

### Dokumentarische Metaevaluation als empirischer Zugang zu einer Reflexion der Wissenschaftspraxis evidenzbasierter Bildungsforschung? Überlegungen am Beispiel der didaktischen Interventionsstudie Serelisk

Werner, Silke

Erstveröffentlichung / Primary Publication

Sammelwerksbeitrag / collection article

#### Empfohlene Zitierung / Suggested Citation:

Werner, S. (2022). Dokumentarische Metaevaluation als empirischer Zugang zu einer Reflexion der Wissenschaftspraxis evidenzbasierter Bildungsforschung? Überlegungen am Beispiel der didaktischen Interventionsstudie Serelisk. In S. Hoffmann, D. Klinge, D. Petersen, & S. Rundel (Hrsg.), *Jahrbuch Dokumentarische Methode. Heft 5/2022* (S. 177-199). Berlin: centrum für qualitative evaluations- und sozialforschung e.V. (ces). <https://doi.org/10.21241/ssoar.85962>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-SA Lizenz (Namensnennung-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-sa/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY-SA Licence (Attribution-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-sa/4.0>

# Dokumentarische Metaevaluation als empirischer Zugang zu einer Reflexion der Wissenschaftspraxis evidenzbasierter Bildungsforschung? Überlegungen am Beispiel der didaktischen Interventionsstudie *Serelisk*.

## 1 Die didaktische Interventionsstudie *Serelisk* als Ausgangspunkt der Überlegungen

Af: [...] ich bin bei dieser- (.) bei diesen Methoden geliebt (1) äh auch selbst=reflektieren (.) dann (.) eben auch in der Gruppenarbeit miteinander arbeiten (.) Problemlöseaufgaben und=so=weiter (.) und äh jetzt sagen sie (.) „ja jetzt ähm: (1) verstehen wir warum sie damals ebe so konsequent ware“ ne? [...] jetzt merke se au das sie mit diesen äh m=m jetzt schon anstehenden Prüfungsaufgaben doch recht gut klarkommen (Francke, GD, Z. 292–296).

Diese Aussage der Lehrerin Frau Azur (Af) entstammt einer Gruppendiskussion (GD) zwischen ihr und einer weiteren Lehrerin der Realschule Francke<sup>1</sup>. Zusammen mit einer dritten Lehrerin hatten sie sich drei Jahre zuvor zur gemeinsamen Teilnahme an einer einjährigen Intervention im Rahmen der *didaktischen Interventionsstudie Serelisk –selbstreflexives Lernen im schulischen Kontext* bereiterklärt, die an Realschulen in Baden-Württemberg von 2007 bis 2009 von der Baden-Württemberg Stiftung gefördert und an der Pädagogischen Hochschule Freiburg unter der Leitung von Katharina Maag Merki, Hans-Georg Kotthoff und Alfred Holzbrecher durchgeführt wurde. Angestrebt

1 Die Personennamen wurden via Farben, die Namen der schulischen Kooperationsgruppen nach historischen Pädagogen pseudonymisiert.

war, die Lehrpersonen zu befähigen, den Mathematikunterricht der 7. und 8. Klassen in kollegialer Teamarbeit so zu gestalten, dass eine Entwicklung des selbstregulierten Lernens<sup>2</sup> bei den Schüler\*innen erfolgt. Die hierfür herangezogenen didaktischen Mittel waren *selbstaktivierende Unterrichtsmethoden* (vgl. Barzel et al. 2007) und *Problemlöseaufgaben* (vgl. Büchter und Leuders 2005),<sup>3</sup> die an zwei Interventionsveranstaltungen vermittelt wurden. Zudem wurden die Lehrpersonen im Rahmen dieser Veranstaltungen zum Aufbau professioneller Lerngemeinschaften (PLG)<sup>4</sup> angeregt, welche die Implementation der didaktischen Mittel unterstützen sollten. Ein solches Vorgehen wird als *kooperative Implementationsstrategie* bezeichnet (vgl. Werner 2022).

Das Zitat von Frau Azur oben verdeutlicht, dass die didaktischen Mittel für Frau Azur anschlussfähig wurden und zu einer intendierten Veränderung ihres Unterrichts sowie zu einer Kompetenzentwicklung der Schüler\*innen geführt haben. Aus dieser subjektiven Perspektive von Frau Azur erscheint die Intervention wirksam. Allerdings kann diese individuelle Erfahrung nicht generalisiert werden. Die didaktische Interventionsforschung, in deren Kontext diese Intervention entwickelt wurde, kommt, wie im Anschluss gezeigt wird, zu einer ähnlichen generalisierungsfähigen Einschätzung.

Bei der *didaktischen Interventionsforschung* handelt es sich um eine *spezifische Form empirischer Unterrichtsforschung*. Ziel dieser ist es, wissenschaftliches Wissen hinsichtlich der Effektivität neuer Formen des Lehrens und Lernens zu generieren. *Didaktische Interventionen* sind wiederum Programme, die auf eine Optimierung der Lehr- und Lernprozesse im Unterricht und damit auf eine Verbesserung der Lernergebnisse von Schüler\*innen ausgerichtet sind. Um etwaige intendierte Effekte zu evaluieren, wird vor allem mit Methoden der pädagogischen Psychologie – im Fall der *Serelisk*-Studie mit Leistungstests und standardisierten Fragebögen – gearbeitet, die beispielsweise und so auch hier in quasiexperimentellen Forschungsdesigns zum Einsatz kommen und vorrangig mittels statistischer Verfahren ausgewertet werden (vgl. Hascher und Schmitz 2010).

Ein Auszug aus den Ergebnissen der statistischen Analyse von *Serelisk*, die auf

- 2 Selbstreguliertes Lernen stellt eine Komponente der überfachlichen Kompetenzen in den Bildungsplänen in Baden-Württemberg dar. Die Zielniveaustufe wird als das „planmäßige Bearbeiten komplexer Gegebenheiten“ angegeben, „um selbstständig zu Lösungen, Begründungen, Folgerungen, Interpretationen und Wertungen zu gelangen“ (Landesinstitut für Schulentwicklung 2009, S. 3).
- 3 Da keine Erläuterung der didaktischen Mittel für ein Verständnis des Artikels erforderlich ist, wird im Folgenden hierauf verzichtet.
- 4 Als PLG wird eine Form kooperativen Arbeitens von Lehrpersonen verstanden; hierzu gehören unterrichtsbezogene Kooperationen hinsichtlich Planung, Vorbereitung und Auswertung von Unterrichtsstunden, gegenseitige Unterrichtsbesuche (Hospitationen), reflektierte Gespräche über Unterrichtseinheiten und Mentoring (vgl. Bensen et al. 2013, S. 109 f.).

den Datensätzen von 64 Lehrpersonen<sup>5</sup> und 1.598 Schüler\*innen<sup>6</sup> basieren, verdeutlicht, dass sich die *kooperativen Aktivitäten zur Unterrichtsgestaltung* der Lehrpersonen der Experimentalgruppe<sup>7</sup> während des Projektverlaufs gesteigert haben, und zwar von im Mittel „ca. 1-2-mal halbjährlich“ (Mittelwert: 2.1) auf annähernd „ca. einmal monatlich“ (Mittelwert: 2.78)<sup>8</sup> (Maag Merki et al. 2009, S. 52f.). In der Kontrollgruppe zeigten sich keine relevanten Entwicklungen. Zwar wird im Forscher\*innenteam angemerkt, dass man von einer höheren Intensitätssteigerung ausgegangen sei (vgl. Werner 2012, S. 138), dennoch werten sie die Ergebnisse so, dass ein Aufbau von PLG „auf den Weg gebracht werden konnte“ (Maag Merki et al. 2009, S. 109). Zudem waren drei Viertel der Lehrpersonen der Experimentalgruppe zum Ende des Projekts davon überzeugt, dass die Arbeit in der Kooperationsgruppe einen Effekt auf die eigene Unterrichtsqualität und das Lernen der Schüler\*innen gehabt habe und somit wirksam war. Des Weiteren ließ sich eine positive Entwicklung der Einstellung zu planmäßigen Formen von Kooperation<sup>9</sup> im Projektverlauf und im Vergleich der Untersuchungsgruppen zueinander feststellen. Insgesamt wurden die Ergebnisse zum Implementationsmittel *Kooperation* von den Forscher\*innen dahingehend interpretiert, dass sich „die praktische Kooperation im Rahmen des Projekts bewährt, d. h. als gewinnbringend erwiesen hat“ (Maag Merki et al. 2009, S. 109).

Hinsichtlich der *Wirkungen der didaktischen Intervention auf die Unterrichtsgestaltung* ist festzuhalten, dass es nicht gelang, den eigentlich von den Forscher\*innen intendierten Effekt, nämlich die Etablierung eines „offen-entdeckenden Unterrichts“<sup>10</sup> (Maag Merki et al. 2009, S. 67) für die

- 5 Davon gehörten 36 Lehrpersonen der Experimentalgruppe und 28 Lehrpersonen der Kontrollgruppe an.
- 6 Davon gehörten 924 Schüler\*innen zur Experimentalgruppe und 674 Schüler\*innen zur Kontrollgruppe.
- 7 Als Experimentalgruppe werden alle Lehrpersonen der Studie bezeichnet, die an den didaktischen Interventionen teilgenommen haben und für die jeweils schulische Kooperationsgruppen gebildet wurden. Die Kontrollgruppe bezeichnet jene Lehrpersonen der Studie, die ausschließlich in die Fragebogenbefragungen involviert waren und ihren Unterricht entsprechend wie bisher weitergestalten sollten.
- 8 Eine exemplarische Aussage im Fragebogen hierzu war: „Wir tauschten Mathematikaufgaben untereinander aus.“; die dazugehörigen Antwortmöglichkeiten reichten von 1 – „ca. 0-1 mal halbjährlich“ bis 5 – „ca. einmal täglich“ (Maag Merki et al. 2009, S. 52).
- 9 Ein exemplarisches Item hierzu lautete: „MathematiklehrerInnen sollten sich regelmäßig wechselseitig im Unterricht besuchen, um voneinander zu lernen.“; die zugehörigen Antwortkategorien reichten von 1 – „trifft gar nicht zu“ bis 4 – „trifft genau zu“ (Maag Merki et al. 2009, S. 49f.).
- 10 Die Autor\*innen verstehen hierunter einen Unterricht der die individuellen Lernvoraussetzungen und Lernbedingungen der Schüler\*innen berücksichtigt. Die Unterrichtsgestaltung weist folgende Merkmale auf: individuelle Lösungswege sind möglich; die Lehrperson hält sich bei fachlichen Diskussionen zurück; Schüler\*innen können Aufga-

Schüler\*innen wahrnehmbar nachzuweisen. Die Einstellung der Lehrpersonen zu ebendieser Unterrichtsform entwickelte sich im Gegensatz zu denjenigen aus der Kontrollgruppe jedoch tendenziell positiv. Zudem verloren deduktiv-schematische Unterrichtselemente<sup>11</sup> im Projektverlauf für die Lehrpersonen der Experimentalgruppe stärker an Bedeutung als für Lehrpersonen der Kontrollgruppe. Diese Entwicklung und die Gruppenunterschiede ließen sich auch für die Wahrnehmung der Unterrichtsgestaltung durch die Schüler\*innen abbilden. Auch wenn diese Ergebnisse nicht ganz den Erwartungen entsprachen, wurden sie als Hinweis gedeutet, dass die Intervention „wirksam ist und wichtige Veränderungen angestoßen worden sind“ (Maag Merki et al. 2009, S. 110). Direkte Wirkungen der Interventionen auf die Kompetenzen der Schüler\*innen ließen sich nicht oder nur in der Tendenz aufzeigen. Plausibilisiert wurde dieser ausbleibende Effekt durch die nur einjährige Implementationsphase, die es möglicherweise nur zuließ, „Einstellungen der Lehrpersonen projektintendiert zu verändern und, auch für die SchülerInnen sichtbar, deduktiv-schematische Unterrichtselemente bei der Gestaltung des Unterrichts abzubauen“ (Maag Merki et al. 2009, S. 110).

Die Forscher\*innen kommen entsprechend zu einer positiven Bilanz der didaktischen Intervention und der Implementationsstrategie im Sinne eines *it works*. Auch in ähnlich angelegten didaktischen Interventionsstudien wie *SI-NUS* (vgl. u. a. Ostermeier 2004) oder *Chemie im Kontext* (vgl. u. a. Gräsel et al. 2006) ist es gemäß den Forscher\*innen gelungen, kooperative Prozesse anzustoßen sowie wahrgenommene Veränderungen auf der Unterrichtsebene nachzuweisen; auch hier ziehen sie eine positive Wirksamkeitsbilanz. Was allen Studien gemein ist, ist, dass in jeder einzelnen auch nichtintendierte Entwicklungen und Effekte festgestellt wurden, die eine zumindest differenziertere Bewertung der Wirksamkeit zuließen (vgl. ausführlicher hierzu Werner 2022, S. 98 ff.).

*Was ist hier bei der Interpretation passiert, dass eine It-works-Bilanz gezogen wird?* Eine Reflexion der Plausibilisierungen nichtintendierter Effekte bei *Serelisk* verdeutlicht, dass unter anderem auf die *Zukunft* verwiesen wird, da davon ausgegangen wird, dass der gewünschte Entwicklungsprozess „auf den Weg gebracht werden konnte“ oder „wichtige Veränderungen angestoßen

ben modifizieren; Schüler\*innen können gegenseitig ihre Lösungsansätze begründen (vgl. Maag Merki et al. 2009, S. 67). Eine Beispielaussage zur Erfassung dieser Unterrichtsform im Fragebogen lautete: „SchülerInnen erarbeiten individuelle Lösungswege, die bisweilen auch stark von Musterlösungen abweichen“; die entsprechenden Antwortitems reichten von 1 – „trifft gar nicht zu“ bis 4 – „trifft genau zu“ (Maag Merki et al. 2009, S. 32).

11 Eine entsprechende Aussage hierzu lautete zum Beispiel: „SchülerInnen orientieren sich bei der Aufgabenbearbeitung vorwiegend an Musterlösungen“; die dazugehörigen Antwortitems reichten von 1 – „trifft gar nicht zu“ bis 4 – „trifft genau zu“ (Maag Merki et al. 2009, S. 32).

worden sind“ (Maag Merki et al. 2009, S. 109.). Diese Prognosen werden aber in der Regel nicht mehr empirisch überprüft (vgl. Emmerich und Werner 2013). Zudem werden *Begründungen in Form von Theorien konstruiert*, beispielsweise indem, wie bereits erwähnt, die einjährige Interventionszeit nachträglich als zu gering markiert und die ursprüngliche Hypothese nachjustiert wird. Eigentlich handelt es sich also hierbei um eine induktive Schlussfolgerung, die darauf basiert, dass sich die gewünschte Wirkung nicht eingestellt hat. Was dann allerdings passiert, ist, dass die Anschlusshypothese ohne Überprüfung (prognostisch) angenommen und die Intervention als wirksam ausgelegt und entsprechend kommuniziert wird.

Was im Fall *Serelisk* eine gewisse Ausnahme darstellt, ist, dass diese Prognose statistisch überprüft wurde. Und so zeigte sich zwei Jahre nach dem Projekt, dass lediglich fünf Lehrpersonen (aus unterschiedlichen Teams!) einschätzen, dass ihre zu diesem Zeitpunkt stattfindenden unterrichtbezogenen kooperativen Aktivitäten in einem kausalen Zusammenhang mit ihrer Teilnahme an der Intervention stehen. Auch das selbstregulierte Lernen spielte bei den Lehrpersonen sowohl der Experimental- als auch der Kontrollgruppe nur noch eine untergeordnete Rolle. Die Conclusio, dass ‚eine Entwicklung auf den Weg gebracht wurde‘, musste und muss – zumindest als generalisierungsfähiges Ergebnis – ex post abgelehnt werden (vgl. Kotthoff und Werner 2011, S. 36).

Gleichzeitig erkennen wir, dass auch standardisierte Nachhaltigkeitsuntersuchungen bei ausbleibender Wirkung nicht das offenbar bestehende *Reflexionsproblem* didaktischer Interventionsforschung durchbrechen können. Denn wir stehen wieder vor der Frage: Warum hat es nicht gewirkt? Oder aber: Hat es vielleicht gewirkt, ohne dass das Instrumentarium die Wirkung erfassen konnte? In der Interventionsforschung und mit ihren Methoden können nicht-intendierte Effekte nur „als Folge eines ‚Beta-Fehlers‘ interpretiert werden“ (Emmerich und Werner 2013, S. 172). Dieser lässt sich empirisch jedoch nicht ‚beweisen‘ und berührt damit auch nicht die grundlegende Frage der Angemessenheit der Methoden (vgl. Emmerich und Werner 2013, S. 172). Stattdessen werden neue Plausibilisierungen generiert, die in immer neue Hypothesen überführbar und testbar sind. Daraus ergibt sich ein Diskussionsstrang der Kritik an evidenzbasierter Bildungsforschung.

Dieser Artikel knüpft hier an und wird im Kapitel 2 auf die Notwendigkeit einer Reflexion der Wissenschaftspraxis evidenzbasierter Bildungsforschung eingehen. In Kapitel 3 werden dann Überlegungen skizziert, wie ein reflexiver Zugang zur Wissenschaftspraxis durch eine dokumentarische Metaevaluation erfolgen könnte. Sodann werden in Kapitel 4 exemplarische Ergebnisse einer solche Analyse im Kontext des Projekts *Serelisk* vorgestellt. Der Artikel schließt mit einem kritischen Ausblick.

## 2 Die Wissenschaftspraxis evidenzbasierter Bildungsforschung in der Kritik

Die oben skizzierte Art der Forschung, die „ein experimentell gesichertes und generalisierbares Wissen über die Wirksamkeit pädagogischer Technologien und bildungspolitischer Maßnahmen“ (Bellmann und Müller 2011, S. 9) verspricht, wird spätestens seit PISA und TIMSS stark gefördert und gefordert, verbunden mit politischen Steuerungserwartungen und auf die Praxis ausgerichteten Rationalisierungshoffnungen (Meseth 2011, S. 12 f.). Es wird entsprechend nicht nur ein Wissen über Wirkungszusammenhänge in Aussicht gestellt, also ein „*Wissen, über das was wirkt*“, sondern auch dessen Übertragbarkeit in die Bildungspraxis und Bildungspolitik, also ein „*Wissen, das wirkt*“ (Bellmann und Müller 2011, S. 9; Herv. i. O.; vgl. Leutner 2010).

Gesprochen wird an dieser Stelle auch von „evidenzbasierter Bildungsforschung“ (Jornitz 2009) und „evidenzbasierter Steuerung“ (Thiel et al. 2013). Anders als das Wort *Evidenz* in seiner deutschen Übersetzung vermuten lässt, meint *evidence* ursprünglich nicht, dass etwas unmittelbar auf der Hand liegt (vgl. Jornitz 2009, S. 68), „sondern die durch empirische Forschungsmethoden gewonnenen Belege oder Hinweise, die für die Wirksamkeit einer bestimmten Methode oder Maßnahme sprechen“ (Bellmann und Müller 2011, S. 11). In der bildungspolitischen und erziehungswissenschaftlichen Praxis gestalten sich das Verständnis und die Praxis der Rezeption jedoch etwas anders. Johannes Bellmann und Thomas Müller (2011) beschreiben, dass zumeist von einem gesicherten Wissen ausgegangen wird, das als Grundlage für praktisches Handeln dienen kann und Hinweise häufig als Beweise verstanden werden (S. 12). Damit liefert die *evidenzbasierte Bildungsforschung* der Bildungspolitik und -praxis, die durch „Evidenzmangel und Handlungszwang“ (Blumenberg 2021, S. 417) gekennzeichnet ist, eine nachgefragte „Ersatzevidenz“ (Bellmann und Müller 2011, S. 13).

Zudem arbeiten sie heraus, dass evidenzbasierte Forschung mit wissenschaftlichem Wissen assoziiert wird, während andere Formen der Forschung als „bloße Meinung, Ideologie, Intuition oder Vorurteil“ gegenübergestellt werden (Bellmann und Müller 2011, S. 13). Somit kommt es zu einer Hierarchisierung dessen, „was Wissenschaft überhaupt ausmacht“ (Bellmann und Müller 2011, S. 14; vgl. ähnlich auch Böttcher et al. 2013). Um dies kenntlich zu machen, weisen die Autoren „evidenzbasierte Pädagogik“ als Paradigma aus. Sie charakterisieren es darüber, dass (1) *evidence* auf ein wissenschaftliches Wissen bezogen wird, das durch „privilegierte wissenschaftliche Methoden, insbesondere randomisierte kontrollierte Studien“, gewonnen wurde; dass (2) *evidence* immer als ein „Wissen über kausale Zusammenhänge“ gedacht wird; dass (3) davon ausgegangen wird, dass sich das Wissen über Kausalzusammenhänge nutzen lässt, um bewusst Effekte in der Praxis herbeizuführen

(Bellmann und Müller 2011, S. 14 f.). Bellmann und Müller kritisieren ein solches Paradigma und sprechen sich demgegenüber *für* eine Pädagogik aus, die sich aus unterschiedlichen Methoden und Disziplinen „informiert“ und „komplexere Verhältnisbestimmungen von wissenschaftlichem Wissen und pädagogischer Praxis“ ermöglicht (Bellmann und Müller 2011, S. 15). Illustrieren lässt sich ein dahingehender Mangel durch Sieglinde Jornitz (2009). Sie benennt den sehr hohen Stellenwert von Metaanalysen in der evidenzbasierten Bildungsforschung (vgl. Jornitz 2009, S. 70). Bei einer Reflexion solcher Arbeiten stellte sie unter anderem fest, dass beispielsweise von der vollständigen Sichtung aller Studien die Rede sei, gleichzeitig aber nur quantitative Studien berücksichtigt worden seien (vgl. Jornitz 2009, S. 72 ff.; für eine positivere Bewertung vgl. Böttcher et al. 2013). Zudem zeigt sie an Beispielen auf, dass das wissenschaftliche Wissen, das aus diesen Studien erwachse, an den „common sense“ anknüpfe, aber „keine erschließende Erkenntnis“ darüber liefere, was in der Praxis geschehe (Jornitz 2009, S. 74). Gleichzeitig, so Jornitz' Kritik, werde in der Praxis evidenzbasierter Bildungsforschung „gar kein wissenschaftlicher Diskurs mehr geführt, welche Methode zu welchem Ergebnis führt, sondern ein Teil der Wissenschaft bleibt von Beginn unberücksichtigt“ (Jornitz 2009, S. 70). Daraus resultiert eine Lenkung und Wertung der Art wissenschaftlichen Wissens in der Gesellschaft (vgl. Jornitz 2009) und führt zu einem weiteren Kritikpunkt.

Es wird ein Wechselseitigkeitsverhältnis von Bildungsforschung und Bildungspolitik ‚diagnostiziert‘. So werden „Forschungsförderprogramme eng an politische Steuerungserwartungen des Erziehungssystems gekoppelt“ und darauf ausgerichtet, „entscheidungsrelevante Daten für die Bildungspolitik und die Lehrerbildung zu generieren“ (Meseth 2011, S. 13). In der Konsequenz führt dies zum einen dazu, dass die Wissenschaft ihre Forschung nach diesen Bedarfen der Bildungspolitik als Geldgeberin ausrichtet (Wunsch nach Evidenz), und zum anderen richtet die Bildungspolitik ihre Praxis nach den Ergebnissen der Wissenschaft aus, indem sie diese als Evidenz interpretiert. Zudem macht „der wissenschaftliche Diskurs selbst von politisch-öffentlicher Rhetorik Gebrauch“, was als Ausdruck eines sogar symbiotischen Verhältnisses zwischen Bildungspolitik und Bildungsforschung verstanden werden kann (Bellmann und Müller 2011, S. 13).

Und gerade diese Art Wechselseitigkeitsverhältnis behindert einen *reflexive turn*, wie ihn Wolfgang Meseth (2011) fordert. Er betrachtet Erziehungswissenschaft „als widerstrebende Einheit von normativ-pädagogischer Reflexionstheorie und analytisch-deskriptiver Wissenschaft“ (Meseth 2011, S. 13). Während die Erziehungswissenschaft als Reflexionstheorie von der pädagogischen Praxis erwartetes Reflexionswissen zur Verfügung stelle bzw. angewandte Forschung betreibe, bestehe ihre Aufgabe als wissenschaftliche Disziplin darin, sich mit den Methoden und Theorien der Wahrheitsfindung auseinanderzusetzen bzw. sich der

Grundlagenforschung zu widmen. Die Ergebnisse und Überlegungen an dieser Stelle müssten „sich nicht selbstverständlich mit den normativen Erwartungen an eine pädagogische Reflexionstheorie decken“, sie könnten ihr auch widersprechen (Meseth 2011, S. 14 f.). Dabei gelte die Erziehungswissenschaft als „Integrationswissenschaft“, die sich aus Theorien und Methoden von Nachbardisziplinen bediene (Meseth 2011, S. 15). Meseth (2011, S. 15) vertritt die Ansicht, dass externe Theorieangebote „auf die normativen pädagogischen Interessen der Reflexionstheorie zugeschnitten“ und „dadurch tendenziell um ihr sozialwissenschaftliches Erkenntnispotential gebracht“ würden. Von ihm gefordert werden entsprechend „[g]rundlagentheoretische Debatten, die das Verhältnis von gegenstandstheoretischer Bestimmung und empirischer Erschließung eines sozialen Phänomens reflektieren“ und – vor allem in der empirischen Forschung zum Unterricht – nur randständig in Erscheinung treten, da „sie zur Bearbeitung pädagogisch-präskriptiver Fragen nicht nur nicht notwendig sind, sondern sogar hinderlich sein können“ (Meseth 2011, S. 16).

Was bei dieser sehr reduzierten Darlegung kritischer Diskurse zur evidenzbasierten Bildungsforschung auffällt, ist, dass sie theoretisch-reflexiv geführt werden und damit möglicherweise ‚eine andere Sprache sprechen‘ als ihre Adressat\*innen, die Vertreter\*innen evidenzbasierter Bildungsforschung und -politik. Hier stehen Wirkungen im Zentrum. Was an dieser Stelle möglicherweise fehlt, sind empirische Forschungen, die die Forschungen selbst nochmals in den Blick nehmen und somit eine „*Forschung 2. Ordnung*“ darstellen (Lambrecht und Rürup 2012, S. 71; Herv. i. O.). Was in Kapitel 1 aber bereits deutlich wurde, ist, dass die quantitativen Methoden bei Falsifizierung der Hypothesen an Erkenntnisgrenzen kommen. Gleichzeitig sind Wirkungen jenseits der Hypothesen aber denkbar sowie Antworten darauf, warum etwas (nicht) gewirkt hat. Darauf aufbauend könnte dann reflektiert werden, was die jeweiligen Interventionen (Beitrag zur Reflexionstheorie) sowie wissenschaftlichen Methoden (Beitrag zur Wissenschaftstheorie) jenseits der normativen Annahmen zu leisten imstande sind, um auch als Wissenschaft lernen zu können. Daher wäre hier ein Methodenwechsel angezeigt, mit dem die Erkenntnislücke geschlossen werden kann. Maike Lambrecht und Matthias Rürup (2012, S. 74) fordern im Kontext einer wissenschaftlichen Reflexion des Steuerungsinstruments Schulinspektion, das selbst als wissenschaftlicher Forschungsprozess angelegt ist, ebenfalls einen solchen Methodenwechsel, um zu einer umfassenderen Bewertung, also jenseits der Wirksamkeitsfrage zu gelangen und stattdessen Wirkungen – also auch jenseits der intendierten Wirkungen – einzuschließen. Sie fordern eine „stärkere Beachtung *qualitativer und rekonstruktiver Verfahren sowie dekonstruktiver Ansätze*“ (Lambrecht und Rürup 2012, S. 74; Herv. i. O.). Auch Meseth (2011, S. 14) plädiert „für eine Forschungsperspektive, die in Distanz zu den normativen Erwartungen der Praxis geht und

Unterricht als ein soziales Phänomen mit sozialwissenschaftlichen Mitteln aufzuklären versucht“.

Maike Lambrecht (2018), Daniel Goldmann (2017) sowie Sigrid Zeitler et al. (2012) legten unter anderem bereits Studien vor, die als exemplarisch für solch einen Methodenwechsel stehen können und einen dahingehend erweiternden Beitrag zur Reflexionstheorie leisten, also zur Reflexion der Frage, *was die jeweiligen Interventionen jenseits der normativen Annahmen zu leisten im Stande sind*. Vor allem die Studie von Zeitler et al. (2012) hätte in ihrer Anlage eventuell sogar darüber hinaus die Möglichkeit für eine Methodenreflexion und damit für einen Beitrag zur Wissenschaftstheorie geboten. Hier wurde in zwei methodisch unterschiedlich angelegten Teilstudien (standardisierte Verfahren und rekonstruktive Verfahren) die Einführung der Bildungsstandards evaluiert. Dass eine solche Reflexion nicht erfolgte, ist vermutlich der zusammenhängenden Projektanlage in zwei klar abgesteckte Teilprojekte mit je eigenem Verantwortungsbereich geschuldet und verweist möglicherweise auf den ‚Tabucharakter‘, der mit dieser selbstreflexiven Blickrichtung auf standardisierte Methoden einhergeht.

Eine Vorgehensweise, wie man zu evidenzbasierter Forschung einen empirisch-reflexiven Bezug herstellen und begründen kann, soll im nächsten Kapitel im Rahmen der Überlegungen zu einer dokumentarischen Metaevaluation skizziert werden.

### **3 Dokumentarische Metaevaluation als empirischer Zugang zu einer Reflexion der Wissenschaftspraxis didaktischer Interventionsforschung**

Die Überlegungen zu einer dokumentarischen Metaevaluation als reflexiven Zugang zur Wissenschaftspraxis evidenzbasierter Bildungsforschung werden nachfolgend am Beispiel der didaktischen Interventionsforschung skizziert. Hierfür ist zunächst zu klären: Was wird hier unter Metaevaluation verstanden? Warum eignet sich die Dokumentarische Methode als reflexiver Zugang zur Wissenschaftspraxis? Abschließend wird ein Modell dokumentarischer Metaevaluation vorgestellt.

Eine *Metaevaluation* wird hier in Anlehnung an Michael Scriven (2007, S. 2) als „evaluation of evaluations“ verstanden. Evidenzbasierte Bildungsforschung und exemplarisch die didaktische Interventionsforschung fokussiert auf eine Wirkungsevaluation, weshalb sie grundsätzlich den Evaluationsstudien zugeordnet werden kann. Es geht im Kern darum, die erwarteten Wirkungen zu evaluieren und auf dieser Grundlage die Intervention zu bewerten (vgl. Döring und Bortz 2016). Bei einer Metaevaluation wird auf eine übergeordnete Ebene gewechselt, um die Wissenschaftspraxis – hier exemplarisch die der didaktischen Interventionsforschung – in den Blick zu nehmen. Dabei ist

es möglich, sowohl (1) die *Methodologie* (z. B. Fragen nach den Möglichkeiten von Evaluationen in bildungspolitischen Prozessen) als auch (2) die *Qualitätsbewertung von Evaluationen* (z. B. Fragen nach der angemessenen Anwendung von Forschungsmethoden) in den Fokus zu rücken (vgl. Böttcher et al. 2013, S. 213).

Eine dokumentarische Metaevaluation bedient sich dabei der Methode der dokumentarischen Interpretation und folgt den methodologischen Ausarbeitungen der *dokumentarischen Evaluationsforschung* (vgl. Bohnsack und Nentwig-Gesemann 2010). Grundlagentheorie einer solchen Forschung bildet die praxeologische Wissenssoziologie. Dieser Bezug bietet sich auch deshalb an, da bereits auf eine kritische Auseinandersetzung mit der gängigen Praxis der Evaluationsforschung rekuriert werden kann.

Zentrale Argumentationslinien der kritischen Auseinandersetzung mit der traditionellen – oder von Ralf Bohnsack (2010) als interpretative Evaluationsforschung benannten – Evaluationsforschung<sup>12</sup> sollen im Folgenden in Gegenüberstellung zur rekonstruktiven Evaluationsforschung, zu der die dokumentarische Evaluationsforschung zählt, nachgezeichnet werden, um deren Eigenung für einen reflexiven Zugang nachvollziehbar zu machen.

Gemäß der praxeologischen Wissenssoziologie könne die Beziehung zwischen Theorie und Praxis, so Bohnsack (2010, S. 23), nicht auf die Beziehung von wissenschaftlichen Konstruktionen zu Common-Sense-Konstruktionen heruntergebrochen werden, auch deshalb nicht, weil Wissenschaft in eine eigene (Forschungs-)Praxis eingebunden sei. Er schlägt vor, zwischen einer theoretischen und einer praktischen Beziehung zur Welt zu unterscheiden. Die interpretative Evaluationsforschung folge dabei einer theoretischen Beziehung zur Welt. Es würden Theorien der Akteur\*innen *über* ihr Handeln erfasst, die Praxis des Handelns selbst bleibe unberücksichtigt. Die Aussagen der Akteur\*innen würden sodann „wörtlich“ genommen“ (Bohnsack 2010, S. 26). Mit dem damit einhergehenden analytischen Bezug auf Common-Sense-Theorien sei ein Verständnis „des zweckrationalen, utilitaristischen Handelns“ verknüpft (Bohnsack 2010, S. 26). Dieses Handlungsverständnis werde aber vor allem deshalb problematisiert, da gerade die Evaluationsforschung mit ihrem hohen Anspruch auf *Praxisrelevanz* „herausgefordert ist, das Verhältnis des Sozialwissenschaftlers zur Praxis der Akteure im Forschungsfeld zu klären“ (Bohnsack 2010, S. 26). Bohnsack (2010, S. 28) stellt hier mit der dokumentarischen Evaluationsforschung eine praxeologische Alternative zur Verfügung, die bei der Handlungspraxis der Erforschten ansetzt und einer „performativen Analyseeinstellung“ folgt. Dabei wird davon ausgegangen,

12 Hier lässt sich die zuvor als evidenzbasierte Bildungsforschung und damit auch die didaktische Interventionsforschung zuordnen.

dass uns nur die Rekonstruktion der Handlungspraxis, das heißt des Handlungsprozesses, des gesellschaftlichen, des interaktiven Prozesses, in dem der Handelnde zum Zeitpunkt des Handels existiert, zum Erfassen jener Bedeutung führt, die der jeweiligen Handlung oder Äußerung, dem jeweiligen Begriff, innerhalb dieses Handlungs- und Interaktionszusammenhanges zukommt (Bohnsack 2010, S. 28).

Demnach erschließe sich die Bedeutung einer Aussage erst, „wenn wir den existentiellen Zusammenhang, den Interaktionsprozess, der den Kontext der Äußerung bildet, beobachten oder wenn uns dieser Kontext – auf dem Wege von Beschreibungen und Erzählungen – mitgeteilt wird“ (Bohnsack 2010, S. 28). Hierüber werde ermöglicht, handlungsleitende Wertorientierungen – ein *implizites Wissen* – zu erschließen. Die Rekonstruktion erfolge über positive und negative Gegenhorizonte, die in Erzählungen und Beschreibungen eingelassen seien und die Werthorizonte einer Gruppe oder eines Milieus konstituieren (vgl. Bohnsack 2010, S. 31). Was im Rahmen einer dokumentarischen Evaluationsforschung rekonstruiert wird, entspricht somit nicht einer *Bewertung* im Sinne eines Urteils, sondern sie zielt auf die Rekonstruktion einer *Wertung*, die eine „Verwurzelung des Denkens im sozialen Raum“ (Mannheim 1952, S. 73) widerspiegelt. Zudem hätten „die EvaluatorInnen selbst sich dem Problem der Bewertung und Beurteilung zu stellen“, und es gelte, „die Werte derjenigen zu rekonstruieren, die Gegenstand der Evaluation“ seien (Bohnsack 2010, S. 30). Dabei sei die Unterscheidung in eine Beobachtung erster und zweiter Ordnung von Relevanz. „Die erkenntnistheoretische Differenz zwischen dem Beobachter erster und demjenigen zweiter Ordnung macht Luhmann daran fest, dass letzterer den ‚blinden Fleck‘, also den impliziten oder latenten Vergleichshorizont des Ersteren, zu identifizieren vermag“ (Bohnsack 2010, S. 33). Dies erfolge darüber, dass die Verwendung der Werte in der Kommunikation rekonstruiert werde, und zwar über einen analytischen Wechsel weg davon, *was* darüber kommuniziert werde, hin zu dem, *wie* darüber kommuniziert werde (Bohnsack 2010, S. 33). Der ‚blinde Fleck‘ der Forschenden – also ihre eigene Standortgebundenheit und Wertgebundenheit – könne jedoch auch hier nicht ohne Weiteres überwunden, so jedoch kontrolliert werden. Dies sei im Rahmen der komparativen Analyse der Dokumentarischen Methode „insoweit möglich, als an die Stelle der im (impliziten) Wissen des Interpreten fundierten Vergleichshorizonte empirisch überprüfbare Gegenhorizonte treten“: „Je mehr empirisch überprüfbare Fallvergleiche in die Analyse einbezogen werden, desto mehr werden die Vergleichshorizonte intersubjektiv überprüfbar, und die an den eigenen Standort gebundenen Vergleichshorizonte werden relativiert“ (beides Bohnsack 2010, S. 34). Dies hat dann zur Folge, dass ich als Forscher\*in während der Analyse auch mich selbst reflexiv als Gegenstand der Analyse in den Blick nehme und somit eine reflexive Haltung

inhärenter Teil einer dokumentarischen (Evaluations-)Forschung ist.<sup>13</sup> Dies ist auch der Grund, warum geschrieben wird, dass bei der dokumentarischen Evaluationsforschung die Grenzen zwischen angewandter Forschung und Grundlagenforschung zuweilen verschwimmen, wohingegen die interpretative Evaluationsforschung der angewandten Forschung zugeordnet wird (vgl. Bohnsack 2010). An dieser Stelle ergibt sich die Option, dass die dokumentarische Evaluationsforschung die interpretative Evaluationsforschung (zum Unterricht) bereichern kann, indem sie deren Wissenschaftspraxis reflexiv in den Blick nimmt.

Wie also kann eine *dokumentarische Metaevaluation* praktisch aufgebaut werden, damit eine empirische Reflexion der Wissenschaftspraxis didaktischer Interventionsforschung möglich ist? Dies soll am konkreten Beispiel der dokumentarischen Evaluation der ‚Evaluation‘ *Sereisk* beschrieben und in einem Modell illustriert werden (siehe Abb. 3.1).



Abbildung 3.1: Dokumentarische Metaevaluation didaktischer Interventionsforschung

13 Beispielsweise wird mir als Forscher\*in während der Analyse durch empirische Gegenhorizonte bewusst, dass ich ohne ebensolche Aussagen immer in eine bestimmte Richtung bzw. aus einer bestimmten Perspektive heraus interpretiert hätte. Damit rekonstruiere ich als Forscher\*in auch meine eigene Handlungspraxis und die hier zugrundeliegenden Handlungs- und Wertorientierungen mit.

- (1) Im Kontext *didaktischer Interventionsforschung* wird zunächst von einer *didaktischen Intervention* ausgegangen, die mehr oder weniger konkrete Ziele verfolgen und mehr oder weniger konkrete Mittel zur Erreichung ebendieser vorgeben kann. Im Fall der *Serelisk*-Intervention – zur Erinnerung – bestand das Ziel darin, das selbstregulierte Lernen der Schüler\*innen mittels Problemlöseaufgaben und selbstaktivierender Methoden zu fördern. Zudem sollten die Lehrpersonen einer Schule – gemäß einer PLG – eine kooperative Arbeitsweise aufbauen, welche die Implementation der didaktischen Mittel unterstützen sollte (siehe Kap. 1). Die didaktische Interventionsforschung fokussiert dann auf die Resultate der *Umsetzung der Intervention* in der *schulischen Handlungspraxis*. Die Bewertung der Umsetzung erfolgt dabei auf der Basis vorab definierter Wirkungsmaßstäbe (Wirksamkeit). In der *Serelisk*-Studie wurden die intendierten Wirkungen mittels standardisierter Fragebögen sowohl bei den Lehrpersonen der Experimental- und Kontrollgruppe als auch bei den Schüler\*innen der Experimental- und Kontrollgruppe, jeweils vor und nach Abschluss der Interventionsphase erfasst. Zusätzlich wurden bei den Schüler\*innen Leistungstests durchgeführt. Auf der Basis dieser Versuchsanordnung entstanden die in Kapitel 1 dargelegten Ergebnisse, die zu einer positiven Bewertung der Wirksamkeit vonseiten der Forscher\*innen führten und die kritisch besprochen wurden.
- (2) Dem kann eine *dokumentarische Evaluationsforschung*, wie sie ebenfalls durchgeführt wurde, gegenübergestellt werden. Eine solche nimmt in Bezug auf die *schulische Handlungspraxis der Umsetzung der Intervention* nicht nur die intendierten Wirkungen in den Blick, sondern auch solche, die als *transintentional* bezeichnet werden können (Wirkungen). Dies ist möglich, da die Datenerhebung bei der Handlungspraxis ansetzt, indem Erzählungen und Beschreibungen zu dieser erfasst werden. Für die dokumentarische Evaluation der *Serelisk*-Intervention konnte einmal auf *authentische Gruppengespräche* (AG; vgl. Przyborski und Wohlrab-Sahr 2014, S. 143 ff.) aus den beiden Interventionsveranstaltungen rekurriert werden, sodass ein unmittelbarer Zugang zur Performanz und Entwicklung der Kooperation sowie zum Umgang mit den Aufgaben und Inputs der Intervention vorliegt. Darüber hinaus wurden *Gruppendiskussionen* (vgl. Przyborski und Wohlrab-Sahr 2014, S. 88 ff.) zwei Jahre nach Abschluss des Interventionsprojekts mit sechs von insgesamt 13 Schulteams geführt. Im Fokus dieser Diskussionen standen die letzten drei Jahre mit Blick auf das Projekt *Serelisk*. Damit ist in die Gruppendiskussionen thematisch alles eingeklammert, was in den Kontext des Interventionsprojekts einzuordnen ist, also zusätzlich zur Handlungspraxis der (trans-)intentionalen Umsetzung der Intervention auch die

*Praxis des Umgangs mit Erhebungsinstrumenten, die Praxis der Entscheidung zur Teilnahme an der Intervention und weitere Aspekte*, die für die Umsetzung eine Rolle spielen, aber von den Forscher\*innen gegebenenfalls noch nicht antizipiert wurden. Im Falle dieser Studie ergab sich noch eine Besonderheit: Weil eine Forscherin, die in die Konzeption der Intervention eingebunden war, die Gruppendiskussionen leitete, bildet sich im Material auch eine *Kommunikationspraxis zwischen einer Forscherin der didaktischen Interventionsforschung und den Lehrpersonen* ab. Diese Aufzählung an Perspektiven auf die schulische Handlungspraxis im Kontext der Intervention unterstreicht die umfassende Sichtweise der dokumentarischen Evaluationsforschung auf den Gegenstand. Gleichzeitig wird deutlich, dass hier Praxen enthalten sind, die nicht nur die Intervention losgelöst von der Wissenschaftspraxis als Initiatorin der Intervention betrachten. Indem beispielsweise die Kommunikationspraxis zwischen Forscherin und Beforschten thematisch Eingang in die Analyse erhält, wird neben der didaktischen Intervention auch die didaktische Interventionsforschung selbst in die Analyse eingeklammert und damit auch der wissenschaftliche Kontext der Wirkungen mitberücksichtigt. Die Daten müssen zeitlich allerdings so gelagert sein, dass Handlungspraxen der Auseinandersetzung mit der didaktischen Intervention(sforschung) enthalten sein können. Die Auswertung erfolgte nach den gängigen Schritten der dokumentarischen Interpretation (vgl. Bohnsack 2014). Jede Datensorte wurde unter Wahrung der Komplementarität der Daten zueinander zunächst für sich ausgewertet und dann im Zuge der Fallbeschreibung aufeinander bezogen. Die Triangulation der beiden Datensorten und der verschiedenen Erhebungszeitpunkte erfolgte somit unter dem ‚Dach‘ der dokumentarischen Methode in Anlehnung an Nora Friederike Hoffmann (2016), wodurch erst der jeweilige Fall hervorgebracht wurde, der dann Eingang in die sinngenetische und soziogenetische Typenbildung erhielt (vgl. ausführlicher zum Vorgehen Werner 2022).

- (3) Die *dokumentarische Metaevaluation* ergibt sich vor diesem Hintergrund einzig mit Bezug auf die didaktische Interventionsforschung, da diese aus eigenen Mitteln heraus über keinen so breiten Zugang zu einer empirisch angelegten Selbstreflexion verfügt. Über eine dokumentarische Evaluation der Evaluation lässt sich, so die Überlegung, hier entsprechend eine ‚Lücke‘ schließen. Bei der dokumentarischen Evaluationsforschung – als einer praxeologisch-empirischen Forschungsrichtung – hingegen ist eine solche Selbstreflexion bereits immer Bestandteil der Analyse und erfordert somit keine separate Begrifflichkeit (vgl. Bohnsack 2017, S. 11).

## 4 Exemplarische Blicke auf die Wissenschaftspraxis von *Serelisk*

Nachfolgend werden Ergebnisse der dokumentarischen Metaevaluation von *Serelisk* vorgestellt, um einen exemplarischen Einblick in die reflexiven Erkenntnismöglichkeiten für die didaktische Interventionsforschung zu geben. Fokussiert wird dabei auf ausgewählte Ergebnisse der Evaluation des Implementationsmittels Kooperation und der Grundannahme einer freiwilligen Projektteilnahme.

### 4.1 *Das Implementationsmittel Kooperation im Blick*

In der didaktischen Interventionsforschung *Serelisk* wurde auf eine kooperative Implementationsstrategie gesetzt, das heißt, die Implementation der didaktischen Mittel sollte über die Etablierung schulischer Kooperationsteams (PLG), die gemeinsam an der Umsetzung der didaktischen Mittel im Unterricht arbeiten, gefördert werden (vgl. Werner 2022). Auf der Basis der quantitativen Analysen kamen die Forscher\*innen zur Einschätzung, dass das Mittel intentional gewirkt habe, da unter anderem die kooperativen Aktivitäten der Experimentalgruppenlehrpersonen zum Unterricht im Projektzeitraum tendenziell gesteigert worden seien und sich deren Einstellung zu verbindlichen Formen der Kooperation zum Unterricht positiv entwickelt habe (siehe Kap. 1). Die dokumentarische Evaluation der Umsetzung der Kooperation hat jedoch verschiedene Wirkungsweisen bzw. -typen des Mittels Kooperation hervorgebracht, von denen nachfolgend zwei vorgestellt werden.

#### 4.1.1. *Kooperation als Implementationsproblem*

Um den ersten Wirkungstyp zu illustrieren, komme ich wieder auf das Eingangszitat zurück. So wie dort für Frau Azur interpretiert, werden die didaktischen Ziele des Interventionsprogramms bei diesem Typ tatsächlich für alle drei Lehrerinnen anschlussfähig. Allerdings gehen deren Vorstellungen von der konkreten praktischen Umsetzung im Unterricht auseinander. So möchte Frau Azur (Af) ihren Unterricht weitgehend gemäß der Projektziele innovieren, während die Kolleginnen Frau Kiwi (Kf) und Frau Rot (Rf) nur eine niederschwellige, aber voneinander abweichende Umsetzung anvisieren. Zurückgeführt wird dies unter anderem bei Frau Kiwi auf ihre Funktion als Klassenlehrerin in der Projektklasse, die bereits mit Zeitverlust im Fachunterricht verbunden sei. Die Lehrerinnen finden keine gemeinsam anschlussfähigen Lösungen für die jeweiligen Umsetzungsprobleme, da sie sich an ihren *individuellen Zielen orientieren*. In der Folge wird die Erwartung des Projekts nach einer



der Gruppen anschlussfähig werden, lassen sie die Umsetzung *passiv versanden* („letschendlich isch=s einfach versandet“; Trapp, GD, Z. 1083). Dennoch erzeugen die Interventionsveranstaltungen und die Erhebungsinstrumente bei den Lehrpersonen einen *moralischen Konflikt mit ihrer Projektzugehörigkeit*, denn die Umsetzung wird „in gewisser Hinsicht so als Hausaufgabe“ (Comenius, GD, Z. 17) verstanden. Zur Lösung dieses moralischen Problems dient ihnen die Kooperationsgruppe. Exemplarisch zeigen lässt sich dies anhand einer kurzen Passage aus dem authentischen Gespräch der zweiten Interventionsveranstaltung (AG 2) der Gruppe Comenius. Die Lehrpersonen der Gruppe – Frau Coralle (Cf), Herr Grün (Gm) und Herr Braun (Bm) – reflektieren dem Auftrag gemäß ihre bisherige Umsetzungspraxis. In diesem Rahmen verhandeln sie die zu erwartende Intensitätsnorm, in deren Rahmen die ‚Hausaufgabe‘ der Implementation erfüllt wäre. Dabei rekurren sie sowohl auf die eigene Gruppe als auch auf andere Gruppen des Projekts. Die Intensität der Umsetzung von Herrn Braun liegt unter der seiner Bezugspersonen, was er kritisch zur Diskussion stellt („Vergleichsch=e=mal-“; Comenius, AG 2, Z. 65). Daraufhin entsteht eine kooperative Praxis der Legitimation ihrer unverbindlichen Umsetzungspraxis („mal die Methode ausprobiert (.) dann die Methode ausprobiert“; Comenius, AG 2, Z. 67 f.) und bildet das gemeinsame Problem der Gruppe. Die Passage schließt mit folgender Konklusion:

Cf: Aber ich glaub (1) dass du automatisch au scho (1) in deim normale Unterricht vielleicht mehr so Problemlösung mit eibausch wenn des mal agfange hasch (.)  
 Bm: Ja (1)  
 Cf: Dass de deine Frage au scho andersch stellsch (.) dass du mit andre Sache |konfrontiersch|  
 Gm: [des machsch|  
 halt öfters wenn=s |dir | bewusst isch  
 Cf: [Ja ] [Genau  
 Gm?: [Ja  
 Cf: [Des |isch des was du ( ) |  
 Bm: [un des- (.) und des isch au| ja  
 genau

(Reflexion der bisherigen Umsetzungspraxis, Comenius, AG 2, Z. 80–90)

Die hier kooperativ – von allen gemeinsam – hergestellte Konklusion der Gruppe besteht darin, dass der Transfer in die Unterrichtspraxis selbstläufig („automatisch“; Comenius, AG 2, Z. 80) erfolgt und damit teilweise auch implizit bleibt. Das bestehende ‚schlechte Gewissen‘ aufgrund einer möglicherweise geringeren Umsetzungsintensität im Vergleich zu anderen Gruppen wird damit als Gemeinschaft nivelliert und der bestehende Modus der Implementation kooperativ legitimiert. Indem die Umsetzung in der Gruppe argumentativ ins Unbewusste gelagert wird, wird eine weitere Thematisierung einer bewusst wahrgenommenen Umsetzungsintensität in der Gruppe hinfällig.

Nun schließen sich hier wiederum Fragen an, so zum Beispiel: Warum entfaltet das Mittel der Kooperation bei der einen Gruppen diese und bei den anderen Gruppen jene Wirkung? Um die Genese zu ergründen, soll hier auf eine

soziogenetische Analyse Bezug genommen werden. Diese wurde auch deshalb ausgewählt, da sie auch einzeln betrachtet einen reflexiven Blick auf ein zentrales Grundprinzip didaktischer Interventionsforschung ermöglicht.

#### *4.2 Das Grundprinzip der freiwilligen Teilnahme im Blick*

Vorgestellt werden ausgewählte Erkenntnisse der soziogenetischen Analyse zum Erfahrungsraum *Projektteilnahmeentscheidung*. Die Annahme einer freiwilligen Projektteilnahme bildet einen der Grundpfeiler von Forschung, speziell von (didaktischer) Interventionsforschung und wird in der Regel im Rahmen von Projektverträgen bzw. Einverständniserklärungen per Unterschrift bestätigt und zählt in der Forschungspraxis eher in den formalen Bereich der Projektakquise. Darauf aufbauend wird von einem Interesse an der Umsetzung der Intervention und von einer Relevanz für den eigenen Unterricht und damit von einer Anschlussfähigkeit ausgegangen. Wie die Teilnahme tatsächlich zustande gekommen ist, wird in den standardisierten Erhebungen didaktischer Interventionsforschung für gewöhnlich nicht mehr betrachtet. Die soziogenetischen Analysen der Gruppendiskussionen zeichnen jedoch ein komplexeres Bild von der Teilnahmeentscheidung, und es wird deutlich, dass die Art und Weise, wie eine Lehrperson zur Projektteilnahme gelangt ist, auch einen Einfluss auf die Umsetzung nehmen kann. So gibt es Gruppen, bei denen nicht alle Lehrpersonen wirklich freiwillig am Projekt teilgenommen haben. Es zeigte sich, dass Frau Beige aus der Gruppe Trapp vom Schulleiter „sozusagen verdonnert“ (GD, Z. 98) wurde, und sie „wollte nicht“ (GD, Z. 92); und Herr Braun aus der Gruppe Comenius hat im Rahmen seiner Projektanmeldung einen Ausfüllfehler im Formular gemacht und ist dadurch statt in die Kontrollgruppe in die Experimentalgruppe gekommen („un denn hab i zuerscht gdacht ohje (.) jetzt bin i da in der: (.) Nichtkontrollgruppe“, GD, Z. 1679 f.). Dieser heterogene Erfahrungsraum der Projektteilnahmeentscheidung in den Gruppen korrespondiert mit einer versandenden Umsetzungspraxis (siehe Kap. 4.1) und wird als Störgröße vor allem durch divergente Diskurstrukturen im Implementationsprozess erkennbar. Während für Frau Beige zu Beginn des Projekts aufgrund ihrer oktroyierten Teilnahme die Frage einer generellen Umsetzbarkeit der Interventionen virulent ist, beschäftigt ihre Kolleginnen, die sich selbst für die Teilnahme eingesetzt haben, die Frage der inhaltlichen Ausgestaltung der Umsetzung. Dieser Konflikt wird in der Gruppe zunächst nicht reflexiv, sondern verdeckt gehalten und die gemeinsame Zieldefinition rituell beendet, ohne dass ein tatsächlich gemeinsames Ziel definiert wird. Im Fall Comenius und im Fall Trapp zeichnet sich die ‚unfreiwillige Projektperson‘ jeweils durch eine geringere Umsetzungspraxis aus, die von ihr selbst problematisiert wird. Das Team begegnet dem jedoch jeweils nicht beispielsweise mit einem Hilfsangebot, sondern mit einer Verschiebung des Themas (Trapp: Fokus auf die

schweren Bedingungen; Comenius: Fokus auf eine unbewusste Umsetzung).

Das eigentliche Problem der geringen Umsetzung, das im Grunde alle Teilnehmer\*innen beider Gruppen teilen, bleibt jedoch im Kern ungelöst und manifestiert sich hingegen in einer kooperativen Legitimation des Versandens, was auf das Problem moralisch entlastend wirkt.

### *4.3 Reflexive Erkenntnisse für die didaktische Interventionsforschung*

Welche Erkenntnisse lassen sich nun aus den vorangegangenen dokumentarischen Analysen über und für die Wissenschaftspraxis didaktischer Interventionsforschung ziehen? Die vorgestellten Analysen zu den Wirkungsweisen des Implementationsmittels Kooperation zeigen, dass (1) die didaktischen Mittel auch ohne das Implementationsmittel Kooperation anschlussfähig und implementiert werden können und dass (2) das Implementationsmittel Kooperation selbst sogar zu einem Problem für die anvisierte Implementation werden kann, wenn keine gemeinsamen Vorstellungen von der Umsetzungspraxis hergestellt werden können. Es wird deutlich, dass das Entwickeln gemeinsamer Vorstellungen zur Umsetzung im Unterricht eine eigene komplexe Herausforderung ist, für die es keine ‚einfachen‘ Implementationslösungen geben kann. In jedem Fall scheint es plausibel, eine solche Problematik als Option für die Implementation mitzudenken und sich Wege zu überlegen, wie Forscher\*innen solche Situationen erkennen und mit diesen umgehen können.

Zudem hat sich gezeigt, dass es im Laufe einer Intervention unter anderem erfahrungsbasiert zu der Einschätzung kommen kann, die angebotenen Mittel nicht weiter oder nur noch reduziert implementieren zu wollen. Der Interventionsaufbau sieht jedoch aus methodischen Gründen in der Regel keine inhaltlichen Anpassungen im laufenden Prozess vor, und ein Ausstieg aus dem Projekt wird von den Lehrpersonen oftmals nicht in Erwägung gezogen. Somit bleibt – trotz fehlender Anschlussfähigkeit der didaktischen Intervention – eine moralische Verpflichtung dem Projekt gegenüber bestehen (u. a. werden die Fragebögen noch ausgefüllt), mit der die Lehrpersonen umgehen müssen. Interessant ist, dass an dieser Stelle das Instrument der Kooperation transintentional zur moralischen Entlastung genutzt wurde. Der Nutzen, der sich hier jedoch entfaltet, steht diametral zur Intention der Wissenschaft. Es ist jedoch offenkundig, dass es diese Verläufe geben kann. Entsprechend gilt es auch hier, diese aufzuspüren. Ersichtlich wurde, dass eine heterogene Ausgangslage von freiwilliger und unfreiwilliger Projektteilnahme in den Gruppen mit diesem Verlauf korrespondiert. Somit könnte es hier eine Option sein, die Prozesse der Projektakquise differenzierter zu gestalten und empirisch in den Blick zu nehmen. Zudem erscheint es sinnvoll, die Entwicklung der Einschätzung der Praxisrelevanz oder die Entwicklung zur Motivation der Umsetzung empirisch zu erfassen.

Vor dem Hintergrund dieser Zusammenhänge wird besonders deutlich, dass die gängige Praxis von einer formalen Angabe zur freiwilligen Projektteilnahme auf die Einschätzung der Praxisrelevanz der Intervention zu schließen, so allgemein nicht gehalten werden kann.

Und letztlich zeigt sich anhand der exemplarischen Analysen, dass es im Rahmen der statistischen Analyse der Wirksamkeit des Implementationsmittels Kooperation Leerstellen gibt, die jenseits der durchschnittlichen Intensitätssteigerung von Kooperationstreffen liegen, aber für die Bewertung der Intervention von nicht unerheblicher Relevanz sind.

## **5 Kritischer Ausblick: Mehrwert mit ‚Risiken und Nebenwirkungen‘**

Es konnte meines Erachtens theoretisch hergeleitet und anhand von empirischen Beispielen gezeigt werden, dass eine dokumentarische Metaevaluation für die didaktische Interventionsforschung als Beispiel einer evidenzbasierten Bildungsforschung die Möglichkeit bietet, analytische Leerstellen jenseits der Intentionen der Forscher\*innen zu füllen und einen reflexiven Bezug zur eigenen Praxis herzustellen, der nötig ist, um die Wissenschaftspraxis mit ihren Methoden weiterzuentwickeln und dem Ziel von *evidence* näher zu kommen. Dies erfordert jedoch die Überwindung von methodologischen Barrieren, also eine wechselseitige Offenheit von Vertreter\*innen standardisierter und rekonstruktiver Methoden.

Zu den ‚Risiken und Nebenwirkungen‘ der Nutzung eines solchen Ansatzes in der evidenzbasierten Forschung wäre möglicherweise das Anerkennen von methodischen Limitierungen und einem damit einhergehenden ‚Hierarchieverlust‘, der dann auch mit einem Verlust des Stellenwerts in der öffentlichen bildungspolitischen Debatte einhergehen könnte, zu nennen. Vor diesem Hintergrund wäre danach zu fragen, wie es möglich werden kann, hier methodisch miteinander ins Gespräch zu kommen – auf Augenhöhe, ohne in erneute methodische Grundsatzstreitigkeiten zu geraten (Stichwort: Positivismusstreit). Denn Analysen wie die vorgestellte bringen ‚Probleme‘ hervor, die es dann zu lösen gilt. Gleichzeitig, das konnte auch gezeigt werden, können solche Analysen auch Ideen zur Bewältigung dieser ‚Probleme‘ hervorbringen, sodass es hier nicht darum gehen soll, die Berechtigung der einen Forschungsrichtung in Abrede zu stellen, sondern vielmehr nach Möglichkeiten zu suchen, diese weiterzuentwickeln. Aber natürlich ist dies dann eine unbekannte ‚Reise‘: und es gilt auch hier: „Nichts ist so beständig wie der Wandel“ (Heraklit). Eine solche offene Entwicklung muss in Kauf genommen, aber auch finanziell unterstützt werden.

Eine Frage, die sich zudem noch anschließt, ist die der praktischen Übertragbarkeit der vorgestellten Überlegungen auf andere Formen evidenzbasierter (Bildungs-)Forschung. Dies bleibt zu prüfen.

## 6 Literatur

- Barzel, B., Büchter, A., & Leuders, T. (2007). *Mathematik Methodik. Handbuch für die Sekundarstufe I und II*. Berlin: Cornelsen.
- Bellmann, J., & Müller, T. (2011). Evidenzbasierte Pädagogik – ein Déjà-vu? Einleitende Bemerkungen zur Kritik eines Paradigmas. In J. Bellmann, & T. Müller (Hrsg.), *Wissen, was wirkt. Kritik evidenzbasierter Pädagogik* (S. 9–32). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blumenberg, H. (2021). *Ästhetische und metaphorologische Schriften*. Frankfurt a. M.: Suhrkamp Verlag.
- Bohnsack, R. (2010). Qualitative Evaluationsforschung und dokumentarische Methode. In R. Bohnsack, & I. Nentwig-Gesemann (Hrsg.), *Dokumentarische Evaluationsforschung. Theoretische Grundlagen und Beispiele aus der Praxis* (S. 23–62). Opladen: Barbara Budrich.
- Bohnsack, R. (2014). *Rekonstruktive Sozialforschung. Einführung in qualitative Methoden*. Opladen: Barbara Budrich.
- Bohnsack, R. (2017). Zur ‚Praxisrelevanz‘ der Praxeologischen Wissenssoziologie. In R. Bohnsack (Hrsg.), *Praxeologische Wissenssoziologie* (S. 324–340). Opladen: Barbara Budrich.
- Bohnsack, R., & Nentwig-Gesemann, I. (2010). *Dokumentarische Evaluationsforschung. Theoretische Grundlagen und Beispiele aus der Praxis*. Opladen: Barbara Budrich.
- Bonsen, M., Hübner-Schwartz, C., & Mitas, O. (2013). Teamqualität in der Schule. Lehrerkooperation als Ausgangspunkt für Schul- und Unterrichtsentwicklung. In M. Keller-Schneider, S. Albisser, & J. Wissinger (Hrsg.), *Professionalität und Kooperation in Schulen. Beiträge zur Diskussion über Schulqualität* (S. 105–122). Bad Heilbrunn: Klinkhardt.
- Böttcher, W., Hogrebe, N., & Schulz, S. (2013). Evidenzgewinn und Generalisierbarkeit durch systematische Auswertungen von Evaluationen? Das Beispiel Erziehungswissenschaft. In J. Hense, S. Rädiker, & T. Widmer (Hrsg.), *Forschung über Evaluation. Bedingungen, Prozesse und Wirkungen* (S. 211–230). Münster: Waxmann.
- Büchter, A., & Leuders, T. (2005). *Mathematikaufgaben selbst entwickeln. Lernen fördern – Leistungen überprüfen*. Berlin: Cornelsen.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin: Springer.
- Emmerich, M., & Werner, S. (2013). Kooperationsgestützte Unterrichtsentwicklung zwischen Interventionslogik und schulischer Eigenrationalität. In M. Keller-

- Schneider, S. Albisser, & J. Wissinger (Hrsg.), *Professionalität und Teamqualität* (S. 169–183). Bad Heilbrunn: Klinkhardt.
- Goldmann, D. (2017). *Programmatik und Praxis der Schulentwicklung. Rekonstruktionen zu einem konstitutiven Spannungsverhältnis*. Wiesbaden: Springer VS.
- Gräsel, C., Fussangel, K., & Parchmann, I. (2006). Lerngemeinschaften in der Lehrerfortbildung. Kooperationserfahrungen und -überzeugungen von Lehrkräften. *Zeitschrift für Erziehungswissenschaft* 9(4), S. 545–561.
- Hascher, T., & Schmitz, B. (Hrsg.) (2010). *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen*. Weinheim: Juventa.
- Hoffmann, N. F. (2016). *Szene und soziale Ungleichheit. Habituelle Stile in der Techno/Elektro-Szene*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Jornitz, S. (2009). Evidenzbasierte Bildungsforschung. *Pädagogische Korrespondenz* 40, S. 68–75.
- Kotthoff, H.-G., & Werner, S. (2011). *Abschlussbericht: Nachhaltigkeit extern initiiert unterrichtszentrierter Kooperation von Lehrpersonen (EiKo). Eine empirische Untersuchung der Teamentwicklung und Institutionalisierung unterrichtszentrierter Lehrerverbände an Realschulen in Baden-Württemberg*. Freiburg im Breisgau: Pädagogische Hochschule (unveröff.).
- Lambrecht, M. (2018). *Steuerung als pädagogisches Problem. Empirische Rekonstruktionen zur Interaktion in Schulinspektions-Interviews*. Wiesbaden: Springer VS.
- Lambrecht, M., & Rürup, M. (2012). Bildungsforschung im Rahmen einer Evidence Based Policy. Das Beispiel ‚Schulinspektion‘. In A. Wacker, U. Maier, & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung. Empirische Befunde und forschungsmethodische Implikationen* (S. 57–77). Wiesbaden: Springer VS.
- Landesinstitut für Schulentwicklung (2009). *Bildungsplan 2004. Grundschule, Hauptschule, Realschule, Allgemein bildendes Gymnasium: niveaunkretisierung für alle Fächer/Fächerverbände/Themenorientierten Projekte*. Landesinstitut für Schulentwicklung. <http://www.bildungsplaene-bw.de/site/bildungsplan/get/documents/lsw/Bildungsplaene/Bildungsplaene-2004/Niveaunkretisierung/Realschule/NK-Rs-M.pdf>. Zugegriffen: 20. November 2022.
- Leutner, D. (2010). Perspektive pädagogischer Interventionsforschung. In T. Hascher, & B. Schmitz (Hrsg.), *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen* (S. 63–72). Weinheim: Juventa.
- Maag Merki, K., Holzbrecher, A., Kotthoff, H.-G., Ehlert, A., & Werner, S. (2009). *Abschlussbericht: Selbstreflexives Lernen im schulischen Kontext. Eine Interventionsstudie in Realschulen in Baden-Württemberg*. Freiburg im Breisgau: Pädagogische Hochschule (unveröff.).
- Mannheim, K. (1952). *Ideologie und Utopie*. Frankfurt a. M.: Schulte-Bulmke-Verlag.
- Meseth, W. (2011). Erziehungswissenschaftliche Forschung zwischen Zweck- und Wertrationalität. Überlegungen zum Normativitätsproblem der empirischen

- Unterrichtsforschung. *Zeitschrift für Soziologie der Erziehung und Sozialisation* 31(1), S. 12–27.
- Ostermeier, C. (2004). *Kooperative Qualitätsentwicklung in Schulnetzwerken*. Münster: Waxmann.
- Przyborski, A., & Wohlrab-Sahr, M. (2014). *Qualitative Sozialforschung. Ein Arbeitsbuch*. München: Oldenbourg.
- Scriven, M., *The Logic of Evaluation* (2007). OSSA Conference Archive 138. S. 1–17. <https://scholar.uwindsor.ca/ossaarchive/OSSA7/papersandcommentaries/138>. Zugegriffen: 4. Juni 2022.
- Thiel, F., Heinrich, M., & Ackeren, I. v. (2013). Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-Förderschwerpunkt Steuerung im Bildungssystem (SteBis). *Die Deutsche Schule* 12. Beiheft, S. 11–18.
- Werner, S. (2012). Wie kommt Kooperation in die Schule? Zum Spannungsverhältnis zwischen Interventionsidee und schulpraktischer Umsetzung. In E. Baum, T.-S. Idel, & H. Ullrich (Hrsg.), *Kollegialität und Kooperation in der Schule. Theoretische Konzepte und empirische Befunde* (S. 135–148). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Werner, S. (2022). *Die kooperative Implementationsstrategie. Praxeologische Rekonstruktionen im Kontext didaktischer Interventionsforschung*. Wiesbaden: Springer VS.
- Zeitler, S., Heller, N., & Asbrand, B. (2012). *Bildungsstandards in der Schule. Eine rekonstruktive Studie zur Implementation der Bildungsstandards*. Münster: Waxmann.