



Andrea Albers | Nina Jude (Hrsg.)

**Blickpunkt
Bildungsmonitoring
– Bilanzen und
Perspektiven**

BELTZ JUVENTA

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Der Text dieser Publikation wird unter der Lizenz **Creative Commons Namensnennung – Nicht kommerziell** - Weitergabe unter gleichen Bedingungen **4.0 International (CC BY-NC-SA 4.0)** veröffentlicht. Den vollständigen Lizenztext finden Sie unter: <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.de> legalcode. Verwertung, die den Rahmen der **CC BY-NC-SA 4.0 Lizenz** überschreitet, ist ohne Zustimmung des Verlags unzulässig. Die in diesem Werk enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Quellenangabe/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.



Dieses Buch ist erhältlich als:
ISBN 978-3-7799-7842-8 Print
ISBN 978-3-7799-7843-5 E-Book (PDF)
ISBN 978-3-7799-8726-0 E-Book (ePub)
DOI 10.3262/978-3-7799-7843-5

1. Auflage 2025

© 2025 Beltz Juventa
in der Verlagsgruppe Beltz · Weinheim Basel
Werderstraße 10, 69469 Weinheim
service@beltz.de
Einige Rechte vorbehalten

Herstellung: Ulrike Poppel
Satz: xerif, le-tex
Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza
Beltz Grafische Betriebe ist ein Unternehmen mit finanziellem Klimabeitrag
(ID 15985-2104-1001)
Printed in Germany

Weitere Informationen zu unseren Autor:innen und Titeln finden Sie unter: www.beltz.de

Inhalt

Blickpunkt Bildungsmonitoring – Bilanzen und Perspektiven nach zwei Jahrzehnten <i>Andrea Albers und Nina Jude</i>	9
---	---

Teil A: Internationale Schulleistungsvergleiche

Warum immer noch PISA? <i>Christine Sälzer</i>	33
Die OECD-Bildungsindikatoren – Ziele und Herausforderungen <i>Abel Schumann</i>	50
Zur Konzeption und Durchführung eines Schulleitungsmonitorings im deutschsprachigen Bildungsraum – Chancen und Grenzen <i>Jana Groß Ophoff, Stefan Brauckmann-Sajkiewicz, Pierre Tulowitzki, Marcus Pietsch und Colin Cramer</i>	61

Teil B: Überprüfung der Umsetzung von Bildungsstandards

Weiterentwicklung der Bildungsstandards für den Primarbereich und die Sekundarstufe I <i>Nada Abouelseoud und Jennifer Schwarze</i>	81
Bildungsstandards im Fach Deutsch – Weiterentwicklung und Perspektiven <i>Michael Krelle, Jörg Jost, Irene Pieper, Norbert Maritzen und Petra Stanat</i>	96

Teil C: Verfahren zur Qualitätssicherung auf Schulebene

Datengestützte Unterrichtsentwicklung mit Vergleichsarbeiten <i>Patrick Hawlitschek, Sofie Henschel, Carola Schnitzler und Petra Stanat®</i>	111
Die VERA3-Vergleichsuntersuchungen an Grundschulen als Teil der Monitoringstrategie in Baden-Württemberg, Nordrhein-Westfalen und Schleswig-Holstein <i>Horst Weishaupt</i>	129
Nutzung von Lernstandserhebungen zur Förderung adaptiven Unterrichts <i>Holger Gärtner</i>	142

Die externe Schulevaluation im Dienst des Bildungsmonitorings <i>Andreas Brunner und Sandy Taut</i>	158
Evidenzinformierte Schulentwicklung in der Einzelschule – welche Rolle kann die Schulaufsicht bei der Unterstützung spielen? <i>Ruth Anna Hejtmanek, Esther Dominique Klein, Stefan Hahn und Klaudia Schulte</i>	169
Datengestützte Qualitätsentwicklung konkret: Das Schuldatenblatt Baden-Württemberg <i>Ulrike Rangel und Günter Klein</i>	185
Welche Story steckt in den Daten einer Schule? <i>Stefan Hahn, Klaudia Schulte, Peter Schulze und Franziska Thonke</i>	199
Die Alltäglichkeit der Daten in der Schule <i>Ben Mayer und Sieglinde Jornitz</i>	213
Teil D: Bildungsberichterstattung	
Herausforderungen und Potenziale einer indikatorengestützten Bildungsberichterstattung auf gesamtstaatlicher Ebene <i>Stefan Kühne</i>	231
Bildungsmonitoring „vor“ und „neben“ der Schule <i>Susanne Kuger und Susanne Lochner</i>	241
Entwicklung eines Monitorings für die politische Bildung in Deutschland <i>Hermann Josef Abs, Tim Engartner, Reinhold Hedtke, Monika Oberle, Marie Heijens, Simon Niklas Hellmich, Valeria Hulkovych, Lucy Huschle und Stella Wasenitz</i>	263
Das Hamburger Sprachfördermonitoring als Instrument zur Steuerung <i>Britta Pohlmann und Yvonne Hoffmann</i>	278

Teil E: Perspektiven Bildungsmonitoring

Die Bedeutung des Nationalen Bildungspanels (NEPS) für das Bildungsmonitoring <i>Michaela Sixt und Cordula Artelt</i>	295
20 Jahre Bildungsmonitoring – mit welchem Erfolg? <i>Dorit Stenke, Janina Roloff und Désirée Burba</i>	311
„Es ist kompliziert“. Über den Beziehungsstatus von Bildungsjournalismus zu Bildungswissenschaft und -politik. <i>Annette Kuhn und Niklas Prenzel</i>	324
Bildungsmonitoring – eine berufliche Tätigkeit im Spannungsfeld unterschiedlicher Handlungslogiken <i>Daniel Kneuper</i>	332
Bildungscontrolling – Lesarten eines noch immer unbeliebten Begriffs <i>Nils Berkemeyer und Philipp Glanz</i>	346
Netzwerk Bildungsmonitoring im Dialog: Potenzial und Perspektiven <i>Martina Diedrich und Günter Klein</i>	363

Teil F: Kommentierung

Bildungsmonitoring – Grundlage für eine datengestützte Qualitätsentwicklung	375
Ein kritischer Kommentar zum Positionspapier „Bildungsmonitoring – Grundlage für eine datengestützte Qualitätsentwicklung“ <i>Norbert Maritzen</i>	390
Bildungsmonitoring – ein Kommentar zum Sammelband <i>Jenny Tränkmann</i>	397

Blickpunkt Bildungsmonitoring – Bilanzen und Perspektiven nach zwei Jahrzehnten

Andrea Albers und Nina Jude

Woher wissen wir, wie gut unser Schulsystem ist? Wie können wir feststellen, an welcher Stelle Bildungsinvestitionen notwendig sind? Wie verändert sich unsere Schüler:innenschaft? Was macht guten Unterricht aus? Auf welche Herausforderungen müssen Schulen in Zukunft reagieren? Um diese und noch viele weitere Fragen zu beantworten, wurden in Deutschland seit den späten 1990er Jahren empirische Instrumente auf den verschiedenen Ebenen des Bildungssystems eingeführt, das sogenannte „Bildungsmonitoring“. Durch eine systematische und regelmäßige Erhebung von Daten zur Leistungsfähigkeit unseres Bildungssystems und deren gezielte Aufarbeitung soll Bildungsmonitoring dazu beitragen, die Qualität im System zu sichern und weiterzuentwickeln.

Die Instrumente des Bildungsmonitorings sind im Bildungssystem Deutschlands inzwischen fest etabliert. Sie dienen den unterschiedlichen Zielgruppen in Bildungspolitik, Bildungsadministration, Bildungspraxis und der breiten Öffentlichkeit zur Informationsgewinnung über den Zustand unseres Bildungssystems. Zielgerichtet sollen sie auch der Weiterentwicklung von Schule und Unterricht dienen. Alle diese Instrumente verbindet die gemeinsame Zielrichtung, auf Einzelschul- und auf Systemebene wissenschaftlich fundierte, empirische Daten für Qualitätsentwicklung und Steuerungsentscheidungen zur Verfügung zu stellen.

Dieser Sammelband hat seinen Ursprung in der mehrteiligen Serie zum Bildungsmonitoring in der Zeitschrift PÄDAGOGIK, die im Jahr 2022 erschien. Die in den dortigen Beiträgen aufgeworfenen Fragen werden hier wieder aufgegriffen und durch vielfältige zusätzliche Perspektiven erweitert. Der vorliegende Band versammelt Stimmen aus Bildungsforschung, -administration und -praxis zu einem aktuellen Blick auf den Status Quo des Bildungsmonitorings in Deutschland. In den vier Abschnitten werden ausgewählte Elemente des Bildungsmonitorings entsprechend der Strategie des Bildungsmonitorings der Kultusministerkonferenz (siehe Abb. 1) näher beleuchtet. Durchgängig werden die vorliegenden Instrumente analysiert, über die Zielsetzung deren jeweiliger Erfolg reflektiert und stellenweise auch hinterfragt. In Frage gestellt wird dabei nicht die generelle Sinnhaftigkeit von Monitoringansätzen für die Systemsteuerung, jedoch durchaus deren aktuelle Ausgestaltung oder bisher fehlende Weiterentwicklungen. In allen Beiträgen schwingt die Frage mit, welche Entwicklungsbedarfe sich für die Zukunft des Bildungsmonitorings ergeben könnten bzw. wie diese mit den be-

stehenden Elementen oder neu zu entwickelnden Ansätzen angegangen werden könnten. Zugleich kann auch dieser Sammelband die große Vielfalt der mit Bildungsmonitoring zusammenhängenden Aspekte nicht erschöpfend darstellen. Durch die Orientierung an den vorhandenen vier Bereichen des Bildungsmonitorings werden ebenfalls relevante Themenbereiche beispielsweise zu kommunalen Ansätzen des Bildungsmonitoring oder Digitalisierung eher am Rande behandelt.

Im Folgenden werden die Grundlagen des Bildungsmonitorings sowie die vier Bereiche der Strategie zum Bildungsmonitoring der Kultusministerkonferenz (KMK) dargestellt, die Rolle der aktuell im Bildungsmonitoring agierenden Akteursgruppen erläutert und abschließend mögliche Perspektiven zur Weiterentwicklung von zukünftigen Ansätzen des Bildungsmonitorings aufgeworfen.

Abb. 1: Vier Bereiche des Bildungsmonitorings entsprechend der Strategie des Bildungsmonitorings der KMK von 2015



Was ist Bildungsmonitoring?

Bildungsmonitoring ist die stetige Beobachtung der Rahmenbedingungen, Verlaufsmerkmale, Ergebnisse und Erträge von Bildungsprozessen mit Hilfe empirisch-wissenschaftlicher Methoden (vgl. z. B. Maritzen/Tränkmann 2015). Bildungsmonitoring zielt darauf ab, Informationen über die Qualität des Bildungssystems auf unterschiedlichen Ebenen zu gewinnen: auf Systemebene, auf Schulebene und auf Unterrichtsebene. Dabei werden mögliche Handlungsbedarfe für die Qualitätsentwicklung an Schulen identifiziert und die gewonnenen

Daten für die politisch-administrative Planung und Steuerung des Bildungswesens („Governance“) aufbereitet (vgl. z. B. Grünkorn et al. 2019).

Neben der Bildungsberichterstattung als Monitoring im eigentlichen Sinne werden unterschiedliche Ansätze der Evaluation für die Gewinnung dieser Daten verfolgt. Dabei stellt Bildungsmonitoring nicht nur Daten über institutionalisierte Bildungsangebote und deren Nutzung zur Verfügung, sondern fragt umfassender nach den Bildungschancen und den Bedingungen für kulturelle Teilhabe aller Menschen, auch über die Lebensspanne hinweg. Bildung wird dabei als individuelle Entwicklung verstanden, entsprechend erfasst Bildungsmonitoring Bildungsvoraussetzungen, Bildungswege und Bildungsergebnisse von Individuen. Diese Informationen werden mit Qualitätsmerkmalen von Institutionen zusammengebracht. Danach werden die Daten zu Indikatoren gebündelt, die auf der Ebene des Gesamtsystems angesiedelt sind. Die Ansätze des Bildungsmonitoring sind daher zunächst systembezogen und im Unterschied beispielsweise zu einer individuellen Diagnostik nicht personenbezogen.

Seit wann gibt es Bildungsmonitoring in Deutschland?

Die Ständige Konferenz der Kultusminister (KMK) in Deutschland beschloss erstmals 1997, die Qualitätssicherung des Bildungssystems als eine gemeinsame Aufgabe der Länder anzusehen (Konstanzer Beschlüsse). Nicht zuletzt verdeutlichte das überraschende Abschneiden Deutschlands in den ersten internationalen Schulleistungsvergleichen zur Jahrtausendwende, dass nur wenige Informationen zur Qualität von Schule und Unterricht vorhanden waren. Wurde bis dato vor allem über die sogenannte Inputsteuerung gelenkt – das heißt über die Zuweisung von Ressourcen – zeigte sich nun die Notwendigkeit, den Output, also Ergebnisse von institutioneller Bildung, zu erheben und zu analysieren. Daraus wurde die Notwendigkeit abgeleitet, wissenschaftlich fundierte Indikatoren im Sinne von vergleichbaren Kennzahlen zu den verschiedenen Aspekten des Bildungssystems zur Verfügung zu stellen.

Eine Gesamtstrategie der Länder zum Bildungsmonitoring als datengestütztes Instrument für Qualitätsentwicklung im Bildungswesen führte die KMK im Jahr 2006 ein und überarbeitete sie im Jahr 2015 (KMK 2015). Die Gesamtstrategie führt unterschiedliche Bestandteile zusammen:

- Verfahren zur Qualitätssicherung auf Schulebene, u. a. Vergleichsarbeiten oder externe Evaluation,
- die Umsetzung und Überprüfung von Bildungsstandards,
- die regelmäßige Teilnahme an international vergleichenden Schulleistungsstudien und

- die nationale Bildungsberichterstattung, die Kernindikatoren aller Bildungsbereiche alle zwei Jahre zur Verfügung stellt, sowie regionale Bildungsberichte.

Diese vier Bereiche basieren auf unterschiedlichen Maßnahmen und Institutionen, deren Aufgabe es ist, aktuelle Indikatoren zum Status Quo zu liefern. Die Indikatoren aller Bereiche zusammengenommen dienen dazu, die Qualitätsentwicklung im Bildungswesen auf den unterschiedlichen Ebenen – vom Unterricht über die Schulebene bis hin zur administrativen Ebene – zu unterstützen. Mit diesen Instrumenten zielt Bildungsmonitoring darauf ab, „Entwicklungen nicht nur zu beschreiben, sondern auch zu erklären und dies mit Hinweisen zu verbinden, wie die festgestellten Probleme gelöst werden können,“ sowie „steuerungsrelevantes Wissen auch tatsächlich für die Entwicklung des Bildungssystems und jeder Schule zu nutzen“ (KMK 2015).

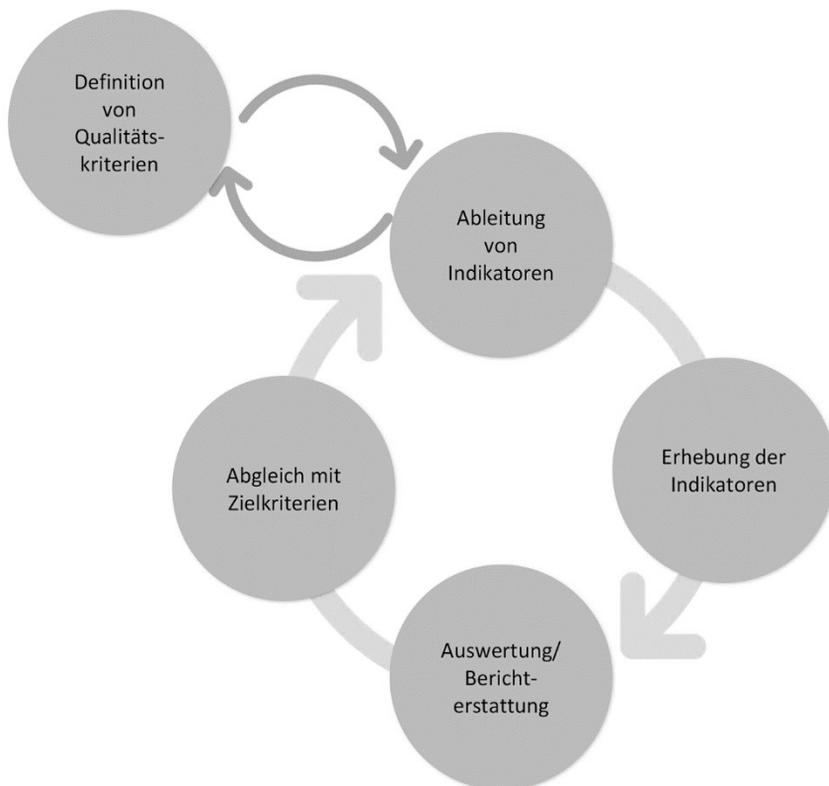
Warum werden regelmäßig Daten erhoben?

Monitoring bedingt eine regelmäßige Darstellung relevanter Kernindikatoren, die sich über die Zeit durchaus verändern können (u. a. Döbert/Weishaupt 2012). Als Beispiel sei hier die Anzahl von Schüler:innen genannt, die jedes Jahr das Schulsystem mit einem Abschluss verlassen. Dieser Indikator gibt darüber Auskunft, wie viele Menschen im berufsbildenden System und an Universitäten zu erwarten sind. Dieser Indikator zeigt auch an, ob sich die Anzahl an Schulabgänger:innen ohne Abschluss verändert. Entsprechend sollte Bildungssteuerung darauf reagieren. Solche Strukturzahlen werden alle zwei Jahre im Bildungsbericht abgebildet. Der Bildungstrend liefert regelmäßig Informationen darüber, wie viele Schüler:innen die vorgesehenen Bildungsstandards erreichen. Dadurch lässt sich feststellen, inwieweit Maßnahmen der Qualitätsentwicklung auf Schul- und Unterrichtsebene wirken und an welchen Stellen möglicherweise Handlungsbedarf besteht. In internationalen Studien werden neben den vorhandenen Basiskompetenzen regelmäßig die Herkunftsländer bzw. die zu Hause gesprochene Sprache der Schüler:innen erhoben. Dies lässt Rückschlüsse darauf zu, wie sich die Heterogenität der Schüler:innenschaft über die Zeit ändert. Nur durch die wiederholte Messung und regelmäßige Berichterstattung dieser und vieler weiterer Indikatoren ist es möglich, Entwicklungen über die Zeit festzustellen – und jene Bereiche zu identifizieren, die von großen Veränderungen betroffen sind oder in denen die bisherigen Maßnahmen keine Wirkung zeigen.

Bildungsmonitoring lässt sich deshalb auch als ein klassischer Evaluationszyklus verstehen (siehe Abb. 2). Er zielt darauf ab, durch die regelmäßige Erfassung des Status Quo über die Zeit Veränderungen im Schulsystem abzubilden. Daher werden standardisierte Informationen wiederholt erhoben und analysiert. Dazu

ist es notwendig, zunächst relevante Qualitätskriterien zu definieren: Welche Informationen benötigen wir, um die Qualitätsentwicklung im Bildungswesen einschätzen zu können? Welche Daten liefern uns Anhaltspunkte dafür, ob unsere Maßnahmen wirken? Aus diesen grundsätzlichen Überlegungen sind dann messbare Indikatoren abzuleiten.

Abb. 2: Bildungsmonitoring als Qualitätszyklus



Vier Elemente des Bildungsmonitorings

Die verschiedenen Bestandteile der deutschen Gesamtstrategie zum Bildungsmonitoring zeigen auf, welche unterschiedlichen Informationsquellen notwendig sind, um ein umfassendes Bild der Qualität unseres Bildungssystems zu erfassen und dessen Entwicklung über die Zeit abzubilden. Dafür bedarf es zentraler, transparenter Maßstäbe zur Bestimmung von Bildungsqualität sowie eines Transfers wissenschaftlich abgesicherter Ergebnisse zur Bildungsqualität für

alle Akteure, um Qualitätssicherung und Qualitätsentwicklung zu verbinden. In Deutschland speist sich das Bildungsmonitoring aus vier Bereichen.

Internationale Schulleistungsvergleiche

Bereits seit den 1960er Jahren nahmen einzelne Länder der damaligen Bundesrepublik an ausgewählten internationalen Schulleistungsvergleichen statt (vgl. Goy/van Ackeren/Schwippert 2008; Drechsel/Prenzel/Seidel 2009). Mit der Entscheidung, im Jahr 1995 eine gesamtdeutsche Teilnahme an der TIMSS-Studie umzusetzen, begann die sogenannte empirische Wende mit einem Fokus auf die Ergebnisse von Bildung im deutschen Schulsystem im internationalen Vergleich (Baumert/Lehmann 1997). Durch die elaborierten Erhebungsinstrumente, die stetige Weiterentwicklung des Erhebungsdesigns durch Erweiterung um innovative Kompetenzbereiche und die regelmäßige Berichterlegung zeigen sie den aktuellen Stand und Veränderungen der Kontextfaktoren und ausgewählter Bildungsergebnisse im internationalen Vergleich auf. Die Ergebnisse der PISA-Studie im Jahr 2001 wurden als sogenannter PISA-Schock auch in der Öffentlichkeit diskutiert und führten zu weitreichenden Qualitätsoffensiven im deutschen Bildungssystem (vgl. Klieme et al. 2010). Die Teilnahme an den internationalen Schulleistungsstudien ist seitdem fester Bestandteil der Gesamtstrategie zum Bildungsmonitoring der Länder in der Bundesrepublik Deutschland. Zu den verbindlich festgelegten Instrumenten zählen jedoch lediglich drei Studien: die internationale Grundschul-Leseuntersuchung (IGLU), die Trend in Mathematik und Naturwissenschaften Studie (TIMSS) sowie das Programme for International Studie Assessment (PISA), die aktuell in der Gesamtstrategie verankert sind (KMK 2015). Darüber hinaus nimmt Deutschland regelmäßig auch an weiteren internationalen Vergleichen teil, deren Ergebnisse in die Bildungsberichterstattung einfließen (Autor:innengruppe Bildungsberichterstattung 2024). Diese internationalen Studien erfassen Basiskompetenzen in unterschiedlichen Altersstufen:

- Die Studien PIRLS mit dem Fokus auf die Lesekompetenz und TIMSS mit dem Fokus auf Mathematik und Naturwissenschaften testen diese Kompetenzen bei Schüler:innen der 4. Klasse alle fünf bzw. vier Jahre (vgl. McElvany et al. 2023; Schwippert et al. 2020).
- PISA erfasst im dreijährigen Rhythmus die Kompetenzen im Lesen, in Mathematik und den Naturwissenschaften sowie einem wechselndem, innovativen Kompetenzbereich bei 15-jährigen Schüler:innen (Lewalter et al. 2023).

Dabei unterscheiden sich die Studien bewusst in der Auswahl der Stichprobe: PIRLS und TIMSS erheben die Kompetenzen von Schüler:innen der 4. Klasse im Klassenverbund mittels einer Clusterstichprobe von Klassen in Schulen, so

dass hier Effekte der Klassenebene analysiert werden können. PISA erfasst die Kompetenzen von Schüler:innen im Alter von 15 Jahren, die sich auf unterschiedliche Klassenstufen verteilen können. Entsprechend können Aussagen über Kontextfaktoren des Unterrichts bspw. nur aggregiert auf Schulebene erfolgen. Alle drei Studien erfassen neben den Kompetenzen der Schüler:innen vielfältige Kontextfaktoren der Bildungsqualität über Fragebögen für Schüler:innen und deren Familien, Lehrpersonen und die Schulleitungen. Diese ermöglichen Rückschlüsse über Bedingungsfaktoren des Kompetenzerwerbs auf unterschiedlichen Ebenen (vgl. Kuger et al. 2016).

Bildungsstandards

Die Entwicklung und Einführung von Bildungsstandards ist ein wesentliches Merkmal des Reformprozesses hin zur Outputsteuerung. Bildungsstandards definieren länderübergreifend, welche Kompetenzen Schüler:innen in bestimmten Fächern zu verschiedenen Bildungsetappen erreichen sollen (vgl. z. B. Köller 2010). Seit 2003 werden Bildungsstandards für verschiedene Fächer und Abschlüsse konzipiert und weiterentwickelt. Sie definieren fachbezogene Kompetenzen in unterschiedlichen Niveaustufen und sollen dadurch die Unterrichtsentwicklung orientiert an verbindlichen Kompetenzstandards unterstützen. Die Bildungsstandards sind Grundlage sowohl für die ländervergleichende Überprüfung als auch für Verfahren zur Qualitätssicherung auf Ebene der Schule. Ob Schüler:innen die Bildungsstandards erreichen, wird auf Ebene der Länder mittels empirischer Studien anhand einer Stichprobe von Schüler:innen der Jahrgangsstufen 4 und 9 regelmäßig überprüft und als sogenannter Bildungstrend veröffentlicht. Diese Berichte zeigen die Entwicklung der Kompetenzen über die Zeit und ermöglichen einen Vergleich zwischen den Bundesländern.

Qualitätssicherung auf Schulebene

Verfahren zur Qualitätssicherung auf Ebene der Schule beinhalten Sprachstandsmessungen für unterschiedliche Altersgruppen, Externe Evaluation und Lernstandserhebungen bzw. Vergleichsarbeiten, welche sich an den Kompetenzen orientieren, die in Bildungsstandards festgelegt sind. Die Ergebnisse der jährlichen Vergleichsarbeiten dienen den Schulen dazu, die eigene Unterrichtsqualität zu reflektieren und weiterzuentwickeln (z. B. Maier et al. 2012). Alle Schüler:innen der 3. und 8. Jahrgangsstufe nehmen jeweils verbindlich an den Vergleichsarbeiten teil. Die Lehrpersonen erhalten so Informationen über den Lernstand ihrer Klasse im Vergleich zu den definierten Bildungsstandards. Die differenziert dargebotenen Ergebnisse können sie für die Schul- und Unterrichtsentwicklung

nutzten. Darüber hinaus fließen sie auch in die Bildungsmonitoringsysteme der Länder ein. Zusätzliche Instrumente auf Schuleben sind beispielsweise Sprachstandsmessungen, landesspezifische Vergleichsstudien oder Ergebnisse aus der externen Evaluation, wie zum Beispiel Schulinspektion (z. B. Warmt et al. 2020).

Bildungsberichterstattung

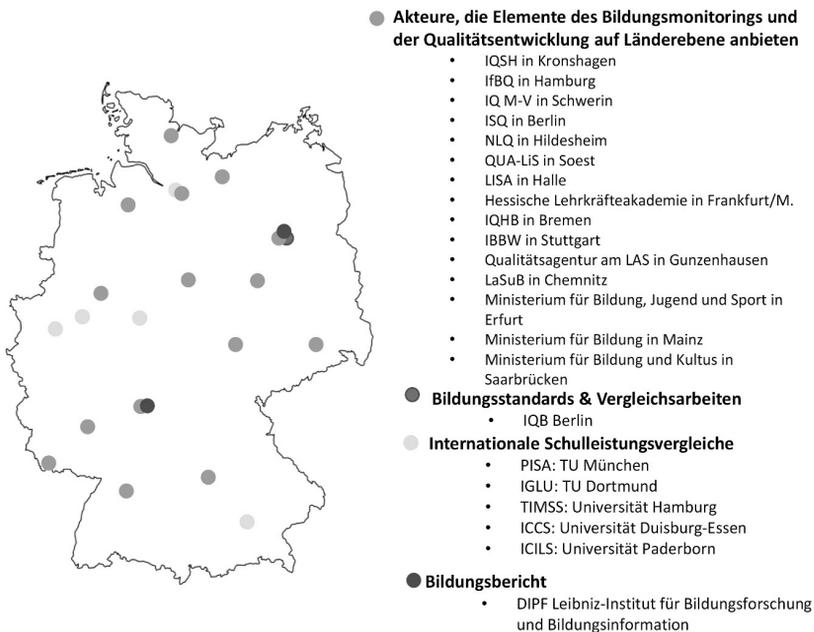
Die Bildungsberichterstattung als zentrales Element des Bildungsmonitorings publiziert seit 2006 alle zwei Jahre eine umfassende empirische Bestandsaufnahme zur Bildung in Deutschland (Autor:innengruppe Bildungsberichterstattung 2024). Dies beinhaltet Informationen zu formalen, non-formalen und informellen Lernsettings über alle Bildungsetappen. Daten aus amtlichen Statistiken und die Indikatoren der drei anderen Bestandteile des Bildungsmonitorings fließen ebenso ein wie die Ergebnisse aktueller sozialwissenschaftlicher Studien. Der nationale Bildungsbericht kann als umfassender Evaluationsbericht zum Bildungswesen angesehen werden. Er gibt regelmäßig einen Überblick über die sich verändernde Bildungslandschaft, die zu erwartenden Kohorten von Schüler:innen und im System vorhandene Ressourcen. Mit unterschiedlicher Schwerpunktsetzung in jedem Erscheinungsjahr werden aktuelle bildungspolitische Themen in den Mittelpunkt gerückt, wie bspw. Digitalisierung, Migration und Inklusion. Ergänzt wird der nationale Bildungsbericht durch kommunales Bildungsmonitoring in vielen Kommunen Deutschlands.

Wer sind die Akteure des Bildungsmonitorings?

Die Instrumente des Bildungsmonitorings basieren auf der Zusammenarbeit zwischen Bildungsforschung, Bildungspraxis und Bildungsadministration. Die Gesamtstrategie zum Bildungsmonitoring obliegt der KMK in Zusammenarbeit mit dem Bund.

- Für die Qualitätssicherung auf Ebene der Einzelschule sind in den meisten Bundesländern extra eingerichtete Qualitätsinstitute oder -abteilungen zuständig.
- Die Entwicklung, Überprüfung und Implementation der Bildungsstandards sowie der Instrumente der Vergleichsarbeiten werden zentral vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) als wissenschaftliches Institut der Humboldt-Universität Berlin verantwortet.
- Internationale Vergleichsstudien werden initiiert von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), die mit ihrem jährlichen Bericht „Bildung auf einen Blick“ regelmäßig Bildungsindikatoren

Abb. 3: Akteure, die Daten zum Bildungsmonitoring liefern



der OECD-Staaten veröffentlichen, sowie von der International Association for the Evaluation of Educational Achievement (IEA) als internationaler Zusammenschluss von Forschungseinrichtungen, die evidenzbasiert Schul- und Unterrichtsqualität untersuchen. Die Durchführung und Auswertung dieser Studien in Deutschland werden von universitären und außeruniversitären Instituten verantwortet.

- Der Bildungsbericht wird am Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF) herausgegeben.

Die internationale Infrastruktur

Die Planung, Durchführung und Auswertung der groß angelegten internationalen Bildungsvergleichsstudien beansprucht in der Regel mehrere Jahre und erfordert umfassende Ressourcen. Daher liegt die Verantwortung bei internationalen Konsortien mit oft jahrzehntelanger Erfahrung in der Konzeption solcher Studien (Wagemaker 2020; Kuger et al. 2016): Die International Association for the Evaluation of Educational Achievement (IEA) mit Sitz in Amsterdam und Hamburg ist eine internationale Vereinigung von nationalen Forschungseinrichtungen in über 60 Ländern der Welt. Gegründet 1958 erforscht sie Bildungsergebnisse

se und -bedingungen weltweit. Die IEA Studien PIRLS und TIMSS, die sich auf schulische Bildung in den Basiskompetenzen beziehen, berücksichtigen in der Regel auch die jeweiligen nationalen Curricula (Schwippert et al., 2020; McElvany et al., 2023). Dies ermöglicht die Analyse des bildungspolitisch intendierten Curriculums im Vergleich zu den im Unterricht implementierten Inhalten und den tatsächlich erworbenen Kenntnissen der Schüler:innen. Darüber hinaus erfasst die International Civic and Citizenship Education Study (ICCS) die politische Bildung und Einstellung von Jugendlichen und die International Computer and Information Literacy Study (ICILS) die computerbezogenen Kompetenzen und unterrichtliche Bedingungen (Abs et al., 2024, Eickelmann et al., 2019).

Die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) mit Sitz in Paris wurde 1960 gegründet und ist eine internationale Organisation, die gemeinsam mit den Regierungen ihrer 38 Mitgliedstaaten gesamtgesellschaftliche Herausforderungen auf Basis evidenzgestützter Analysen bearbeitet. Für den Bereich der Bildung als Grundlage für gesellschaftlichen Wohlstand pflegt die OECD ein umfassendes Indikatorensystem, das zentrale Daten zu Strukturen, der Finanzierung sowie Ergebnisfaktoren von Bildungssystemen jährlich im Sinne eines internationalen Monitorings veröffentlicht („Bildung auf Blick“, OECD 2023). Mit den international vergleichenden Bildungsstudien PISA, dem auf die Kompetenzen Erwachsener abzielenden Programme for the International Assessment of Adult Competencies (PIAAC) und der Befragung von Lehrpersonen (TALIS) implementiert die OECD seit dem Jahr 2000 Instrumente, die Ergebnisse und Kontexte von Bildungsprozessen erfassen (<https://www.oecd.org>).

Die internationale Organisation dieser Studien erfolgt durch wissenschaftliche Einrichtungen als Auftragnehmer der IEA bzw. OECD sowie in Zusammenarbeit mit Bildungswissenschaftler:innen in aller Welt. Die nationale Umsetzung in den Teilnehmerstaaten erfolgt ebenfalls durch wissenschaftliche Institutionen, in Deutschland durch Universitäten oder außeruniversitäre Forschungsinstitute.

Die nationale Infrastruktur

Jede der hier vorgestellten internationalen Bildungsstudien wird in Deutschland von einem eigenen Study Center koordiniert und geleitet. Dieses verantwortet die Vorbereitungen, die nationale Stichprobenziehung und die Durchführung in den Schulen sowie die Berichterlegung und vertiefende wissenschaftliche Publikationen, in der Regel gemeinsam mit einem Konsortium nationaler Expert:innen. Die Überprüfung der Qualität von Stichprobe, Übersetzung der Testmaterialien sowie der Implementation der Studie erfolgt wiederum durch die internationalen Auftragnehmer und wird öffentlich dokumentiert (vgl. die sogenannten „Technical Reports“ der Studien).

So wird die ICILS-Studie aktuell an der Universität Paderborn koordiniert, die ICCS-Studie an der Universität Duisburg-Essen und PIAAC vom Leibniz-Institut für Sozialwissenschaften in Mannheim. Die nationale Umsetzung von PIRLS erfolgt durch die Technische Universität Dortmund und für TIMSS durch die Universität Hamburg. Für die nationale Umsetzung der TALIS-Video Studie war das DIPF|Leibniz-Institut für Bildungsforschung und Bildungsinformation in Frankfurt am Main verantwortlich, für TALIS Starting Strong das Deutschen Jugendinstitut in München.

Für die Implementation der PISA-Studie sowie darüberhinausgehende Begleitforschung wurde in Deutschland eine eigene Infrastruktur geschaffen. Nachdem die ersten vier Erhebungen von PISA in Deutschland vom Max-Planck-Institut in Berlin (PISA 2000), dem Institut für die Pädagogik der Naturwissenschaften IPN in Kiel (PISA 2003 und PISA 2006) und dem DIPF|Leibniz-Institut für Bildungsforschung und Bildungsinformation in Frankfurt am Main (PISA 2009) umgesetzt wurden, wurde 2010 das Zentrum für internationale Vergleichsstudien (ZIB) gegründet. Diese gemeinsame wissenschaftliche Einrichtung aller 16 Länder und des Bundesministeriums für Bildung und Forschung (BMBF) institutionalisiert den Aspekt der internationalen Vergleichsstudien in der Gesamtstrategie zum Bildungsmonitoring durch einen Verbund der Technischen Universität München (TUM), des DIPF und IPN. Dazu wurde an diesen drei Standorten jeweils eine einschlägige Stiftungsprofessuren eingerichtet. Zu den Aufgaben des ZIB gehören neben der Planung, Durchführung und Auswertung der PISA-Studien in Deutschland vor allem innovative Methodenforschung im Kontext der internationalen Vergleichsstudien sowie Anwendungsorientierung in der Distribution der Ergebnisse dieser Studien, bspw. über Forschungssynthesen.

Das ZIB arbeitet dabei zusammen mit dem Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin, das für die Entwicklung und Messung der nationalen Bildungsstandards verantwortlich ist.

Perspektiven für die nächsten zwei Jahrzehnte Bildungsmonitoring

Die Beiträge in diesem Sammelband regen in unterschiedlichen Aspekten zum Nachdenken an: Sie reflektieren die etablierten Instrumente des Bildungsmonitoring und öffnen den Blick auf die Frage, was bei einer zukünftig anstehenden Überarbeitung der Gesamtstrategie zum Bildungsmonitoring zu berücksichtigen wäre. Eine in den letzten Jahren lauter werdende Kritik richtet sich an die bisherige Engführung des Bildungsmonitorings auf einzelne schulische Aspekte der Bildungslaufbahn. Für die aktuellen Ansätze des Bildungsmonitorings kann durchaus konstatiert werden, dass es sich im wesentlichen um ein „schulisches Bil-

„bildungsmonitoring“ handelte, das im Besonderen auf den Outcome bei den Schüler:innen im Bereich der fachlichen Kompetenzen von ausgewählten Fächern fokussierte, als um ein Monitoring, das mehrere Etappen einer Bildungslaufbahn erfasst oder umfassend außerschulische Aspekte berücksichtigt, die zur Bildung dazu gehören. Wenn es nun um die weitere Ausrichtung des Bildungsmonitoring geht, gilt es sowohl inhaltliche und methodische als auch technische und professionelle Aspekte zu berücksichtigen.

Inhaltliche Weiterentwicklung

In der Fachcommunity wird schon länger über die Erweiterung des Bildungsmonitorings um ergänzende inhaltlich-fachliche sowie überfachliche Aspekte diskutiert. Dazu zählt beispielsweise eine Erweiterung der bisherigen Kompetenzfeststellung um Aspekte der Demokratiefähigkeit oder um Aspekte der Bildung für nachhaltige Entwicklung (vgl. z. B. Holst et al. 2020). Zugleich wird auch um die Erweiterung im Bereich des Wohlbefindens, der psychischen Gesundheit oder der Selbstkonzepte von Lernenden nachgedacht (vgl. z. B. Reiß et al. 2023). Aktuell fokussieren die Instrumente der Gesamtstrategie zum Bildungsmonitoring in den meisten Bereichen auf Schüler:innen, obwohl bereits vielfältige Ansätze existieren, Monitoring auf andere Bildungsbereiche auszuweiten; beispielsweise auf den Bereich der frühkindlichen Bildung (Kuger et al. in diesem Band sowie z. B. BMFSJ 2023) und den Bereich der Erwachsenenbildung (vgl. z. B. Schrader/Kuper 2016; Müller et al. 2023). So fokussiert das Schwerpunktthema des 2024 erschienenen Bildungsberichts für Deutschland die berufliche Bildung.

Ziel wäre ein Bildungsmonitoring, das nicht nur Ergebnisse in Bereichen der formalen schulischen Bildung berücksichtigt, sondern auch non-formale und informelle Aspekte der Bildung, wie beispielsweise im schulischen Ganztags, sowie die Übergänge zwischen KiTa, Schule und Ausbildung in den Blick nimmt. Um diesem Ziel näher zu kommen, wäre bedenkenswert, bei der nächsten Überarbeitung der bisherigen KMK-Strategie des Bildungsmonitorings eine Abstimmung oder sogar Zusammenarbeit mit der Jugend- und Familienministerkonferenz (JFMK) herbeizuführen. Ebenso wie das BMBF veröffentlicht das BMFSJ Monitoringberichte im Bildungsbereiche; exemplarisch sei der „Gute-KiTa-Bericht“ genannt (u. a. BMFSJ 2021). Wünschenswert wäre, dass die Monitoringbestrebungen der KMK und der JFMK sich möglichst produktiv aufeinander beziehen und ergänzen, was eine strategische Abstimmung der Gremien voraussetzt. Die Vorzeichen könnten gut stehen: Im Oktober 2023 fand erstmals eine gemeinsame Sitzung der KMK und der JFMK statt, bei der auch die Aspekte der Übergänge von frühkindlicher Bildung in den Primarbereich der Schulen und die Qualität des Ganztags thematisiert wurden (KMK Pressemitteilung 2023). Für diese Bereiche wäre eine gemeinsame Monitoringstrategie begrüßenswert, die mit

einem gemeinsamen abgestimmten Indikatorensystem und Datenmanagement beginnt, hin zu einer gemeinsamen Berichtslegung und Diskussion der daraus abzuleitenden Konsequenzen.

Neben der inhaltlichen Erweiterung des Bildungsmonitorings um weitere Aspekte der Bildungsverläufe von Lernenden, wäre zudem vielversprechend, neben Lernenden auch die Lehrenden in das Tableau der Bildungsmonitoring-Daten aufzunehmen. Momentan liegen beispielsweise kaum Monitoringdaten über professionelles Wissen oder Einstellungen der Pädagog:innen in unterrichtlichen oder außerunterrichtlichen Angeboten vor, die wiederum wesentliche Informationen über die Voraussetzung für der Unterrichtsqualität bieten könnten. Gerade, um das Ziel der Gesamtstrategie zu erreichen, „Entwicklungen nicht nur zu beschreiben, sondern auch zu erklären“ (KMK 2015), ist es nötig, auch Indikatoren über den Professionalisierungsstand der Lehrenden und zur Qualität der Bildungsangebote in den Monitoringkanon aufzunehmen und diese systematisch auf die vorhandenen Daten über Lernstände zu beziehen.

Darüber hinaus wäre eine inhaltliche Weitung des Monitorings hin zur Prognostik vielversprechend. Derweil will Bildungsmonitoring den jeweiligen Status Quo abbilden und Entwicklungen über die Zeit analysieren. Dies erfolgt in der Regel als Rückschau und ermöglicht Aussagen dazu, wie sich beispielsweise Bevölkerungszahlen, die Anzahl an Schüler:innen mit Förderbedarfen oder die Basiskompetenzen seit der letzten Messung verändert haben. Bemängelt wird an diesen Ansätzen, dass eine tatsächliche Steuerung auf ein Ziel hin ausschließlich in der Retrospektive und ohne Vorausschau auf ein konkretes Ziel schwer möglich ist. Bisher werden die vorliegenden Daten jedoch kaum für prognostische Modelle verwendet, wie sie beispielsweise in der Ökonomie regelmäßig für die Abschätzung von Veränderungen zum Einsatz kommen. Erste Ansätze der Modellierung von Vorhersagemodellen mit Daten aus Bildungsstudien finden sich unter anderem bei Kaplan (u. a. Kaplan/Huang 2021; Kaplan/Harra/Stampka/Jude, im Druck).

Methodische Weiterentwicklung

Die Erkenntnisse über Lernstände und Qualität von Schulen aus der bisherigen Strategie zum Bildungsmonitoring basieren überwiegend auf quantitativen Methoden der empirischen Bildungsforschung. Aus diesen Bemühungen sind hilfreiche und sinnvolle Erkenntnisse entstanden. Die überarbeitete Strategie aus dem Jahr 2015 fordert allerdings mehr „Wirkungs- und Erklärungswissen“.

Vor dem Hintergrund der umfassenden und elaborierten quantitativen Zugänge des aktuellen Bildungsmonitorings, die auf summative Indikatoren abzielen, wäre mit Blick auf die Frage nach der Erfassung der Qualität von Bildungsprozessen über die Erweiterung der forschungsmethodischen Ansätze zu diskutie-

ren. Der Mehrwert qualitativer Zugangsweisen, die zum Beispiel durch die Analysen von Dokumenten der Bildungssteuerung, von Maßnahmen der Implementation oder der Interessen unterschiedlicher Akteursgruppen auch Prozesse in den Blick nehmen, wird bisher erst in Ansätzen erkannt und noch nicht in Form von Kennzahl in Ansätzen des Monitorings erfasst.

Bedenkenwert wäre, neben der längsschnittlichen Begleitung von Leistungsdaten der Schüler:innen ebenfalls längsschnittlich die Handlungen, Einstellungen und Entwicklungen zentraler Akteure einzelner Schulen zu erfassen: Das Monitoring also entsprechend um kasuistische Betrachtungsweisen im Sinne qualitative Fallarbeit zu ergänzen (vgl. u. a. Bender/Dietrich/Silkenbeumer 2021). Fälle können sowohl Organisationen (z. B. Schulen, Kindertageseinrichtungen oder Ausbildungsbetriebe) als auch Akteure wie Pädagog:innen, Leitungen oder Schüler:innen sein. Die könnten anhand qualitativer Stichprobenpläne (u. a. Kelle und Kluge 2010) unter bestimmten festgelegten Kriterien ausgewählt und parallel zum bestehenden Monitoring über mehrere Jahre begleitet werden. Ebenso wie für die bereits vorhandenen Aspekte könnten Indikatoren festgelegt werden, zu denen Informationen über ein Monitoring gesammelt wird – hierbei wären die vorhanden numerischen Darstellungen der Informationen um qualitative Darstellungsformen wie Netzwerkkarten oder Texte nachzudenken. Indikatoren könnten beispielsweise die Schulentwicklungskapazität (u. a. Maag Merki, Wullschleger und Rechsteiner 2021), die Kommunikations- und Kooperationsnetzwerke unter Bildungsakteuren oder die Common Sense-Theorien – also die Alltagstheorien – der handelnden Pädagog:innen und Schüler:innen (u. a. Bohnsack 2017) sein, die im Monitoring Auskunft über den Entwicklungsstands des Systems an sich aufzeigen könnten. Qualitativ-rekonstruktive Zugänge (Bohnsack 2021) in der Auswertung, wie beispielsweise die dokumentarische Organisationsforschung, haben das Potenzial neben expliziten Regeln, Programmen und Regulationsprozessen im Zusammenwirken der Akteure im Bildungssystem auch deren implizite Programme und Routinen zugänglich zu machen (u. a. Nohl 2017, Nohl und Somel 2017), die zentrale Informationen darüber geben, auf welche (gelebte) Struktur und Alltagspraxis beispielsweise politisch initiierte Bildungsprogramme treffen. Multimethodische Datenerhebungen (u. a. Vogd 2009) ermöglichen neben den quantitativen Daten auch authentische Daten aus dem Feld ins Monitoring einzubeziehen. Die notwendige Voraussetzung, um das bisherige Monitoring um derartige Erkenntnisse zu erweitern, wäre die methodische Professionalisierung aller Akteursgruppen in diesem Bereich, da an den Qualitätsinstituten derweil die Methodenkenntnisse im quantitativen Forschungsbereich überwiegen – entsprechende Stellenprofile für die qualitative Fallarbeit müssten organisiert werden oder neue Formen der Kooperation mit Universitäten im Monitoring eingegangen werden.

Neben der Erweiterung der methodischen Zugänge gehört auch ein stärkerer Fokus auf die Theoriebildung auf Basis der Erkenntnisse der letzten 25 Jahre.

Inbesondere die in den letzten Jahren zunehmend entstandenen systematischen Forschungssynthesen und Review-Studien im Bereich des Bildungsmonitorings, die empirische Erkenntnisse unterschiedlicher Studien auch über die Zeit vergleichen und diskutieren, bieten inzwischen gute Ansatzpunkte für die Weiterentwicklung von Theorien der bildungspolitischen Steuerung, der Schulentwicklung sowie der Frage nach Qualitätsmerkmalen von Unterricht.

Technische Weiterentwicklung

So wie in allen gesellschaftlichen Bereichen befindet sich auch das Bildungsmonitoring in einem digitalen Transformationsprozess. Das betrifft Datenerhebung, -aufbereitung, -auswertung und -haltung: Kurzum alle Bereiche des Datenmanagements. Neben der Einführung von Datenbanken zählen auch die verfahrensübergreifende Variablenbezeichnung oder die Einführung einer bundesweiten übergreifenden Bildungs-ID zu diesem Bereich (u. a. Brändle 2024). Ziel all dieser Bemühungen sollte sein, die Datenqualität zu erhöhen sowie zusätzliche Möglichkeiten der längsschnittlichen Auswertungen zu ermöglichen, um der Forderungen nach der Erklärung von Wirkungszusammenhängen näher zu kommen. Darüber hinaus verspricht die durchgängige Digitalisierung, Analysen zeitlich zu beschleunigen und damit – unter gleichbleibender Ressource – mehr Zeit für vertiefende Analysen zu haben. Die digitale Erfassung bietet Potenzial für die Analyse von Prozessen zusätzlich zu Outcomes, beispielsweise durch die kleinschrittige Aufzeichnung von Verhaltensdaten (Anghel et al. 2024). Diverse Lernplattformen, die unabhängig von Akteuren des und Bemühungen im Bildungsmonitoring entstehen, bieten Pädagog:innen in Lehr-Lernsituationen unmittelbar Daten zum Lernstand der Schüler:innen in bestimmten Bereichen (u. a. Hartong/Decuyper 2023). Die Entwicklung wirft Fragen unter anderem bezüglich des Verhältnisses dieser Daten zu Daten aus Lernstandserhebungen des „klassischen“ Bildungsmonitorings, bezüglich der Datenqualität, der „Datifizierung“ der Schule (u. a. Bock et al. 2023) und der Dateninfrastruktur auf (u. a. Hartong/Nikolai 2021). Eine konzeptionelle Klärung steht bisher aus. Zugleich gilt es auch im Bildungsmonitoring, nicht den Anschluss zu digital- und KI-gestützten Systemen zu verlieren. Hier eröffnen sich in den nächsten Jahren mannigfaltige Entwicklungsräume, wie die Programmierung von einer Bildungsmonitoring-GPT, die auf Grundlage der bereits vorhandenen Erkenntnisse und Daten trainiert werden kann, oder die inhaltliche Auswertung von größeren Textmengen in der Codierung von offenen Antworten bei Lernstandserhebungen und Befragungen (vgl. z. B. Andersen et al. 2022). Zugleich ist es in dieser Diskussion unerlässlich, Fragen der Datenhoheit und Datensicherheit zu berücksichtigen.

Professionelle Weiterentwicklung

Nach knapp zwei Jahrzehnten Erfahrung im Bildungsmonitoring stellt sich auch die Frage nach dem Professionalisierungsbedarf der Akteure, die im Bildungsmonitoring arbeiten (vgl. z. B. Ackerman et al. 2023). In viele Fällen sind sie diejenigen, die Ansätze des Monitorings konzipieren, die Daten des Monitorings erheben, aufbereiten und kommunizieren. Exemplarisch seien drei Fragen aufgeworfen: Wäre es an der Zeit eine Fachgesellschaft für Bildungsmonitoring zu gründen? Wäre es ebenso an der Zeit, eine Aus- und Fortbildungsstrategie für Mitarbeiter:innen im Bildungsmonitoring zu konzipieren und vor allem umzusetzen? Und: Wer evaluiert eigentlich das Bildungsmonitoring? Dabei sollte auch diskutiert werden, wie mit den unterschiedlichen Akteursgruppen eine Datenkompetenz aufgebaut werden kann.

- *Fachgesellschaft Bildungsmonitoring*

Bisher sind mehrere Fachgesellschaften in Teilen Versammlungsorte für Bildungsmonitorende, wie beispielsweise die Kommission Bildungsorganisation, Bildungsplanung, Bildungsrecht (KBBB) als Sektion der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE) sowie deren Arbeitsgruppe für empirische Pädagogische Forschung (AEPF), die Gesellschaft für empirische Bildungsforschung (GEBF), oder die Deutsche Gesellschaft für Bildungsverwaltung (DGBV) und das Netzwerk zur empiriegestützten Schulentwicklung (EMSE). Die Teilnahme an Tagungen und Diskursen von mehreren Fachgesellschaften fördern selbstverständlich die Vernetzung von Bildungsmonitoren zu ihren zentralen Schnittstellenpartnern in Wissenschafts- und Bildungsadministration, haben jedoch auch den Nachteil, dass ein gemeinsamer Diskurs zu originären Themen des Bildungsmonitorings nur schwer über mehrere Jahre breit geführt und aufrechterhalten werden kann.

Zudem etablierten sich über die letzten Jahre „Teilnetzwerke“ entlang einzelner Instrumente des Monitorings, wie die VERA-Steuergruppe im Bereich der Lernstandserhebungen und KODEX als Konferenz der externen Evaluation für Schulinspektionen sowie auf internationaler Ebene die jährlichen Forschungskonferenzen der International Association for the Evaluation of Educational Achievement (IEA IRC). Während derartige Netzwerke eine hohe Bedeutung für den fachlichen Diskurs zum jeweiligen Instrument haben, bergen sie auch die Gefahr – wenn sie nicht um andere Netzwerke ergänzt werden – dass sich die Instrumente des Monitorings zu sehr vereinzeln und nicht vernetzt gedacht werden.

Das „Netzwerk Bildungsmonitoring“, in dem die Leitungen von acht Qualitätsinstituten sich vernetzen, ist eine Möglichkeit, das gemeinsame systemische Denken und Sprechen über das Zusammenspiel der Instrumente des Monitorings zu fördern. Derartige fachliche Diskurse sind auch für die Mit-

arbeiter:innen im Bildungsmonitoring zentral. Eine Fachgesellschaft könnte den Charme entfalten, institutionsübergreifende Standards zu verhandeln und festzulegen, wie beispielsweise für die Qualifikation sowie Aus- und Fortbildung von Mitarbeiter:innen im Bildungsmonitoring, für Instrumente und Methoden, für Kommunikationswege oder Produkte. Selbst wenn sich eine Fachgesellschaft gegen die Einigung auf Standards in diesen oder weiteren Bereichen einigen würde, birgt allein der Diskurs das Potenzial einen Beitrag zur Qualitätssicherung im Bildungsmonitoring zu leisten.

- *Aus- und Fortbildungsstrategie für Bildungsmonitorende*

An den Qualitätsinstituten, die die Instrumente des Bildungsmonitorings entwickeln und einsetzen, arbeiten überwiegend statistisch- und datenaffine Akademiker:innen aus den Fachrichtungen Psychologie, Soziologie, empirische Bildungsforschung und verwandten sozialwissenschaftlichen Disziplinen. Eine systematische Auswertung, wer mit welcher Qualifikation an den Instituten arbeitet ist uns nicht bekannt und wäre durchaus – im Sinne einer Erkenntnisgewinnung über das Monitoring an sich – sinnvoll. Für ihre berufliche Aus- und Weiterbildung sind die Qualitätsinstitute selbst zuständig. Für methodisch und inhaltliche Professionalisierung können die Angebote von Fachgesellschaften, wie beispielsweise Summer Schools oder Tagungen besucht werden. Zugleich sind die Institute meistens an Weiterbildungsstrukturen der Bildungsverwaltung angebunden mit Fortbildungen zum Projekt- oder Führungsmanagement. Einheitliche Standards für die Qualifikation von Mitarbeiter:innen im Bildungsmonitoring bestehen allerdings noch nicht und sind für die Weiterentwicklung des Bildungsmonitorings als Fachdisziplin nicht nur wünschenswert, sondern notwendig. Entsprechende Diskurse entwickeln sich derweil in den USA, die bereits Standards diskutieren (Ackerman et al. 2024).

- *Evaluation des Bildungsmonitorings*

Einzelne Elemente des Bildungsmonitorings wurden bereits teilweise und punktuell evaluiert, wie beispielsweise Teile von Schulinspektionen in einzelnen Bundesländern oder – aktuell in einzelnen Bundesländern geplant – der Sozialindex. Was bisher aussteht ist eine Evaluation des Zusammenspiels der Elemente des Bildungsmonitorings und, inwiefern die Bildungsmonitoring-Strategie der KMK ihre Ziele erreicht. Eine derartige Evaluation könnte Hinweise auf Stärken und Optimierungsbedarfe geben. Geklärt werden müsste allerdings, wer Auftraggeber:in der Evaluation sein könnte; dies könnte sowohl der Bund als auch eine Vereinigung der Länder sein. Im Grundsatz könnte wiederum die Evaluation des Bildungsmonitorings an sich auch als Ziel in der nächsten Weiterentwicklung der Gesamtstrategie festgehalten werden.

Fragenimpulse zur Weiterentwicklung des Bildungsmonitorings

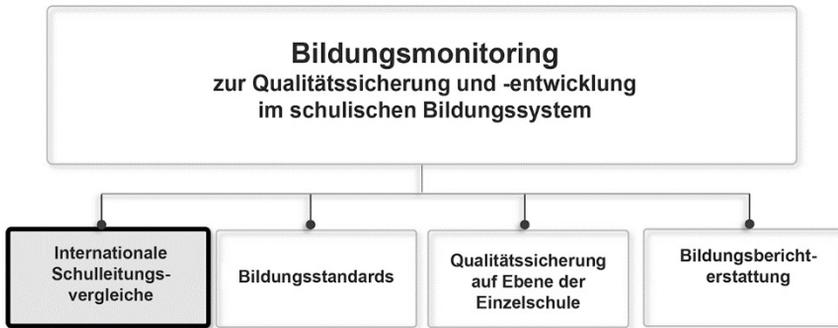
Die Idee einer Gesamtstrategie zum Bildungsmonitoring begann vor fast 20 Jahren und es ist absehbar, dass eine Überarbeitung der 2015er Strategie der KMK ansteht. Die vielfältigen Beiträge dieses Sammelbandes können als Impulse für eine Weiterentwicklung dienen. Als Kondensat aus den Anregungen, die in den Beiträgen für die einzelnen Teilbereiche des Bildungsmonitorings vorhanden sind, könnten folgende zehn Fragenimpulse für die Weiterentwicklung dienlich sein:

- Welchen Mehrwert hätte ein Bildungsmonitoring, das sich auf lebenslanges Leben bezieht?
- Inwiefern können neben den schulischen Aspekten im Bildungsmonitoring weitere Bildungsbereiche in den Blick genommen werden, wie frühkindliche und Erwachsenenbildung sowie non-formale und informelle Bildungsbereiche?
- In welchen Monitoringbereichen lassen sich durch längsschnittliche Stichproben Bildungsverläufe abbilden?
- Inwiefern können neben den bisherigen Outcomes im Sinne von Basiskompetenzen überfachliche und Zukunftskompetenzen in das Tableau aufgenommen werden?
- Welche quantitativ und qualitativ erfassbaren Indikatoren wollen wir im Monitoring abbilden?
- Welche vorhandenen Indikatoren sollten weiterverfolgt, welche zukünftig entfallen und welche ergänzt werden?
- Wie können Prozesse der Qualitätsentwicklung in den Bildungsorganisationen in Monitoringdaten abgebildet werden?
- Wie kann ein Monitoring aussehen, dass die Ergebnisse für unterschiedliche Ebenen des Bildungssystems nutzbar machen lässt (System, Organisation, Individuum)?
- An welchen Stellen können Informationen aus den kommunalen Bildungsmonitoring nutzbar gemacht werden für eine länderübergreifende Betrachtung?
- Braucht es eine Evaluation des Bildungsmonitorings als solche?

Literatur

- Abs, Hermann J./Hahn-Laudenberg, Katrin/Deimel, Daniel/Ziemes, Johanna F. (Hrsg.) (2024): ICCS 2022. Schulische Sozialisation und Politische Bildung von 14-Jährigen im internationalen Vergleich. Münster, New York: Waxmann.
- Ackerman, Terry A./Bandalos, Deborah L./Briggs, Dere C./Everson, Howard T./Ho, Andrew D./Lottridge, Susan M./Madison, Matthew J./Sinharay, Sandip/Rodriguez, Michael C./Russell, Michael/von Davier, Alian A./Wind, Stefanie A. (2024): Foundational Competencies in Educational Measurement. Educational Measurement: Issues and Practice.

Teil A:
Internationale
Schulleistungsvergleiche



Mit der Teilnahme an der Third International Mathematics and Science Study (TIMSS) und dem Programme for International Student Assessment (PISA) um die Jahrtausendwende erhielt Deutschland zum ersten Mal standardisierte Erkenntnisse über die Bildungsqualität von Schulen im internationalen Vergleich. Seitdem ist die regelmäßige Teilnahme an internationalen Schulleistungsstudien ein wesentlicher Bestandteil der Gesamtstrategie zum Bildungsmonitoring. Erfasst werden die Kompetenzen in Mathematik, Naturwissenschaften und Lesen in der Grundschule durch die Teilnahme an TIMSS und der Internationalen Grundschul-Lese-Untersuchung (IGLU/PIRLS) sowie in der Sekundarstufe I durch die Teilnahme an PISA. Der Vergleich von Bildungsqualität auf Systemebene betrachtet Bildungsergebnisse aus einer globalen Perspektive: Verfügen Schüler:innen in Deutschland über grundlegende Kompetenzen für die Partizipation in einer internationalen ausgerichteten Gesellschaft und erfüllt unser Bildungssystem das Ziel der qualitativ hochwertigen Bildung für alle?

Die drei Beiträge in diesem Abschnitt diskutieren den Status Quo dieses Monitoringansatzes mit Blick auf die Systemebene. *Christine Sälzer* beschreibt die Entwicklungen von Schulleistungsstudien und das grundsätzliche Erkenntnispotenzial der internationalen Perspektive. *Abel Schumann* erläutert die Bedeutung von internationalen Bildungsindikatoren und die Herausforderungen des Datenmanagements bei der Generierung aussagekräftiger, internationaler Daten. *Jana Groß Ophoff, Stefan Brauckmann-Sajkiewicz, Pierre Tulowitzki, Marcus Pietsch und Colin Cramer* eröffnen die Perspektive für die konzeptionellen Grundlagen der Entwicklung eines internationalen Monitorings, das Schulleitungen in den Blick nimmt.

Warum immer noch PISA?

Internationale Schulleistungsstudien als Kernelement des Bildungsmonitorings in Deutschland

Christine Sälzer

Zusammenfassung

Die PISA-Studie und der sogenannte „PISA-Schock“ stehen exemplarisch für einen prüfenden Blick auf die Leistungsfähigkeit unseres Schulsystems im Vergleich mit Bildungssystemen weltweit. Seit mittlerweile mehr als zwei Jahrzehnten werden funktionale Grundkompetenzen von Schüler:innen gemessen, die Entwicklung der durchschnittlichen Kompetenzen und weiterer Maße für die Qualität von Bildungssystemen ist durchwachsen und wird auch im öffentlichen Diskurs zunehmend kritisiert: Das deutsche Bildungssystem steckt in einer veritablen Krise. Angesichts dessen lohnt sich die Betrachtung der Frage, welchen Mehrwert diese internationale Perspektive konkret für Bildungspolitik und Bildungspraxis in Deutschland bringt und ob sich eine Beteiligung an PISA weiterhin lohnt. Der Beitrag greift auf, was die zentralen Funktionen internationaler Schulleistungsstudien sind und welches die für Deutschland wichtigsten Herausforderungen im Bildungssystem sind, die daraus abgeleitet werden können. Zum Anspruch von PISA und Co. gehört auch deren kontinuierliche methodische und konzeptuelle Weiterentwicklung, ohne die eine Erfassung funktionaler Grundbildung nicht möglich wäre. Entsprechend wird herausgearbeitet, inwieweit sich diese Studien an dynamische Entwicklungen sowohl im Bereich der Bildungsinhalte als auch im Bereich der Bildungsforschung angepasst haben. Den Abschluss bildet ein Plädoyer für langfristiges, regelmäßiges Bildungsmonitoring.

Wie alles begann

Die Veröffentlichung der ersten Ergebnisse des *Programme for International Student Assessment* (PISA) der OECD im Jahr 2001 brachte nach bereits schwachen Ergebnissen in der *Trends in International Mathematics and Science Study* (TIMSS) vier Jahre zuvor (Baumert/Lehmann 1997) das Selbstverständnis des deutschen Bildungswesens gehörig ins Wanken: Sowohl das unterdurchschnittliche Abschneiden der Schüler:innen in Deutschland sorgte für ungläubige und distanzierende Reaktionen als auch die systematischen Disparitäten innerhalb Deutschlands (Baumert 2001; Baumert et al. 2002). Kompetenzunterschiede wurden besonders im Lesen und in der Mathematik zwischen Jungen und Mädchen offensichtlich, aber auch zwischen Schüler:innen mit und ohne Zuwanderungshintergrund oder aus unterschiedlichen Bundesländern und mit unterschiedlichem sozioökono-

mischem Hintergrund (Baumert et al. 2002; Baumert et al. 2006; Baumert et al. 2003). Unter dem Begriff „PISA-Schock“ (Roeder 2003) schrieben diese Ergebnisse ein Stück Bildungsgeschichte. Für sich genommen sind unterschiedlich ausgeprägte Kompetenzen ganz normal und kein Grund zur Sorge; allerdings sind diese Unterschiede dann problematisch, wenn sie systematisch auftreten, d. h. in Verbindung mit bestimmten unveränderlichen Merkmalen der Schüler:innen. Denn dann liegt nahe, dass Schüler:innen mit bestimmten Merkmalen in der Schule weniger erfolgreich lernen als andere, also möglicherweise aufgrund dieser Merkmale benachteiligt sein können (Sälzer 2021). Solche unveränderlichen Merkmale sind beispielsweise der sozio-ökonomische Hintergrund von Schüler:innen oder ihre Zuwanderungsgeschichte.

Gut 20 Jahre nach dem PISA-Schock sind zahlreiche differenzielle Entwicklungen im Bereich des Bildungsstandes der Jugendlichen in Deutschland nachgezeichnet worden. Es gab Erfolge zu vermelden, etwa im Rahmen von PISA 2012, als die Kompetenzmittelwerte erstmals in allen drei Domänen Lesen, Mathematik und Naturwissenschaften über dem OECD-Durchschnitt lagen (Prenzel et al. 2013). Zugleich verdeutlichte nicht nur die PISA-Studie, sondern regelmäßig auch andere Large-Scale Assessments wie TIMSS oder der Bildungstrend des Instituts zur Qualitätsentwicklung im Bildungswesen (kurz: IQB-Bildungstrend), dass wesentliche Herausforderungen nach wie vor bestehen und beispielsweise die Digitalisierung der Schulen und vor allem des Fachunterrichts weit hinter den Entwicklungen unserer Nachbarländer zurückbleibt (Eickelmann et al. 2019; Reiss et al. 2019; Schwippert et al. 2020). In der jüngsten PISA-Studie aus dem Jahr 2022 wurden für die Schüler:innen in Deutschland gar die niedrigsten Kompetenzmittelwerte überhaupt in allen bisherigen PISA-Runden erfasst, ein neuer Tiefpunkt also (Lewalter et al. 2023), der etwa in der Berichterstattung der Süddeutschen Zeitung als neuer PISA-Schock rezipiert wurde (dpa 2023).

Die Ständige Konferenz der Kultusminister in Deutschland (KMK) reagierte 2002 auf den damaligen PISA-Schock mit der Verabschiedung von sieben zentralen Handlungsfeldern (KMK 2002). In diesen sieben zentralen Handlungsfeldern wurden Maßnahmen identifiziert, mit deren Hilfe die Qualität des Bildungswesens in Deutschland langfristig und nachhaltig verbessert werden sollte. Insofern initiierte die KMK mit den sieben zentralen Handlungsfeldern eine umfangreiche Reform-Agenda und damit eine langfristige Strategie zum Bildungsmonitoring in Deutschland, die gezielt Schwerpunkte auf wesentliche Defizite im deutschen Bildungssystem setzte (etwa Lehrkräftebildung, frühe sprachliche Förderung oder den Ausbau gebundener Ganztagschulen). Die regelmäßige Beteiligung Deutschlands an nationalen und internationalen Schulleistungsstudien ist in den Handlungsfeldern festgeschrieben und damit ein Kernelement des Bildungsmonitorings in Deutschland. Neben der bekanntesten internationalen Bildungsvergleichsstudie, PISA, nimmt Deutschland an mehreren weiteren na-

tionalen und internationalen Schulleistungsuntersuchungen teil. Tabelle 1 gibt einen Überblick über diese Studien.

Tab. 1: Nationale und internationale Schulleistungsstudien mit Beteiligung Deutschlands.

	PISA	TIMSS	PIRLS/IGLU	ICILS	IQB-Bildungstrend	NEPS
Initiierende Organisation	OECD	IEA	IEA	IEA	KMK	BMBF
Bisherige und geplante Erhebungs- runden	2000, 2003, 2006, 2009, 2012, 2015, 2018, 2022, 2025	1995, 1999, 2003, 2007, 2011, 2015, 2019, 2023	2001, 2006, 2011, 2016, 2021, 2026	2013, 2018, 2023	2008/2009, 2011, 2012, 2015, 2016, 2018, 2021, 2022, 2024	seit 2010, abhängig von der Kohorte
Population/ Zielgruppe	15-jährige Schüler:innen	Schüler:innen der vierten Jahrgangsstufe und/oder Schüler:innen der achten Jahrgangsstufe bzw. am Ende Sekundarstufe I	Schüler:innen der vierten Jahrgangsstufe	Schüler:innen der achten Jahrgangsstufe	Schüler:innen kurz vor einem bestimmten Schulabschluss (Primarstufe, Sekundarstufe I, Sekundarstufe II)	Sechs Kohorten: Kleinkinder (7 Monate), 4-Jährige, Schüler:innen der Klassenstufen 5 und 9, Studienanfänger:innen sowie Erwachsene der Jahrgänge 1944 bis 1986
Design	Querschnitt, Trend	Querschnitt, Trend	Querschnitt, Trend	Querschnitt, Trend	Querschnitt, Trend	Längsschnitt mit mehreren Kohorten

Es handelt sich dabei um eine breit orchestrierte, einander ergänzende Palette an wissenschaftlich fundiert durchgeführten Studien. Die Studien verfolgen unterschiedliche Zugänge und Schwerpunkte, sind nicht unnötig redundant und differenzieren gezielt curriculare Validität (Abdeckung von Lehrplänen durch die Tests) und Kompetenzmodelle wie das Literacy-Konzept (funktionale Grundbildung, vgl. etwa (Sälzer 2016)). Im Folgenden werden die wichtigsten Ziele, Funktionen und einige exemplarische Erkenntnisse aus PISA und anderen Large-Scale Assessments skizziert.

Ziele von PISA & Co.

Die PISA-Studie verfolgt seit mittlerweile mehr als 20 Jahren das Ziel, festzustellen, inwieweit Jugendliche gegen Ende der Pflichtschulzeit in der Lage sind, das (mutmaßlich) in der Schule gelernte Wissen in alltäglichen Situationen anzuwenden (OECD 2019a). Wie gut Jugendliche auf die Anforderungen einer sich ständig wandelnden Wissensgesellschaft vorbereitet sind, zeigen mit jeder PISA-Erhebungsrunde zahlreiche Berichtsbände der OECD, aber auch der nationalen Projektzentren in den beteiligten Bildungssystemen. Alle drei Jahre werden Schüler:innen im Alter von 15 Jahren in den drei Domänen Lesen, Mathematik und Naturwissenschaften getestet; zusätzlich geben sie Auskunft in einem modular aufgebauten Fragebogen, der unter anderem zur Modellierung der theoretisch angenommenen Einflüsse von Lernumgebungen und familiären Umfeldern dient (Kuger et al. 2017). Neben den Jugendlichen werden in einem solchen Fragebogen auch deren Eltern, Schulleitungen und Lehrkräfte befragt. Dabei ist die Stichprobe auf nationaler Ebene (also: für Deutschland) repräsentativ, nicht jedoch auf der Ebene der Bundesländer. Beteiligt sind alle Schulformen, an denen (potenziell) Schüler:innen im Alter von 15 Jahren zu finden sind (OECD 2020b). Nachdem die Schulen in einer geschichteten Wahrscheinlichkeitsstichprobe ausgewählt worden sind, erfolgt die Ziehung der Schüler:innen zur Teilnahme an PISA innerhalb der gezogenen Schulen nach dem Zufallsprinzip. Dieses Prozedere der Stichprobenziehung läuft in allen Bundesländern gleich ab und berücksichtigt die jeweils existierenden Schulformen, die potenziell von 15-Jährigen besucht werden. Das Gymnasium ist hierbei die einzige Schulform, die in allen 16 Bundesländern vorhanden ist. Daneben gibt es unterschiedlich viele weitere Schulformen, an denen 15-Jährige zu finden sind und sie alle werden bei der Stichprobenziehung berücksichtigt. Neben Basiskompetenzen in den genannten drei Domänen erfasst PISA jeweils auch eine sogenannte lehrplanübergreifende, extracurriculare Kompetenz. Bisher waren dies unter anderem das selbstregulierte Lernen (Baumert 2001), Problemlösen (Klieme et al. 2001), Problemlösen im Team (Zehner et al. 2019), Global Competence (Sälzer/Roczen 2018) oder Creative Thinking (OECD 2019d).

Durchführung von PISA

Über die Jahre hat sich der Teilnehmerkreis von PISA kontinuierlich erweitert. Waren es bei der ersten Runde im Jahr 2000 noch 32 Staaten, darunter 28 OECD-Staaten, so sind es in PISA 2022 bereits 40 OECD-Staaten und 46 Partnerstaaten bzw. Partnerregionen, die sich für eine Teilnahme an PISA entschieden haben. Die Studie wird initiiert und geleitet von der OECD in Paris. Die beteiligten Bildungssysteme haben je ein nationales Projektmanagement, das an Einrich-

tungen der Bildungsadministration oder Bildungsforschung angesiedelt ist und aus Teams mit unterschiedlichen Rollen besteht. So verfügt jedes nationale Projektzentrum über Expertise im Bereich Projektkoordination, Datenmanagement, Stichprobenziehung, Übersetzung sowie in den in PISA erfassten Domänen.

Grundbildungskonzept

Konzeptuell legt PISA ein funktionales Grundbildungskonzept zu Grunde, welches als Literacy bezeichnet wird (Baumert 2001). Zentral ist dabei, dass die erfassten Inhalte und Prozesse sowohl anwendbar sind für die aktuelle und zukünftige Teilhabe an einer Kultur als auch anschlussfähig für nachfolgendes Lernen (Sälzer 2016). Dahinter steckt die Annahme, dass eine solide Grundbildung die Jugendlichen dazu befähigt, auch in neuen, unbekanntenen Situationen durch erfolgreichen Transfer handlungs- und teilhabefähig zu sein, informierte Entscheidungen treffen zu können und ein eigenständiges, selbstbestimmtes (Erwerbs-)Leben zu führen. Konkret misst der PISA-Test also, inwieweit die 15-Jährigen bis zum Testtag Lerngelegenheiten genutzt haben, um in bestimmten Bereichen des Lebens gut zurechtzukommen. Dabei wird die Qualität vorangegangener Bildungsprozesse als Grundlage für die Prognose künftiger Bildungsprozesse herangezogen. Die Validität des Assessments konnte beispielsweise durch einen Längsschnitt im Rahmen der kanadischen Youth in Transition Study gezeigt werden (Borgonovi et al. 2017). Mit dem Anspruch der Lebensweltnähe und Anschlussfähigkeit geht auch der Anspruch einher, dass die konkreten Inhalte der zu messenden Grundbildung nicht statisch ist, sondern mit der Zeit an die Anforderungen der Gesellschaft, insbesondere des Arbeitsmarktes, angepasst werden müssen (OECD 2019a).

Mehr als eine Klassenarbeit

Ein zentraler Punkt bei der Bewertung der PISA-Studie ist der Hinweis, dass es sich beim PISA-Test nicht um eine Klassenarbeit oder Klausur handelt: Der PISA-Test folgt einer gänzlich anderen Logik. Während in einer Klassenarbeit der Unterrichtsstoff vergangener Wochen geprüft wird, wird der in PISA erfasste Kompetenzbegriff der Literacy in einer theoretischen Rahmenkonzeption beschrieben und eingegrenzt (OECD 2019a; Sälzer 2016). Darin werden die getesteten Domänen strukturiert und in Subdimensionen untergliedert, denen dann im PISA-Test entsprechende Aufgaben (Items) zugeordnet werden. Beispielsweise lässt sich die Domäne Mathematik in Inhalte, Prozesse und Kontexte unterteilen, wobei die Inhalte etwa Raum und Form (Geometrie) oder Veränderung und Beziehungen (Algebra) heißen, die Prozesse grundlegende mathematische Aktivitäten wie die

Interpretation mathematischer Ergebnisse umfassen und die Kontexte verschiedene Lebensbereiche der befragten Jugendlichen abbilden (Reiss et al. 2016). PISA strebt explizit keine curriculare Validität an, d. h. die Kompetenzmessung bezieht sich nicht darauf, inwieweit die Lehrpläne in einzelnen Bildungssystemen erfolgreich umgesetzt werden (Sälzer 2016). Zwar weisen die Curricula im internationalen Vergleich Überschneidungen auf, jedoch sind sie bei weitem nicht identisch. Bezugsnorm ist daher die theoretische Rahmenkonzeption für jede erfasste Domäne. Die Ergebnisse pro Domäne, d. h. Mittelwerte und Verteilungen auf inhaltlich beschriebene Kompetenzstufen, lassen sich als repräsentatives Bild aller 15-Jährigen in den jeweiligen Teilnehmerstaaten darstellen. Eine Individualdiagnostik oder Rückmeldung an die Teilnehmenden ist jedoch nicht möglich: Da der PISA-Test in der Regel einen Vormittag umfasst, kann jede:r Schüler:in lediglich einen Bruchteil aller verfügbaren Testaufgaben bearbeiten. Somit erhält man jeweils nur ein unvollständiges Bild der Aufgaben, die ein:e Schüler:in lösen kann — von allen Aufgaben, die sie theoretisch hätten bekommen können. Aus diesem Grund ist eine Einzeldiagnostik der Kompetenz je Schüler:in anhand von PISA nicht zulässig. Die Testaufgaben werden jedoch systematisch auf alle teilnehmenden Jugendlichen verteilt, so dass jede Aufgabe von ausreichend vielen Teilnehmenden bearbeitet wird und an unterschiedlichen Stellen im Test vorkommt (so genanntes Multi-Matrix-Design (vgl. Sälzer 2016)). Auf dieser Basis lassen sich Angaben für das Gesamtbild aller Teilnehmenden machen, ohne dass individuelle Ergebnisse ermittelt werden.

Funktionen von Large-Scale Assessments

Internationale Large-Scale Assessments haben zwei Hauptfunktionen: Monitoring und Benchmarking (Seidel/Prenzel 2008). Monitoring bedeutet, dass anhand der Studien eine Bestandsaufnahme von Stärken und Schwächen erfolgt und damit eine gezielte Identifikation von Problemen im Bildungswesen. Ergänzend dazu liefert das Benchmarking einen Vergleich mit anderen Staaten oder Regionen, Einordnung der Befunde in einen größeren Kontext. Damit sind nationale wie internationale Vergleichsstudien weit mehr als ein bloßes Ranking. Die Studien sind dabei häufig als Trenduntersuchung mit mehreren Durchgängen angelegt, so dass Entwicklungen von Kennwerten wie Kompetenzen oder des sozialen Gradienten, der den Zusammenhang zwischen dem sozioökonomischen Hintergrund und der Kompetenz in den PISA-Domänen angibt, über die Zeit darstellbar sind. Konkret die PISA-Studie untersucht Kompetenzen und Bildungsergebnisse auf verschiedenen Ebenen: Bildungssystem, Schule, Klasse sowie Schüler:innen. Insgesamt betrachtet geben diese Ebenen ein sehr detailreiches Bild davon, wo spezifische Stärken und Schwächen von Bildungssystemen liegen — sowohl im zeitlichen Verlauf (Trends) als auch im Ver-