

Bericht zur Expertise im Kontext des Forschungsverbunds der Initiative „Leistung macht Schule“ zur Vorbereitung der Transferphase

Wie können wissenschaftliche Generalisierung und pädagogische Praxisrelevanz in LemaS verbunden werden?

Juli 2020

Dr. Kenneth Horvath

Kernpunkte im Überblick

- Der Forschungsverbund LemaS betritt in seiner Gesamtkonzeption konzeptuelles und methodologisches Neuland. Er ist dabei auch im internationalen Vergleich außergewöhnlich.
- Ein Kernproblem dieser neu ausgerichteten Bildungsforschung ist die Frage nach dem Verhältnis von wissenschaftlicher Generalisierung und pädagogischer Praxisrelevanz. Diese noch stärker in den Blick zu nehmen, ist eine der zentralen Herausforderung für den Forschungsverbund LemaS.
- Die vorliegende Expertise stellt in diesem Zusammenhang die These ins Zentrum, dass die Forderung nach *Praxisrelevanz* von Sozial- und Bildungsforschung keine Einschränkung, sondern eine Vertiefung und Verbreiterung der wissenschaftlich-empirischen Fundierung verlangt.
- Um diesen Anforderungen gerecht zu werden, muss der Forschungsverbund notwendigerweise multimethodisch arbeiten und, vor allem, drei Formen der wissenschaftlichen Generalisierung zueinander in Verbindung setzen: Aussagen zu den Effekten einzelner Instrumente („Wirksamkeit“), zur Übertragbarkeit von Instrumenten in die Breite der Schullandschaft („Repräsentativität“) und zu den kontextuellen Bedingungen einer erfolgreichen Nutzung („Transferabilität“). Diese drei Generalisierungsansprüche entsprechen methodologisch den drei Grunddesigns sozialwissenschaftlicher Forschung (Experiment, Erhebung und Fallstudie).
- Über den gesamten Forschungsverbund hinweg betrachtet spielen alle drei Grunddesigns eine wichtige Rolle. Allerdings ist kein Teilprojekt für sich genommen in der Lage, mehr als eine dieser Anforderungen zu erfüllen. Zudem kann in der ersten Förderphase keines der Grunddesigns in Reinform umgesetzt werden. Es wird daher angeregt, Wege zur Verschränkung und gegenseitigen Validierung und Plausibilisierung zwischen den Teilprojekten zu entwickeln.
- Vor diesem Hintergrund sollte der Fokus nicht auf der Aufweichung von disziplinären Ansprüchen liegen. Ganz im Gegenteil sollten die Teilprojekte bewusst auf die je eigene

wissenschaftliche Profilbildung achten. Für den Forschungsverbund als ganzen geht es darum, Ergebnisse und Befunde aus den Teilprojekten zueinander in Beziehung zu setzen und miteinander zu verschränken.

- Nicht zuletzt, um Befunde aus verschiedenen Teilprojekten transparent und reflektiert miteinander verschränken zu können, sollte der Forschungsverbund wissenschaftliche Standards und Leitlinien für die Gestaltung und Dokumentation praxisrelevanter Forschung formulieren. Den Teilprojekten wird so gegenseitige Orientierung und zielgerichtete Kommunikation ermöglicht. Gleichzeitig kann so die Nachvollziehbarkeit und Bewertung der empirischen Befunde nach außen gewährleistet werden.
- Für die Gestaltung der formativen Evaluation wird empfohlen, die kombinierten Anforderungen der wissenschaftlichen Generalisierung und der pädagogischen Praxisrelevanz in Leitfragen zu übersetzen, die typischen Entscheidungsprozessen in der pädagogischen Praxis folgen. Ein solches an Leitfragen ausgerichtetes Modell würde auch eine einheitliche Aufbereitung der verschiedenen Produkte aus den Teilprojekten am Ende der ersten Förderphase erlauben.

1. Einleitung

Das vorliegende Diskussionspapier fasst die wesentlichen Ergebnisse einer Expertisetätigkeit zusammen, die im Zeitraum September 2019 bis Mai 2020 im Auftrag des Bundesministeriums für Bildung und Forschung durchgeführt wurde. Thema der Expertise war die Frage, wie die Ziele der pädagogischen Praxisrelevanz und der wissenschaftlichen Generalisierbarkeit im Kontext des Forschungsverbunds „Leistung macht Schule“ miteinander verbunden werden können. Dazu sollen unter anderem – vorbereitend auf die Transferphase (Förderphase 2) – Empfehlungen zur formativen Evaluation und zur Präsentation der in den Teilprojekten erarbeiteten Praxisprodukte ausgesprochen werden.

Die Expertise beruht auf einer intensiven und mehrgliedrigen Auseinandersetzung mit dem Forschungsverbund LemaS. Das Vorgehen und die inhaltlichen Schwerpunkte wurden in enger Abstimmung mit der Verbundkoordinatorin, Prof. Dr. Gabriele Weigand, festgelegt und bei Bedarf angepasst. Neben der Sichtung aller verfügbaren Unterlagen zu den Forschungstätigkeiten der einzelnen Teilprojekte (Roadmaps, Unterlagen aus Datenschutzverfahren, im Forschungsverbund offen zugängliche Forschungsinstrumente) wurden ausgewählte Teilprojekte vertiefend in den Blick genommen; mit diesen wurden unter anderem ausführliche Interviews geführt (die Auswahl der Teilprojekte zielte darauf, die disziplinäre, methodische und thematische Vielfalt von LemaS abzudecken). Die in der Beschäftigung mit den Teilprojekten identifizierten Fragen und Herausforderungen wurden (1) vor dem Hintergrund aktueller Debatten in der internationalen Bildungsforschung und (2) mit Blick auf Erwartungen aus Bildungspolitik und pädagogischer Praxis in eine Reihe konkreter Empfehlungen übersetzt. Diese Empfehlungen wurden im Forschungsverbund diskutiert und auf dieser Grundlage weiter konkretisiert. Sie sollen nun für die weitere Arbeit in LemaS

genutzt werden, insbesondere für die Gestaltung der verbundübergreifenden formativen Evaluation der in den Teilprojekten erarbeiteten Produkte.

Das zentrale Anliegen der Expertisetätigkeit bestand darin, die Erarbeitung einer gemeinsamen Position und Vorgehensweise im Forschungsverbund zu ermöglichen. Dementsprechend sollen die im Folgenden formulierten Empfehlungen in erster Linie eine offene Verständigung über die angesprochenen Fragen im Forschungsverbund fördern und vorantreiben. Sie sind daher zum jetzigen Zeitpunkt als vorläufig zu verstehen. Sie werden aber mit dem Anspruch präsentiert, eine in den Teilprojekten häufig als komplex und überfordernd empfundene Konstellation auf die wesentlichen Aspekte zu kondensieren. Das Papier definiert eine forschungslogische Perspektive und präsentiert insofern ideale Orientierungspunkte, die in der Forschungspraxis nur eingeschränkt realisiert werden können.

Die folgende Darstellung beginnt mit einer groben Skizze des Ausgangsproblems (Abschnitt 2). Für diese Skizze werden unter anderem Schlaglichter auf Diskussionen in der internationalen Bildungsforschung geworfen. Der häufig eher diffuse Anspruch nach „mehr Praxisrelevanz“ lässt sich auf dieser Grundlage als methodologische Anforderung fassen, mehrere Formen der wissenschaftlichen Fundierung aufeinander zu beziehen und miteinander zu verschränken. Abschnitt 3 konkretisiert dieses Argument und benennt drei Generalisierungsansprüche, vor denen der Forschungsverbund LemaS steht. Abschnitt 4 bespricht auf dieser Grundlage die Aufgabe der formativen Evaluation der in den Teilprojekten von LemaS erarbeiteten Praxisprodukte und diskutiert Grundlinien ihrer angemessenen Präsentation. Alle drei Abschnitte münden jeweils in eine Reihe von Empfehlungen für den Forschungsverbund. Abschnitt 5 schlägt nächste Arbeitsschritte zur konkreten Umsetzung über die kommenden Projektphasen vor. Das Dokument schließt mit einem kurzen Ausblick auf die mögliche Ausgestaltung der Begleitforschung in der Transferphase (Abschnitt 6).

2. Ausgangsproblem: Empirische Forschung und Praxisrelevanz

Der Forschungsverbund LemaS betritt in einigen Hinsichten Neuland, die gerade für die empirische Forschung relevant und herausfordernd sind. Im Kern geht es darum, in der Bildungsforschung zu mehr pädagogischer Praxisrelevanz und (damit eng zusammenhängend) auch zu mehr wissenschaftlicher Erklärungskraft zu kommen. Dieser Anspruch und der daraus folgende innovative Zuschnitt von LemaS ergibt sich aus der Bilanzierung der „empirischen Wende“ in der deutschen Bildungslandschaft, wie sie in der KMK Gesamtstrategie zum Bildungsmonitoring¹ in ihrer aktuellen Fassung (und im darauf aufbauenden gemeinsamen Positionspapier der Landesinstitute und Qualitätseinrichtungen der Länder²) sowie im Rahmenprogramm Empirische Bildungsforschung des

1 https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf, abgerufen am 2. März 2020.

2 https://bildung-rp.de/fileadmin/user_upload/bildung-rp.de/pl/Newsletter/2018/8/12/Positionspapier_Transfer_31.10.18.pdf, abgerufen am 2. März 2020.

BMBF³ Ausdruck findet. Damit verbunden ist die Vorstellung, von einem asymmetrischen Experten-Novizen-Verhältnis zwischen Wissenschaft und Praxis zu einem „Dialog auf Augenhöhe“ zu kommen, in dem der Komplexität pädagogischer Praxis angemessen Rechnung getragen wird. Allerdings fehlen im deutschsprachigen Kontext gegenwärtig explizit ausformulierte Vorstellungen dazu, wie eine diesen veränderten Erwartungshaltungen entsprechende praxisrelevante Bildungsforschung eigentlich aussehen kann und soll – wie sie vorgehen soll, an welchen Standards sie sich orientieren soll, welche Fragen sie stellen und wie sie ihre Antworten präsentieren soll.

An dieser Stelle setzt die vorliegende Expertise an. Ihre zentrale Aussage ist, dass die Forderung nach mehr Praxisrelevanz entgegen weit verbreiteter Vorstellungen nicht bedeutet, dass wissenschaftliche Ansprüche aufgeweicht werden müssen. Vielmehr und ganz im Gegenteil ergeben sich verschärfte und verschränkte Anforderungen an die empirische Fundierung wissenschaftlicher Befunde. Diese Konstellation wird im Forschungsverbund zwar auf breiter Basis wahrgenommen; gleichzeitig zeigt sich in der Auseinandersetzung mit einzelnen Teilprojekten eine gewisse Verunsicherung, wie den gleichzeitigen Anforderungen von Praxisrelevanz und Generalisierbarkeit entsprochen werden kann.

Der Forschungsverbund LemaS bietet vor diesem Hintergrund auch die Möglichkeit, die Ziele und Probleme praxisrelevanter Forschung präziser als bislang üblich zu benennen und in konkrete Forschungsstrategien zu übersetzen. Aufgrund seiner auch international gesehen einzigartigen Ausrichtung ermöglicht er, neuartige Ansprüche und bewährte Vorgehensweisen bewusst aufeinander zu beziehen und produktive Strategien für den Dialog zwischen Wissenschaft und Praxis zu erarbeiten. Die Herausforderung besteht dabei weniger in der Notwendigkeit, neue Formen empirischer Forschung entwickeln zu müssen. Der Forschungsverbund kann sich an vielfältigen und seit Jahrzehnten bewährten methodologischen Grundlagen empirischer Forschung orientieren. Die Schwierigkeit liegt vielmehr darin, Methoden, Methodologien und Forschungsstile aktiv aufeinander zu beziehen, die in der traditionellen deutschen Bildungsforschung lange Jahre als strikt voneinander getrennt gehandhabt wurden.

Wichtige Inspirationen können dazu aus aktuellen Diskussionen in der internationalen Bildungsforschung gewonnen werden. So gibt es im gesamten englischsprachigen Raum seit einigen Jahren eine vielfältige und vielschichtige bildungswissenschaftliche Auseinandersetzung mit der Frage des Verhältnisses von Wissenschaft und Praxis (Hargreaves, 1997; Hammersley, 2002; Connolly et al., 2018; Wrigley, 2018; Ladwig, 2018; Wiliam, 2019). Ausgangspunkt dieser Debatten ist die Kritik an einseitig verstandenen Modellen der Evidenzbasierung pädagogischer Praxis, wie sie seit Ende der 1990er Jahre speziell in den USA und Großbritannien etabliert wurden (Thomas, 2016). Bezweifelt wird dabei nicht das Ziel einer sozialwissenschaftlich informierten Reflexion und Gestaltung pädagogischer Praxis, sondern die Beschränkung auf die zu eng verstandene Prüfung der Wirksamkeit von vereinzelt Maßnahmen – unter „Laborbedingungen“ und im Hinblick auf einzelne, isoliert betrachtete Zielindikatoren, die der Komplexität pädagogischer Praxis nicht gerecht werden.

3 <https://www.empirische-bildungsforschung-bmbf.de/de/211.php>, abgerufen am 2. März 2020.

Diskutiert werden in diesem Kontext gegenwärtig beispielsweise weit verbreitete Fehlinterpretationen von statistischen Effektstärkemaßen als pädagogische Bedeutsamkeit (Simpson, 2018, 2019), die Notwendigkeit, von einer einseitig verstandenen „evidence-based practice“ zu „practice-based evidence“ zu gelangen (Bryk, 2015; Kvernbekk, 2019) oder auch der Mehrwert einer angemessenen Interpretation von medizinischen und naturwissenschaftlichen Metaphern, die neben reinen Laboreffekten etwa auch zu Fragen nach der „Dosierung“, den „Wechselwirkungen“ und der „Wirksamkeitsdauer“ von Maßnahmen etc. führen sollten (Wrigley & McCusker, 2019; Zhao, 2017).

In anderen nationalen Kontexten werden ähnliche Fragen in anderer Zuspitzung diskutiert. So gibt es in Frankreich eine stärker soziologisch geprägte Auseinandersetzung mit dem Wechselverhältnis von Politik, Pädagogik und Forschung, die beispielsweise die unhintergehbare Pluralität (und damit auch inhärente Widersprüchlichkeit) von pädagogischen Zielsetzungen und Handlungslogiken fokussiert (Normand, 2016; Normand et al., 2019; Derouet, 2019). In Dänemark wird seit einigen Jahren eine vergleichbare Debatte geführt, die auch international Resonanz findet (Bjerg Petersen et al., 2014). Diese ist für den deutschsprachigen Raum auch deswegen interessant, weil sowohl die Ausrichtung der Bildungs- und Erziehungswissenschaften als auch das Verhältnis von Bildungspolitik und Bildungspraxis besser vergleichbar ist als das US-amerikanische oder das britische Beispiel.

Dass diese Diskussionen aufkommen, mag auf den ersten Blick schon deswegen überraschen, weil das Versprechen von Praxisrelevanz gewissermaßen die ursprüngliche Motivation für die Förderung von „Evidence-Based Education“ lieferte (wie in der programmatischen Frage „What works?“ zum Ausdruck kommt). Der Stand der internationalen Debatte läuft auf die Diagnose hinaus, dass sich das Problem der mangelnden Praxisrelevanz dieser Programme aus einer zu engen und einseitigen Auslegung der quasi-medizinischen Wirksamkeitsmetapher ergibt (McKenney & Schunn, 2018). So sagen durchschnittliche Effektstärken wenig über die reale praktische Bedeutsamkeit einer Maßnahme aus (Simpson, 2017). Vor allem erlauben sie für sich alleine genommen nicht, förderliche und hinderliche Faktoren für die erfolgreiche Umsetzung zu erschließen und zu verstehen (Biesta, 2010; Wiliam, 2019). Für pädagogische Maßnahmen ist typisch, dass sie in manchen Kontexten stark, in anderen kaum, im einen Kontext positiv, im anderen neutral oder gar negativ wirken. Aus einer Praxisperspektive ist hier durchaus Evidenz gefragt, aber eher zu den Faktoren, die über Schulpopulationen hinweg und in divergierenden Kontexten für Varianz und Variation sorgen (Cartwright, 2019; Joyce, 2019). Entscheidend ist in diesem Kontext eine Differenzierung in Aspekte, die im Einflussbereich der Lehrkraft liegen, und solche, die nicht beeinflussbare Randbedingungen definieren. Auch die Definition des eigentlich zu lösenden Problems und die Abstimmung zwischen vielfältigen gleichzeitigen Zielsetzungen sind für Praxiskontexte unweigerlich zentral. Zhao (1997) weist in diesem Zusammenhang auf die Relevanz von „Nebenwirkungen“ hin, die im Bereich der Medizin ganz selbstverständlich ein zentrales Element der wissenschaftlichen Begleitforschung und Evidenzbasierung sind. Keine pädagogische Maßnahme ist ohne Risiken. Nicht intendierte oder von der Forschung nicht berücksichtigte Veränderungen im schulischen Alltag sind eine berechnete Sorge von Lehrkräften. Welche Auswirkungen hat beispielsweise ein neues didaktisches Konzept auf das Selbstbild, die Selbstwirksamkeit oder das Sozialverhalten von Schüler/innen? Nebenwirkungen

können negativer, ebenso gut aber positiver Art sein (zum Beispiel, wenn die Einführung eines neuen Unterrichtsformats nebenher die Kommunikation im Kollegium intensiviert). Die Klärung solcher Nebenwirkungen ist aus einer Praxisperspektive für die Entscheidung für oder gegen die Umsetzung einer Maßnahme wesentlich, weil sie Unsicherheiten beseitigt, Konflikten vorbeugt und das Verständnis für die Funktionsweise von Maßnahmen und Instrumenten fördert.

Ausgehend von diesen facettenreichen Diskussionen rund um die Rolle von Evidenz in der und für die pädagogische Praxis wurde eine Reihe von innovativen Konzepten und neuartigen Forschungsprogrammen entwickelt, von Design-Based-Research (Reinmann, 2005; The Design-Based Research Collective, 2003) über Research-Practice-Partnerships (Coburn & Penuel, 2016) und Improvement Studies (Bryk, 2015) hin zu Mixed-Methods-Meta-Reviewing (Gough, 2015; Harden & Thomas, 2005, 2010). Gemeinsam ist diesen Modellen das Bestreben, wissenschaftliche Evidenz „realistischer“ zu machen, indem die institutionellen, organisatorischen und situativen Bedingungen pädagogischer Praxis berücksichtigt werden.

Bei existierenden Modellen der Forschung-Praxis-Kooperation (Networked Improvement Communities etc.) handelt es sich im Gegensatz zu LemaS allerdings überwiegend um Konzepte für klar umgrenzte Projekte mit einzelnen Schulen und mit diesen verbundenen lokalen Akteuren (Eltern, regionale Schulverwaltung etc.). Damit noch einmal deutlicher erkennbar, dass LemaS auch international gesehen einen außergewöhnlichen Entwicklungs- und Forschungszusammenhang darstellt.

Zusammenfassend ergeben sich zwei zentrale Schlussfolgerungen:

(1) Zugespitzt formuliert erfordert Praxisrelevanz nicht weniger, sondern mehr (und teilweise auch andersartige) wissenschaftliche Evidenz. Es geht daher auch für LemaS nicht so sehr darum, einen „Kompromiss“ zwischen Wissenschaftlichkeit und Praxisrelevanz zu finden, sondern um eine systematische Verbreiterung der theoretisch unterfütterten und empirisch verankerten Wissensbasis. Auch geht es nicht einfach um die vereinfachende, quasi didaktisierte Darstellung von Forschungsbefunden für Praktiker/innen (es gibt keine Anhaltspunkte dafür, dass die Praxisrezeption von Forschungsbefunden an deren „Komplexität“ scheitert – eher im Gegenteil wird von Pädagog*innen in der Regel eher deren simplifizierender Blick auf Bildungspraxis kritisiert). Wissenschaft wird praxisrelevant, wenn sie Probleme ernst nimmt, wie sie sich in der Praxis stellen, und die situativen Konstellationen empirisch und theoretisch erschließt, die pädagogische Praxis kennzeichnen. Ihre spezifische Relevanz liegt aber nichtsdestotrotz darin, über Praxis hinauszugehen, indem sie von lokalen Spezifitäten abstrahiert und in diesem Sinne durchgängig generalisiert.

(2) Aus internationalen Modellen für die Organisation von Forschung-Praxis-Kooperationen ergeben sich wichtige Erfahrungen und Anregungen. Zum Beispiel wird durchgängig die Notwendigkeit betont, zwischen Wissenschaft, Politik und Praxis an gemeinsamen Problemdefinitionen zu arbeiten, bevor „Lösungen“ entwickelt und implementiert werden. Allerdings können bestehende Modelle nicht eins-

zu-eins auf den Kontext von LemaS übertragen werden. Erstens sind dafür Bildungssysteme und Bildungsprozesse in verschiedenen nationalen Kontexten zu unterschiedlich organisiert. Zweitens ist LemaS mit seinen insgesamt 300 beteiligten Schulen und seinen spezifischen Kooperationsstrukturen zwischen Bildungspolitik, Schulverwaltung, pädagogischer Praxis und Forschung nicht mit lokal klar abgegrenzten Projekten und Kooperationen an einzelnen Schulen vergleichbar, wie sie etwa im Zusammenhang von Design-Based-Research üblich sind.

Empfehlungen für den Forschungsverbund LemaS

- Im Sinne der nachhaltigen Verankerung der in LemaS erarbeiteten innovativen Forschungszugänge muss neben der Praxisrelevanz auch der wissenschaftlichen Profilierung und Positionierung des Forschungsverbunds und der Teilprojekte angemessen Beachtung geschenkt werden. Immerhin kann der Forschungsverbund LemaS aufgrund seines einzigartigen Zuschnitts wichtige Beiträge zur internationalen Diskussion leisten und so die internationale Relevanz und Resonanz der deutschsprachigen Bildungsforschung erhöhen.
- Aufgrund ihrer großen Bedeutung für die tägliche Arbeit, das Profil und die Nachhaltigkeit von LemaS sollten aktuelle internationale Debatten zum Verhältnis von Bildungsforschung und pädagogischer Praxis gezielt aufgearbeitet und im Forschungsverbund breit diskutiert werden. Das kann beispielsweise im Rahmen von Verbundtreffen oder in Form gemeinsamer Publikationsprojekte geschehen.

3. Drei Generalisierungsansprüche und drei Grunddesigns praxisrelevanter Bildungsforschung

Programme praxisrelevanter Forschung sehen sich routinemäßig mit dem Einwand konfrontiert, Abstriche im Hinblick auf wissenschaftliche Generalisierbarkeit machen zu müssen. Nimmt man die in Abschnitt 2 benannten Anforderungen an Praxisrelevanz ernst, erscheint dieser Einwand in entscheidenden Hinsichten verkürzt. Als zentrale Herausforderung ergibt sich ganz Gegenteil die Schwierigkeit, dass praxisrelevante Forschung mehrere Formen der wissenschaftlichen Fundierung miteinander verzahnen muss: Sie muss, unter anderem, Aussagen zur Wirksamkeit von Maßnahmen treffen, Bedingungen ihrer erfolgreichen Implementation identifizieren, die Übertragbarkeit auf verschieden gestaltete pädagogische Kontexte prüfen, die komplexen Problemlagen pädagogischer Situationen rekonstruieren oder auch möglichen nicht intendierten Wirkungen sowohl in spezifischen lokalen Kontexten als auch in der Breite der Schullandschaft nachgehen.

Aus einer methodologischen Sicht sind alle diese Formen der empirischen Fundierung mit einem Anspruch der Generalisierbarkeit verbunden (unter Generalisierung wird hier ein begründeter und gültiger Geltungsanspruch von empirischen Erkenntnissen über die jeweils konkret

vorliegenden/untersuchten Einzelfälle hinaus verstanden, Firestone, 1993; Little, 1993; Polit & Beck, 2010). Das Problem der Generalisierung wird aber auf je spezifische Art und Weise gefasst und bearbeitet. Diesem Verständnis folgend ist der Forschungsverbund LemaS mit drei Generalisierungsansprüchen konfrontiert, die zwar teilweise implizit bleiben, aber doch klar identifizierbar sind: (1) Fragen der *Wirksamkeit* von konkreten Produkten/Maßnahmen, (2) Fragen der *Repräsentativität* bzw. der Übertragbarkeit in die Breite der Schullandschaft sowie (3) Fragen der Umsetzbarkeit in diversen praktischen Kontexten bzw. der *Transferierbarkeit* von Maßnahmen über verschiedene lokale pädagogische Situationen hinweg.

Auf Grundlage der hier zusammengefassten Expertisetätigkeit lässt sich die zentrale These formulieren, dass in der *kombinierten Erfüllung dieser drei Generalisierungsanforderungen* der wesentliche Ankerpunkt für pädagogische Praxisrelevanz empirischer Forschung liegt. Alle drei Arten von Generalisierung erfordern theoretische fundierte Fragestellungen, die empirisch gestützt beantwortet werden. Für alle drei sind je spezifische Voraussetzungen und Qualitätsanforderungen zu erfüllen. Je nach Situation und konkreter Problemlage spielt mal die eine, mal die andere Anforderung eine prominentere Rolle. Für die bildungspolitische Ebene der Steuerung von Bildungssystemen geht es häufig in erster Linie darum, zu klären, welche Veränderungen im Durchschnitt und über alle lokalen Variationen hinweg zu erkennbaren Veränderungen und identifizierbaren Verbesserungen in einem Bildungssystem führt (es geht also zentral um Generalisierung im Sinne von Repräsentativität). Für die Implementierung einer Maßnahme an einer konkreten Schule oder die Neugestaltung von Unterricht durch eine Lehrkraft ist die Frage nicht so sehr, was im Durchschnitt über alle Kontexte hinweg zu einer Veränderung führt, sondern was „hier und jetzt“ auf welche Art und Weise funktionieren kann (und demnach Generalisierung im Sinn von Transferabilität).

Um von in einer empirischen Studie untersuchten Fällen auf andere lokale Situationen oder auf die Breite der Schullandschaft schließen zu können, muss pädagogische Praxis selbst vom *Kontext* zum *Gegenstand* der Forschung werden: Während Generalisierung im Sinne von Wirksamkeitsstudien idealerweise unter Laborbedingungen (und damit unter systematischer Ausklammerung realer pädagogischer Kontextbedingungen), erfordern Repräsentativität und Transferabilität Studien unter „Realbedingungen“, um hinderliche und förderliche Bedingungen systematisch eruieren und in ihren Wechselbeziehungen analysieren zu können. Die Analyse kann dabei etwa auf die realistische Einschätzung der Kosten einer Maßnahme, auf ihre möglichen (positiven wie negativen) Nebenwirkungen oder auf in konkreten schulischen Settings erwartbare Erfolge und Veränderungen abzielen. Es geht also die Bandbreite und das Ausmaß von erwartbaren Veränderungen unter Realbedingungen sowie um praktisch feststellbare Indikatoren für Erfolg und Misserfolg. Diese Anforderungen können methodologisch nur auf Basis der Verschränkung von verschiedenen Forschungsdesigns und Forschungsmethoden eingelöst werden.

Die drei genannten Generalisierungsansprüche an LemaS (und allgemeiner an die Produktion anwendungsbezogenen Wissens) entsprechen *grosso modo* den drei Grunddesigns sozialwissenschaftlicher Forschung (De Vaus, 2001): Experiment, Erhebung und Fallstudie. Diese drei

Designs zeichnen sich durch ihre je eigene Strategie aus, zu verallgemeinerbaren Aussagen zu kommen (Polit & Beck, 2010). Jedes der Designs erfordert, dass bestimmte Vorannahmen getroffen werden und bestimmte Voraussetzungen erfüllt sind. Jedes der Designs geht mit je eigenen Qualitätskriterien und Geltungsbedingungen einher. Und jedes trägt auf seine Art zur Herstellung eines Anwendungsbezugs von Forschung bei:

- Experimente verfolgen als Generalisierungsziel die Quantifizierung von Effekten (Kirk, 2013). In der Isolation von Effekten besteht die methodologische Kernherausforderung. Sie erfordert unter Laborbedingungen durchgeführte, vollständig randomisierte Kontrollgruppendesigns. Experimentelle Designs lassen für sich genommen keine Schlüsse von den untersuchten Fällen auf eine breitere Population, keine Aussagen über die konkreten Mechanismen und Faktoren, die zur Wirksamkeit einer Intervention führen, und keine Einschätzung der realen Effekte einer Maßnahme unter Realbedingungen zu.
- Erhebungen verfolgen als Generalisierungsziel den Schluss von einer Stichprobe auf eine Grundgesamtheit (Dillman et al., 2014). Um gültig zu sein, muss dieser Schluss auf Basis von unter Realbedingungen und randomisiert gezogenen Fällen erfolgen. Erhebungen erlauben für sich genommen keine Aussagen über Kausalitäten. Sie können Informationen zu relevanten hinderlichen und förderlichen Faktoren liefern, aber nicht die zugrundeliegende Wirkrichtung und auch nicht die involvierten Mechanismen benennen.
- Fallstudien verfolgen als Generalisierungsziel die umfassende Bestimmung der Faktoren und Mechanismen, die in ihrem Wechselspiel ein Phänomen konstituieren, über den Weg der systematischen Analyse eines Problems in seiner komplexen Ganzheit (Yin, 2014; Ylikoski, 2018). Um tragfähige Aussagen machen zu können, muss erstens die Auswahl der untersuchten Fälle theoretisch und empirisch begründet sein und muss zweitens durch kombinierten Methodeneinsatz dafür gesorgt werden, dass (dem Anspruch nach: alle) wesentlichen Aspekte identifiziert und in der Analyse berücksichtigt werden. Fallstudien erlauben für sich genommen keine Aussagen über das Ausmaß von Effekten und über die Häufigkeit von bestimmten Konstellationen.

In seinem aktuellen Zuschnitt deckt der Forschungsverbund LemaS als ganzer zwar alle drei Designs ab. Allerdings kann aufgrund interner und externer Gegebenheiten keines der Grunddesigns in methodologischer Reinform umgesetzt werden. Zudem kann kein Teilprojekt für sich alleine genommen alle Design- und Generalisierungsanforderungen abdecken. Im bewussten und produktiven Umgang mit dieser Konstellation liegt einer der Dreh- und Angelpunkte für die Verbindung von wissenschaftlicher Generalisierbarkeit und pädagogischer Praxisrelevanz im Rahmen von LemaS.

Natürlich ist die kombinierte Erfüllung von mehreren Generalisierungsformen für sich alleine genommen nicht *hinreichend*, um Bildungsforschung praxisrelevant zu machen. Sie ist aber *notwendig*, weil erst in der Bereitstellung von Erkenntnissen, die lokale Kontexte überschreiten, der genuine Beitrag wissenschaftlicher Forschung liegt. Um als praxisrelevant zu gelten, muss Forschung die jeweils relevanten Generalisierungsansprüche von Fall zu Fall im Austausch mit den beteiligten

Akteuren konkret bestimmen und die eigenen wissenschaftlichen Erkenntnisziele im Verhältnis zu den in einer gegebenen Praxiskonstellation aktuell zu lösenden Probleme explizit festlegen.

Empfehlungen für den Forschungsverbund LemaS:

- Alle Teilprojekte sollten die für die Bewertung von Praxistauglichkeit und Transferierbarkeit ihrer Produkte notwendigen Formen der wissenschaftlichen Fundierung im Auge behalten und Vorschläge formulieren, wie eventuelle Leerstellen zu einzelnen Generalisierungsansprüchen gefüllt werden können.
- Die Vielfalt an Forschungsdesigns und -methoden in den Teilprojekten sollte systematisch für gegenseitige Bezugnahme genutzt werden. Nur über eine solche Verschränkung können aus partikularen Befunden der Teilprojekte praxisrelevante Ergebnisse werden.
- Es sollten verbundübergreifende Standards und Leitlinien zur Gestaltung und Dokumentation der verschiedenen Formen empirischer Forschung formuliert werden. Das ist die Voraussetzung dafür, dass produktive und transparente Bezugnahmen zwischen den Teilprojekten erfolgen können.
- Der Forschungsverbund sollte frühzeitig Vorschläge zur Gestaltung der Begleitforschung in der Transferphase (2. Förderphase) erarbeiten, die den verschränkten Generalisierungs- und Designanforderungen praxisrelevanter Forschung gerecht werden. Es geht also darum, Wirksamkeitsanalysen, Monitoring und Fallstudien von Anfang an gezielt zu verzahnen.

4. Was bedeutet formative Evaluation im Rahmen von LemaS?

Die formative Evaluation der in der ersten Förderphase entwickelten Instrumente bildet eine zentrale Schnittstelle zwischen „praktischer“ und „wissenschaftlicher“ Dimension von LemaS. In der Evaluationsforschung gibt es kein einheitliches und abschließendes Verständnis davon, wodurch sich formative Evaluation letztlich definiert und wie diese von anderen Varianten der wissenschaftlichen Prozessbegleitung (wie Implementations- oder Entwicklungsforschung) abzugrenzen ist (Bartsch et al., 2016; Giel, 2013; von Werthern, 2020). Es können aber für die spezifische Situation des Forschungsverbunds LemaS einige zentrale Punkte festgehalten werden:

- Im Gegensatz zu summativen Wirksamkeitsstudien steht in formativen Evaluationen die Frage der *praxistauglichen Realisierung* und damit auch die *Transferierbarkeit* einer Maßnahme im Fokus.
- Als *Evaluationsstrategie* zielt formative Evaluation nicht nur auf die reflektierende Begleitung, sondern auch auf die *Bewertung* einer Maßnahme. Daraus folgt ein Doppelcharakter von Prozess und Überprüfung: Einerseits dient formative Evaluation der stetigen Entwicklung und Verbesserung von Maßnahmen im Feld der Praxis, andererseits liefert sie (wenn auch vorläufige) Urteile bezüglich der realistisch (nicht unter Laborbedingungen) zu erwartenden Wirkungen, zu Bedingungen, die Erfolgsaussichten verbessern oder auch zu den Prozessen,

über die Maßnahmen ihre Wirksamkeit entfalten. Formative Evaluationen streben danach, nachzuvollziehen, wie und warum ein Instrument eigentlich wirkt oder eben nicht.

- Diese Bewertung kann von den involvierten Akteuren selbst vorgenommen werden – es handelt sich dann um ein Format der systematischen und kriterienbasierten Praxisreflexion, die sich aber zwangsläufig auf den jeweiligen konkreten und lokalen Kontext konzentriert. Der spezifische Beitrag einer *wissenschaftlichen* formativen Evaluation liegt in der Identifikation allgemeiner (jenseits der vorliegenden Einzelkontexte und in diesem Sinne abstrakter) Bedingungen einer erfolgreichen Einführung und Umsetzung einer Maßnahme. Die wissenschaftliche Bewertung einer Maßnahme zielt also auch in formativen Evaluationen auf Generalisierbarkeit.
- Während summative Evaluationen definitionsgemäß primär auf experimentelle Forschungsdesigns setzen, wird in formativen Evaluationen mit verschiedenen Methoden und Designs gearbeitet. Es geht dabei darum, fundierte Einsichten zu Aspekten zu gewinnen, die unter den künstlich kontrollierten Laborbedingungen eines (quasi-)experimentellen Vergleichs nicht zu erschließen sind. Konkret dienen (a) vorwiegend qualitativ arbeitende Fallstudien der Rekonstruktion der komplexen Wechselwirkungen zwischen Faktoren und Prozessen und (b) Formen der standardisierten Erhebung der Identifikation von Gelingensbedingungen über viele Praxiskontexte hinweg.
- Summative Evaluationen liefern in der Regel Effektstärken – also quantitative Maßzahlen. Formative Evaluationen müssen bezüglich der Form ihrer Bewertung andere Wege gehen. Zentral für die systematische Bewertung sind in diesem Fall inhaltlich bestimmte Kriterien, die numerischer Art sein können, üblicherweise aber eher diskursiv sind. Ein übliches Format zur Operationalisierung dieser Kriterien sind theoriegestützte Leitfragen, die den Prozess der reflexiven Entwicklung ebenso begleiten wie sie die „abschließende“ Bewertung strukturieren. Diese Fragen simulieren idealerweise die Entscheidungsprozesse, die mit der Ein- und Durchführung einer Maßnahme verbunden sind.
- Angemerkt sei, dass die wesentlich auf die komplexen Gegebenheiten eines realen Praxiseinsatzes von Maßnahmen abzielenden Fragen einer formativen Evaluation auch in späteren Entwicklungsphasen nicht an Bedeutung verlieren. Eher im Gegenteil stellen sich diese noch einmal neu, sobald eine Maßnahme für sich alleine ohne (quasi beaufsichtigende) Begleitung ein- und umgesetzt werden soll.

Die Formulierung von Leitfragen bietet eine Möglichkeit, formative wissenschaftliche Fundierung und Praxisrelevanz aktiv aufeinander zu beziehen. Der Katalog an Leitfragen sollte so konzis und vollständig wie möglich die Entscheidungen abbilden, die Schulen und/oder Lehrkräfte im Zuge der ernsthaften Implementation eines LemaS-Produkts treffen müssen. Im Folgenden werden Vorschläge für solche Leitfragen präsentiert, die sich aus den bisherigen Diskussionen im Forschungsverbund (unter anderem im Rahmen und am Rande der letzten Verbundtreffen) ergeben:

Auf welches praktische Problem gibt das Produkt eine Antwort und welche Ziele verfolgt es?
Praxisrelevant sind LemaS-Produkte, wenn sie für tatsächliche Praxisprobleme eine Lösung bieten.

Um den praktischen Nutzen einzelner Produkte zu unterstreichen, aber auch um einen unangemessenen (aus unpassenden Problemlagen motivierten) Praxiseinsatz zu vermeiden, sollten daher sowohl Probleme als auch Zielsetzungen für jedes Produkt klar benannt werden. Das eine (Probleme) hängt mit dem anderen (Ziele) zusammen, ist aber nicht zwingend deckungsgleich; das wäre von Einzelfall zu Einzelfall zu entscheiden. So kann eine Lehrkraft beispielsweise mit Unruhe im Klassenverband konfrontiert sein (= Problem) und zum Umgang mit diesem auf neue Arten der Differenzierung und Individualisierung im Unterricht setzen (= Ziel). Zu bedenken ist auch die Möglichkeit, dass das Ziel einer Maßnahme auch darin bestehen kann, für ein Problem zu sensibilisieren, dem in der pädagogischen Praxis bislang aus Sicht der Forschung zu wenig Beachtung geschenkt wird. Die Reflexion der Passung von Problemdefinitionen und Zielsetzungen wird in der internationalen Literatur jedenfalls durchgängig als entscheidend hervorgehoben. Um die Praxisrelevanz eines Produkts aufzuzeigen, ist es wichtig, dass hier auf mögliche Divergenzen Rücksicht genommen wird.

Welcher Erfolg ist realistisch zu erwarten? / Welche Wirkungen soll das Produkt haben? Neben den abstrakten Zielsetzungen zu unterscheiden sind die realen Erfolge, die im Verlauf der Produktnutzung erwartet werden können. Hier geht es also um die realistische Einschätzung des Mehrwerts – auch in Verbindung mit oder im Vergleich zu alternativen Strategien. Antworten auf diese Frage können sich aus aktueller Theorie und aus bereits existierenden Befunden ergeben. Zumindest exemplarisch sollten aber zum Ende der ersten Förderphase alle LemaS-Teilprojekte auch eigene empirische Befunde zu dieser Frage vorweisen können.

Welche Kosten sind mit Einführung und Einsatz des Produkts verbunden? Kosten und Nutzen einer pädagogischen Maßnahme sind nur schwierig umfassend und abschließend zu quantifizieren, es sollte aber zu einem erwartbaren Erfolg auch immer eine möglichst konkrete Einschätzung gegeben werden, welche Ressourcen für den Produkteinsatz aufzuwenden sind. Insbesondere Zeit- und Geldbedarf sollten insbesondere dann möglichst exakt benannt werden, wenn mit „versteckten“ Kosten zu rechnen ist (beispielsweise Aufwand für Folgebetreuung, für begleitende Kommunikation mit SuS oder im Kollegium etc.). Für die Beantwortung dieser Frage scheint ein Austausch zwischen den Teilprojekten in jedem Fall sinnvoll, und sei es nur, um an möglichst alle relevanten Kostenpunkte zu denken. Auch sollten hier die Erfahrungen und Einschätzungen der Schulleitungen (Informationen aus den Begleiterhebungen) mit jenen der Lehrkräfte und den Beobachtungen aus den empirischen Forschungstätigkeiten der Teilprojekte kombiniert betrachtet werden.

Woran erkenne ich die „Wirksamkeit“ des Produkts im Schulalltag bzw. wie stelle ich sie fest? Die (intendierten) Wirkungen eines Produkts können im pädagogischen Alltag nicht immer unmittelbar beobachtet werden. Für Lehrkräfte praktisch beobachtbare Erfolgsindikatoren können einerseits helfen, sich den Nutzen eines Produkts noch einmal aus einer anderen Perspektive konkret vorzustellen, andererseits können sie zu einem reflektierteren Einsatz beitragen. Die Antwort auf diese Frage hängt stark von der Art des Produkts ab, um das es geht, und muss daher für jedes Teilprojekt

individuell formuliert werden. Entscheidend ist, dass die Indikatoren auf ihre Praxistauglichkeit geprüft sind, sich also in der Arbeit mit den LemaS-Schulen bewährt haben.

Welche Faktoren fördern/hindern die erfolgreiche Implementation eines Produkts? / Welche Rahmenbedingungen müssen erfüllt sein? Hier geht es um notwendige Rahmen- und Gelingensbedingungen. Wie auch bei den anderen Fragen der Fall ist, kann in der ersten Förderphase keine abschließende empirische Prüfung dieser Bedingungen erfolgen. Daher geht es um Hinweise auf relevante Faktoren, die sich aus der Arbeit mit Schulen oder aus anderen Forschungszusammenhängen ergeben. Für diese Frage ist eine Kombination von verschiedenen Formen der empirischen Unterfütterung entscheidend. Manche Faktoren können am besten aus einer wissenschaftlichen Außenperspektive identifiziert werden, manche erfordern eher eine Innenperspektive. Manche sind offensichtlich, andere wirken unter der Oberfläche. Daher scheint einerseits eine Verschränkung von Beobachtungen etc. der Teilprojekte mit Daten aus den Begleiterhebungen nötig. Andererseits kann ein Austausch zwischen den Teilprojekten und eine Verschränkung von Befunden produktiv sein.

Welche (positiven wie negativen) Nebenwirkungen sind möglich? Jeder Eingriff in das komplexe Gefüge des schulischen Regelunterrichts hat mutmaßlich Folgen, die über die unmittelbar angestrebten Wirkungen hinausgehen. Diese Nebenfolgen können positiver wie negativer Art sein. Die Entscheidung für oder gegen ein Produkt wird in der Praxis seriöserweise auch mit Blick auf die zu erwartenden Nebenwirkungen getroffen (positiv zum Beispiel ein Mehr an Kommunikation im Kollegium oder die Möglichkeit, das innovative Profil einer Schule zu unterstreichen; negativ die Gefahr, dass bestimmte Schüler*innen abgehängt werden, Lehrkräfte sich überlastet fühlen etc.). Dabei kann es durchaus auch darum gehen, unbegründete Ängste vor solchen Nebenwirkungen zu entkräften und bislang nicht beachtete positive Aspekte ins Bewusstsein zu rücken. Nachdem Nebenwirkungen meist unintendiert/ungeplant auftreten, führt an möglichst vielfältigen Informationsquellen und das Herstellen von Querbezügen zwischen Projekten in diesem Fall kein Weg vorbei.

Empfehlungen für den Forschungsverbund LemaS:

- Um am Ende der ersten Förderphase die Instrumente aus verschiedenen Teilprojekten einheitlich und systematisch hinsichtlich ihrer Praxistauglichkeit bewerten zu können, sollten zeitnah Leitfragen für die formative Evaluation vereinbart werden. Diese Leitfragen werden aufgrund der Diversität der Instrumente relativ abstrakt sein müssen.
- Diese formativen Leitfragen spiegeln Bewertungskriterien wider, die sich an der Praxistauglichkeit und Transferierbarkeit orientieren. Sie sollen demnach die Entscheidungs- und Gestaltungsprozesse aufgreifen, die in der Praxis und für die Praxis auftreten, sobald eines der Instrumente implementiert werden soll. Auf Grundlage der bisherigen Erfahrungen in den Teilprojekten sollte dabei der konkreten Problemstellung, auf die mit einem LemaS-Instrument geantwortet werden soll, ausreichend Beachtung geschenkt werden.
- Die Berechnung isolierter Effektstärken ist in der aktuellen Phase weder möglich noch entspricht sie den Zielsetzungen (realer Praxistransfer, nicht Laborbedingungen). Die wissenschaftliche Bewertung erfolgt in der ersten Förderphase sinnvollerweise in Form theoretisch gestützter und empirisch plausibilisierter Einschätzungen. Der Forschungsverbund sollte zeitnah beginnen, dafür geeignete Vorlagen zu entwickeln, die sich auch für die Präsentation in Form einer Maßnahmendatenbank („Toolbox“) eignen. Die Einigung auf Leitfragen für die formative Evaluation der Produkte ist dafür einer der zentralen nächsten Arbeitsschritte.
- Je nach Zuschnitt eines Teilprojekts sind verschiedene Aspekte gut, andere nicht abgedeckt. Der Forschungsverbund sollte daher frühzeitig Formen und Wege festlegen, über die Verzahnungen und Verschränkungen zwischen Teilprojekten hergestellt werden können. Je nach Art des Produkts und den Orten und Formen seines Einsatzes sind Erkenntnisse aus einem Teilprojekt ceteris paribus auf andere Zusammenhänge übertragbar.
- Der dezidierten Praxisorientierung zum Trotz besteht die Aufgabe des Forschungsverbunds in der wissenschaftlich fundierten (methodisch kontrollierten, validen und generalisierbaren) Bewertung von Instrumenten. Die wissenschaftliche Bewertung setzt die Definition von Standards voraus. Diese sollten im FV diskutiert und aufgearbeitet werden.
- Um mit nach wie vor bestehenden Unsicherheiten zur Gestaltung der formativen Evaluation in einzelnen Teilprojekten umzugehen, sollten Forschungsstrategien, Methoden etc. breit im Forschungsverbund kommuniziert und geteilt werden.

5. Nächste Schritte

Auf Grundlage des präsentierten Problemaufrisses ergeben sich eine Reihe von konkreten Anregungen für die Gestaltung der kooperativen Forschungsaktivitäten in LemaS über die kommenden Projektphasen. Die folgende Liste ist als Sammlung von Vorschlägen zu verstehen – die Entscheidung über die Umsetzung einzelner Punkte und über die Prioritätensetzung zwischen ihnen muss im Forschungsverbund getroffen werden:

1. Aufbereitung aktueller internationaler Debatten: Um eine Positionierung und Profilierung der innovativen Forschung-Praxis-Bezüge in LemaS voranzutreiben, sollten die international aktuell diskutierten Fragen aktiv im Forschungsverbund aufgegriffen werden (beispielsweise in thematisch einschlägigen Workshops oder auch in Form von Publikationsprojekten).
2. Definition von Standards für die Gestaltung und Dokumentation von Forschungsaktivitäten: Um Teilprojekten den produktiven und kohärenten Umgang mit teilweise noch unvertrauten Forschungszugängen anderer Teilprojekte zu erleichtern und die Validität der wissenschaftlichen Fundierung zu gewährleisten, sollte der Forschungsverbund Leitlinien formulieren. Diese Leitlinien können einerseits als Hilfestellung für die Gestaltung von Forschungsprozessen dienen, andererseits sollten sie Vorgaben zur Dokumentation von Forschungsaktivitäten machen. Das scheint vor allem für gegenseitige Bezugnahmen und Verschränkungen entscheidend – nur auf Basis transparenter Dokumentation kann die Übertragbarkeit von Befunden aus den Teilprojekten angemessen bewertet werden.
3. Methodologische Reflexion und Systematisierung von begleitenden Aktivitäten: Im FV finden vielfältige Forschungsaktivitäten statt, die bislang nicht vollständig dokumentiert, methodologisch reflektiert und systematisiert sind. Um vorhandene Informationen und Daten sinnvoll und ausgiebig zu nutzen, sollten FV-übergreifend relevante Tätigkeiten an Schulen festgehalten und systematisch reflektiert werden. Zu denken ist beispielsweise an die Rolle von verschiedenen Arten von Feedback auf die Arbeit der Teilprojekte (von Lehrkräften, Schulleiter/innen, Schüler/innen etc.) oder auch an aktuell noch unstrukturierte Beobachtungen im Zuge von Schulbesuchen. Eine solche Dokumentation kann für die formative Evaluation von LemaS-Produkten hochrelevant sein – etwa wenn es um die Identifikation von „Wechsel- und Nebenwirkungen“ oder um die Benennung von möglichen Gelingensbedingungen geht.
4. Einigung auf Leitfragen für die formative Evaluation: Um eine praxisrelevante, fundierte und einheitliche Bewertung und Präsentation der Produkte aus den Teilprojekte zum Ende der ersten Förderphase zu ermöglichen, sollte der FV zeitnah Kriterien festlegen, nach denen evaluiert werden soll. Eine erste Auswahl an möglichen Fragen wird in Abschnitt 4 besprochen. Diese Fragen sind nicht zuletzt deswegen wichtig, weil in ihnen zwangsläufig die Brücke zwischen Praxisrelevanz und wissenschaftlicher Fundierung geschlagen wird: Sie sollen

empirisch fundierte Einschätzungen zu den wesentlichen Aspekten geben, die sich für die Praxis bei der Nutzung von LemaS-Produkten real ergeben.

5. Formate und Formen der Verschränkung von Forschungsbefunden aus den TP: Da jedes Teilprojekt für sich alleine genommen nur einen Ausschnitt der evaluationsrelevanten Aspekte abdecken kann, sollten zeitnah Wege der gegenseitigen Bezugnahme und der TP-übergreifenden wissenschaftlichen Bewertung gesucht werden. Relativ zeitnah sind beispielsweise Workshops zu Forschungsmethoden und -designs denkbar, in denen mit Problemstellungen und Beispielen aus den Teilprojekten gearbeitet wird. Aus mehreren Teilprojekten wurde bereits ein entsprechender Bedarf signalisiert. Im Rahmen der jährlichen Verbundtreffen könnte zudem ein Raum zur Präsentation von ersten empirischen Ergebnissen geschaffen werden. Entscheidend ist, dass diese Formen des Austauschs systematisch dokumentiert und wieder in den Forschungsverbund zurückgespielt werden. Auch in diesem Zusammenhang können die vorgeschlagenen Leitfragen für die formative Evaluation eine wichtige strukturierende Rolle spielen: Sie erlauben, Ergebnisse aus Einzelprojekten im Hinblick auf ihre breitere Relevanz für den Forschungsverbund einzuschätzen, sie ermöglichen, Befunde aus verschiedenen Projekten sinnvoll inhaltlich zu bündeln, und sie geben Teilprojekten Anhaltspunkte für die Aufbereitung ihrer Forschungsarbeiten für die Präsentation und Diskussion im Forschungsverbund.
6. Schulportraits als eine Möglichkeit der Verschränkung auf Schulebene: Kollaborativ erarbeitete individuelle Portraits der LemaS-Projektschulen stellen eine weitere Möglichkeit der gezielten Verschränkung von Informationen aus verschiedenen Teilprojekten dar. In solche Schulportraits können etwa Angaben zum Projektverlauf, zu Eigenheiten einer Schule, zu ihrer Ressourcenausstattung oder zu ihrer sozialen Zusammensetzung etc. einfließen. Für die formative Evaluation können solche Darstellungen vielfältig nutzbar sein (beispielsweise für die Identifikation relevanter Rahmenbedingungen oder für Beurteilung die Passung zwischen Problemwahrnehmung auf Schulebene und den Zielsetzungen von LemaS-Produkten).
7. Fokussierung der Zwischenerhebung auf Fallstudienelemente: Nachdem die Ausgangserhebung ein breites Spektrum an Aspekten über alle Projektschulen hinweg geliefert hat, sollte der Schwerpunkt der Zwischenerhebung auf der qualitativen Vertiefung mit Fokus auf hinderliche und förderliche Faktoren für die erfolgreiche Implementation von LemaS-Produkten liegen.

6. Ausblick zur Gestaltung der Begleitforschung in der Transferphase

In seinem aktuellen Zuschnitt erweist sich für die praxisrelevante Gestaltung des Forschungsverbunds die rigide Kopplung von einzelnen Teilprojekten und einzelnen Praxisprodukten als herausfordernd. Diese Struktur erschwert die mehrperspektivische Auseinandersetzung mit einzelnen Produkten ebenso wie den forschenden Vergleich zwischen Produkten. Aus heutiger Sicht liegt es daher nahe,

(1) mögliche zukünftige Projektstrukturen von Anfang an interdisziplinär und multimethodologisch auszurichten, (2) einzelne Teilstrukturen/Teilprojekte weniger ausschließlich an isolierte Instrumente/Produkte zu koppeln, (3) erhebungsgestütztes Monitoring, experimentelle Wirksamkeitsstudien und die fallstudienbasierte Begleitung der konkreten Praxisimplementierung als drei Kernsäulen praxisrelevanter Forschung miteinander zu verbinden und (4) Formen der organisierten Begleitung von Multiplikatorenschulen und Schulnetzwerken vorzuschlagen, in denen Wissenschaft, Praxis und begleitende administrative Strukturen mit klar definierten Rollen in Dialog treten und zusammenarbeiten (zum Beispiel in Form von „Entwicklungskonferenzen“).

Zitierte Quellen

- Bartsch, S., Beywl, W., Farrokhzad, S., Gutknecht-Gmeiner, M., Jelitto, M., Lück-Filsinger, M., Niestroj, M., Schmidt, S., Schobert, B., Schott, D., Schröder, U. B., Sperling, M., & Uzunova, E. (2016). *Evaluationspraxis: Professionalisierung - Ansätze - Methoden* (S. Giel, K. Klockgether, & S. Mäder, Hrsg.; 2., korrigierte und ergänzte Auflage). Waxmann.
- Biesta, G. J. J. (2010). Why 'What Works' Still Won't Work: From Evidence-Based Education to Value-Based Education. *Studies in Philosophy and Education*, 29(5), 491–503.
<https://doi.org/10.1007/s11217-010-9191-x>
- Bjerg Petersen, K., Reimer, D., Qvortrup, A., @Aarhus University, & Department of Education. (2014). *Evidence and evidence-based education in Denmark: The current debate*. Department of Education, @Aarhus University.
- Bryk, A. S. (2015). 2014 AERA Distinguished Lecture: Accelerating How We Learn to Improve. *Educational Researcher*, 44(9), 467–477. <https://doi.org/10.3102/0013189X15621543>
- Cartwright, N. (2019). What is meant by “rigour” in evidence-based educational policy and what’s so good about it? *Educational Research and Evaluation*, 25(1–2), 63–80.
<https://doi.org/10.1080/13803611.2019.1617990>
- Coburn, C. E., & Penuel, W. R. (2016). Research–Practice Partnerships in Education: Outcomes, Dynamics, and Open Questions. *Educational Researcher*, 45(1), 48–54.
<https://doi.org/10.3102/0013189X16631750>
- Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research*, 60(3), 276–291. <https://doi.org/10.1080/00131881.2018.1493353>
- Derouet, J. L. (2019). Die Soziologie der Konventionen im Bereich der Bildung. Wissenschaft, Politik und Gesellschaftskritik in Frankreich am Übergang vom 20. Ins 21. Jahrhundert. In C. Imdorf, R. J. Leemann, & P. Gonon (Hrsg.), *Bildung und Konventionen* (S. 47–90). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23301-3_2
- De Vaus, D. A. (2001). *Research design in social research*. SAGE.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th edition). Wiley.

- Firestone, W. A. (1993). Alternative Arguments for Generalizing From Data as Applied to Qualitative Research. *Educational Researcher*, 22(4), 16–23.
<https://doi.org/10.3102/0013189X022004016>
- Giel, S. (2013). *Theoriebasierte Evaluation: Konzepte und methodische Umsetzungen*. Waxmann.
- Gough, D. (2015). Qualitative and mixed methods in systematic reviews. *Systematic Reviews*, 4(1), 181, s13643-015-0151-y. <https://doi.org/10.1186/s13643-015-0151-y>
- Hammersley, M. (2002). *Educational research, policymaking and practice*. P. Chapman.
- Harden, A., & Thomas, J. (2005). Methodological Issues in Combining Diverse Study Types in Systematic Reviews. *International Journal of Social Research Methodology*, 8(3), 257–271.
<https://doi.org/10.1080/13645570500155078>
- Harden, A., & Thomas, J. (2010). Mixed Methods and Systematic Reviews: Examples and Emerging Issues. In A. Tashakkori & C. Teddlie, *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (S. 749–774). SAGE Publications, Inc.
<https://doi.org/10.4135/9781506335193.n29>
- Hargreaves, D. H. (1997). In Defence of Research for Evidence-based Teaching: A rejoinder to Martyn Hammersley. *British Educational Research Journal*, 23(4), 405–419.
<https://doi.org/10.1080/0141192970230402>
- Joyce, K. E. (2019). The key role of representativeness in evidence-based education. *Educational Research and Evaluation*, 25(1–2), 43–62. <https://doi.org/10.1080/13803611.2019.1617989>
- Kirk, R. E. (1993). *Experimental design: Procedures for the behavioral sciences* (2. ed., [Nachdr.]). Brooks/Cole.
- Kvernbekk, T. (2019). Practitioner tales: Possible roles for research evidence in practice. *Educational Research and Evaluation*, 25(1–2), 25–42. <https://doi.org/10.1080/13803611.2019.1617988>
- Ladwig, J. G. (2018). CRITIQUE: On the Limits to Evidence-Based Learning of Educational Science. In G. E. Hall, L. F. Quinn, & D. M. Gollnick (Hrsg.), *The Wiley Handbook of Teaching and Learning* (S. 639–659). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118955901.ch27>
- Little, D. (1993). On the scope and limits of generalizations in the social sciences. *Synthese*, 97(2), 183–207.
- McKenney, S., & Schunn, C. D. (2018). How can educational research support practice at scale? Attending to educational designer needs. *British Educational Research Journal*, 44(6), 1084–1100. <https://doi.org/10.1002/berj.3480>
- Normand, R. (2016). *The changing epistemic governance of European education: The fabrication of the homo academicus Europeanus?* <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1203682>
- Normand, R., Liu, M., Carvalho, L. M., Oliveira, D. A., & Levasseur, L. (Hrsg.). (2019). *Education policies and the restructuring of the educational profession: Global and comparative perspectives* (1st edition 2019). Springer Singapore.
- Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, 47(11), 1451–1458.
<https://doi.org/10.1016/j.ijnurstu.2010.06.004>
- Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. *Unterrichtsforschung*, 33, 52–69.
- Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy*, 32(4), 450–466.
<https://doi.org/10.1080/02680939.2017.1280183>

- Simpson, A. (2018). Princesses are bigger than elephants: Effect size as a category error in evidence-based education. *British Educational Research Journal*, 44(5), 897–913. <https://doi.org/10.1002/berj.3474>
- Simpson, A. (2019). Separating arguments from conclusions: The mistaken role of effect size in educational policy research. *Educational Research and Evaluation*, 25(1–2), 99–109. <https://doi.org/10.1080/13803611.2019.1617170>
- The Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32(1), 5–8. <https://doi.org/10.3102/0013189X032001005>
- Thomas, G. (2016). After the Gold Rush: Questioning the “Gold Standard” and Reappraising the Status of Experiment and Randomized Controlled Trials in Education. *Harvard Educational Review*, 86(3), 390–411. <https://doi.org/10.17763/1943-5045-86.3.390>
- von Werthern, A. (2020). *Theoriebasierte Evaluation: Entwicklung und Anwendung eines Verfahrensmodells zur Programmtheoriekonstruktion*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-27579-2>
- William, D. (2019). Some reflections on the role of evidence in improving education. *Educational Research and Evaluation*, 25(1–2), 127–139. <https://doi.org/10.1080/13803611.2019.1617993>
- Wrigley, T. (2018). The power of ‘evidence’: Reliable science or a set of blunt tools? *British Educational Research Journal*, 44(3), 359–376. <https://doi.org/10.1002/berj.3338>
- Wrigley, T., & McCusker, S. (2019). Evidence-based teaching: A simple view of “science”. *Educational Research and Evaluation*, 25(1–2), 110–126. <https://doi.org/10.1080/13803611.2019.1617992>
- Yin, R. K. (2014). *Case study research: Design and methods* (Fifth edition). SAGE.
- Ylikoski, P. (2018). Mechanism-based theorizing and generalization from case studies. *Studies in History and Philosophy of Science Part A*, S0039368117303175. <https://doi.org/10.1016/j.shpsa.2018.11.009>
- Zhao, Y. (2017). What works may hurt: Side effects in education. *Journal of Educational Change*, 18(1), 1–19. <https://doi.org/10.1007/s10833-016-9294-4>