptanz der Maßnahmen und deren Wirksamkeit in der pädagogischen Praxis untersuchen, sind dabei ein integraler Bestandteil der Qualitätsentwicklung und -sicherung und wesentliche Grundlage für den Transfer der Maßnahmen in die Fläche (Schneider et al., 2012). Gleichzeitig verbessern Evaluationen in diesem Bereich auch die Infrastruktur für Forschung durch die Neu- und Weiterentwicklung sowie empirische Überprüfung geeigneter Erhebungsverfahren.

Im folgenden Kapitel wird anhand von Konzepten der Sprachbildung und -förderung exemplarisch herausgearbeitet, welche Funktionen und Formen von Evaluation unterschieden werden. Dabei wird betrachtet, wie Sprachbildungs- und Sprachförderkonzepte im Verlauf ihrer Entwicklung und Umsetzung evaluiert werden können. Abschließend werden zentrale Qualitätsmerkmale sowie Chancen und Grenzen von Evaluationsvorhaben im Bereich der Sprachbildung und -förderung zusammengefasst.

2. Ziele und Funktionen von Evaluationen

Im Vergleich zur Grundlagenforschung, bei der das Ziel primär darin besteht, eine Theorie empirisch zu überprüfen und weiterzuentwickeln, sind Evaluationsvorhaben anwendungsorientiert. Das heißt, mit einer Evaluation sind verschiedene praxisbezogene Ziele bzw. Funktionen verbunden, um konkrete Maßnahmen weiterzuentwickeln bzw. Entscheidungsprozesse für zukünftiges Handeln zu unterstützen. In der Literatur werden typischerweise vier Funktionen unterschieden, die im Evaluationsprozess eng miteinander verbunden sind (Wottawa & Thierau, 1998):

- *Erkenntnisfunktion*: Evaluationen sollen Erkenntnisse über Eigenschaften und Wirkungen einer Maßnahme liefern, also etwa aufzeigen, wie akzeptiert, praktikabel, wirksam und nachhaltig ein Fortbildungskonzept zur fachintegrierten Sprachförderung im Sachunterricht ist. Die Erkenntnisfunktion hat einen übergeordneten Charakter, da alle weiteren Evaluationsfunktionen darauf basieren und die gewonnenen Erkenntnisse für unterschiedliche Zwecke nutzen.
- *Optimierungsfunktion*: Erkenntnisse über die Alltagstauglichkeit, Durchführbarkeit und Akzeptanz eines Fortbildungs- oder Förderansatzes oder einer ausgewählten Maßnahme (z.B. Frageverhalten der Lehrkraft) liefern Hinweise darauf, wie diese an lokale Bedingungen angepasst werden sollte, um wirksam zu sein. Zum Beispiel könnte die Befragung von Lehrkräften, die einen Sprachförderansatz im Sachunterricht umsetzen, darauf hinweisen, dass das Anspruchsniveau der Fördermaterialien für Schüler:innen mit nicht-deutscher Herkunftssprache zu hoch ist. Folglich sollte es überarbeitet und ggf. differenzierter gestaltet werden, damit auch diese Schüler:innen von der Förderung

profitieren können und die Lehrkräfte die Maßnahme als sinnvoll erachten, akzeptieren und systematisch umsetzen.

- *Entscheidungsfunktion:* Erkenntnisse über die Wirksamkeit und Effizienz dienen als Grundlage für eine Entscheidung darüber, ob eine Maßnahme fortgeführt oder ausgedehnt werden sollte. Kriterien für die Fortführung könnten z. B. sein, dass ein Fortbildungskonzept zur fachintegrierten Sprachförderung sowohl das Wissen und die Sprachförderkompetenz der Lehrkräfte als auch die Unterrichtsqualität verbessert und das sprachliche und fachliche Lernen der Schüler:innen fördert. Liegen zwei alternative und bereits positiv evaluierte Fortbildungskonzepte mit der gleichen Zielsetzung vor, kann anhand der Evaluationsergebnisse das Konzept ausgewählt werden, das mit einem geringeren Fortbildungsaufwand verbunden ist und eine vergleichbare oder stärkere bzw. nachhaltigere Wirksamkeit aufweist.
- *Legitimationsfunktion:* Positive Evaluationsergebnisse können z. B. gegenüber der Bildungsverwaltung rechtfertigen, dass eine aufwendige Professionalisierungsphase von Lehrkräften, zusätzlicher Personalbedarf oder die Anschaffung von teuren Fördermaterialien sinnvoll ist. Dies könnte der Fall sein, wenn Evaluationsergebnisse zeigen, dass eine Sprachfördermaßnahme kompensatorische Effekte hat und dazu beiträgt, dass sich Bildungsdisparitäten in den fachlichen und bildungssprachlichen Kompetenzen zwischen Schüler:innen mit deutscher und nicht-deutscher Herkunftssprache verringern oder sogar vollständig beseitigen lassen.

3. Formen und Fragestellungen von Evaluationen

Konzepte der Sprachbildung und -förderung können in verschiedenen Stadien ihrer Entwicklung, Erprobung und Umsetzung anhand unterschiedlicher Fragestellungen evaluiert werden. Die dabei genutzten Evaluationsformen können sich, entsprechend der Rahmenkonzeption zur Evaluation psychologischer Interventionsmaßnahmen von Mittag und Hager (2000), auf die Konzeption, Durchführung und Wirksamkeit beziehen.

Prinzipiell können alle Evaluationsformen extern oder intern durchgeführt werden. Während *externe Evaluationen* von Außenstehenden (z. B. Wissenschaftler:innen) umgesetzt werden, erfolgt dies bei *internen Evaluationen* zum Beispiel durch Lehrkräfte des Kollegiums, die das Konzept nicht selbst durchführen. In beiden Fällen handelt es sich um *Fremdevaluationen*. Eine *Selbstevaluationen* ist ein Spezialfall der internen Evaluation, die von den Personen durchgeführt wird, die das zu evaluierende Konzept selbst umsetzen (z. B. Lehrkräfte, die eine Sprachfördermaßnahme im Sachunterricht durchführen). Während bei Fremdevaluationen eine neutrale Bewertung von außen im Vordergrund steht, geht es bei der Selbstevaluation vor allem um die Beobachtung, Reflexion und Weiterentwicklung der eigenen Arbeit. Das Vorgehen und die Ergebnisse sind dabei stärker subjektiv geprägt und die Methoden weniger komplex

und standardisiert als bei Fremdevaluationen. Mit entsprechender professioneller Kompetenz lassen sich Selbstevaluationen häufig mit überschaubarem Aufwand auf die eigenen Ziele und Fragestellungen anpassen und in den pädagogischen Alltag integrieren. Auch zur Evaluation von Sprachbildungs- und Sprachförderkonzepten werden sie eingesetzt, vor allem wenn die Konzepte aus der Praxis heraus (bottom-up) von Lehrkräften selbst entwickelt und angewendet werden und im Entwicklungsprozess weiter optimiert werden sollen (Giel, Klockgether & Mäder, 2018).

Insgesamt sind externe Fremdevaluationen, die von Wissenschaftler:innen durchgeführt werden und eine Kontrollgruppe einbeziehen (siehe Abschnitt 2.3), in der die Intervention nicht umgesetzt wird, am besten geeignet, um möglichst belastbare Erkenntnisse über die Qualität eines Sprachbildungs- oder Sprachförderkonzepts zu gewinnen. Bei allen anderen Evaluationsformen ist die Erkenntnisfunktion eingeschränkt, u. a. weil keine Kontrollgruppen und selten standardisierte Erhebungsverfahren verwendet werden. Die mit der Entscheidungs- und Legitimationsfunktion verbundenen Ziele können mit diesen Evaluationsformen deshalb nicht oder nur sehr eingeschränkt erreicht werden. Im pädagogischen Alltag sind interne Evaluationen und Selbstevaluationen jedoch eine sinnvolle Ergänzung externer Fremdevaluationen, um Maßnahmen im Rahmen von Schul- und Unterrichtsentwicklungsprozessen vor dem Hintergrund lokaler Bedingungen zu optimieren.

In welchen Phasen der Entwicklung und Umsetzung Sprachbildungs- und -förderkonzepte evaluiert werden können, welche Fragestellungen dabei leitend sind und welche Methoden dabei zum Einsatz kommen können, wird nachfolgend beschrieben.

3.1 Konzeptevaluation

Eine Konzeptevaluation wird vor der eigentlichen Erprobung oder Umsetzung durchgeführt. Dabei wird die Wahrscheinlichkeit der intendierten Wirksamkeit anhand von theoretischen Annahmen und empirischen Befunden bewertet. Im Mittelpunkt steht somit die Frage, auf welcher theorie- oder evidenzbasierten Grundlage anzunehmen ist, dass das Konzept bzw. eine einzelne Maßnahme des Konzepts wirksam sein wird. So könnte erwartet werden, dass sich der allgemeine bildungssprachliche Wortschatz von Schüler:innen im Sachunterricht der Grundschule durch den gezielten Einsatz von Sprachförderstrategien im Unterricht verbessern lässt. Das Förderkonzept könnte z. B. sprachförderliches Frageverhalten der Lehrkraft beinhalten, das u. a. Fragen umfasst, die explizit die Sprache fokussieren („Wie kann man noch dazu sagen?“), die implizit komplexe Äußerungen anregen („Was konntest du dabei beobachten?“) und die kognitiv anregend formuliert sind und eine Verbindung zwischen sprachlichem und fachlichem Lernen herstellen („Was vermutest du, woran das liegt?“; Archie, Rank & Franz, 2014). In diesem Fall bildet der *Scaffolding-Ansatz* (Gibbons, 2002) eine zentrale theoreti-

sche Grundlage für die Planung und Durchführung sprachförderlicher Aktivitäten im Fachunterricht. Das Frageverhalten (z.B. offene Fragen oder Fragen mit explizit sprachförderlichem Fokus) ist dabei ein integraler Bestandteil des Scaffoldings (Hardy, Hettmannsperger & Gabler, 2019), das sich unter bestimmten Bedingungen als wirksam erwiesen hat (van de Pol, Volman, Oort & Beishuizen, 2015). Das Konzept bzw. dieser Teil des Konzepts ist also nicht nur theoretisch fundiert, sondern auch evidenzbasiert.

Darüber hinaus wird in der Konzeptevaluation bewertet, inwieweit Zielsetzungen für die jeweilige Zielgruppe angemessen und Inhalte sowie Vermittlungsstrategien unter den gegebenen lokalen Bedingungen (z.B. Zusammensetzung der Schüler:innen) geeignet und gut aufeinander abgestimmt sind. Dies umfasst auch die Bewertung der diagnostischen Verfahren (z.B. Handelt es sich um valide standardisierte oder informelle bzw. selbstentwickelte Instrumente?), mit denen die Zielerreichung anhand messbarer Kriterien (z.B. bildungssprachlicher Wortschatz) und Bewertungsmaßstäbe überprüft werden kann (z.B. Ist ein mittelgroßer statistischer Effekt im Prä-Post-Vergleich gegenüber einer Kontrollgruppe realistisch?). Die Ergebnisse liefern Hinweise darauf, in welchen Bereichen Konzepte noch vor einer Erprobung oder Durchführung bzw. Implementation angepasst werden sollten, um voraussichtlich wirksam zu sein. Die Optimierungsfunktion ist somit zentral für diese Evaluationsform.

Beispiel

Im Rahmen einer externen Fremdevaluation im BiSS-Programm zeigten Schmitz, Zeuch, Karstens, Meudt, Jost & Souvignier (2021), dass sowohl bottom-up von Lehrkräften selbst entwickelte als auch bildungsadministrativ top-down bereitgestellte Konzepte zur Leseförderung in der Sekundarstufe I theoretisch sinnvoll und zielgerichtet aufgebaut waren. Gleichzeitig wurden Optimierungspotenziale festgestellt, etwa dass die Umsetzung durch Lehrkräfte mit Hilfe konkreter Nutzungshinweise verbessert werden könnte (z. B. zu welchem Zeitpunkt im Schuljahr welche Fördermaterialien zu welchem Zweck eingesetzt werden sollten). Die Erkenntnisse aus der Konzeptevaluation wurden durch Ergebnisse der formativen Evaluation untermauert. Dies zeigt, dass es sinnvoll ist, eine Konzeptevaluation durchzuführen und die durchführenden Personen (z. B. Lehrkräfte) möglichst bereits in die Konzeptentwicklung einzubeziehen. Dadurch kann verhindert werden, dass erst bei der Erprobung bzw. Durchführung festgestellt wird, dass einzelne Komponenten nicht ausreichend konkretisiert oder aufeinander abgestimmt sind.

3.2 Formative Evaluation

Formative Evaluationen werden begleitend während des Erprobungs- bzw. Interventionsprozesses durchgeführt. Bei einer *formativen Prozessevaluation* werden u.a. die Akzeptanz und Zufriedenheit der Beteiligten, wahrgenommene (Nutzungs-)Barrieren sowie die Qualität der Umsetzung insgesamt (*global formativ*) oder einzelner Komponenten (z.B. eine Fördereinheit; *analytisch formativ*) bezogen auf Praktikabilität, Vollständigkeit und Wiedergabetreue analysiert. Eine *formative Ergebnisevaluation* kann bereits in einer frühen Entwicklungsphase erste Anhaltspunkte auf die intendierte Wirksamkeit einzelner Konzeptkomponenten liefern.

Ziel der formativen Evaluation ist es, vorhandene Schwachstellen in der Maßnahmenkonzeption (z.B. Anforderungsniveau oder Verständlichkeit von Fördermaterial) oder in deren Umsetzung (z.B. Dauer, Umfang, Ablauf) anhand von Zwischenergebnissen aufzudecken und das Konzept auf dieser Grundlage zu optimieren. Mitunter werden neue Konzepte wiederholt formativen Prozess- und Ergebnisevaluationen unterzogen, um sie im Entwicklungsprozess zu erproben und schrittweise unter den situativen Bedingungen zu optimieren (Design-Based Research; Cobb, Confrey, diSessa, Lehrer & Schauble, 2003). Aufschlussreich können formative Evaluationen aber auch sein, um etablierte Konzepte an veränderte Bedingungen (z.B. eine leistungsschwächere Lerngruppe) anzupassen und die Durchführbarkeit zu gewährleisten oder unter den gegebenen Bedingungen zu verbessern. Häufig werden dabei sowohl informelle und qualitative als auch niedrigschwellige quantitative Verfahren eingesetzt (z.B. Gruppenreflexionen, Befragungen).

Beispiel

In einer formativen Prozessevaluation im BiSS-Programm beobachteten Schmitz et al. (2021), wie Lehrkräfte Förderansätze zum selbstregulierten Lesen im Deutsch- und Fachunterricht der Sekundarstufe I in verschiedenen Praxisverbünden umsetzten. Es zeigte sich, dass sowohl bottom-up entwickelte als auch top-down bereitgestellte Materialien zur Leseförderung im Deutsch- und im Fachunterricht kaum verwendet wurden. Eine Lehrkräftebefragung wies darauf hin, dass dies mit gering ausgeprägter Akzeptanz, gering wahrgenommener Praktikabilität und geringer Unterstützung im Kollegium sowie fehlendem Wissen über die Fördermaterialien und fehlenden zeitlichen Ressourcen zur Auseinandersetzung damit einherging. Anhand der Ergebnisse konnten die Autor:innen der Studie konkrete Vorschläge zur Konzeptoptimierung ableiten und an die Konzeptentwickler:innen (z.B. exemplarische Einbindung der Fördermaterialien in konkrete Unterrichtssequenzen) und Schulen (z.B. Bereitstellung von zeitlichen und räumlichen Ressourcen sowie Unterstützung des kollegialen Austauschs) zurückmelden.

3.3 Summative Evaluation

Eine summative Evaluation sollte erst durchgeführt werden, wenn begründet anzunehmen ist, dass ein Konzept mit hoher Wahrscheinlichkeit wirksam sein wird. Dazu sollte es nicht nur ausgearbeitet vorliegen, sondern möglichst bereits erprobt und formativ evaluiert sowie auf Basis dieser Ergebnisse optimiert worden sein. Im Mittelpunkt der summativen Evaluation steht die zusammenfassende Analyse und Bewertung der Qualität des Konzepts nach Abschluss der Durchführung bzw. Implementation. Untersucht werden Fragen zur Wirksamkeit (Effektivität) und zum Nutzen (Effizienz), wobei das gesamte Konzept (*global summativ*) oder einzelne Komponenten (*analytisch summativ*) betrachtet werden können. Die Entscheidungs- und Legitimationsfunktion stehen im Vordergrund, die Ergebnisse liefern aber auch Hinweise zur Optimierung. Die summative Evaluation stellt im Vergleich zur Konzept- und formativen Evaluation die höchsten Qualitätsansprüche, sowohl an das Design der Studie und die eingesetzten Erhebungsverfahren als auch an die Auswertungsmethoden.

Während in formativen Prozessevaluationen nur die Interventionsgruppe begleitend untersucht wird, die das Konzept erprobt bzw. umsetzt, sind für die Prüfung der Wirksamkeit und Effizienz in einer summativen Evaluation Vergleichsgruppen erforderlich. Geht es um die prinzipielle Wirksamkeit (*isolierte Evaluation*), wird eine Interventionsgruppe, die die Maßnahme erhält bzw. umsetzt (z.B. Grundschullehrkräfte, die an einer Sprachförderfortbildung teilnehmen und ihren Sachunterricht anschließend sprachförderlich gestalten) mit einer Kontrollgruppe verglichen. Die Kontrollgruppe erhält entweder keine besondere Maßnahme (z.B. regulärer Unterricht), nimmt an einer Maßnahme mit anderer Zielsetzung teil (z.B. Förderung mathematischer Kompetenzen) oder sie erhält die Maßnahme der Interventionsgruppe – ggf. in optimierter Form – zu einem späteren Zeitpunkt nach Abschluss der Evaluation (*Warte-Kontrollgruppe*). Haben sich mehrere Maßnahmen mit der gleichen Zielsetzung (z.B. Fortbildung zur fachintegrierten Sprachförderung) aber unterschiedlichen Vorgehensweisen (z.B. Dauer, Vermittlungsstrategien) jeweils in isolierten Evaluationen gegenüber einer Kontrollgruppe als wirksam erwiesen, können sie in einer *vergleichenden Evaluation* bezogen auf Wirksamkeit und Effizienz gegenübergestellt werden.

Die Herausforderung besteht bei der Prüfung der Wirksamkeit darin, die Varianzanteile in den insgesamt beobachteten Veränderungen bzw. Unterschieden in dem interessierenden Kriterium (z.B. bildungssprachlicher Wortschatz) zu ermitteln, die ausschließlich auf das Konzept (z.B. Sprachförderfortbildung) zurückzuführen sind. Alternative Erklärungen für beobachtete Veränderungen (z.B. fachliche Ausgangsleistungen der Schüler:innen) müssen durch angemessene Untersuchungsdesigns und statistische Verfahren möglichst vollständig ausgeschlossen werden. Die höchste Zuverlässigkeit bieten *experimentelle Designs*, bei denen Personen randomisiert, also zufällig einer Interventions- oder Kontrollbedingung zugewiesen werden (z.B. ob Lehrkräfte an einer Sprachförderfortbildung teilnehmen oder nicht). Störeinflüsse wie z.B. Interesse oder Vorkenntnisse der Lehr-

kräfte, die sowohl mit der Teilnahme an der Sprachförderfortbildung als auch mit dem Kriterium (z. B. bildungssprachlicher Wortschatz der Schüler:innen) zusammenhängen könnten, werden dabei minimiert, weil die potenziell konfundierten Merkmale zufällig bzw. gleichmäßig über die Gruppen verteilt werden. Folglich sind beobachtete Unterschiede zwischen der Interventions- und der Kontrollgruppe mit hoher Wahrscheinlichkeit auf die Intervention zurückzuführen.

Im Bereich der Sprachbildungs- und -förderung werden allerdings häufiger *quasiexperimentelle Designs* genutzt. Das heißt, Schüler:innen verschiedener Klassen werden nicht zufällig auf die Lehrkräfte verteilt, die ein Sprachförderkonzept im Unterricht umsetzen, sondern alle Schüler:innen verbleiben in ihren Klassen. Dabei kann nicht ausgeschlossen werden, dass sich die Klassen in relevanten Hintergrundmerkmalen (z. B. Anteil mono- und bilingualer Schüler:innen) und Ausgangsleistungen bedeutsam voneinander unterscheiden und dies einen Einfluss auf die Wirksamkeit des Konzepts hat. In diesem Fall wäre die interne Validität, also die kausale Interpretierbarkeit der Ergebnisse, eingeschränkt. Um dies zu vermeiden, müssen potenzielle Störvariablen (z. B. Ausgangsleistung, Geschlecht oder Sprachhintergrund der Schüler:innen) in quasiexperimentellen Designs erfasst und statistisch kontrolliert werden oder es werden mittels statistischer Methoden (z. B. Propensity Score Matching; Rosenbaum & Rubin, 1983) nachträglich vergleichbarere Gruppen gebildet. Dabei kann allerdings nie vollständig ausgeschlossen werden, dass ebenfalls relevante Faktoren unberücksichtigt geblieben sind.

Von Interesse ist bei Wirksamkeitsevaluationen auch, wie nachhaltig beobachtete Effekte sind. Dazu werden längsschnittliche Studien mit mehr als zwei Messzeitpunkten durchgeführt, die nicht nur vor (Prätest) und unmittelbar nach (Posttest), sondern erneut einige Zeit nach dem Abschluss der Intervention stattfinden (ein oder mehrere Follow-up-Tests). Ein generelles Problem ist bei längsschnittlichen Untersuchungen der (ggf. selektive) Stichprobenausfall im Verlauf der Studie, weil dies zu Effizienzeinschränkungen bei der Prüfung der Wirksamkeit bzw. zu verzerrten Parameterschätzungen führen kann. Ein fallweiser Ausschluss der Daten, bei dem Fälle aus der Analysestichprobe entfernt werden, wenn sie auf mindestens einer der Analysevariablen einen fehlenden Wert aufweisen, geht allerdings mit erheblichen Validitätseinschränkungen der Ergebnisse einher. Deshalb werden in der Regel statistische Verfahren für den Umgang mit fehlenden Werten genutzt, die zu weitgehend unverzerrten Parameterschätzungen führen (z. B. Multiple Imputation, Full-Information-Maximum-Likelihood-Verfahren; Lüdtke, Robitzsch, Trautwein & Köller, 2007).

Beispiel

Bei der Wirksamkeitsevaluation von Sprachbildungs- bzw. Sprachförderkonzepten können unterschiedliche Evaluationsebenen betrachtet werden. Wenn die Wirksamkeit von Fortbildungskonzepten untersucht wird, kann dies beispielsweise folgende Aspekte betreffen (Lipowsky, 2010):

- 1) die Reaktionen der Lehrkräfte auf die Fortbildung (z. B. Zufriedenheit),
- 2) Kompetenzen und Überzeugungen (z. B. Veränderungen im Wissen, Können oder der Selbstwirksamkeit),
- 3) das unterrichtspraktische Handeln (z. B. Ausmaß an sprachförderlichem Frageverhalten) und
- 4) die sprachlichen Kompetenzen der Schüler:innen

Um die Wirksamkeit eines Fortbildungskonzepts bei den Schüler:innen (Ebene 4) zu beobachten, sollte sich dieses auf den Ebenen 1–3 als effektiv erwiesen haben.

Für Fortbildungskonzepte zur fachintegrierten Sprachförderung zeigten Kalinowski, Egert, Gronostaj und Vock (2020) in einer Metaanalyse, dass die Evaluationsforschung in diesem Feld noch am Anfang steht. Insgesamt liegen kaum Ergebnisse für Ebene 4 vor und für die Ebenen 2 und 3 fallen die berichteten Effekte sehr unterschiedlich aus, was teilweise mit methodischen Merkmalen der Studien zusammenhing. Gemischte Befunde aus summativen Evaluationen zeigten sich auch in den Evaluations- sowie Forschungs- und Entwicklungsprojekten im BiSS-Programm (Gentrup, Henschel, Beck & Stannat, 2021; Titz et al., 2020):

Rank et al. (2021) beobachteten in einer externen Fremdevaluation verschiedener bottom-up entwickelter Ansätze zur Sprachbildung im Mathematikunterricht zunächst eine hohe Zufriedenheit von Grundschullehrkräften mit dem Sprachbildungsansatz. Allerdings konnte weder ein Zuwachs der Sprachförderkompetenz der Lehrkräfte noch der (bildungs-)sprachlichen oder mathematikbezogenen Kompetenzen der Schüler:innen im Vergleich mit einer Kontrollgruppe festgestellt werden. Schmitz et al. (2021) konnten in ihrer externen Fremdevaluation von Konzepten zur Förderung von Lesestrategien nicht feststellen, dass Lehrkräfte diese im Unterricht häufiger anwendeten als Lehrkräfte der Kontrollgruppe. Auch bei den Schüler:innen waren keine Effekte zu erkennen, da in der längsschnittlichen Untersuchung im Vergleich mit Schüler:innen der Kontrollgruppe keine Unterschiede in der Leseflüssigkeit, Lesekompetenz oder Lesefreude beobachtet wurden.

Im BiSS-Forschungs- und Entwicklungsprojekt ProSach, wurden Lehrkräfte dafür fortgebildet, implizit bedeutungsfokussierte sowie fachlich integrierte Unterstützungsstrategien der sprachlichen Bewusstmachung und Sprachmodellierung anzuwenden. Die Evaluation wies auf positive Effekte auf den ersten drei Evaluationsebenen hin zusammenfassend (Henschel, Heppt, Hardy,

Mannel & Stanat, 2020). So bewerteten die Lehrkräfte die von den Wissenschaftlerinnen entwickelten Fortbildungsinhalte als praxisrelevant und gaben an, diese in ihrem Unterricht umsetzen zu wollen (Ebene 1). Zudem verfügten die Lehrkräfte über bessere Sprachförderkompetenzen (Ebene 2) und gestalteten ihren Unterricht sprachförderlicher und fachlich herausfordernder (Ebene 3) als Lehrkräfte in der Kontrollgruppe (Heppt et al., 2022). Dennoch zeigten sich keine stärkeren Kompetenzzuwächse im Bereich fachlicher und bildungssprachlicher Leistungen bei Schüler:innen im Verlauf der 3. Klasse, deren Lehrkräfte an der Sprachförderfortbildung teilgenommen hatten (Ebene 4).

4. Qualitätsmerkmale von Evaluationen

Wie bei allen empirischen Studien stellen die Gütekriterien der eingesetzten Erhebungsverfahren (Tests, Befragungen, Beobachtungen) zentrale Qualitätsmerkmale von Evaluationsstudien dar:

- *Objektivität*: Die erfassten Daten sollen unabhängig davon sein, wer die Erhebung durchführt (Durchführungsobjektivität), wer die Daten auswertet (Auswertungsobjektivität) und wer die ausgewerteten Daten interpretiert (Interpretationsobjektivität). Die Interpretationsobjektivität der Daten steigt, wenn Verfahren zuvor an einer repräsentativen Stichprobe normiert wurden.
- *Reliabilität*: Die Daten sollen das Konstrukt genau erfassen und der Messfehler sollte möglichst klein sein.
- *Validität*: Die erfassten Daten sollen ausschließlich das intendierte Konstrukt abbilden.

Ferner sind zwei Qualitätsmerkmale für das verwendete Untersuchungsdesign maßgeblich:

- *Interne Validität*: Die Evaluationsergebnisse sind eindeutig auf die Intervention zurückzuführen und erlauben eine kausale Interpretation. Zum Beispiel ist eine Evaluationsstudie intern valide, wenn die Entwicklung des bildungssprachlichen Wortschatzes bei Drittklässler:innen im Sachunterricht eindeutig auf die Fortbildung der Lehrkräfte zur fachintegrierten Sprachförderung zurückzuführen ist und nicht auf andere Faktoren (z. B. fachliche Ausgangsleistungen der Schüler:innen im Vergleich zur Kontrollgruppe).
- *Externe Validität*: Die Evaluationsergebnisse sind auf andere Kontexte generalisierbar. Beispielsweise wäre die externe Validität eingeschränkt, wenn die o. g. Ergebnisse nur auf Berliner Grundschüler:innen generalisiert werden können, die ein Jahr länger in der flexiblen Schulanfangsphase waren.

Darüber hinaus hat die Deutsche Gesellschaft für Evaluation vier Evaluationsstandards festgelegt, zu denen Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit zählen (DeGEval, 2016). Diese Standards können als Nebengütekriterien be-

trachtet werden, die die Qualität von Evaluationen und deren Fairness gegenüber allen Beteiligten sicherstellen und bei der Durchführung von Evaluationen ebenfalls zu berücksichtigen sind.

Evaluationen, die die genannten Aspekte erfüllen, liefern evidenzbasierte Erkenntnisse über die Qualität und den Nutzen von Konzepten der Sprachbildung und -förderung. Dadurch bieten sie die Chance, die Akzeptanz neu eingeführter Maßnahmen zu fördern, Erkenntnisse über Weiterentwicklungsmöglichkeiten zu generieren und Erfolge sichtbar zu machen sowie evidenzbasierte Entscheidungen zu treffen und Vorgehensweisen zu legitimieren.

5. Chancen und Herausforderungen von Evaluationen

Evaluationen tragen erheblich dazu bei, die Qualität von Konzepten der Sprachbildung und -förderung sowie die Forschungsinfrastruktur in diesem Bereich (z. B. durch die Validierung von Erhebungsverfahren) zu verbessern (Gentrup et al., 2021). Gleichzeitig sind die Anzahl vorliegender Studien zur (fachintegrierten) Sprachbildung und -förderung sowie deren Belastbarkeit und Aussagekraft mitunter aber erheblich eingeschränkt (Kalinowski et al., 2020). Denn häufig werden lediglich formative Prozessevaluationen durchgeführt oder die Wirksamkeit wird nur auf den ersten drei der oben genannten vier Wirksamkeitsebenen untersucht, sodass vor allem Befunde zur Effektivität bei Schüler:innen oftmals fehlen (Gentrup et al., 2021). Ein Grund für die unbefriedigende Befundlage ist, dass zur Prüfung von Wirksamkeitshypothesen hohe wissenschaftliche Standards erfüllt sein müssen, die unter alltagspraktischen Bedingungen in der Schule häufig nicht oder nur eingeschränkt gegeben sind. Zum Beispiel ist es aus ethischen aber auch aus praktischen Gründen im schulischen Kontext mitunter nicht möglich, ein experimentelles Design zu realisieren, bei denen die Schüler:innen zufällig einer Interventions- und einer Kontrollgruppe zugewiesen werden. Die Belastbarkeit der Erkenntnisse ist dadurch auf den lokalen schulischen Kontext begrenzt, was die Entscheidungs- und Legitimationsfunktion der Evaluation einschränkt.

Zwar sind Evaluationen von (neuen) pädagogischen Konzepten in der schulischen Praxis in Deutschland in einigen Ländern sogar im Schulgesetz verankert und verpflichtend vorgesehen, allerdings greifen wiederholte Erhebungen über einen längeren Zeitraum teilweise stark in den schulischen Alltag ein. Von Lehrkräften und Schulleitungen werden Evaluationen deshalb häufig eher als Störfaktor wahrgenommen. Insbesondere wenn Wissenschaftler:innen Konzepte evaluieren, an deren Durchführung sich die Schulen freiwillig beteiligen sollen, muss der Mehrwert der Teilnahme für Schulen, Lehrkräfte und Schüler:innen sowie Eltern klar erkennbar sein und gut kommuniziert werden. Trotz vielfältiger Rekrutierungsbemühungen und -anreize wird die benötigte Stichprobengröße für die angemessene Überprüfung eines Konzepts in komplexen Prä-Post-Follow-up Kontrollgruppendesigns nicht immer erreicht. Oder es kommt im Verlauf der Studie, insbesondere nach Abschluss einer Fortbildung an die sich mehrere Da-

tenerhebungen anschließen, zu einem (selektiven) Dropout der Teilnehmenden. Dies kann die Belastbarkeit der Ergebnisse deutlich einschränken oder die Prüfung der Wirksamkeit sogar unmöglich machen. Bei ausbleibenden Effekten kann zudem oft nicht genauer geklärt werden, ob dies darauf zurückzuführen ist, dass das Konzept unwirksam ist oder die Umsetzung nicht in der intendierten Weise erfolgte. Folglich ist auch die Überprüfung der Implementationsqualität – insbesondere bezogen auf Vollständigkeit und Wiedergabetreue bei der Umsetzung der Maßnahme – ein weiteres wichtiges Qualitätsmerkmal von Evaluationsvorhaben, um die beobachteten Effekte oder deren Ausbleiben nachzuvollziehen.

6. Forschungsdesiderata und Ausblick

Ähnlich wie bei Evaluationen anderer pädagogischer Konzepte sind diese auch im Bereich der Sprachbildung und -förderung mit zwei zentralen Herausforderungen konfrontiert: Zum einen ist es kaum möglich, die hohen wissenschaftlichen Standards, insbesondere für die Prüfung von Wirksamkeitshypothesen, aufgrund der Komplexität von Lehr-Lernprozessen in der schulischen Praxis ohne Einschränkungen zu erfüllen. Die Belastbarkeit der Ergebnisse im Sinne der internen Validität ist somit oft eingeschränkt. Zum anderen untersuchen Evaluationen im Bereich der Sprachbildung und -förderung in der Regel die prinzipielle Wirksamkeit. Die Ergebnisse zeigen also beispielsweise, dass ein Fortbildungs- und Förderkonzept zur fachintegrierten Sprachförderung im Sachunterricht zur Entwicklung des bildungssprachlichen Wortschatzes bei Schüler:innen im Verlauf der dritten Klassenstufe beiträgt. Das Gesamtkonzept wird dabei als wirksam bewertet, welche Komponenten und Mechanismen (z. B. das sprachförderliche Frageverhalten der Lehrkräfte, Fördermaterialien), diese Wirkung evoziert haben, bleibt aber unbekannt. Wenngleich die Untersuchung von Wirkmechanismen nicht das primäre Ziel von Evaluationen ist, erschwert die fehlende Befundlage den Transfer evidenzbasierter Konzepte auf veränderte schulische Bedingungen. Da der „wirksame Kern“ des Konzepts nicht isoliert wurde, bleibt unklar, welche Komponenten an veränderte Kontexte angepasst werden können und welche die Wirkung ausmachen und deshalb konzepttreu umzusetzen sind. Die Anpassung wirksamer Konzepte an veränderte Bedingungen erfordert also eigentlich eine erneute Evaluation, um die Robustheit des Konzepts in diesen veränderten Kontexten zu prüfen. Aufgrund des großen Transferbedarfs von Konzepten der Sprachbildung und -förderung wäre es wünschenswert, auch Wirkmechanismen der jeweiligen Konzepte in Evaluationsstudien genauer zu untersuchen. Andernfalls sollte der Transfer wirksamer Konzepte auf veränderte schulische Bedingungen jeweils durch eine externe Fremdevaluation begleitet werden, die idealerweise sowohl formative als auch summative Komponenten in einem längsschnittlichen Kontrollgruppendesign umfasst.

Literatur

- Archie, C., Rank, A. & Franz, U. (2014). Sprachbildung im und durch Sachunterricht. In A. Hartinger & K. Lange (Hrsg.), *Sachunterricht: Didaktik der Grundschule* (S. 198–206). Berlin: Cornelsen.
- Bortz, J. & Döring, N. (2016). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (5. Aufl.). Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32 (1), 9–13. <https://doi.org/10.3102/0013189X032001009>
- DeGEval – Deutsche Gesellschaft für Evaluation (2016). *Standards für Evaluation. Erste Revision 2016*. Mainz: DeGEval.
- Gentrup, S., Henschel, S., Beck, L. & Stanat, P. (Hrsg.). (2021). *Sprach- und Schriftsprachförderung wirksam gestalten: Evaluation umgesetzter Konzepte*. Stuttgart: Kohlhammer.
- Gibbons, P. (2002). *Scaffolding language. Scaffolding learning. Teaching second language learners in the mainstream classroom*. Portsmouth: Heinemann.
- Giel, S., Klockgether, K. & Mäder, S. (2018). Möglichkeiten der Selbstevaluation in der Initiative Bildung durch Sprache und Schrift. In C. Titz, S. Weber, A. Ropeter, S. Geyer & M. Hasselhorn (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung umsetzen und überprüfen* (S. 93–114). Stuttgart: Kohlhammer.
- Hardy, I., Hettmannsperger, R. & Gabler, K. (2019). Sprachliche Bildung im Fachunterricht: Theoretische Grundlagen und Förderansätze. In J. Ziehm, B. V. Cornelli, B. Menzel & M. Goßmann (Hrsg.), *Schule migrationssensibel gestalten: Impulse für die Praxis* (S. 31–61). Weinheim: Beltz.
- Henschel, S., Heppt, B., Hardy, I., Mannel, S. & Stanat, P. (2020). *Professionalisierungsmaßnahmen zur bedeutungsfokussierten Sprachförderung im Sachunterricht der Grundschule „ProSach“ (Verbundvorhaben): Forschungs- und Entwicklungsvorhaben im Rahmen der Bund-Länder-Initiative „Bildung durch Sprache und Schrift“. Schlussbericht 2020*. Berlin: Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen.
- Heppt, B., Henschel, S., Hardy, I., Hettmannsperger-Lippolt, R., Gabler, K., Sontag, C., Mannel, S. & Stanat, P. (2022). Professional development for language support in science classrooms: Evaluating effects for elementary school teachers. *Teaching and Teacher Education*, 109, 103518. <https://doi.org/10.1016/j.tate.2021.103518>
- Kalinowski, E., Egert, F., Gronostaj, A. & Vock, M. (2020). Professional development on fostering students' academic language proficiency across the curriculum – A meta-analysis of its impact on teachers' cognition and teaching practices. *Teaching and Teacher Education*, 88, 1–15. <https://doi.org/10.1016/j.tate.2019.102971>
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2015). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Beschluss der 350. Kultusministerkonferenz vom 11.06.2015*. Verfügbar unter: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf
- Lipowsky, F. (2010). Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. In F. H. Müller (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 51–70). Münster: Waxmann.

- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58 (2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch* (S. 102–128). Bern: Huber.
- Rank, A., Deml, I., Lenske, G., Merkert, A., Binder, K., Schulte, M., Schilcher, A., Wildemann, A. & Bien-Miller, L. (2021). Eva-Prim: Evaluation von Sprachförderkompetenz und (bildungs)sprachlichen Leistungen von Schülerinnen und Schülern in Mathematik. In S. Gentrup, S. Henschel, K. Schotte, L. Beck & P. Stanat (Hrsg.), *Sprach- und Schriftsprachförderung gestalten: Evaluation von Qualität und Wirksamkeit umgesetzter Konzepte* (S. 105–124). Stuttgart: Kohlhammer.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Schmitz, A., Zeuch, N., Karstens, F., Meudt, S.-I., Jost, J. & Souvignier, E. (2021). Leseförderung im Schul- und Unterrichtsalltag implementieren – Erste Erkenntnisse des Evaluationsprojekts BiSS-EiLe. In S. Gentrup, S. Henschel, K. Schotte, L. Beck & P. Stanat (Hrsg.), *Sprach- und Schriftsprachförderung wirksam gestalten: Evaluation umgesetzter Konzepte* (S. 201–222). Stuttgart: Kohlhammer.
- Schneider, W., Baumert, J., Becker-Mrotzek, M., Hasselhorn, M., Kammermeyer, G., Rauschenbach, T., Roßbach, H.-G., Roth, H.-J., Rothweiler, M. & Stanat, P. (2012). Expertise „Bildung durch Sprache und Schrift (BiSS)“: Bund-Länder-Initiative zur Sprachförderung, Sprachdiagnostik und Leseförderung. Verfügbar unter: <https://www.biss-sprachbildung.de/pdf/biss-website-biss-expertise.pdf>.
- Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 60–91). München: R. Pieper & Co. Verlag.
- Stockmann, R. & Meyer, W. (2010). *Evaluation. Eine Einführung*. Opladen: Barbara Budrich.
- Titz, C., Weber, S., Wagner, H., Ropeter, A., Geyer, S. & Hasselhorn, M. (Hrsg.). (2020). *Bildung durch Sprache und Schrift; Band 4: Sprach- und Schriftsprachförderung wirksam gestalten: Innovative Konzepte und Forschungsimpulse*. Stuttgart: Kohlhammer.
- van de Pol, J., Volman, M., Oort, F. & Beishuizen, J. (2015). The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instructional Science*, 43 (5), 615–641. <https://doi.org/10.1007/s11251-015-9351-z>
- Watras, J. (2006). The Eight-Year Study: From evaluative research to a demonstration project, 1930–1940. *Education Policy Analysis Archives*, 14. <https://doi.org/10.14507/epaa.v14n21.2006>
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Bern: Hans Huber.