

# **IQB und SWK und ihr Beitrag zur Reform der Grundschule – Befindet sich die evidenzbasierte Reform in der Krise oder führt sie aus ihr heraus?**

**Torsten Eckermann**

Angenommen: 28. Mai 2024 / Online publiziert: 14. Juni 2024  
© The Author(s) 2024

**Zusammenfassung** Am Beispiel der IQB-Bildungstrend 2021 wird die Rolle des nationalen Bildungsmonitorings für das Initiieren von Reformen untersucht. Entfaltet wird die These, dass sich das IQB nicht allein als Datenlieferant begreift, sondern gemeinsam mit dem Beratungsgremium (SWK) als Reformakteur in Erscheinung tritt. Die Analyse des anvisierten Reformprogramms legt nahe, dass die Maßnahmen einer evidence-based-policy folgen. Es wird eine Expansion des Monitorings über die verschiedenen Bildungsetappen (von frühkindlicher Bildung bis zur Lehrkräfteprofessionalisierung) angestrebt. Für die Grundschulforschung bieten die Reformabsichten wichtige Impulse und Diskussionsanstöße. Sie sind zugleich einer theoretisch und empirisch informierten Analyse, insbesondere im Hinblick auf nicht-intendierte Nebenfolgen, zu unterziehen. Die Reform strebt eine disparitätsmindernde Wirkung der Grundschule an und imaginiert einen zentralen Hebel (Evidenz, Diagnostik, Standardisierung) und eine pädagogische Monokultur. Unterstellt wird eine Homogenität (alle Kinder profitieren von mehr Diagnostik und Testung). Entgegen der Intention zur Verringerung des Anteils an „Risikokinder“ beizutragen, besteht jedoch die Gefahr ungewollt mehr „Risikokinder“ zu produzieren. Damit läuft die anvisierte Reform Gefahr, die eigentliche „Risikogruppe“ aus dem Auge zu verlieren, die begründet Anlass für Reformen gibt.

**Schlüsselwörter** Bildungsmonitoring · Evidenzbasierte Reform · Grundschule · Chancengleichheit

---

✉ Prof. Dr. Torsten Eckermann  
Institut für Erziehungswissenschaft, Pädagogische Hochschule Freiburg, Freiburg, Deutschland  
E-Mail: [Torsten.Eckermann@ph-freiburg.de](mailto:Torsten.Eckermann@ph-freiburg.de)

## **IQB and SWK and their contribution to primary school reform – Is the evidence-based reform in crisis or is it leading out of it?**

**Abstract** The paper examines the role of the national school monitoring system in initiating educational reforms. It is assumed that the IQB is not only a data provider but rather appears as a reform actor together with the advisory board (SWK). The analysis of the reform program suggests that the measures follow an evidence-based policy. Expansion of monitoring across the different stages of education (from early childhood education to teacher professionalization) are envisioned. For primary school research, the reform intentions offer important impulses and stimuli for discussion. At the same time, they must be subjected to a theoretically and empirically informed analysis, especially with regard to unintended side effects. The reform aims to reduce disparities in primary school and imagines a central lever (evidence, diagnostics, standardisation) and a pedagogical monoculture. Homogeneity is assumed (all children benefit from more diagnostics and testing), which can unintentionally produce more “at-risk children”. However, the envisaged reform runs the risk of losing sight of the actual “at-risk group”, which is a justified reason for reform.

**Keywords** Educational monitoring · Evidence-based reform · Primary school · Equal opportunities

### **1 Einleitung**

Die Veröffentlichung der Ergebnisse des nationalen und internationalen Bildungsmonitorings hat sich mittlerweile fest etabliert. Mit der kontinuierlichen Bildungsberichterstattung wird das Ziel verfolgt, empirisch gesicherte Informationen über die Qualität des Bildungssystems zu gewinnen. Seit der PISA-Studie im Jahr 2001 sind verschiedene Aspekte Gegenstand der Debatte: Forschungsmethodik (z. B. Repräsentativität: Stanat und Pant 2022), der Qualitätsbegriff (Sauerwein und Klieme 2016) sowie das Verhältnis von Bildungsforschung und Politik. Kontrovers wird diskutiert, inwieweit die Ansprüche des Bildungsmonitorings datengestützt relevantes Handlungswissen für Politik und pädagogische Praxis bereitzustellen und Innovationen zu befördern, auch eingelöst werden (vgl. Maritzen 2018; Prenzel et al. 2008; Stanat et al. 2013). Angesichts der Ergebnisse aus dem IQB-Bildungstrend 2021 erfährt diese Frage erneut eine besondere Relevanz, weshalb sich im vorliegenden Diskussionsbeitrag der Frage näher gewidmet werden soll.

Der Beitrag beleuchtet am Beispiel der IQB-Bildungstrend 2021 und dem sich daran anschließenden Gutachten der Ständigen Wissenschaftlichen Kommission (SWK) die durch das datengestützte Bildungsmonitoring anvisierten Reformmaßnahmen für die Grundschule. Zentral ist dabei die These, dass sich das Bildungsmonitoring nicht allein auf die deskriptive Beschreibung der Daten („*Beschreibungswissen*“: Bromme et al. 2014) und das Generieren von Evidenz beschränkt, sondern gemeinsam mit dem Beratungsgremium (SWK) als Reformakteur („*Veränderungswissen*“: Bromme et al. 2014) in Erscheinung tritt. Daraus resultieren jedoch andere Anforde-

rungen an die Aussagekraft der Daten und deren Reichweite: Wenn das Anliegen nicht allein darin besteht, Ergebnisse als ‚nüchterne‘ Bestandsaufnahme in einen *akademischen Diskurs* einzuspeisen, sondern auch *bildungspolitische* Maßnahmen evidenzbasiert abzuleiten, besteht die Gefahr, wissenschaftliche Kriterien (wie z. B. die Vorläufigkeit und fehlende Eindeutigkeit der Ergebnisse) zugunsten einer für die bildungspolitische Verwertung der Ergebnisse notwendigen Eindeutigkeit und Konsistenz zu vernachlässigen. Der Beitrag diskutiert vor diesem Hintergrund die intendierten Wirkungen und nicht-intendierten Nebenwirkungen einer evidenzbasierten Reform der Grundschule. Hierzu erfolgt zunächst eine kurze Skizze zur Evidenzorientierung (2.), bevor dann das anvisierte Reformprogramm von IQB und SWK der Analyse unterzogen wird (3.). Der Beitrag schließt mit einer Diskussion der Befunde (4.).

## 2 Evidenzorientierung und Reform der (Grund-)Schule

### 2.1 „Follow the Science“? – Evidenzorientierung als Leitmaxime zwischen Wissenschaft und Alltag

Die Orientierung an wissenschaftlichem Wissen nimmt in vielen Lebensbereichen eine zentrale Rolle ein (z. B. medizinisches Wissen zu Gesundheit, Ernährung und Familienplanung, Zillien 2013). Dies zeigen auch die Debatten um den Klimawandel oder die Pandemie. Was als hinreichend ‚evident‘, d. h. wissenschaftlich gesichert gilt, ist nicht nur Gegenstand des akademischen Diskurses; vielmehr nimmt es auch im Alltag eine wichtige Rolle ein, um Entscheidungen treffen und Risiken bewerten zu können (z. B. *Kauf eines Unkrautvertilgungsmittel*), die wissenschaftliches Wissen voraussetzen (z. B. *Erhöht Glyphosat das Krebsrisiko?*, vgl. Bromme und Kienhues 2017).

Auch innerwissenschaftlich hat der Evidenzbegriff Konjunktur, gleichwohl keineswegs von einem einheitlichen Verständnis gesprochen werden kann (Besa et al. 2023). Differente Evidenzarten (z. B. „*strong evidence*“, „*hard evidence*“), deren forschungsmethodische Hervorbringungen (z. B. experimentelle Studien, Meta-Analysen, Leistungsvergleichsstudien) sowie die sich aus ihnen ableitbaren Schlussfolgerungen (z. B. handlungsleitend vs. handlungsunterstützend) sind Gegenstand wissenschaftlicher Kontroversen (Wilkes und Stark 2023; Biesta 2011). In der empirischen Bildungsforschung wird unter dem Evidenzbegriff jenes wissenschaftliche Wissen subsumiert, das mit Hilfe nachvollziehbarer methodischer Vorgehensweisen produziert wird und sich somit von der anekdotischen Evidenz des Alltags, einem personengebundenen Erfahrungswissen sowie den sich darum kreisenden (Bildungs-)Mythen abgrenzt (Bauer und Kollar 2023; Scharlau 2019). Evidenz stellt in diesem Sinne ein Wissen dar, welches nicht notwendigerweise allgemein zugänglich ist, sondern erst mit Hilfe wissenschaftlicher Methoden erzeugt wird und auf diese Weise die Legitimität wissenschaftlichen Wissens sichert (Forster 2014).

Insbesondere seit der empirischen Wende in der Bildungspolitik (Marttila 2014) soll dieses Wissen das Fundament bildungspolitischer Reformmaßnahmen sein. Es trägt dabei das Versprechen der Aufklärung, des Fortschritts und der Innovation in

sich (Heinrich 2015) und stellt eine „transnational verbreitete Form der Regierung der Bildung“ (Marttila 2014, S. 259) dar. Insbesondere in Großbritannien und den USA haben Reformen, die unter Nutzung von wissenschaftlicher Evidenz initiiert werden, Konjunktur (vgl. *standard-based education reform*, Kim 2018; *data-driven reform*, Bertrand und Marsh 2021); aber auch in Deutschland ist die Evidenznutzung integraler Bestandteil schulischer Qualitätssicherung und erfüllt eine wichtige Steuerungsfunktion (Bellmann 2017).

## 2.2 Evidenzbasierte Reform der (Grund-)Schule

Mit dem Ausbau des nationalen und internationalen Bildungsmonitorings<sup>1</sup> hat im Hinblick auf das Initiieren von Schulreformen ein bedeutsamer Wandel stattgefunden (Ditton 2019; Tillmann 2017). Auf der Grundlage systematisch und nach wissenschaftlichen Qualitätsstandards generierten Daten sollen der Bildungspolitik belastbare, d. h. wissenschaftlich abgesicherte Informationen bereitgestellt werden, die politische Entscheidungen über Reformmaßnahmen empirisch untermauern und zielgerichtet anleiten können (vgl. Bellmann 2017). Nach dem Vorbild der Medizin sollen bildungspolitische Maßnahmen aus nachweisbaren empirischen Belegen („*what works*“) abgeleitet werden. Diese datengestützte (Weiter-)Entwicklung des Bildungssystems ist politisch mit der Gesamtstrategie der KMK (2015) besiegelt. Das Monitoring erfolgt in regelmäßigen Abständen, international etwa durch die OECD (PISA), die IEA (IGLU) und national durch das IQB (IQB-Bildungstrend). Diese transnational agierenden Akteure kennzeichnet das gemeinsame Steuerungsverständnis „*governance by numbers*“ (Hartong 2014). Ziel ist es demnach, evidenzbasiert, d. h. mit Hilfe von Daten und Indikatoren, Aussagen zur Qualität und Leistungsfähigkeit des Bildungssystems und zur Wirksamkeit bildungspolitischer Maßnahmen zu treffen, wobei die Leistungstestdaten dabei die zentrale Grundlage der Steuerung durch Standards darstellen (Grünkorn et al. 2019)<sup>2</sup>.

Trotz ihrer mittlerweile festen Etablierung sind die Studien des Bildungsmonitorings nicht unumstritten. Kontrovers werden forschungsmethodische Voraussetzungen (vgl. Maritzen 2014; Brügelmann 2017), ihre Anschlussfähigkeit an das Konzept der Allgemeinbildung (Frohn und Heinrich 2018) sowie der Anspruch, ein handlungsrelevantes Wissen für Politik und pädagogische Praxis bereitzustellen, diskutiert (Tillmann 2017). Befürworter:innen sehen den Mehrwert des Monitorings darin, dass für mehr Transparenz gesorgt wird, eine systematische Gesamtschau über das Bildungssystem vorliegt und durch sorgfältige Evidenzabschätzung der Anspruch realisiert wird, steuerungsrelevantes Wissen für Bildungspolitik zu generieren (Schrader et al. 2020; Bromme et al. 2014). Es lässt sich im Hinblick auf das Monitoring aber auch Kritik vernehmen. Beispielsweise wird moniert, dass das Monitoring aufgrund der Aggregationsebene der Daten in der Vogelperspektive ver-

<sup>1</sup> Bildungsmonitoring wird hier als Sammelbegriff für die systematische und kontinuierliche Datenerhebung wie bei der IGLU-Studie (international) oder den IQB-Bildungstrends (national) verwendet.

<sup>2</sup> Für die Koordination der standardbasierten Leistungsmessung ist das IQB zuständig, das offiziell eine Organisation des politischen Ländergremiums (KMK) und gleichzeitig eine wissenschaftliche Einrichtung der Humboldt-Universität zu Berlin ist.

harrt und damit nicht zur Mikroebene des Unterrichts und den Lernenden vordringt (Oelkers und Reusser 2008). Zudem wird in Frage gestellt, ob sich mit Hilfe empirischer Daten *eindeutige* Reformmaßnahmen ableiten lassen, da aus „jeder Messung sich unzählige Konsequenzen“ (Ladenthin 2013, S. 63) ziehen lassen und empirisch gewonnenes Wissen notwendigerweise *relativ* sei, da es stets davon abhängt, wie es *interpretiert* werde (vgl. Interpretationen von PISA-Daten: Klemm 2017, S. 168).

Bislang fehlt es jedoch weitgehend an empirischen Arbeiten zu dem im Rahmen des nationalen Bildungsmonitorings wissenschaftlich gewonnenen Wissen, das für die Ableitung von Reformmaßnahmen genutzt wird (vgl. Hermstein et al. 2018). Hier knüpft nun die nachfolgende Ex-ante-Programmanalyse an.

### 3 IQB und SWK und ihr Beitrag zur Reform der Grundschule – eine Programmanalyse

#### 3.1 Methodische Einordnung: Ex-ante-Programmanalyse

Die nachfolgenden Analysen stützen sich auf eine *Ex-ante-Programmanalyse* (vgl. Stockmann und Ertl 2021). Dieses Evaluationsinstrument findet ursprünglich in der Stadtplanung, bei Gesetzgebungsverfahren oder auch in der Technikentwicklung Anwendung (vgl. Sackmann 2021). Dort wird es mit dem Ziel eingesetzt, eine Folgeabschätzung über intendierte und nicht-intendierte Wirkungen von Maßnahmen bereits im Planungsprozess, d. h. vor ihrer Implementation, vorzunehmen. Die Programmanalyse zielt darauf ab, die Passung der Bedarfe und Programmziele sowie die aus dem Programm abgeleiteten Maßnahmen im Hinblick auf intendierte und nicht-intendierte Wirkungen zu analysieren (SWK 2022). Gegenstand der Analyse sind dabei verschriftlichte und veröffentlichte Dokumente (z. B. Berichte, Gutachten), die nicht nur Informationen dokumentieren, sondern als methodisch gestaltete Kommunikationszüge auf spezifische Weise organisieren, um sie als realitätsangemessene Beschreibungen und mit Absichten ausstatten zu können (Wolff 2000). Für die nachfolgende Analyse des Reformprogramms wurden die zentralen Dokumente, d. h. die IQB-Bildungstrend 2021 und das darauf Bezug nehmende Gutachten der Ständigen Wissenschaftlichen Kommission (SWK) herangezogen.

##### 3.1.1 Ergebnisse IQB-Bildungstrend 2021: Problembeschreibung und angestrebte Reformziele

Der IQB-Bildungstrend 2021 wurde am 17. Oktober 2022 vorgestellt. Ein darauf Bezug nehmendes Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK)<sup>3</sup> wurde wenig später am 09. Dezember 2022 ver-

---

<sup>3</sup> Bei der SWK handelt es sich um ein „unabhängiges Beratungsgremium der Kultusministerkonferenz“, dem 16 Bildungsforscher:innen angehören. Zu den Aufgaben gehören – wie auch auf der Homepage der KMK nachzulesen ist – die Weiterentwicklung des Bildungswesens, was die Identifikation bestehender Probleme sowie deren Lösung durch evidenzbasierte Empfehlungen beinhaltet. Die Empfehlungen der SWK basieren u. a. auf Erkenntnissen des Bildungsmonitorings (z. B. durch das IQB).

öffentlich. Im Ergebnisbericht wird als zentraler Befund hervorgehoben, dass der Anteil der Viertklässler:innen, die „nicht einmal die Mindeststandards erreichen“ bei „fast 19 % im *Lesen*, gut 18 % im *Zuhören*, etwa 30 % in der *Orthographie* und fast 22 % im Fach *Mathematik*“ liege (Stanat et al. 2022, S. 281). Der IQB-Bildungstrend erfüllt damit die Funktion des Überbringers unangenehmer Wahrheiten. Ihre politische Brisanz, die im Text deutlich artikuliert wird (z. B. „besorgniserregend“, S. 281) entfalten die statistischen Kennwerte nicht nur für sich, sondern vor allem *in Relation* (Trendanalysen). Damit verfährt das IQB – analog zur OECD – nach dem Prinzip „*governance by comparison*“ (Martens und Niemann 2010), d. h. über den Vergleich, der häufig auch durch Rankings veranschaulicht wird. Der Trend, hinter dem sich ein Kohortenvergleich verbirgt, deutet auf ungünstige Entwicklungen hin. Eindeutige Erklärungen liefert das IQB dazu nicht. Stattdessen wird ein Bündel heterogener Erklärungen aufgezählt (z. B. pandemiebedingte Effekte, veränderte Zusammensetzung der Schüler:innenschaft, Weiterentwicklungen curricularer Vorgaben). Trotz fehlender Ursachenklärung wird ein politisches Handeln nahelegt:

„Da die Resultate ‚nicht hinnehmbar‘ seien, müssten ‚gezielte Anstrengungen unternommen werden, um die Bildungsqualität in der Breite zu erhöhen, mit besonderem Augenmerk auf die Sicherung der Mindeststandards bei denjenigen Schüler:innen, die aufgrund ungünstigeren Lernausgangslagen und Lernbedingungen einem besonderen Risiko ausgesetzt sind, abgehängt zu werden““ (ebd., S. 282).

Aus der empirischen Evidenz wird ein Reformziel abgeleitet. Ziel ist die Niveauanhebung in der *Breite* und die gezielte Förderung der *Risikogruppe*. Ein Zielkonflikt wird damit vermieden (z. B. einseitige Bevorzugung der „Risikogruppe“). Das Reformziel erweist sich aber nur bedingt passförmig zum alarmistisch dargestellten Problembedarf (fehlende Basiskompetenzen bei fast 20 % der Schüler:innen). Dieser Bedarf würde eine bildungspolitische Priorisierung der „Risikogruppe“ (z. B. Aufholprogramm für Risikoschüler:innen)<sup>4</sup> legitimieren, um langfristige Folgen abzuwenden. Denn geht man von den kumulierenden, d. h. als Folgeprobleme über die Bildungsbiographie erwachsenden Risiken für Kinder mit nicht hinreichenden Basiskompetenzen aus, liegt – unter einer politischen Optik betrachtet – eine Priorisierung nahe. Das Ausbleiben einer solchen Gewichtung lässt sich vermutlich aber damit erklären, dass die durch das Bildungsmonitoring initiierte Reform – so die These – im Dienst der Leistungsfähigkeit des Bildungssystem steht. Mit anderen Worten: Es setzt sich das meritokratische Leistungsprinzip durch, wonach Leistungsunterschiede zwischen „Risikogruppe“ und „Nicht-Risikogruppe“ legitimiert sind, solange sie durch das meritokratische Leistungsprinzip gedeckt werden.

Ein detaillierter Blick in die Empfehlungen der Ständigen Wissenschaftlichen Kommission (SWK) soll hierzu aber noch nähere Informationen liefern.

<sup>4</sup> In Hamburg wird eine solche gezielte Förderung für *Risiko-Schüler:innen* mit dem Förderprogramm „Mathe sicher können“ (<https://mathe-sicher-koennen.dzlm.de/>) durchgeführt.

**Tab. 1** Übersicht über die Handlungsfelder des Reformprogramms

Empfehlung	Reformziel	Kategorie
<i>1. Handlungsfeld</i>		
1	Fortbildungen stärker an evidenzbasierten Ansätzen ausrichten	Evidenz/ Qualifizierung
3	Implementation einer frühen flächendeckenden Diagnostik	Evidenz/ Diagnostik
6	Verbindliche Verankerung eines Konzepts zur systematischen Diagnose und Förderung basaler Kompetenzen	Evidenz/ Diagnostik
7	Erhöhung der Unterrichtsqualität durch die Nutzung von VERA-Daten für die Unterrichtsentwicklung	Evidenz/ Ressourcen
8	Konsequente Umsetzung und Weiterentwicklung der Gesamtstrategie der KMK zum Bildungsmonitoring	Evidenz
9	Bereitstellung von wissenschaftlich fundierten, qualitätsgesicherten diagnostischen Instrumenten	Evidenz/ Diagnostik
10	Systematische Diagnose mit bedarfsorientierten, forschungs-basierten Maßnahmen zur Förderung sozialer Integration und sozial-emotionaler Kompetenzen	Evidenz/ Diagnostik
16	Implementation forschungsbasierter Fortbildungsprogramme zur diagnosebasierten Förderung	Evidenz/ Diagnostik
17	Koordination der datenbasierten Formulierung von Entwicklungszielen	Evidenz/ Kooperation
18	Entwicklung von Strukturen und Gewährung von zeitlichen Ressourcen für eine datenbasierte Schulentwicklung	Evidenz/ Ressourcen
19	Indexbasierte Zuweisung zusätzlicher Ressourcen auf allen Ebenen	Evidenz/ Ressource
<i>2. Handlungsfeld</i>		
4	Stärkung der Teilhabe an frühkindlicher Bildung für alle Kinder	Kita/Ressourcen
2	Verbesserung des Fachkraft-Kind-Schlüssels	Ressourcen
<i>3. Handlungsfeld</i>		
15	Gezielte Gewinnung und Qualifizierung von Fachleiter:innen und Mentor:innen in der zweiten Phase	Qualifizierung/ Lehrkräftebildung
14	Ländergemeinsame Entwicklung eines kohärenten phasenübergreifenden Kerncurriculums für Lehrkräfte	Qualifizierung/ Lehrkräftebildung
<i>4. Handlungsfeld</i>		
12	Entwicklung eines im Schulprogramm verankerten Konzepts für die Zusammenarbeit mit Eltern	Kooperation/ Eltern
5	Verankerung elternbildender Maßnahmen der Zusammenarbeit von Kindertageseinrichtungen und Familien	Kooperation/ Eltern
13	Verankerung des Kooperationsauftrags von Lehrkräften	Kooperation/ Lehrkräfte
11	Unterstützung bei der Diagnose von Auffälligkeiten in der sozial-emotionalen Entwicklung durch Schulpsychologie bzw. sonderpädagogischen Zentren	Kooperation/ Diagnostik
<i>5. Handlungsfeld</i>		
20	Entwicklung von Strategien zur Reduzierung von Segregationstendenzen	Segregation

## 3.2 Analyse des Reformprogramms

Die nachfolgende Ex-ante-Programmanalyse widmet sich den Empfehlungen der SWK, die in ihrem Gutachten auf die IQB-Bildungstrend unmittelbar Bezug nimmt. Um sich einen Überblick über das anvisierte Reformprogramm verschaffen zu können, wurden in einem ersten Schritt die Reformziele kategorisiert (vgl. Tab. 1). Im Folgenden werden die ersten beiden Handlungsfelder kurz skizziert. Aufgrund des begrenzten Raums wird dabei eine zusammenfassende Darstellung zentraler Aspekte gewählt.

### 3.2.1 Evidenz

Betrachtet man die für die Grundschule formulierten Reformziele, so lässt sich der Eindruck gewinnen, dass ein ausführlicher Maßnahmenkatalog präsentiert wird, der nicht die isolierte Einzelmaßnahme in den Vordergrund rückt. Bei genauerem Hinsehen offenbart sich jedoch eine eindeutige Dominanz: Über die Hälfte der Maßnahmen zielen auf eine stärkere *Evidenzorientierung* – womit die SWK dem Anspruch einer *evidence-based policy* folgt (Heinrich 2010; Jornitz 2008). Es wird eine Reform der Grundschule anvisiert, die sich gegenüber Ideologisierung zu immunisieren versucht: *Daten*, nicht *pädagogische Ideale*, sollen Entscheidungsgrundlage für die Reform sein.

Implizit wird auch das schlechte Abschneiden der Grundschüler:innen mit einem *Evidenzdefizit* in Verbindung gebracht – andernfalls wäre dies als Reformempfehlung obsolet. Bemerkenswert ist diese Defizitdiagnose vor dem Hintergrund, dass das Bildungsmonitoring mit dem Anspruch angetreten ist, durch die regelmäßige Erfassung über längere Zeiträume (IQB-Bildungstrends: *10 Jahre*, IGLU: *20 Jahre*) zur Beseitigung des Mangels an Evidenz beizutragen. Müsste nach diesem Zeitraum nicht hinreichend Evidenz vorliegen? Diese Reformempfehlung liefert aber auch aus einem anderen Grund Anstoß zur Kritik, denn sie erscheint empirisch nur wenig belastbar. So wird in den IQB-Bildungstrend 2021 von den Autor:innen selbst problematisiert, dass die Daten eindeutige Ursachenzuschreibungen nicht zulassen. Beispielhaft zeigt sich diese Problematik bei der Einordnung zum guten Abschneiden Hamburgs:

„Ob dies mit der Strategie einer datengestützten Schul- und Unterrichtsentwicklung zu tun hat, die das Land in mehr als 20 Jahren etabliert und konsequent weiterentwickelt hat, lässt sich nicht mit Sicherheit feststellen. Es erscheint jedoch plausibel, dass in einem System, in dem die Entwicklung zentraler Rahmenbedingungen, Verläufe und Ergebnisse schulischer Bildungsprozesse auf den verschiedenen Akteursebenen kontinuierlich beobachtet wird, auf Veränderungen gezielter reagiert und bei sich abzeichnenden Problemlagen frühzeitiger interveniert werden kann“ (Stanat et al. 2022, S. 282).

Bei der erhofften Evidenz handelt es sich somit weniger um eindeutige Belege als vielmehr um *Evidenzerwartungen*. Die als „*plausibel*“ ausgewiesene Evidenz baut demnach auf Erwartungen auf, deren Überprüfung jedoch in die Zukunft verschoben wird (vgl. Emmerich 2019).



### 3.2.2 Frühkindliche Bildung als Frühwarnsystem

Einen weiteren Schwerpunkt nimmt in den SWK-Reformempfehlungen die *vorschulische, frühkindliche Bildung* ein. Bereits vor Eintritt in die Grundschule soll die Beseitigung der Chancenungleichheit ansetzen. Dies steht im Einklang mit der international zu beobachtenden Ausweitung des Monitorings für die frühkindliche Bildung (vgl. Böttcher et al. 2014; Hartong 2020). Die frühkindliche Bildung soll als Frühwarnsystem dem *präventiv* vorbeugen, was die Grundschule *nicht kompensiert*.

Diese politisch wohlwollende Reformabsicht ist bei näherem Hinsehen äußerst voraussetzungs-voll – und zwar aus unterschiedlichen Gründen: *Empirisch* ist die Forschungslage inkonsistent. Eine gelingende Verzahnung von Diagnostik und Förderung ist beim Übergang von KiTa in die Grundschule nur in Ansätzen empirisch untersucht – was auch daran liegt, dass unterschiedliche Disziplinen beteiligt sind (vgl. Moser Opitz 2022). Zum anderen sind psychometrisch zuverlässige Diagnostikinstrumente für die Gewinnung *diagnostischer Informationen* hilfreich, für eine *gezielte Förderung* sind sie dies häufig nicht. Um eine Reduktion der Anteile in der sog. „Risikogruppe“ zu erzielen, wäre die bisherige diagnostische Praxis und auch die Implementationsbedingungen (z. B. Akzeptanz) diagnosegeleiteter Förderkonzepte erst einmal genauer zu untersuchen und zu fragen: *Wird zu wenig diagnostiziert (Quantität)? Sind es nicht-adäquate Diagnoseinstrumente (Qualität)? Baut die ‚falsche‘ Förderung auf der ‚richtigen‘ Diagnose auf (Inkongruenz)?* Dies sind Fragen, die es wissenschaftlich zu überprüfen gilt. Der Reformvorschlag, wie er von der SWK formuliert wird, erwägt jedoch den Eindruck, bereits eindeutige Antworten darauf gefunden zu haben.

Zusammenfassend lässt sich bilanzieren, dass die SWK eine Expansion des Monitorings über die Grundschule hinaus im Bereich der frühkindlichen Bildung anstrebt. Dies scheint mit der Hoffnung verbunden zu sein, ein *präventives Frühwarnsystem* zu installieren – was auf den ersten Blick auch plausibel erscheint. Bereits im frühkindlichen Vorschulbereich soll mit einer diagnosegeleiteten Förderung Kompetenzunterschieden begegnet werden (vgl. die BiKS-Studie). Was die SWK – auch mit Bezug auf die IQB-Bildungstrend – mit ihrem Fokus auf einen kumulativen Kompetenzaufbau allerdings zu übersehen scheint, ist der Umstand, dass von *Kompetenzen* nicht auf *Entwicklungs- bzw. Lernverläufe* geschlossen werden kann (vgl. Weinert 2007; zu einen *pädagogischen Lernbegriff*: Göhlich et al. 2014). Es besteht keine einfache Kontinuität im Hinblick auf die Anreicherung von Wissen und nicht bei allen Kindern ein linearer Entwicklungsverlauf; vielmehr vollzieht sich die Entwicklung von Kindern auch in *qualitativen Sprüngen*, d. h. Inhalte werden gelernt, obwohl die Grundlagen noch nicht vorhanden sind (vgl. am Beispiel Sprache: Weinert 2007).

## 4 Fazit und Diskussion der Befunde

Ziel des Beitrags war es, die Rolle des nationalen Bildungsmonitorings (IQB) und dem sich auf die IQB-Ergebnisse beziehenden Beratungsgremium (SWK) für die Initiierung der Reform der Grundschule näher zu beleuchten. Die Ex-ante-Programm-

analyse weist mit Blick auf das anvisierte Reformprogramm darauf hin, dass als zentrales Reformziel eine *Expansion des Bildungsmonitorings* über die verschiedenen Bildungsetappen (beginnend bei der frühkindlichen Bildung bis hin zur Lehrkräfteprofessionalisierung) anvisiert wird. Über das *externe* Bildungsmonitoring durch das IQB und die Unterstützung der SWK (z. B. hinsichtlich der Konkretion von Reformzielen) hinaus, legt jedoch das Reformprogramm einen *schulinternen* Ausbau des Monitorings durch Unterrichtsdiagnosen der Lehrkräfte und Diagnosen der Pädagog:innen im frühkindlichen Bereich nahe – ein Aspekt, auf den unten erneut einzugehen sein wird.

Evidenz wird damit zu einer zentralen *gouvernementalen Reformstrategie* (vgl. Forster 2014). Eine solche Reform folgt, mit Charles Dickens gesprochen, der *Maxime*: „*Take nothing on its looks; take everything on evidence*“ (2008, S. 307). Sie bricht mit früheren Reformtraditionen, deren Ausgangspunkt nicht Daten, sondern bildungspolitische Ideale waren und immunisiert sich gegenüber der Kritik, Reformen ausgehend von wissenschaftlich nicht haltbaren Ideologien zu initiieren. Insofern weist eine evidenzorientierte Reform der Grundschule auf den ersten Blick entscheidende Vorzüge auf: Bildungspolitische Forderungen lassen sich auf wissenschaftliche Grundlagen stützen, hitzige Debatten versachlichen und pädagogische Missstände (z. B. fehlende Sicherung der Mindeststandards) transparent machen. Als *Seismograph* für Anzeichen einer *Bildungs-* oder besser: *Kompetenzkrise* vermögen die Studien des Bildungsmonitoring bereits einen ersten Schritt aus der Krise darzustellen – auch wenn sie unerwünschte, empirische Wahrheiten überbringen (z. B. über den Anteil an Schüler:innen, die Mindeststandards nicht erfüllen).

Man würde jedoch die Leistungen des nationalen Bildungsmonitorings überschätzen, wenn man annehme, dass sich weitreichende Entscheidungen über die Reform der Grundschule aus den Daten empirisch ableiten ließen. Die Daten mögen Veränderungen beschreiben, inwieweit diese aber das Ergebnis von Reformen sind oder Anlass für Reformen geben, lässt sich wissenschaftlich nicht seriös beantworten. Das IQB operiert mit dem Begriff „*Trend*“, der andere Assoziationen wecken mag und verschleiert, dass sich vor dem Hintergrund der Datengrundlage weder eindeutig entscheiden lässt, ob die Grundschule nun *reformbedürftig* ist – z. B. aufgrund der zunehmend gesellschaftlichen Herausforderungen (Digitalisierung, Inklusion, Migration) oder der „*Trend*“ doch eher darauf hindeutet, dass die Grundschulen von *Reformen überfrachtet* werden. Dies hat auch damit etwas zu tun, dass sich die anvisierten Reformen auf *Evidenzerwartungen* (Emmerich 2019) stützen, die weder das Scheitern noch Gelingen einer Reform empirisch präzise nachweisen (siehe das Beispiel Hamburg). Evidenz als zentrale *gouvernementale Reformstrategie*, wie sie von IQB und SWK forciert wird, erzeugt demnach einen steigenden Bedarf an zuverlässigen Informationen, der jedoch das Nicht-Wissen nicht beseitigt.

Man dürfte aber einem Trugschluss aufsitzen, wenn man daraus voreilig schlussfolgert, dass die Reformabsichten von IQB und SWK schnell im Sande verlaufen und ihre Steuerungswirkungen verfehlen. Stattdessen ist eher davon auszugehen, dass die Steuerungsbemühungen *transintentionale Wirkungen* entfalten:

1. *Eindimensionale Verengung des Forschungsgegenstands* „*Schule*“: Es ist zu vermuten, dass ein erhöhter Forschungsbedarf erzeugt wird – insbesondere in den

- fachlichen Domänen (Deutsch, Mathematik), die auch Gegenstand des Bildungsmonitorings sind. Grundsätzlich stellt eine Erhöhung des Forschungsoutputs auch für die empirische Grundschulforschung einen Gewinn dar, jedoch wird dieser durch eine eindimensionale Verengung des Forschungsgegenstands „Schule“ erlangt, indem Schule auf das fachliche Lernen in den Hauptfächern reduziert wird.
2. *Reduktionistisches Bildungsverständnis*: Für die Unterrichtspraxis sind folgenreiche Nebenwirkungen im Hinblick auf die Verteilung von Unterrichtsstunden zu erwarten (z. B. Erhöhung des Stundenkontingents in den Hauptfächern, Kürzung in den Nebenfächern). Dies geht zu Lasten eines umfassenden Bildungsanspruchs der Grundschule, der sich nicht in der Konzentration auf einzelne Unterrichtsfächer erschöpft. Erste Maßnahmen einer solcher Grundschulreform erfolgen bereits in Bayern (sog. „PISA-Offensive“).
  3. *Pädagogische Monokultur und Unterricht als teaching-to-the-test*: IQB und SWK imaginieren eine pädagogische Monokultur, die zwar nicht das Ziel der Reform ist, sie aber als Voraussetzung hat. Denn vorausgesetzt wird, dass *alle* Kinder von mehr Testung und Diagnostik profitieren. Die Komplexität pädagogischer Wirklichkeit wird auf ein sparsames Modell reduziert – was aus Sicht der Forschung äußerst legitim – man könnte sogar sagen: *normal science* ist. Bildungspolitisch birgt dies jedoch Gefahren: Denn Unterschiede zwischen Schüler:innen werden nicht nur vermessen, sie werden auch sichtbar – was Etikettierung, Ausgrenzung und Konkurrenzwettkämpfe zur Folge haben kann. Zudem wird ein Unterricht befördert, der sich an einer *teaching-to-the-test-Kultur* ausrichtet und die eigentlichen Unterrichtsinhalte ins Abseits geraten lässt. Auch ohne high-stakes-Testung besteht – im Sinne „*Campbells Law*“ (1975) – das Risiko, dass mit einer starken Konzentration auf Test- und Diagnosergebnisse, andere, nicht-gemessene, aber für das Lernen wichtige Aspekte vernachlässigt werden.

Diese *Trias der transintentionalen Wirkungen* mögen für IQB und SWK vertretbare Nebenwirkungen sein, die im Interesse der anvisierten Reform in Kauf zu nehmen sind. Allerdings bleibt die Frage offen – und diese betrifft dann den Kern der Reform – inwieweit durch die Reform die eigentliche „*Risikogruppe*“ profitiert, die tatsächlich Anlass für Reformen gibt. Der Ausbau des *internen* Monitorings (mehr Testung und Diagnostik) verspricht mehr Transparenz und frühzeitige Prävention für Kinder *„unter dem Standard“*. Dies scheint auf den ersten Blick ein pädagogisch lohnendes Unternehmen, was jedoch nicht darüber hinwegtäuschen kann, dass ein Ausbau der diagnostischen Infrastruktur in pädagogischen Institutionen auch Gefahr läuft, nicht zum *Abbau*, sondern *Ausbau* der „*Risikogruppe*“ beizutragen. Mit der Erhöhung der Testung – so auch eine Erfahrung aus der Corona-Pandemie – entstehen „*Verdachtsfälle*“, die sich an der Grenze bewegen und je nach Test einige Risikofaktoren aufweisen, andere aber nicht; zudem kann nicht ausgeschlossen werden, dass Kinder fälschlicherweise der „*Risikogruppe*“ zugeordnet werden – sowohl die fehlende Sensitivität des Tests, aber auch die Durchführung und Auswertung des Tests sowie die Untersuchungssituation (z. B. Instruktionen der Testleiter:in, emotional-motivationale Situation der Proband:innen) sind nur einige mögliche Gründe. Auch lassen sich Erwartungseffekte (wie *Pygmalion-* bzw. *Andorraeffekt*) durch die Schüler:innen nicht ausschließen, die sich den entgegengebrachten Erwartungen an-

passen. Für das pädagogische Personal in Schule und KiTa entsteht mit einer solchen evidenzbasierten Reform und dem Ausbau des *internen* Monitorings eine neue Unübersichtlichkeit, die nicht nur das Arbeitsaufkommen erhöhen dürfte, sondern mit der man auch die eigentliche „Risikogruppe“ ungewollt aus dem Auge zu verlieren droht. Kinder dieser „Risikogruppe“ können jedoch schlecht darauf warten, bis hinreichend viel Evidenz aus der Testung vorliegt – wie auch die *BRISE-Studie* zeigt: Schütte et al. (2024). Zudem erfolgt mit dem Ausbau des *internen* Monitorings eine Vereinnahmung, die zwar beabsichtigt, die Reform auf mehrere Schultern zu verteilen, zugleich aber auch Lehrkräfte, Erzieher:innen und andere pädagogische Fachkräfte zu *Agenten des Wandels* macht, d. h. als mitverantwortlich für das Gelingen und Scheitern einer Reform, über die sie selbst nicht mitbestimmt haben (Brinkmann 2023).

Abschließend lässt sich festhalten, dass die Ergebnisse des nationalen Bildungsmonitorings und die daraus abgeleiteten Reformmaßnahmen Anlass für eine Debatte innerhalb der Fachcommunity geben, da sie das Selbstverständnis der Grundschule (Vermittlung grundlegender Bildung) sowie grundlegend auch das Verhältnis von Wissenschaft und Politik berühren. Ziel des vorliegenden Diskussionsbeitrags war es einige Fragen aufzuwerfen und insbesondere die intendierten Wirkungen und nicht-intendierten Nebenwirkungen einer evidenzbasierten Reform der Grundschule zu beleuchten – obgleich die Aufzählung hier keineswegs erschöpfend ist. Wie sich nicht nur im innerwissenschaftlichen Diskurs zeigt, stellt *Evidenz eine unkämpfte Ressource* dar, die Verhältnisse stabilisiert, aber auch Transformationen (z. B. „Klimakrise“) anzeigen kann; die „Krisen“ ausruft oder auch für nichtig erklärt. Eine – auch in der Grundschulpädagogik und -forschung noch offensiver geführte Debatte über die Notwendigkeit empirischer Evidenz und ihrer Grenzen, erscheint daher wünschenswert.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

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

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

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

## Literatur

Bauer, J., & Kollar, I. (2023). (Wie) kann die Nutzung bildungswissenschaftlicher Evidenz Lehren und Lernen verbessern? Thesen und Fragen zur Diskussion um evidenzorientiertes Denken und Handeln von Lehrkräften. *Unterrichtswissenschaft*, 51, 123–147.

- Bellmann, J. (2017). Datengetrieben und/oder evidenzbasiert? In J. Baumert & K.-J. Tillmann (Hrsg.), *Empirische Bildungsforschung. Der kritische Blick und die Antwort auf die Kritiker*. ZfE Sonderheft 31. (S. 147–161).
- Bertrand, M., & Marsh, J. (2021). How data-driven reform can drive deficit thinking. *Phi Delta Kappan*, 102(8), 35–39.
- Besa, K.-S., Demski, D., Gesang, J., & Hinzke, J.-H. (2023). *Evidenz- und Forschungsorientierung in Lehrer\*innenbildung, Schule, Bildungspolitik und -administration. Neue Befunde zu alten Problemen*. Wiesbaden: Springer VS.
- Biesta, G. (2011). Warum „What works“ nicht funktioniert: Evidenzbasierte pädagogische Praxis und das Demokratiedefizit der Bildungsforschung. In J. Bellmann & T. Müller (Hrsg.), *Wissen, was wirkt. Kritik evidenzbasierter Pädagogik*. Wiesbaden: VS.
- Böttcher, W., Hogrebe, N., & Strietholter, R. (2014). Frühe Hilfen und frühe Bildung. Chancen zur Reduktion von Chancenungleichheit? In U. Bauer & al (Hrsg.), *Expansive Bildungspolitik – Expansive Bildung?* (S. 107–125). Wiesbaden: Springer VS.
- Brinkmann, M. (2023). *Datengesteuerte Leistungsmessung und evidenzbasierte Bildungsforschung – von den perversen Effekten Neuer Steuerung in Schule und Unterricht*. Vortrag in der Vortragsreihe „Fachdidaktik kontrovers“ des Arbeitsbereichs Psychologie-Philosophie, Universität Wien.
- Bromme, R., & Kienhues, D. (2017). Gewissheit und Skepsis: Wissenschaftskommunikation als Forschungsthema der Psychologie. *Psychologische Rundschau*, 68(3), 167–171.
- Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. Eine Analyse von Anforderungen an die Darstellung, Interpretation und Rezeption empirischer Befunde. *Zeitschrift für Erziehungswissenschaft*, 17, 3–54.
- Brügelmann, H. (2017). Was sagen uns IQB-Bildungstrend, TIMSS, PISA und andere Ländervergleiche? *Lehren und Lernen*, 43(2), 4–9.
- Campbell, D.T. (1975). Assessing the impact of planned social change. In G. Lyons (Hrsg.), *Social research and public policies*. The Dartmouth/OECD Conference. (S. 3–45). Hanover: Dartmouth College, The Public Affairs Center.
- Dickens, C. (2008). *Great expectations*. Oxford: Oxford University Press.
- Ditton, H. (2019). Schulreformen seit 1995: Ein Überblick. In N. Berkemeyer, W. Bos & B. Hermstein (Hrsg.), *Schulreform. Zugänge, Gegenstände, Trends* (S. 248–259). Weinheim: Beltz.
- Emmerich, M. (2019). (N)Ever change a running system! Schulreform aus systemtheoretischer Sicht. In N. Berkemeyer & al (Hrsg.), *Schulreform. Zugänge, Gegenstände, Trends* (S. 57–68). Weinheim: Beltz.
- Forster, E. (2014). Kritik der Evidenz. Das Beispiel evidence-informed policy research der OECD. *Zeitschrift für Pädagogik*, 60(6), 890–907.
- Frohn, J., & Heinrich, M. (2018). Inkompetente Kompetenzorientierung? Das verkürzte Verständnis der Kompetenzorientierung und die Konsequenzen für die Lehrkräfteausbildung und Lehrkräftefortbildung. *Die Deutsche Schule*, 110(1), 65–74.
- Göhlich, M., Wulf, C., & Zirfas, J. (2014). *Pädagogische Theorien des Lernens*. Weinheim: Beltz.
- Grünkorn, J., Klieme, E., & Stanat, P. (2019). Bildungsmonitoring und Qualitätssicherung. In O. Köller (Hrsg.), *Das Bildungswesen in Deutschland. Bestand und Potenziale* (S. 263–323). Bad Heilbrunn: Klinkhardt.
- Hartong, S. (2014). Neue Bildungsregulierung im Zeitalter der „governance by numbers“. Das Beispiel standardisierter Bildungsreformen in Deutschland und den USA. *Leviathan*, 42(4), 606–634.
- Hartong, S. (2020). Zum Optimierungsdrang des Bildungsmonitorings. *Zeitschrift für Pädagogik*, 66(1), 64–71.
- Heinrich, M. (2010). Bildungsgerechtigkeit durch Evidence-based-policy? In W. Böttcher, J. Dicke & N. Hogrebe (Hrsg.), *Evaluation, Bildung und Gesellschaft. Steuerungsinstrumente zwischen Anspruch und Wirklichkeit* (S. 47–68). Münster: Waxmann.
- Heinrich, M. (2015). Zur Ambivalenz der Idee evidenzbasierter Schulentwicklung. Das Beispiel „Schulinspektion“ – fortschrittlicher Rückschritt oder Innovation? *Zeitschrift für Pädagogik*, 61(6), 778–792.
- Hermstein, B., Berkemeyer, N., & Abendroth, S. (2018). Indikatorengestütztes Bildungsmonitoring zwischen Institutionsgestaltung und Verständigung: Kritisch-konstruktive Überlegungen. In K. Drossel & B. Eickelmann (Hrsg.), *Does ‚What works‘ work? Bildungspolitik, Bildungsadministration und Bildungsforschung im Dialog* (S. 175–190). Münster: Waxmann.
- Jornitz, S. (2008). Was bedeutet eigentlich ‚evidenzbasierte Bildungsforschung‘? Über den Versuch, Wissenschaft für Praxis verfügbar zu machen am Beispiel der Review-Erstellung. *Die deutsche Schule*, 100(2), 206–216.
- Kim, J. (2018). School accountability and standard-based education reform: The recall of social efficiency movement and scientific management. *International Journal of Educational Development*, 60, 80–87.

- Klemm, K. (2017). Die PISA-Studien: Ihre Präsentation und Interpretation im Lichte der Evidenzbasierung. *Zeitschrift für Erziehungswissenschaft*, 31(S), 163–177.
- KMK (2015). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*
- Ladenthin, V. (2013). Grenzen empirischer Forschung in der Pädagogik. In S. Lin-Klitzing, D. Di Fuccia & G. Müller-Frerich (Hrsg.), *Zur Vermessung von Schule. Empirische Bildungsforschung und Schulpraxis* (S. 53–74). Bad Heilbrunn: Klinkhardt.
- Maritzen, N. (2014). Glanz und Elend der KMK-Strategie zum Bildungsmonitoring. Versuch einer Bilanz und eines Ausblicks. *DDS – Die Deutsche Schule*, 106(4), 398–413.
- Maritzen, N. (2018). Was heißt und zu welchem Ende studiert man Daten? Anspruch und Wirklichkeit einer Strategie des Bildungsmonitorings. In Drossel & B. Eickelmann (Hrsg.), *Does 'What works' work? Bildungspolitik, Bildungsadministration und Bildungsforschung im Dialog* (S. 37–54). Münster: Waxmann.
- Martens, K., & Niemann, D. (2010). *Governance by comparison: how ratings & rankings impact national policy-making in education*. Bremen: Working Papers No. 139.
- Marttila, T. (2014). Die wissensbasierte Regierung der Bildung – Die Genese einer transnationalen Governmentalität in England und Schweden. *Berliner Journal für Soziologie*, 24, 257–287.
- Moser Opitz, E. (2022). Diagnostisches und didaktisches Handeln verbinden: Entwicklung eines Prozessmodells auf der Grundlage von Erkenntnissen aus der pädagogischen Diagnostik und der Förderdiagnostik. *Journal für Mathematik-Didaktik*, 43(1), 205–230.
- Oelkers, J., & Reusser, K. (2008). *Qualität entwickeln – Standards sichern – mit Differenz umgehen*. Berlin: BMBF.
- Prenzel, M., Baumert, J., & Klieme, E. (2008). *Steuerungswissen, Erkenntnisse und Wahlkampfmunition: Was liefert die empirische Bildungsforschung? – Eine Antwort auf Klaus Klemm*. <https://doi.org/10.25656/01:1489>.
- Sackmann, R. (2021). Beziehungen zwischen Ex-ante-Evaluation und formativer Evaluation. Beispiele aus der wissenschaftlichen Begleitung zum Pilotprojekt „Entwicklung und Erprobung eines Konzepts zur Förderung von Anerkennungskosten“. In R. Stockmann & H. Ertl (Hrsg.), *Evaluation und Wirkungsforschung in der beruflichen Bildung*. Berichte zur beruflichen Bildung. Bonn: AG BFN.
- Sauerwein, M., & Klieme, E. (2016). Anmerkungen zum Qualitätsbegriff in der Bildungsforschung. *Schweizerische Zeitschrift für Bildungswissenschaften*, 38(3), 459–478.
- Scharlau, I. (2019). Sich verständigen. Überlegungen zur Frage der Evidenzbasierung. In T. Jenert, G. Reinmann & T. Schmohl (Hrsg.), *Hochschulbildungsforschung. Theoretische, methodologische und methodische Denkanstöße für die Hochschuldidaktik* (S. 105–123). Wiesbaden: Springer VS.
- Schrader, J., Hasselhorn, M., Hetfleisch, P., & Goeze, A. (2020). Stichwortbeitrag Implementationsforschung: Wie Wissenschaft zu Verbesserungen im Bildungssystem beitragen kann. *Zeitschrift für Erziehungswissenschaft*, 23, 9–59.
- Schütte, K., Hasselhorn, M., & Köller, O. (2024). Bildungsungleichheit besser verstehen und vermindern. Ansatz und Befunde der Bremer Initiative zur Stärkung frühkindlicher Entwicklung. *Frühe Bildung*, 13(2), 65–70.
- Stanat, P., & Pant, H. (2022). Hat PISA die Schulen besser gemacht? In B. Spinath (Hrsg.), *Empirische Bildungsforschung* (S. 21–38). Wiesbaden: Springer.
- Stanat, P., Pant, H., Richter, D., Pöhlmann, C., & Kuhl, P. (2013). Was kann das IQB leisten? In S. Lin-Klitzing, D. DiFuccia & G. Müller-Frerich (Hrsg.), *Zur Vermessung von Schule: Empirische Bildungsforschung und Schulpraxis* (S. 125–152). Bad Heilbrunn: Klinkhardt.
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). *IQB-Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Münster: Waxmann.
- Stockmann, R., & Ertl, H. (2021). *Evaluation und Wirkungsforschung in der beruflichen Bildung*. Berichte zur beruflichen Bildung. Bonn: AG BFN.
- SWK (2022). *Entwicklung von Leitlinien für das Monitoring und die Evaluation von Förderprogrammen im Bildungsbereich*. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz.
- Tillmann, K.-J. (2017). Empirische Bildungsforschung in der Kritik – ein Überblick über Themen und Kontroversen. In J. Baumert & K. J. Tillmann (Hrsg.), *Empirische Bildungsforschung. Der kritische Blick und die Antwort auf die Kritiker* (S. 5–22). Wiesbaden: Springer VS.
- Weinert, S. (2007). Kompetenzzentwicklung und Kompetenzstruktur im Vorschulalter. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik*. Zeitschrift für Erziehungswissenschaft Sonderheft 8. (S. 89–106).

- Wilkes, T., & Stark, R. (2023). Probleme evidenzorientierter Unterrichtspraxis. Anregungen und Lösungsvorschläge. *Unterrichtswissenschaft*, 51, 289–313.
- Wolff, S. (2000). Dokumenten- und Aktenanalyse. In U. Flick, E. v. Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung: Ein Handbuch* (S. 502–514). Reinbek: Rowohlt.
- Zillien, N. (2013). Laien als Experten. Ungleichheiten des Wissens von reproduktionsmedizinischer Profession und Patientenschaft. In O. Berli & M. Endreß (Hrsg.), *Wissen und soziale Ungleichheit* (S. 177–200). Weinheim: Beltz Juventa.

**Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.