



Qualitätsorientierte Perspektiven: Welche Instrumente und Verfahren sollen zur Qualitätssicherung und Qualitätsentwicklung im Schulsystem beitragen?

Eine staatliche Verantwortung für die *Qualität* im Schulsystem gibt es seit jeher, allerdings hat der Qualitätsbegriff in den beiden vergangenen Jahrzehnten in der öffentlichen, bildungspolitischen und auch bildungswissenschaftlichen Diskussion erheblich an Bedeutung gewonnen. Seit dem unerwartet schwachen Abschneiden der Schülerinnen und Schüler in Deutschland bei internationalen Leistungsvergleichen zu Beginn des Jahrtausends haben im deutschen Schulsystem tiefgreifende Veränderungen stattgefunden, dazu zählen u.a. auch die Einführung verschiedener Instrumente und Verfahren der *Qualitätssicherung* und *Qualitätsentwicklung* im schulischen Mehrebenensystem (vgl. Abschn. 4.1), verbunden mit dem Ziel, die Qualität des Schulsystems (*Systemqualität*), der Einzelschule (*Schulqualität*) und des Unterrichts (*Unterrichtsqualität*) zu sichern und weiterzuentwickeln.

Als Ausdruck des neuen outputorientierten Steuerungsparadigmas (vgl. Abschn. 4.3.1) bildet die *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring* (Kultusministerkonferenz 2006, 2015) die ländergemeinsame Grundlage für eine evidenzbasierte, d. h. auf wissenschaftlichen Erkenntnissen basierende Steuerung in Bildungspolitik und Bildungsadministration.

- ▶ Beim *Bildungsmonitoring* handelt es sich um die kontinuierliche Beobachtung der Rahmenbedingungen, Verlaufsmerkmale, Ergebnisse und Erträge von Bildungsprozessen (vgl. Abschn. 4.3.3) mithilfe empirisch-wissenschaftlicher Methoden (vgl. Maritzen und Tränkmann 2015, S. 233). Dadurch ist sowohl eine Beurteilung aktueller Zustände als auch eine Beschreibung bildungsbezogener Entwicklungen im Zeitverlauf möglich.

Grundsätzlich kann sich Bildungsmonitoring auf unterschiedliche Ebenen des Schulsystems beziehen (vgl. Grünkorn et al. 2019):

- Monitoring auf der *Ebene des Schulsystems*: Im *engeren* Sinne bezeichnet Bildungsmonitoring das Monitoring des Gesamtsystems (*Systemmonitoring*) – im Fokus steht dabei ein *Schulsystem als Ganzes* (z. B. das Schulsystem in Deutschland oder das Schulsystem eines einzelnen Bundeslandes) und nicht etwa einzelne Schülerinnen und Schüler, Lehrkräfte oder eine bestimmte Schule. Systemmonitoring stellt umfassende Informationen über das Schulsystem zur Verfügung, zeigt Stärken und Schwächen in unterschiedlichen Bereichen des Schulsystems auf und zielt darauf ab, administrative und politische Handlungsbedarfe zu identifizieren – es richtet sich also primär an Verantwortliche in Bildungspolitik und Bildungsadministration.
- Monitoring auf der *Ebene einzelner Schulen* (und ggf. Klassen): Bildungsmonitoring kann sich auch auf die Schul- und Klassenebene beziehen. Das Ziel dieses Monitorings im *weiteren* Sinne liegt in der Bereitstellung von Informationen, die für organisatorische, pädagogisch-didaktische und/oder curriculare Entscheidungen und damit die Weiterentwicklung von Schule und Unterricht genutzt werden können – es richtet sich also an Verantwortliche in den Schulen (Schulleitung, Lehrkräfte). Häufig wird für dementsprechende Verfahren anstelle von Monitoring der Begriff *Schulevaluation* verwendet, da der Evaluationsbegriff immer dann verwendet wird, wenn der „Erfolg einer *bestimmten* Maßnahme bzw. die Leistungsfähigkeit einer *einzelnen* Institution (z. B. einer Schule) zu beurteilen ist“ (Klieme et al. 2003, S. 100, Herv. i. O.).

Die Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring wurde erstmals im Jahr 2006 vorgelegt und im Jahr 2015 in einer überarbeiteten Fassung veröffentlicht. Sie umfasst mehrere Maßnahmen auf verschiedenen Ebenen des Bildungssystems:

1. Teilnahme an internationalen Schulleistungsstudien
2. Überprüfung und Umsetzung von Bildungsstandards für die Primarstufe, die Sekundarstufe I und die Allgemeine Hochschulreife
3. Verfahren zur Qualitätssicherung auf Ebene der Schulen
4. Bildungsberichterstattung

Das übergeordnete Ziel dieser Gesamtstrategie liegt in der Beschaffung von Informationen, die für die Steuerung des Bildungssystems und für die Schul- und Unterrichtsentwicklung an jeder einzelnen Schule relevant sind. Dabei sind die Verfahren und Instrumente unter (1), (2) und (4) dem Bildungsmonitoring im engeren Sinne zuzuordnen – sie sind Gegenstand von Abschn. 5.1 (*Schulqualität erfassen, sichern und weiterentwickeln auf der Ebene des Schulsystems*).

Unter (3) sind Verfahren und Instrumente des Monitorings im weiteren Sinne (Schulevaluation) zusammengefasst, mit denen die Qualitätssicherung und Qualitätsentwicklung auf Ebene der Einzelschule gewährleistet werden soll. Neben den s.g. *Vergleichsarbeiten* (vgl. Abschn. 5.2.1) gibt es in fast allen Ländern sowohl *Verfahren zur externen Evaluation der einzelnen Schule* (auch bekannt unter dem Begriff *Schulinspektion*) als auch zur *schulinternen Evaluation*, für die s.g. *Referenzrahmen für Schulqualität* als Grundlage zur Verfügung stehen (vgl. Kultusministerkonferenz 2015, S. 13) – sie werden in Abschn. 5.2 (*Schulqualität erfassen, sichern und weiterentwickeln auf der Ebene der Einzelschule*) dargestellt.

Insgesamt vermittelt Ihnen dieses Kapitel, welche neuen Verfahren und Instrumente in den vergangenen Jahren vor dem Hintergrund des veränderten Steuerungsparadigmas durch die Bildungspolitik zur Qualitätssicherung und -entwicklung eingeführt und wie diese weiterentwickelt wurden. Auf dieser Basis können Sie Wirkungen für das System und Ihr eigenes Handeln in der Schule einschätzen.

5.1 Schulqualität erfassen, sichern und weiterentwickeln auf der Ebene des Schulsystems

Mit der 2006 verabschiedeten und 2015 überarbeiteten Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring sollen die Leistungsfähigkeit und die Qualität des deutschen Schulsystems gesichert und datengestützt weiterentwickelt werden.

Dazu beteiligt sich Deutschland an *internationalen Leistungsvergleichen* (5.1.1) und führt *ationale, bundesländervergleichende Schulleistungstudien* (5.1.2.2) durch, für die mit den *Bildungsstandards* (5.1.2.1) ein bundesweit verbindlicher Referenzrahmen vorliegt, der auch die Grundlage für einen *ländergemeinsamen Pool von Abiturprüfungsaufgaben* (5.1.2.3) ist. Die gemeinsame *Bildungsberichterstattung* von Bund und Ländern (5.1.3) komplementiert die systembezogene Gesamtschau und ermöglicht auf der Grundlage verschiedenster Daten- und Informationsquellen ein umfassendes Bild der Qualität der Rahmenbedingungen, Prozesse und Wirkungen des deutschen Schul- und Bildungssystems. Die verschiedenen Verfahren und Instrumente dienen dazu, das Bildungssystem in Deutschland bzw. bestimmte Teilbereiche/-aspekte aus systemischer Perspektive dauerhaft *datengestützt* zu beobachten und zu beschreiben – sie liefern primär s. g. *Beschreibungswissen*.

5.1.1 Internationale Schulleistungstudien

Die Leistung von Schülerinnen und Schülern gilt als wichtiger Indikator für die Qualität eines Bildungssystems. In *internationalen Schulleistungstudien* werden mittels standardisierter Testverfahren die Leistungen von tausenden Schülerinnen und Schülern an vielen hundert Schulen in einer Vielzahl an Ländern gemessen. So beteiligten sich beispielsweise an der letzten PISA-Studie – der wohl bekanntesten internationalen Schulleistungstudie – etwa 690.000 Schülerinnen und Schüler in 81 Staaten, darunter Schulen aus ganz Deutschland mit rund 6000 Schülerinnen und Schülern, so der OECD PISA-Bericht 2023.

- ▶ Internationale Schulleistungstudien werden aufgrund ihres großflächigen Charakters auch als *Large Scale Assessments* bezeichnet, da sie über die Leistungserhebung an der einzelnen Schule hinausgehen und in großem Umfang schulische Institutionen und die an Schule beteiligten Personen (Schülerinnen und Schüler, und je nach Studie auch deren Eltern sowie Lehrkräfte und Schulleitungen) einbeziehen (vgl. Goy et al. 2008, S. 79).

Die Gesamtstrategie zum Bildungsmonitoring sieht vor, dass sich Deutschland regelmäßig an drei internationalen Schulleistungstudien beteiligt (Kultusministerkonferenz 2015, S. 7):

- *Progress in International Reading Literacy Study (PIRLS) bzw. (übersetzt) Internationale Grundschul-Leseuntersuchung (IGLU)*
Die Studie PIRLS/IGLU testet in einer repräsentativen Stichprobe (s.u.) das Leseverständnis von Schülerinnen und Schülern der Jahrgangsstufe 4. Die Studie wird in einem fünfjährigen Rhythmus durchgeführt und erlaubt so Aussagen darüber, ob sich Testleistungen über mehrere Erhebungsrunden verbessert oder verschlechtert haben. An der Studie 2021 waren mehr als 400.000 Viertklässlerinnen und Viertklässler aus mehr als 60 Staaten und Regionen beteiligt, darunter knapp 4600 Schülerinnen und Schüler von Grund- und Förderschulen aus allen deutschen Bundesländern (McElvany et al. 2023).
- *Trends in International Mathematics and Science Study (TIMSS)*
In TIMSS wird das mathematische und naturwissenschaftliche Grundverständnis von Schülerinnen und Schülern am Ende von Jahrgangsstufe 4 in einem vierjährigen Rhythmus erfasst. An der Studie 2019 nahmen beispielsweise 58 Staaten mit mehr als 300.000 Schülerinnen und Schülern teil, darunter knapp 4900 Viertklässlerinnen und Viertklässler aus Deutschland (Schwippert et al.

2020). Im Gegensatz zu der Mehrheit der anderen Teilnehmerstaaten beteiligt sich Deutschland bislang nicht am äquivalenten Testverfahren für die Jahrgangsstufe 8.

- *Programme for International Student Assessment (PISA)*

In PISA werden im dreijährigen Rhythmus Basiskompetenzen in den Bereichen Lesen, Mathematik und Naturwissenschaften untersucht. Im Fokus stehen 15-Jährige, da in diesem Alter in vielen Staaten die Pflichtschulzeit endet.

PIRLS/IGLU, TIMSS und PISA sind *internationale Schulleistungsstudien mit deutscher Beteiligung*. Davon abzugrenzen sind *nationale, bundesländerübergreifende Schulleistungsstudien*, also nationale Leistungsvergleiche, deren Ergebnisse auf Ebene der Schulsysteme einzelner Bundesländer ausgewertet werden (vgl. Abschn. 5.1.2.2).

Die Studien sind auf die Primarstufe und die Sekundarstufe I konzentriert – eine deutsche Beteiligung an Schulleistungsvergleichen bezogen auf die Sekundarstufe II bzw. einen nationalen Vergleich der Leistungen von Oberstufenschülerinnen und -schülern gibt es aktuell nicht (vgl. Abschn. 5.1.2.3). Nicht zuletzt gibt es auch noch *landesweite Schulleistungsstudien*, die einmalig oder regelmäßig in einzelnen oder mehreren Bundesländern durchgeführt werden, z. T. auch in der Sekundarstufe II. Bekannte länderinterne Studien sind die Hamburger Untersuchungen *Kompetenzen ermitteln (KERMIT)* sowie *Kompetenzen und Einstellungen von Schülerinnen und Schülern (KESS)*, und *Transformation des Sekundarschulsystems und akademische Karrieren (TOSCA)* u. a. in Baden-Württemberg und Sachsen, oder *LISA – Lesen in der Sekundarstufe* in Schleswig-Holstein.

Internationale Schulleistungsstudien stellen regelmäßig vergleichbare Informationen über Bildungssysteme bereit und liefern den Verantwortlichen in Bildungspolitik und Bildungsadministration in den beteiligten Staaten auf der Basis empirischer Befunde eine auf den Output bezogene Rückmeldung über die Leistungsfähigkeit fünfzehnjähriger Schülerinnen und Schüler bzw. eines Schulsystems und können so Stärken und Schwächen eines Schulsystems oder Teilsystems aufzeigen und administrative und politische Handlungsbedarfe identifizieren helfen.

- ▶ Vergleichende Schulleistungsstudien sind also Systemmonitoring-Studien, d.h. im Fokus steht das Schulsystem als Ganzes und es geht nicht darum, das Leistungsniveau einzelner Schülerinnen und Schüler, Klassen oder Schulen zu beschreiben – die erbrachten Leistungen fließen auch nicht in schulinterne Benotungen ein, sondern bleiben anonym.

Schulleistungsstudien analysieren folglich, was ein Schulsystem (nicht) leistet. Wie man z. B. an der Rezeption der PISA-Ergebnisse in Deutschland erkennen kann, machen großflächig angelegte Schulleistungsstudien Probleme sichtbar (z. B. zu Aspekten sozialer Bildungsungleichheit), die gesellschaftlich relevant sind und deshalb hohe Aufmerksamkeit – sowohl seitens der Bildungspolitik als auch der Öffentlichkeit – erfahren. Sie dienen folglich auch der regelmäßigen Rechenschaftslegung gegenüber Politik und Öffentlichkeit und werden als Legitimation für Veränderungen im Schulsystem herangezogen.

So wurden v. a. in Folge der Ergebnisse der ersten PISA-Studie eine Vielzahl bildungspolitischer Reformen diskutiert und zeitnah umgesetzt, u. a. die Einführung von verbindlichen Standards sowie eine ergebnisorientierte Überprüfung (vgl. Abschn. 5.1.2), Maßnahmen zur wirksamen Förderung bildungsbenachteiligter Kinder, zur Förderung der Sprachkompetenz und den Ausbau von Ganztagsangeboten (Kultusministerkonferenz 2001). Neben dem Systemmonitoring ermöglicht die Beteiligung Deutschlands an internationalen Schulleistungsstudien ein internationales *Benchmarking*, also eine nationale Standortbestimmung im Vergleich mit anderen Staaten oder Regionen als Einordnung der deutschen Befunde in einen größeren Kontext (Sälzer 2022).

5.1.1.1 Anlage und Durchführung von internationalen Schulleistungsstudien

Vergleichsstudien produzieren enorme Datenmengen, was auf ihre großflächige Anlage und das häufig verfolgte Ziel, eine Vielzahl an Ländern vergleichen zu können, zurückzuführen ist. Untersuchungsstichproben werden so ausgewählt, dass sie der Gesamtheit der abzubildenden *Grundpopulation* (z.B. fünfzehnjährige Schülerinnen und Schüler, die in den an PISA beteiligten Staaten eine Schule besuchen) möglichst ähnlich sind.

- ▶ Als *Stichprobe* bezeichnet man eine unter bestimmten Gesichtspunkten ausgewählte Teilmenge einer *Grundgesamtheit* (Menge aller potenziellen Untersuchungsobjekte), da es oft nicht möglich ist, mittels einer Vollerhebung zu Aussagen über eine Population zu kommen. „Um mithilfe einer Stichprobenerhebung (anstelle einer Vollerhebung) gültige Aussagen über eine Population treffen zu können, muss die Stichprobe *repräsentativ* sein, d.h. sie muss in ihrer Zusammensetzung der Population möglichst stark ähneln“ (Bortz und Döring 2009, S. 401).

Die PISA-, TIMSS- und IGLU-Stichproben sind repräsentativ für Deutschland, aber nicht für die einzelnen Bundesländer.

Für die standardisierte Stichprobenziehung in den Teilnehmerstaaten, die einen möglichst belastbaren Ländervergleich zulässt, werden die Schulsysteme in der Regel zunächst nach zentralen Merkmalen unterteilt, wie zum Beispiel nach Regionen (Bundesländer, Provinzen, Kantone etc.) und nach Arten von Schulen, z. B. Schulformen innerhalb der Bundesländer. Innerhalb dieser Auswahl werden dann die Schulen und die in einem zweiten Schritt getroffene Auswahl der Schülerinnen und Schüler innerhalb dieser Schulen nach einem Zufallsverfahren ausgewählt.

Für alle Schulleistungsstudien gilt, dass die *Beteiligungsquoten* der ausgewählten Schulen sowie der Schülerinnen und Schüler bestimmte Grenzen nicht unterschreiten dürfen. Ansonsten könnte man plausibel annehmen, dass sich insbesondere engagierte Schulen beteiligen, die möglicherweise ein höheres Leistungspotenzial haben, oder nur besonders leistungsstarke Schülerinnen und Schüler teilnehmen. Dies kann zu erheblichen *Verzerrungen* hinsichtlich der erhobenen Daten führen. So wurden z. B. im ersten PISA-Erhebungszyklus im Jahr 2000 die Niederlande nachträglich aus der Ergebnisdarstellung ausgeschlossen, da die vorab festgelegte Teilnahmequote von 80 % auf der Ebene der ausgewählten Schülerinnen und Schüler nicht erreicht wurde. Dieses Problem taucht in internationalen Vergleichsstudien immer wieder auf. Die teils in den Ergebnisdarstellungen mitgeteilten, aber aufgrund ihres eingeschränkten Informationsgehaltes kenntlich gemachten Befunde sind entsprechend vorsichtig zu interpretieren.

5.1.1.2 Untersuchungsbereiche

Internationale Schulleistungsstudien erfassen nicht isoliertes Wissen, sondern die bis zu einem spezifischen Zeitpunkt (z.B. Ende der Pflichtschulzeit) erworbenen fachlichen und überfachlichen *Kompetenzen*.

- ▶ **Kompetenz:** Kompetenzen werden als basale Kulturwerkzeuge zur verständigen und verantwortungsvollen Teilnahme am gesellschaftlichen Leben verstanden, die sich in variierenden alltäglichen und komplexen authentischen Anwendungssituationen bewähren müssen (PISA-Konsortium 2001). Der Kompetenzbegriff geht also über den Wissensbegriff hinaus und bezeichnet nach Weinert (2001, S. 27) „die bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“

Aktuelle internationale Schulleistungsstudien fokussieren insbesondere *kognitive* Kompetenzen, die erforderlich sind, um Aufgaben oder Probleme eines bestimmten Inhaltsbereichs in alltäglichen oder komplexen authentischen Anwendungssituationen erfolgreich bewältigen zu können. Sie folgen also einer gänzlich anderen Logik als Klassenarbeiten oder Klausuren, die zumeist nur den Unterrichtsstoff vergangener Wochen prüfen (Sälzer 2022).

- ▶ Internationalen Schulleistungsstudien liegt ein deutlich breiter angelegtes Grundbildungskonzept (international=*Literacy*) zugrunde. „Dabei stehen weniger abgegrenzte Lerninhalte im Vordergrund als vielmehr Kompetenzen, die von den Schülerinnen und Schülern erreicht werden sollen“ (Klieme et al. 2003, S. 36).

Die internationalen Schulleistungsstudien basieren auf einem *anwendungsorientierten Kompetenzbegriff*, der auf die Erfassung von Basiskompetenzen in verschiedenen Anwendungssituationen abzielt: Schülerinnen und Schüler sollen in möglichst authentischen Aufgaben ihre in der Schule erworbenen Kompetenzen anwenden.

Im Vordergrund stehen zumeist die Lesekompetenz sowie Kompetenzen in Mathematik und Naturwissenschaften, die offensichtlich über die Länder hinweg als besonders zentral für die private und berufliche Lebensführung angesehen werden, gerade auch im Hinblick auf den wirtschaftlichen Erfolg eines Landes. Fremdsprachenkenntnisse werden im Rahmen der internationalen Schulleistungsvergleiche *nicht* erfasst, wohl aber im Rahmen des IQB-Bildungstrends, einer bundesweiten Leistungsvergleichsstudie in Deutschland (siehe dazu Abschn. 5.1.2.2). Darüber hinaus wurden und werden auch internationale Studien zur politischen und staatsbürgerlichen Bildung (ICCS: *International Civic and Citizenship Education Study*) sowie zu computer- und informationsbezogenen Kompetenzen (ICILS: *International Computer and Information Literacy Study*) mit deutscher Beteiligung durchgeführt – im Gegensatz zu PIRLS/IGLU, TIMSS und PISA ist die Teilnahme an diesen (und anderen) internationalen Vergleichsstudien *nicht* in der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring festgeschrieben.

Je nach Studie werden neben den *fachbezogenen* Kompetenzen auch *fächerübergreifende* Kompetenzen (z. B. Lernstrategien, Problemlösefähigkeiten) erfasst. Die Kompetenzen der Schülerinnen und Schüler werden über domänen-spezifische *Testaufgaben* (*Domäne* = Fach- oder Wissensgebiet) ermittelt, die in enger Zusammenarbeit zwischen internationalen und nationalen Expertengruppen entwickelt werden. Vor dem Einsatz in der Hauptuntersuchung werden sie in allen

Staaten an Stichproben umfassend erprobt. Damit die einzelnen Aufgaben kein Land benachteiligen, wird zum Beispiel geprüft, ob

- die Aufgaben das Leistungsspektrum in angemessener Weise abdecken,
- unabhängige Kodierer bei der Bewertung von Schülerantworten auf offene Fragen mit freien Antwortformaten zu vergleichbaren Ergebnissen kommen,
- die relativen Schwierigkeiten der Items für Jungen und Mädchen sowie für Gruppen mit unterschiedlichem sozioökonomischem Hintergrund vergleichbar sind,
- der Test den angestrebten relativen Anteil von offenen und geschlossenen (Multiple-Choice) Antwortformaten enthält,
- die Aufgaben kulturell und im Hinblick auf ihren Bezug zu den Curricula der Teilnehmerstaaten international vergleichbar sind und ob
- die Aufgaben ein breites, ausgewogenes Spektrum an Themen abdecken (Baumert et al. 2001, S. 40).

Hinweis: Beispielaufgaben

Einzelne Testaufgaben sind auch online verfügbar, z. B. aus der PISA-Studie unter <https://www.pisa.tum.de/beispielaufgaben/>. Ein Teil der computerbasierten Aufgaben (s. u.) kann auf der Website der OECD interaktiv und in allen PISA-Sprachen bearbeitet werden, siehe: <https://www.oecd.org/pisa/test/other-languages/>. Die kompletten Testhefte aus Leistungsstudien dürfen aus Gründen der Testsicherheit nicht veröffentlicht werden. Um Veränderungen in den erfassten Kompetenzen untersuchen zu können, werden viele Aufgaben bei mehreren Erhebungswellen einer Studie wiederholt eingesetzt und können nicht zur Veröffentlichung freigegeben werden.

Um solche Aufgaben zu konstruieren und die Ergebnisse inhaltlich interpretieren zu können, arbeiten die an Leistungsstudien beteiligten Forschenden mit sogenannten *Kompetenzstufenmodellen*. Bei PISA 2018 werden beispielsweise für die Lesekompetenz acht Stufen unterschieden (vgl. Tab. 5.1).

So ist zum Beispiel ein Schüler oder eine Schülerin, der/die die Expertenstufe im Lesen erreicht hat (Stufe VI), in der Lage, tief in einem Text eingebettete Informationen zu lokalisieren, auch wenn Inhalt und Form des Textes unvertraut sind und indirekt erschlossen werden muss, welche Informationen zur Lösung der Aufgabe relevant sind. Jugendliche hingegen, die nur die unterste Kompetenz-

Tab. 5.1 Überblick über die typischen Anforderungen der acht Kompetenzstufen der Lesekompetenz

Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
VI	<p>Jugendliche auf dieser Stufe können lange und abstrakte Texte verstehen, in denen die relevanten Informationen tief eingebettet sind und nur indirekt mit der Aufgabe zusammenhängen. Sie können Informationen, die mehrere und potenziell widersprüchliche Perspektiven enthalten, vergleichen, gegenüberstellen und integrieren. Sie können über die Quelle des Textes in Bezug zum Textinhalt reflektieren, indem sie Kriterien nutzen, die außerhalb des Textes liegen (z. B. eigenes Wissen). Sie können Widersprüche und Konflikte zwischen verschiedenen Texten zum selben Thema identifizieren und auflösen, indem sie Schlüsse über die Quelle ziehen. Sie bedenken dabei auch persönliche Interessen der Autorin bzw. des Autors und nutzen andere Hinweise, um die Glaubwürdigkeit der Informationen einschätzen zu können. Aufgaben dieser Stufe beinhalten ein oder mehrere komplexe Texte, die mehrere und möglicherweise widersprüchliche Perspektiven und Positionen enthalten. Die Informationen, die zur Beantwortung der Frage erschlossen werden müssen, können dabei auch in Form von Details vorliegen, die tief eingebettet sind und möglicherweise durch konkurrierende Informationen überdeckt werden</p>
V	<p>Jugendliche auf dieser Stufe können lange Texte verstehen und herausfinden, welche Informationen im Text relevant sind, auch wenn die relevanten Informationen leicht übersehen werden könnten. Sie können auf der Grundlage eines tiefen Verständnisses textbezogen argumentieren. Sie können den Zusammenhang zwischen einer Frage und einer oder mehreren Informationen, die innerhalb oder über mehrere Texte und Quellen verteilt sind, erschließen. Aufgaben zum Reflektieren erfordern das Aufstellen oder das kritische Bewerten von Hypothesen anhand spezifischer Informationen. Die Jugendlichen können bei komplexen oder abstrakten Aussagen zwischen Inhalt und Zweck sowie zwischen Tatsache und Meinung unterscheiden. Sie können Neutralität und Befangenheit auf der Grundlage expliziter oder impliziter Hinweise beurteilen, welche den Inhalt und/oder die Quelle der Informationen betreffen. Außerdem können sie Schlüsse über die Zuverlässigkeit der Behauptungen oder Folgerungen ziehen, die in einem Text angeboten werden. Aufgaben dieser Stufe beinhalten typischerweise den Umgang mit abstrakten oder kontraintuitiven Konzepten und das Durchlaufen mehrerer Schritte bis zur Zielerreichung. Zudem können Aufgaben dieser Stufe von der Leserin oder dem Leser verlangen, mehrere lange Texte zu bearbeiten und zwischen den Texten hin- und herzuwechseln, um Informationen zu vergleichen und gegenüberzustellen</p>

(Fortsetzung)

Tab. 5.1 (Fortsetzung)

Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
IV	<p>Jugendliche auf dieser Stufe können umfangreiche Abschnitte aus Texten mit einer oder mehreren Quellen verstehen. Sie interpretieren die Bedeutung von sprachlichen Feinheiten in einem Textabschnitt, indem sie den gesamten Text berücksichtigen. Sie können Perspektiven vergleichen und Schlüsse auf der Grundlage mehrerer Quellen ziehen. Sie können eingebettete Informationen suchen, lokalisieren und integrieren. Sie sind in der Lage, auf der Grundlage der Aufgabenstellung Schlüsse zu ziehen, um die Relevanz von Zielinformationen zu beurteilen. Sie können Aufgaben bewältigen, die es erfordern, sich den vorherigen Aufgabenkontext zu merken. Außerdem können die Jugendlichen auf dieser Stufe den Zusammenhang zwischen bestimmten Aussagen und der allgemeinen Haltung oder Folgerung einer Person zu einem Thema bewerten. Sie können anhand hervorstechender Textmerkmale wie Überschriften und Abbildungen über die Strategien reflektieren, die eine Autorin oder ein Autor verwendet, um (Schreib-) Ziele zu erreichen. Sie können Behauptungen, die in mehreren Texten explizit vorgebracht werden, vergleichen und gegenüberstellen sowie die Zuverlässigkeit einer Quelle anhand hervorstechender Kriterien beurteilen. Texte dieser Stufe sind oft lang oder komplex und ihr Inhalt oder ihre Form sind möglicherweise unüblich. Die Texte und die Aufgaben enthalten indirekte oder implizite Hinweise</p>

(Fortsetzung)

Tab. 5.1 (Fortsetzung)

Kompetenzstufe	<p>Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind</p> <p>III</p> <p>Jugendliche auf dieser Stufe können die wörtliche Bedeutung von Texten mit einer oder mehreren Quellen wiedergeben, ohne dass explizite Inhalte oder organisatorische Hinweise vorhanden sind. Sie können Inhalte integrieren und sowohl grundlegende als auch anspruchsvollere Schlussfolgerungen ziehen. Sie können auch mehrere Teile eines Textes miteinander verknüpfen, um die Grundidee zu identifizieren, einen Zusammenhang zu verstehen oder die Bedeutung eines Wortes oder eines Satzes zu interpretieren, wenn die erforderlichen Informationen auf einer Seite dargestellt sind. Sie können nach Informationen suchen und Zielinformationen lokalisieren, die nicht leicht erkennbar sind und/oder sich in der Nähe von Ablenkungen befinden. In manchen Fällen erkennen sie den Zusammenhang zwischen mehreren Informationen. Jugendliche auf dieser Stufe können über einen Textteil oder über eine begrenzte Anzahl an Texten reflektieren und die Standpunkte mehrerer Autorinnen und Autoren auf der Grundlage expliziter Informationen vergleichen und gegenüberstellen. Aufgaben zum Reflektieren können es erfordern, Vergleiche anzustellen, Erklärungen zu erstellen oder ein Merkmal des Textes zu bewerten. Manche der Aufgaben erfordern ein detailliertes Verständnis eines Textes, der sich mit einem vertrauten Thema beschäftigt, während andere ein globales Verständnis von ungewohnten Inhalten erfordern. Aufgaben dieser Stufe verlangen, dass Jugendliche beim Vergleichen, Gegenüberstellen oder Kategorisieren von Informationen mehrere Merkmale berücksichtigen. Die geforderten Informationen sind oft nicht leicht erkennbar oder es kann eine hohe Anzahl konkurrierender Informationen geben. Die Texte dieser Stufe können außerdem unerwartete oder negativ formulierte Ideen enthalten</p>
II	<p>Jugendliche auf dieser Stufe können die Grundidee eines Textes von mittlerer Länge erkennen. Sie können Zusammenhänge verstehen oder die Bedeutung eines begrenzten Textteils interpretieren, wenn die Informationen nicht leicht erkennbar sind und/oder wenn einige wenige ablenkende Informationen vorhanden sind, indem sie grundlegende Schlussfolgerungen ziehen. Sie können, wenn sie explizit darauf hingewiesen werden, über den Gesamtzweck oder den Zweck spezifischer Details von Texten mittlerer Länge reflektieren. Sie können über einfache visuelle oder typografische Merkmale reflektieren. Sie können außerdem Behauptungen vergleichen und deren Begründungen anhand kurzer, expliziter Aussagen, bewerten. Bei Aufgaben dieser Stufe müssen unter Umständen auf der Grundlage eines einzigen Textbestands Vergleiche und Gegenüberstellungen vorgenommen werden oder es müssen, ausgehend von eigenen Erfahrungen oder Standpunkten, Vergleiche an- gestellt oder Zusammenhänge zwischen dem Text und nicht im Text enthaltenen Informationen erkannt werden</p>

(Fortsetzung)

Tab. 5.1 (Fortsetzung)

Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
Ia	<p>Jugendliche auf dieser Stufe können die wörtliche Bedeutung von Sätzen oder kurzen Abschnitten verstehen. Sie können das Hauptthema oder die Absicht des Autors in einem Textteil zu einem vertrauten Thema erkennen und einen einfachen Zusammenhang zwischen benachbarten Informationsteilen oder zwischen den vorliegenden Informationen und dem eigenen Vorwissen herstellen. Sie können eine relevante Seite aus einer begrenzten Auswahl, auf der Grundlage von einfachen Aufforderungen auswählen sowie eine oder mehrere unabhängige Informationsteile aus kurzen Texten lokalisieren. Lesende auf dieser Stufe können über den Gesamtzweck, den Kernpunkt und die Zusatzinformationen von einfachen Texten mit expliziten Hinweisen reflektieren. Die meisten Aufgaben auf dieser Stufe weisen auf entscheidende Elemente in der Aufgabe und im Text hin</p>
Ib	<p>Jugendliche auf dieser Stufe können die wörtliche Bedeutung von Sätzen oder kurzen Abschnitten verstehen. Sie können das Hauptthema oder die Absicht des Autors in einem Textteil zu einem vertrauten Thema erkennen und einen einfachen Zusammenhang zwischen benachbarten Informationsteilen oder zwischen den vorliegenden Informationen und dem eigenen Vorwissen herstellen. Sie können auf der Grundlage von einfachen Aufforderungen eine relevante Seite aus einer begrenzten Auswahl auswählen sowie eine oder mehrere unabhängige Informationsteile aus kurzen Texten lokalisieren. Sie können über die gesamte Textfunktion sowie über die relative Wichtigkeit von Informationen (z. B. Hauptidee vs. unwichtige Details) von einfachen Texten mit expliziten Hinweisen reflektieren. Die meisten Aufgaben auf dieser Stufe weisen auf entscheidende Elemente in der Aufgabe und im Text hin</p>
Ic	<p>Jugendliche auf dieser Stufe können die Bedeutung kurzer, syntaktisch einfacher Sätze auf wörtlicher Ebene verstehen und bestätigen. Sie können für einen eindeutigen und einfachen Zweck innerhalb eines begrenzten Zeitraums lesen. Die Aufgaben auf dieser Stufe umfassen lediglich einfache Wörter und syntaktische Strukturen</p>
Unter Ic	[hierzu werden keine Angaben gemacht]

Quelle: Weis et al. (2019, S. 53 f.)

stufe (Stufe I) erreichen, können nur sehr einfache Leseaufgaben bewältigen, z.B. kurze Sätze auf wörtlicher Ebene verstehen.

- ▶ Anhand von *Kompetenzstufen* wird beschrieben, welche kognitiven Anforderungen Schülerinnen und Schüler in der Regel bewältigen können, wenn sie eine bestimmte Kompetenzstufe erreicht haben.

In Schulleistungsstudien werden also in erster Linie kognitive Kompetenzen mithilfe von standardisierten Tests erfasst und verglichen. Neben den eigentlichen Tests kommen weitere Methoden der Datenerhebung, v.a. *Fragebögen*, zum Einsatz, um das Bedingungsgefüge der gemessenen Leistungen zu beleuchten. Insbesondere werden z.B. Schulleitungen, Lehrkräfte sowie die beteiligten Schülerinnen und Schüler und teilweise auch deren Eltern zu bestimmten individuellen, familiären, unterrichtlichen und schulischen Faktoren befragt. So werden z.B. Schülerinnen und Schüler zu familiären Hintergrundmerkmalen (z.B. Bildungsstand der Eltern, Migrationshintergrund) mithilfe von Fragebögen befragt. Auf diese Weise kann z.B. analysiert werden, inwieweit *Merkmale der sozialen und kulturellen Herkunft* mit Unterschieden in der Kompetenz verbunden sind (vgl. Abschn. 3.3).

Auch werden sie zu ihrer *Wahrnehmung der schulischen und unterrichtlichen Situation* befragt, etwa im Hinblick auf die methodische Gestaltung des Unterrichts (z. B. Arbeitsformen), die Wahrnehmung der Unterrichtsqualität oder das Schulklima (u. a. Verhältnis zwischen Lehrenden und Lernenden, Verhältnis der Lernenden untereinander etc.).

Exkurs: Papier- und computerbasierte Erhebungsmodi

Seit einigen Jahren kommen in internationalen Schulleistungsstudien anstelle der klassischen Papierfragebögen und Papier-Bleistift-Tests, d. h. handschriftlich zu bearbeitenden Kompetenztests (sog. *Paper-Pencil-Format* bzw. *paper-based Assessment*) auch computerbasierte Erhebungsformate bzw. technologiebasierte Tests zum Einsatz (*technologiebasiertes Assessment*, TBA bzw. *computer-based Assessment*, CBA). Auch in weiteren standardisierten Testverfahren in Deutschland (vgl. Abschn. 5.1.2.2 und 5.2.1) ist die Umstellung von papierbasierten auf computer- bzw. technologiebasierte Testverfahren geplant bzw. wird teilweise auch schon umgesetzt.

Durch die Befragung der Lehrkräfte und Schulleitungen gewinnt man zudem Informationen über die konkreten *Bedingungen der Einzelschule* (z. B. Ressourcenausstattung, Schulklima, Lehrkräftekooperation etc.) und über die dort tätigen Lehrkräfte (z. B. formale Qualifikation, Fortbildungsaktivitäten).

Die Befragung der Eltern kann auch die Wahrnehmung von Schule und Unterricht betreffen (z. B. Zusammenarbeit der Eltern mit der Schule, elterliches Engagement in der Schule), umfasst aber auch die häusliche Lernumgebung (z. B. die im Elternhaus verfügbaren Ressourcen für das Lernen, Förderaktivitäten im Elternhaus).

Schulleistungsstudien erfassen teils sensible Personendaten und müssen deshalb die gesetzlichen *Vorgaben zum Datenschutz* erfüllen, um Individuen vor Missbräuchen beim Umgang (Erhebung, Verarbeitung, Nutzung) mit personenbezogenen Informationen zu schützen und *Persönlichkeitsrechte* zu wahren. Die Befunde aus internationalen Vergleichsstudien werden *aggregiert* (zusammengefasst auf einem bestimmten Niveau, z. B. auf der Ebene der Schulform) berichtet und zudem *anonym* erhoben, d. h. sie sollen einzelnen Personen nicht zugeordnet werden können. Weder Schul- noch Schulaufsichtspersonal oder andere Dritte dürfen Einsicht in die Daten erhalten. Dadurch, dass Leistungsstudien in ihrer Intention keine individuelle Kompetenzdiagnostik darstellen und die Daten zudem nicht auf einzelne Personen rückbezogen werden können, kann jedoch auch *keine individuelle Rückmeldung* an die Schülerinnen und Schüler sowie die Lehrkräfte erfolgen.

Aufgrund des KMK-Beschlusses zur regelmäßigen Beteiligung an den internationalen Schulleistungsstudien PISA, IGLU und TIMSS sind alle staatlichen Schulen in Deutschland zur Teilnahme *verpflichtet*. Innerhalb der Schulen, die Teil der repräsentativen Stichprobe sind, sind Schülerinnen und Schüler verpflichtet, die Kompetenztests zu bearbeiten. Die Bearbeitung des Hintergrundfragebogens für Schülerinnen und Schüler ist in einigen Bundesländern verpflichtend, in anderen freiwillig. Auch das Ausfüllen der Lehrkräfte- und Schulleitungsfragebögen ist je nach Bundesland verpflichtend, teilverpflichtend (d. h. nur bestimmte Fragen müssen beantwortet werden und andere können freiwillig beantwortet werden) oder freiwillig; das Ausfüllen der Elternfragebögen ist immer freiwillig. Die nationalen Berichtsbände zu PISA, TIMSS und IGLU sind im Übrigen kostenfrei online verfügbar.

5.1.1.3 Aktueller Stand

Mittlerweile liegt ein enorm reichhaltiger Datenbestand vor, mit dem sich über mehrere Messzeitpunkte und Studien Entwicklungstrends abbilden lassen. Seit dem unerwartet schwachen Abschneiden deutscher Schülerinnen und Schüler bei internationalen Leistungsvergleichen zu Beginn der 2000er Jahre hat sich einiges

getan; insgesamt deutete sich ein positiver Entwicklungstrend an, der sich ausweislich der Daten von 2022 nicht fortgesetzt hat. Die Autorinnen und Autoren des deutschen PISA-Berichtbandes kommen etwa mit Blick auf die Lesekompetenz zu dem Ergebnis, dass die durchschnittliche Lesekompetenz der Jugendlichen in Deutschland signifikant gesunken ist; 2018 hatte sich das Land noch auf einem hohen Niveau über dem OECD-Mittelwert etabliert (während es in der ersten PISA-Studie noch unter dem OECD-Mittelwert lag).

Dabei zeigt sich in Deutschland nach wie vor eine *große Leistungsstreuung*, d. h. es gibt eine große Bandbreite der Leseleistungen Fünfzehnjähriger in Deutschland mit erheblichen Leistungsunterschieden zwischen lesestarken und leseschwachen Jugendlichen. Zugleich ist der Anteil der lesestarken Schülerinnen und Schülern auf den obersten Kompetenzstufen gesunken. Auch ist der Anteil der leseschwachen Jugendlichen auf den untersten Kompetenzstufen wieder gestiegen; im internationalen Vergleich weist Deutschland einen *hohen Anteil leseschwacher Schülerinnen und Schüler* auf (Lewalter et al. 2023). Ähnliche Befunde zeigen sich auch für Mathematik und Naturwissenschaften, sowohl in der Sekundarstufe I als auch im Primarbereich. Es zeigt sich dringender Handlungsbedarf, um

- Kompetenzen insgesamt auf ein höheres Niveau zu entwickeln,
- die leistungsstarke Spitzengruppe auszubauen und vor allem
- die Anteile leistungsschwacher Schülerinnen und Schüler zu reduzieren.

Gerade denjenigen, denen am Ende der Grundschulzeit bzw. am Ende der Pflichtschulzeit grundlegende Kompetenzen fehlen, um den weiteren Bildungs- bzw. Berufsweg erfolgreich durchlaufen zu können, müssen stärker als bislang gezielt gefördert werden. Nicht zuletzt stellt weiterhin der systematische Zusammenhang zwischen dem Kompetenzerwerb und sozialen Herkunftsmerkmalen eine zentrale Herausforderung für die Weiterentwicklung des Bildungssystems dar. So ist eine Reduktion der herkunftsbedingten Leistungsdisparitäten bislang nicht gelungen (siehe auch Abschn. 3.3). Dabei gehen die Autorinnen und Autoren der PISA 2022-Studie davon aus, dass die Pandemie eher verstärkend auf bereits existierende Probleme gewirkt habe (Lewalter et al. 2023).

5.1.2 Überprüfung und Umsetzung von Bildungsstandards

Das unerwartet schwache Abschneiden Deutschlands bei internationalen Leistungsvergleichen zu Beginn des Jahrtausends (vgl. Abschn. 5.1.1) hat ins-

besondere seit PISA zu einer bis dahin in dieser Intensität nicht gekannten Auseinandersetzung um bildungspolitisch angemessene Reaktionen geführt. Auf der Ebene des Schulsystems wurde eine Vielzahl bildungspolitischer Maßnahmen bundesländerübergreifend beschlossen (Kultusministerkonferenz 2001) und umgesetzt, darunter auch die Einführung sogenannter *Bildungsstandards*.

5.1.2.1 Bildungsstandards als Referenzmaßstab

Bildungsstandards beschreiben Leistungserwartungen in Form fachspezifischer Kompetenzen, die Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt ihrer Schullaufbahn erworben haben sollen. Sie gelten als ein entscheidendes Element der Outputorientierung (vgl. Abschn. 4.3.1) und sind als *Leistungsstandards* (auch: *performance standards* oder *output standards*) ergebnisbezogen und kompetenzorientiert formuliert. Sie unterscheiden sich damit von inhaltlichen Standards (*content standards* oder *curriculum standards*), die konkrete Unterrichtsgegenstände bzw. Lerninhalte festlegen. Leistungsstandards können nach ihrem Zielniveau unterschieden werden:

- *Mindeststandards* (auch: Minimal- oder Basisstandards) beschreiben ein definiertes Minimum an Kompetenzen, das alle Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt ihrer Schullaufbahn erreichen sollen.
- *Regelstandards* bezeichnen eine durchschnittliche Leistungserwartung, also Kompetenzen, die in der Regel von den Schülerinnen und Schülern einer bestimmten Jahrgangsstufe erreicht werden sollen.
- *Maximalstandards* (auch: Optimalstandards) entsprechen dem höchsten Leistungsniveau und definieren Kompetenzen, die die besten Schülerinnen und Schüler erreichen sollen (Kultusministerkonferenz 2004).

In Deutschland sind die Bildungsstandards in Form von Könnensbeschreibungen (*Can-do-Statements*) und als Regelstandards konzipiert (Henschel und Stanat 2019). Sie werden von der Kultusministerkonferenz verabschiedet und alle Länder haben sich verpflichtet, die Bildungsstandards in ihren Lehrplänen zu implementieren und in den Schulen umzusetzen, d. h. die Bildungsstandards sind bundesweit gültig und verbindlich. Sie sind jedoch nur als Rahmen zu verstehen: Bei weiterhin geltender Bildungshoheit entstehen in den 16 Bundesländern jeweils eigene Ausgestaltungen, die entweder die äußere Form der Bildungsstandards aufnehmen, eigene Formen verwenden oder diese neben die traditionellen Lehrpläne stellen.

Aufgrund der Heterogenität der Schulsysteme in den 16 Ländern (vgl. Kap. 2) sind die Bildungsstandards nicht schulform-, sondern *abschlussbezogen* formuliert, d. h. sie legen fest, welche Kompetenzen Schülerinnen und Schüler zum

Ende eines *Bildungsganges* in der Regel erreicht haben sollen. Sie gelten für die Bildungsgänge des *allgemeinbildenden* Schulsystems: den Bildungsgang in der *Grundschule* sowie die Bildungsgänge, die zum *Ersten Schulabschluss* (vormals: *Hauptschulabschluss*), zum *Mittleren Schulabschluss* und zur *Allgemeinen Hochschulreife* führen.

Bundesweit geltende Bildungsstandards, die von der Kultusministerkonferenz verabschiedet wurden, gibt es derzeit für die Fächer Deutsch, Mathematik, die Erste Fremdsprache (Englisch, Französisch) und die Naturwissenschaften (Biologie, Chemie, Physik) – diese Fächer werden von der Politik als grundlegend für die Lebensbewältigung und die weitere schulische und berufliche Ausbildung angesehen. Tab. 5.2 gibt einen Überblick über die derzeit vorliegenden Bildungsstandards der Kultusministerkonferenz:

Die Konzentration der KMK-Standards auf *kognitive* Kompetenzen in *ausgewählten* Fächern wurde und wird vielfach kritisiert, da sie allenfalls einen Ausschnitt dessen abbilden, was schulische Bildung ausmacht. Insofern gibt es eigene Entwicklungen von Bildungsstandards in weiteren Fächern (z.B. Geschichte, Religion, Sport), die jedoch *nicht* von der Kultusministerkonferenz verabschiedet werden.

Bildungsstandards sollen verschiedene Funktionen erfüllen (vgl. Grünkorn et al. 2019, S. 278):

Tab. 5.2 Bundesweit geltende Bildungsstandards im Überblick ©

	Primarstufe	Sekundarstufe I		Allgemeine Hochschulreife
		Erster Schulabschluss	Mittlerer Schulabschluss	
Deutsch	x*	x*	x*	x
Mathematik	x*	x*	x*	x
Erste Fremdsprache (Englisch, Französisch)		x	x	x
Naturwissenschaften (Biologie, Chemie, Physik)			x	x

* Die ursprünglich im Jahr 2004 eingeführten Bildungsstandards für die Primarstufe sowie für die Fächer Deutsch und Mathematik in der Sekundarstufe I wurden im Jahr 2022 in einer überarbeiteten Variante verabschiedet.

- Sie sind ein gemeinsam vereinbarter Maßstab und damit ein verbindlicher Referenzrahmen für alle Akteure im Schulsystem (z. B. Lehrkräfte, Schülerinnen und Schüler sowie deren Eltern, aber auch Schulverwaltung, Schulaufsicht, Lehrplankommissionen und Bildungspolitik). Für alle Schulen in allen Bundesländern gelten die gleichen Anforderungen bzw. bestehen die gleichen Ziele – insofern sorgen Bildungsstandards für „mehr Klarheit, größere Objektivität, höhere Verbindlichkeit“ (Blum et al. 2010, S. 16). Sie dienen also der Orientierung, etwa bezogen auf die Vergleichbarkeit schulischer Abschlüsse, die an unterschiedlichen Schulformen erworben werden können (*Orientierungsfunktion*).
- Auch sind sie gemeinsame Verständigungsgrundlage über das, was unter zeitgemäßen Bildungszielen in den jeweiligen Fächern zu verstehen ist (*Klärungsfunktion*).
- Um zu überprüfen, inwieweit Schülerinnen und Schüler die in den Bildungsstandards formulierten Anforderungen erreichen, werden in Deutschland seit mehreren Jahren werden regelmäßig zentral-administrierte, standardisierte Leistungsmessungen durchgeführt (vgl. Abschn. 5.1.2.2 und 5.2.1) (*Überprüfungsfunktion*). Der Abgleich der Leistungsergebnisse mit den in den Bildungsstandards formulierten Erwartungen ermöglicht folglich eine Analyse der Stärken und Schwächen von Schülerinnen und Schülern in verschiedenen Kompetenzbereichen und zu unterschiedlichen Zeitpunkten in der Schullaufbahn, woraus Veränderungs- und Optimierungsprozesse im Schulsystem und an einzelnen Schulen abgeleitet werden können.
- Das wichtigste Ziel, das mit der Einführung der Bildungsstandards verbunden war und ist, ist gleichwohl die Weiterentwicklung des Unterrichts im Sinne der Kompetenzorientierung, d. h. der Unterricht soll sich weniger an Inhalten, sondern am Kompetenzaufbau orientieren (*Entwicklungsfunktion*). Damit ist *keine* Standardisierung der Unterrichtsgestaltung verbunden – die Bildungsstandards ermöglichen Handlungsspielräume, um zur Erreichung der Standards eigenständige Wege gehen zu können (Grundgedanke: *gleiche* Kompetenzen an *unterschiedlichen* Inhalten und auf *unterschiedlichen* Wegen erwerben können). Dabei sollen nicht nur die Standards an sich, sondern auch die Ergebnismeldungen aus den standardisierten Leistungsmessungen Orientierungspunkte für die Weiterentwicklung des Unterrichts bieten.

Im Rahmen der Gesamtstrategie zum Bildungsmonitoring hat die Kultusministerkonferenz nicht nur die Einführung von Bildungsstandards beschlossen, sondern auch vereinbart, dass in der Primarstufe und in der Sekundarstufe I regelmäßig von zentraler Stelle überprüft werden soll, inwieweit die in den Bildungs-

standards beschriebenen Kompetenzen von den Schülerinnen und Schülern erreicht werden. Dies erfolgt

- zum einen in Form einer *zentralen Überprüfung des Erreichens der Bildungsstandards im Ländervergleich* – dem *IQB-Bildungstrend*, einer nationalen, innerdeutschen Schulleistungsstudie auf Basis der KMK-Bildungsstandards (vgl. Abschn. 5.1.2.2),
- und zum anderen in Form von landesweiten *Vergleichsarbeiten* (vgl. Abschn. 5.2.1).

In der Sekundarstufe II gibt es hingegen bislang kein vergleichbares zentral-administriertes Testverfahren, um zu überprüfen, inwieweit die Standards für die Allgemeine Hochschulreife von Schülerinnen und Schülern in der gymnasialen Oberstufe erreicht werden. Stattdessen einigte sich die Kultusministerkonferenz auf ein anderes Instrument: die *Gemeinsamen Abituraufgabenpools der Länder* (vgl. Abschn. 5.1.2.3). Eine zentrale Rolle für die Weiterentwicklung und Überprüfung der Bildungsstandards nimmt das *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB) ein.

Exkurs: Das Institut zur Qualitätsentwicklung im Bildungswesen (IQB)

Das Institut zur Qualitätsentwicklung im Bildungswesen ist eine im Jahr 2004 etablierte wissenschaftliche Einrichtung aller 16 deutschen Bundesländer an der Humboldt-Universität zu Berlin. Das IQB unterstützt die Länder bei der Qualitätsentwicklung und Qualitätssicherung im allgemeinbildenden Schulsystem. Zentrale Aufgabe des IQB sind v.a. die (*Weiter-) Entwicklung der Bildungsstandards* für den Primarbereich, die Sekundarstufe I und die Sekundarstufe II. Ferner soll das IQB Aufgaben entwickeln, die auf den KMK-Bildungsstandards basieren, und zwar *Testaufgaben*, die geeignet sind, um das Erreichen der Bildungsstandards für die Primarstufe und die Sekundarstufe I in Ländervergleichsstudien zu überprüfen, sowie *Aufgaben für den Unterricht*, die die Bildungsstandards konkretisieren und illustrieren, wie die beschriebenen Kompetenzen im Unterricht entwickelt werden können. Weitere Aufgaben des IQB sind die *Koordination der Entwicklung eines Pools von Abiturprüfungsaufgaben*, die auf den Bildungsstandards der KMK für die Allgemeine Hochschulreife basieren (vgl. Abschn. 5.1.2.3), sowie die *Unterstützung der Länder bei Maßnahmen zur Implementation der Bildungsstandards*, z.B. bei der Entwicklung von Fortbildungskonzepten und Unterrichtsmaterialien (vgl. Klein o.J.).

5.1.2.2 Zentrale Überprüfung des Erreichens der Bildungsstandards im Ländervergleich (IQB-Bildungstrends)

Das Institut zur Qualitätsentwicklung im Bildungswesen ist verantwortlich für die Durchführung von Ländervergleichsstudien in der Primarstufe und in der Sekundarstufe I, um zu untersuchen, inwieweit in den einzelnen Ländern die in den Bildungsstandards formulierten Kompetenzanforderungen erreicht werden. Dieser Ländervergleich ist eine *nationale, innerdeutsche Schulleistungstudie*, die erstmals im Jahr 2009 durchgeführt wurde. Zuvor wurden Ländervergleiche als nationale Erweiterungs-/Ergänzungsstudie IGLU-E und PISA-E im Kontext der internationalen Schulleistungsuntersuchungen IGLU und PISA durchgeführt, woran sich allerdings nicht alle Bundesländer beteiligten.

Seit 2009 erfolgt der innerdeutsche Schulleistungsvergleich nicht mehr als nationale Erweiterung der internationalen Vergleichsstudien, sondern als zentrale Überprüfung des Erreichens der Bildungsstandards im Ländervergleich (vgl. KMK o. J.). Dabei hat sich die Bezeichnung dieses zentralen Testverfahrens geändert: Die Ergebnisse des ersten Testzyklus (2009–2012) wurden unter dem Titel *Ländervergleich* (Synonym: IQB-Ländervergleich) veröffentlicht. Seit dem zweiten Testzyklus, beginnend mit der Erhebung im Jahr 2015, werden die Ergebnisse der Ländervergleiche unter dem Titel *Bildungstrend* (Synonym: IQB-Bildungstrend) veröffentlicht, da mit dem zweiten Testzyklus Trendaussagen über die Zeit möglich sind (vgl. Schipolowski o. J.). Genau wie bei den internationalen Schulleistungsstudien (vgl. Abschn. 5.1.1) werden *nicht* alle Schülerinnen und Schüler in Deutschland getestet, sondern nur eine Stichprobe (statt einer Vollerhebung). Diese ist sowohl bundesweit als auch für jedes der 16 Bundesländer *repräsentativ*.

Der Bildungstrend ist ein nationales Bildungsmonitoring auf Basis der Bildungsstandards und dient dem *Systemmonitoring auf der Ebene der Länder*. Die Ergebnisse werden also auf der Ebene der Schulsysteme der Länder berichtet: Die Berichte stellen dar (z. B. Stanat und Schipolowski 2022, S. 14),

- inwieweit die in den Bildungsstandards formulierten Kompetenzziele von den Schülerinnen und Schülern in den einzelnen Bundesländern erreicht werden (*kriteriale Vergleichsperspektive*),
- zeigen Entwicklungstrends auf (*ipsative Vergleichsperspektive*),
- ermöglichen einen Ländervergleich (*soziale Vergleichsperspektive*),
- identifizieren länderspezifische Stärken und Schwächen und
- zeigen den Verantwortlichen in Bildungspolitik und -administration auf, wo Handlungs- oder Optimierungsbedarf besteht.

In der Primarstufe findet der IQB-Bildungstrend in der vierten Jahrgangsstufe alle fünf Jahre in den Fächern Deutsch und Mathematik statt (zuletzt 2021, vgl. Stanat et al. 2022). Die nächste Erhebung ist für 2027 geplant, wobei dann schon das Erreichen der überarbeiteten Bildungsstandards (siehe Tab. 5.1) überprüft wird. In der Sekundarstufe I findet der IQB-Bildungstrend alle drei Jahre in der Jahrgangsstufe 9 statt, wobei im Wechsel die sprachlichen Fächer sowie Mathematik und Naturwissenschaften getestet werden, d.h. je Fach bzw. Fächergruppe findet nur 6 Jahre eine Überprüfung des Erreichens der Bildungsstandards in Form des IQB-Bildungstrends statt (vgl. Schipolowski o.J.).

Ähnlich wie in den internationalen Schulleistungsstudien basiert der IQB-Bildungstrend auf Kompetenztests für Schülerinnen und Schüler sowie auf Fragebögen für Schülerinnen und Schüler, deren Eltern, Fachlehrkräfte und die Schulleitung zur Erfassung von schulischen und außerschulischen Lernbedingungen und der familiären Herkunft. Es besteht eine Pflicht zur Teilnahme an den Kompetenztests; das Ausfüllen der Hintergrundfragebögen ist hingegen nur zum Teil verpflichtend, teilweise aber auch freiwillig. Die Ergebnisse werden in einem Bericht zusammengefasst, der etwa anderthalb Jahre nach der Testdurchführung veröffentlicht wird.

Hinweis: Zugang zu Ergebnisberichten des IQB

Über die Homepage des IQB (<https://www.iqb.hu-berlin.de/bt>) können alle bislang verfügbaren Ergebnisberichte kostenfrei online abgerufen werden; zudem findet man dort Erläuterungen und illustrierende Aufgaben zu den Bildungsstandards sowie (auch zum Download) praxisorientierte Veröffentlichungen für den Primarbereich und die Sekundarstufe I (<https://www.iqb.hu-berlin.de/bista/teach/>).

Die in den Bildungsstandards formulierten Kompetenzen werden durch *Testaufgaben*, zu deren Bearbeitung bestimmte Kompetenzen erforderlich sind, konkretisiert. Für die standardbasierten Kompetenztests wird folglich eine große Anzahl an Testaufgaben benötigt, welche die Bildungsstandards angemessen abbilden. Die Entwicklung von Testaufgaben zur Überprüfung des Erreichens von Bildungsstandards erfolgt in einem mehrstufigen, komplexen und ressourcenintensiven Prozess, an dem eine Vielzahl an beteiligten Akteuren (u.a. erfahrene Lehrkräfte, fachdidaktische Expertinnen und Experten) beteiligt ist.

Zentral für die Interpretation der Testergebnisse sind empirisch validierte *Kompetenzstufenmodelle*, die mehrere Kompetenzstufen umfassen,

welche jeweils ein bestimmtes Leistungsniveau repräsentieren (siehe auch Abschn. 5.1.1). In diesen Modellen wird folglich konkret beschrieben, welche kognitiven Anforderungen Schülerinnen und Schüler in der Regel bewältigen können, wenn sie eine bestimmte Kompetenzstufe erreicht haben.

Zudem sind die Stufen mit dem Erreichen von Mindest- und Regelstandards verknüpft, d.h. je nach Kompetenzstufe kann man eine Aussage darüber machen, ob die von der Kultusministerkonferenz formulierten Kompetenzziele erreicht, verfehlt oder überschritten werden. Von besonderem Interesse sind regelmäßig die Anteile an Schülerinnen und Schülern, die die Regelstandards erreichen, die Mindeststandards verfehlen und die Optimalstandards erreichen.

Hinweis: Zugang zu Kompetenzstufenmodellen der Bildungsstandards

Die Kompetenzstufenmodelle zu den Bildungsstandards, die die Basis für die Darstellung der Ergebnisse aus den IQB-Bildungstrends und auch aus den Vergleichsarbeiten (vgl. Abschn. 5.2.1) sind, können auf der Homepage des IQB eingesehen werden: <https://www.iqb.hu-berlin.de/bista/ksm/>

Der IQB-Bildungstrend 2021 überprüfte beispielsweise zum dritten Mal das Erreichen der Bildungsstandards in den Fächern Deutsch und Mathematik am Ende der vierten Jahrgangsstufe. Nachdem in internationalen Schulleistungsstudien in den letzten Jahren insgesamt ein positiver Entwicklungstrend zu verzeichnen war (vgl. Abschn. 5.1.1), waren die Ergebnisse alarmierend, da nur etwas mehr als die Hälfte der Viertklässlerinnen und Viertklässler den KMK-Regelstandard erreichen (z.B. deutschlandweit knapp 58 % im Bereich Lesen und 55 % in Mathematik); nur wenige erreichen den Optimalstandard (7,8 % im Bereich Lesen und 10,5 % in Mathematik).

- ▶ Ein steigender Anteil an Schülerinnen und Schülern hingegen *verfehlt die Mindeststandards* (z.B. 19 % im Bereich Lesen und 22 % in Mathematik) und verfügt damit am Ende der Grundschulzeit nicht über grundlegende sprachliche und mathematische Kompetenzen, die zentral für das Weiterlernen in der Sekundarstufe I sind.

Dabei gelingt es den Bundesländern unterschiedlich gut, Regel- und Mindeststandards zu sichern – so zeigt sich etwa für Bayern ein besonders gutes Ergebnismuster (z.B. im Bereich Lesen: 67,7 % erreichen den Regelstand, 11,4 % erreichen den Optimalstandard, 14,1 % verfehlen den Mindeststandard), während

die Ergebnisse für Nordrhein-Westfalen deutlich ungünstiger ausfallen (z.B. im Bereich Lesen: 52,7 % erreichen den Regelstand, 6,4 % erreichen den Optimalstandard, 21,6 % verfehlen den Mindeststandard). Im Kontext der Trendanalysen (2011-2021) sprechen die Autorinnen und Autoren von einem durchgängig negativen Entwicklungstrend – besonders besorgniserregend ist die wachsende Zahl an Grundschülerinnen und Grundschulern, die die Mindeststandards in Deutsch und Mathematik nicht erreicht. Dies betrifft insbesondere auch Kinder aus sozial weniger privilegierten Familien und Kinder mit Migrationshintergrund (vgl. Stanat et al. 2022).

Erklärungsansätze für die Länderunterschiede liefert der Ergebnisbericht nicht – lediglich konnte empirisch belegt werden, dass ungünstigen Entwicklungen in keinem Land durchgängig auf Veränderungen in der Schülerschaft (z. B. hinsichtlich der sozialen und zuwanderungsbezogenen Herkunft) zurückgeführt werden können (Stanat et al. 2022, S. 114). Welche Faktoren also ursächlich für diese Entwicklungen sind, bleibt offen.

Für die Sekundarstufe I zeigten die 2023 veröffentlichten Befunde des IQB – vermutlich auch pandemiebedingt – besorgniserregende Ergebnisse für das Fach Deutsch; für das Fach Englisch hingegen sehr positive Ergebnisse, was möglicherweise auf vielfältige außerschulische Lerngelegenheiten zurückzuführen ist. Die Befunde dieser nationalen Leistungsvergleichsstudie haben in Bildungspolitik, Wissenschaft und Öffentlichkeit große Aufmerksamkeit erfahren. Die *Ständige Wissenschaftliche Kommission der Kultusministerkonferenz* (vgl. Abschn. 4.2.5) hat in einem Gutachten zentrale Maßnahmen zur Weiterentwicklung der Grundschule formuliert und dabei v. a. eine Konzentration auf die Förderung grundlegender sprachlicher und mathematischer Kompetenzen empfohlen, damit mehr Schülerinnen und Schüler die Mindeststandards in Deutsch und Mathematik in der Grundschule erreichen können (vgl. Ständige Wissenschaftliche Kommission 2022a). Inwiefern diese Empfehlungen auch umgesetzt und wirksam werden, bleibt abzuwarten.

5.1.2.3 Gemeinsame Abituraufgabenpools der Länder

Anders als in der Primarstufe und der Sekundarstufe I erfolgt in der Sekundarstufe II *kein* nationales Bildungsmonitoring auf Basis der KMK-Bildungsstandards für die Allgemeine Hochschulreife. Stattdessen gibt es ein anderes Instrument: die *Gemeinsamen Abituraufgabenpools der Länder*. Es handelt sich dabei um *schriftliche Abiturprüfungsaufgaben*, die auf Basis der Bildungsstandards für die Allgemeine Hochschulreife von *allen* Bundesländern gemeinsam entwickelt werden.

Dass unter Beteiligung aller Bundesländer gemeinsam Prüfungsaufgaben für das Abitur entwickelt werden, ist angesichts der Verantwortung der einzelnen

Bundesländer für die Abiturprüfungsverfahren (vgl. Abschn. 2.2.3) ein beachtenswertes Novum. Für die Fächer Deutsch, Englisch, Französisch und Mathematik stehen den Ländern bereits seit 2017 Abituraufgabenpools zur Verfügung. Ab 2025 werden auch für die naturwissenschaftlichen Fächer Biologie, Chemie und Physik ländergemeinsame Pools bereitgestellt. Die ländergemeinsame Entwicklung von standardbasierten Abiturprüfungsaufgaben ist ein komplexer Prozess, der durch das IQB *koordiniert* wird (vgl. Hoffmann et al. 2022b):

- Die Abiturkommissionen aller Länder erarbeiten für jedes Fach *Aufgabenvorschläge*, die an das IQB übermittelt werden.
- Im Anschluss erfolgt eine *Bewertung* der eingereichten Aufgabenvorschläge durch eine Arbeitsgruppe (*AG Aufgaben*), die für die Erstellung der ländergemeinsamen Abiturprüfungsaufgaben zuständig ist und die von einem *Fachkoordinator* des IQB begleitet wird. Mitglieder der AG Aufgaben sind jeweils ein Experte oder eine Expertin je Bundesland (zumeist Lehrkräfte für das jeweilige Fach, die über umfangreiche Erfahrung und Kompetenz in der Konzeption von Abituraufgaben verfügen), wobei in jeder Arbeitsgruppe auch die beruflichen Gymnasien durch mindestens eine Lehrkraft vertreten sind. Teil der Arbeitsgruppen sind auch Wissenschaftlerinnen und Wissenschaftler der jeweiligen Fachdidaktik und des jeweiligen Fachs. Sie haben eine beratende Funktion, d. h. sie geben aus der Perspektive der jeweiligen Fachdidaktik bzw. Fachwissenschaft Rückmeldungen im Prozess der Aufgabenentwicklung. Bei der Entwicklung der Gemeinsamen Abituraufgabenpools der Länder werden folglich die Perspektiven aller Länder sowie aktuelle Erkenntnisse der jeweiligen Fachdidaktik und Fachwissenschaft berücksichtigt.
- Die eingereichten Aufgabenvorschläge, die von der AG Aufgaben positiv eingeschätzt werden, werden dann von den Mitgliedern der AG Aufgaben *weiterentwickelt*.
- Die so in einem mehrstufigen Prozess gemeinsam entwickelten Abituraufgaben werden dann als Abituraufgabenpool den Ländern zur Verfügung gestellt, wobei den Ländern pro Fach eine große Anzahl an Aufgaben zur Wahl steht (z. B. ca. 25 Abiturprüfungsaufgaben im Fach Deutsch).

Entsprechend einer KMK-Vereinbarung entnehmen alle Länder Aufgaben aus den gemeinsam entwickelten Abituraufgabenpools, wobei eine jeweils landeseigene Kommission darüber entscheidet, welche Aufgaben aus den Pools für die Abiturprüfung verwendet und in welchem Umfang diese durch landeseigene Aufgaben ergänzt werden.

Folglich umfasst die Abiturprüfung eines Bundeslandes jeweils *ländergemeinsame* Aufgaben aus den Abituraufgabenpools und – je nach Abiturprüfungsorganisation des Bundeslandes (vgl. Abschn. 2.2.3) – dezentral oder zentral gestellte *landeseigene* Aufgaben. Die Verantwortung für die Abiturprüfung eines Landes liegt also weiterhin beim Land selbst (vgl. Hoffmann et al. 2022b).

Hinweis: Veröffentlichung der Poolaufgaben

Auf der Homepage des IQB (<https://www.iqb.hu-berlin.de/abitur/>) sind die Aufgaben aus den Abituraufgabenpools der vergangenen Prüfungsjahre veröffentlicht. Mit Blick auf die Einführung von Abituraufgabenpools für die naturwissenschaftlichen Fächer Biologie, Chemie und Physik ab dem Prüfungsjahr 2025 stehen dort zur Orientierung für Lehrkräfte und Schülerinnen und Schüler auch Beispielaufgaben zur Verfügung, die exemplarisch zeigen, wie auf der Grundlage der Bildungsstandards für die Allgemeine Hochschulreife entwickelte Abiturprüfungsaufgaben in den Naturwissenschaften gestaltet sein können (<https://www.iqb.hu-berlin.de/abitur/sammlung/>).

Mit den Gemeinsamen Abituraufgabenpools der Länder sind mehrere Ziele verbunden (Hoffmann et al. 2022b; Kultusministerkonferenz 2015):

- Sie sollen eine hohe Qualität der Abiturprüfung gewährleisten (*Qualitätssicherung*) sowie
- die *Vergleichbarkeit* der Abituranforderungen der 16 Bundesländer erhöhen und langfristig ein gemeinsames Anforderungsniveaus der Abituraufgaben sichern.
- Sie sollen dazu beitragen, dass Aufgabenstellungen im Abitur grundsätzlich und einheitlich an den Bildungsstandards für die Allgemeine Hochschulreife ausgerichtet werden (*Standardbasierung*) und sind somit ein wesentliches Instrument für die Implementation der Bildungsstandards in der Breite.
- Nicht zuletzt sollen die mit den Pools zur Verfügung gestellten Aufgaben sowohl den Abituraufgabenkommissionen der Länder als auch den Lehrkräften an den Schulen als Orientierung bei der Entwicklung eigener Prüfungs- oder Klausuraufgaben dienen (*Weiterentwicklung*) – sie sollen somit normierend auf länderspezifische Prüfungsaufgaben wie auch die Klausuren in der gymnasialen Oberstufe wirken, und damit letztlich auch auf den Unterricht in der gymnasialen Oberstufe.

Die Einführung der Gemeinsamen Abituraufgabenpools der Länder war und ist angesichts der *unterschiedlichen Prüfungstraditionen im Abitur* (z.B. zentrale, teilzentrale und dezentrale Prüfungssysteme), *länderspezifischen Aufgabenkulturen* (z.B. Aufgabenarten, zugelassene Hilfsmittel, Verständnis von ‚guten‘ Aufgaben) und *unterschiedlichen prüfungsorganisatorischen Rahmenbedingungen* (z.B. keine einheitlichen Prüfungstermine, unterschiedliche Bearbeitungszeiten im schriftlichen Abitur) mit vielen Herausforderungen verbunden.

Mit dem Beginn der Arbeiten an den Gemeinsamen Abituraufgabenpools der Länder wurde ein umfangreicher Standardisierungs- und Annäherungsprozess der Länder zu den Regelungen und Vorgaben für Abiturprüfungen und Abiturprüfungsaufgaben ausgelöst, der bis heute andauert. Es handelt sich dabei um einen sehr aufwendigen, langwierigen und teilweise auch schwierigen Prozess, bei dem sich die Ländervertretungen von den eigenen Traditionen lösen und auf gemeinsame Lösungen einigen mussten.

So wurden u. a. allgemeine Vereinbarungen zur Gestaltung der Aufgaben in den Pools getroffen (v. a. hinsichtlich der Struktur von Abiturprüfungsaufgaben sowie den grundsätzlichen Anforderungen für die Erstellung von Aufgaben, Erwartungshorizonten und Bewertungshinweisen), gemeinsame Abiturprüfungstermine festgelegt, Anpassungen von landesspezifischen Vorgaben und Regelungen vorgenommen – viele getroffene Entscheidungen benötigten einen zeitlichen Vorlauf von mehreren Jahren. Für die Übergangphase wurde vereinbart, dass die Länder Modifikationen der Poolaufgaben vornehmen durften, solange der Kern der Aufgaben erhalten und das Anspruchsniveau unverändert bleiben (vgl. Hoffmann et al. 2022b).

Mittlerweile sind die notwendigen Anpassungen der landesspezifischen Vorgaben und Regelungen bereits erfolgt oder zumindest angekündigt. Dementsprechend dürften die Länder nur noch in wenigen Fällen gezwungen sein, die von ihnen aus den Pools entnommenen Aufgaben vor dem Einsatz in der Abiturprüfung zu modifizieren – der Anteil an vorgenommenen Modifikationen ist im Laufe der Jahre zurückgegangen (Hoffmann et al. 2022a).

Unklar ist, in welchem Umfang die Länder Aufgaben aus den Abituraufgabenpools in ihren länderspezifischen Abiturprüfungen nutzen – entsprechende Daten dazu sind nicht öffentlich verfügbar. Allerdings fällt die Akzeptanz der Abituraufgabenpools länderspezifisch sehr unterschiedlich aus (Groß und Schmid-Kühn 2022, 2023), woraus ein unterschiedlicher Umgang mit den Poolaufgaben resultiert.

Die KMK hat vereinbart, dass ab 2023 (bzw. 2025 für die naturwissenschaftlichen Fächer) in allen Ländern (mindestens) 50 % aller schriftlichen Abituraufgaben aus den gemeinsamen Aufgabenpools entnommen werden müssen, sodass

die Abituranforderungen weiter angeglichen und so Qualität und Vergleichbarkeit im Abitur gesteigert werden sollen. Perspektivisch sollen die Abituraufgabenpools so weiterentwickelt werden, dass die Abiturprüfungen der Länder auch vollumfänglich, d. h. zu 100 % aus Poolaufgaben bestehen *können* (Kultusministerkonferenz 2020).

Standardbasierte Testverfahren im Sinne eines nationalen Bildungsmonitorings sind in der Sekundarstufe II aktuell nicht geplant – verlässliche wissenschaftliche Erkenntnisse darüber, inwieweit Abiturientinnen und Abiturienten in Deutschland die in den Bildungsstandards für die Allgemeine Hochschulreife formulierten Kompetenzziele erreichen, wird es also weiterhin nicht geben. Länderspezifische Befunde weisen allerdings darauf hin, dass ein erheblicher Anteil der Schülerinnen und Schüler die standardbasierten Erwartungen der gymnasialen Oberstufe (im Sinne des Erreichens von Regelstandards) *nicht* erfüllt (Leucht et al. 2016).

5.1.3 Bildungsberichterstattung

Die gemeinsame Bildungsberichterstattung von Bund und Ländern ist ein weiteres zentrales Element der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Der seit 2006 alle zwei Jahre erscheinende nationale Bildungsbericht *Bildung in Deutschland* ist ein umfassender Statusbericht über das gesamte Bildungswesen in Deutschland (vgl. Kühne 2022).

- ▶ Ein *Bildungsbericht* ist eine datengestützte Bestandsaufnahme der Rahmenbedingungen, Verlaufsmerkmale, Ergebnisse und Erträge von Bildung, die regelmäßig, in festen Zyklen und nach einem einmal entwickelten, weitgehend unveränderlichen Schema vorgelegt wird. Neben dem nationalen Bildungsbericht liegen auch internationale, bundesländerspezifische, regionale und kommunale Bildungsberichte vor.

Der nationale Bildungsbericht folgt (wie viele nationale und internationale Bildungsberichterstattungen) dem Leitgedanken *Bildung im Lebenslauf* – er ist nicht allein auf das Schulwesen konzentriert, sondern berücksichtigt alle Bildungsetappen

- von der *frühen Bildung* im Kindesalter
- über die Bereiche *Schule, Berufsausbildung und Hochschule*

- bis zu den verschiedenen Formen der *Weiterbildung im Erwachsenenalter* (vgl. Kühne 2022) einschließlich *außerschulischer Bildungsangebote* bzw. Bildungsaktivitäten wie etwa Kinder- und Jugendarbeit in Vereinen oder das freiwillige Engagement von Jugendlichen.

Die Berichterstattung folgt weitgehend dem in allen Bildungsbereichen etablierten Kontext-Input-Prozess-Output/Outcome-Modell (vgl. Abschn. 4.3.3), d. h. es werden sowohl ausgewählte

- (bildungs-)politisch nicht oder kaum beeinflussbare gesellschaftliche und ökonomische *Kontextfaktoren* (z. B. demographische Aspekte),
- politisch gesetzte *Rahmenbedingungen/Inputs* (z. B. Bildungsausgaben, Personalressourcen),
- *Bildungsprozesse* selbst (z. B. Übergänge und Wechsel im Bildungswesen) sowie
- deren zentrale Wirkungen – differenziert nach unmittelbarem *Output* (z. B. Kompetenzen im Schulalter, erworbene Abschlüsse) und langfristigem *Outcome* (z. B. Beschäftigungschancen, subjektives Wohlbefinden, Gesundheit, soziale Teilhabe und politische Partizipation) – berichtet.

Der Bildungsbericht ist eine *Zusammenschau bereits vorliegender Daten und Informationen*, u. a. fließen Daten aus der amtlichen Statistik (u. a. Kinder- und Jugendhilfestatistik, Schulstatistik, Berufsbildungsstatistik etc.) und regelmäßig erhobene und verfügbare Forschungsdaten (u. a. aus nationalen und internationalen Schulleistungsstudien) in diesen Bericht ein, zudem werden auch Bezüge und Querverweise zu anderen Informationsquellen (z. B. gesetzliche Vorgaben, institutionelle Regularien, eingeleitete bildungspolitische Maßnahmen o.ä.) aufgezeigt.

Die Bildungsberichterstattung erfolgt *indikatorengestützt*. Indikatoren sind quantitative erfassbare Größen, die als Stellvertretergrößen komplexe, in der Regel mehrdimensionale Zusammenhänge (möglichst) einfach und verständlich beschreiben sollen (vgl. Döbert 2007). Indikatoren bilden einen thematisch eindeutig, aber oft mehrdimensional definierten Aspekt von Bildung mit verschiedenen statistischen *Kennziffern* (i. d. R. ausgedrückt als ‚Anzahl der ...‘ oder ‚prozentualer Anteil an...‘ oder ‚Quote der...‘) ab. So fallen unter den Indikator ‚Übergänge in die Schule‘ u. a. Daten zum vorzeitigen (=Anteil der Kinder mit vorzeitigem Schulbesuch) oder verspäteten (=Anteil der Kinder, die verspätet eingeschult werden) Schuleintritt.

Im nationalen Bildungsbericht werden *Kernindikatoren* zu zentralen steuerungsrelevanten Aspekten im Bildungswesen regelmäßig aufgenommen (z. B. Bildungsausgaben, Kompetenzen im Schulalter). Die Indikatoren werden dabei in jedem Bildungsbereich so ausgewählt, dass möglichst ein zentraler Indikator für Input, für

Prozessaspekte und für Wirkungen enthalten ist. *Ergänzungsindikatoren* greifen aktuelle Themen, die für Bildungspolitik und Öffentlichkeit zeitweilig von Interesse sind (z. B. Inklusion, Digitalisierung, Bildungspersonal), auf – sie werden jedoch nicht regelmäßig berichtet (vgl. Deutsches Institut für Internationale Pädagogische Forschung 2007).

- ▶ Die Bildungsberichterstattung dient dem *Systemmonitoring* – der nationale Bildungsbericht macht Aussagen über das Bildungssystem in Deutschland (und teilweise auch in einzelnen Bundesländern), aber nicht über einzelne Bildungseinrichtungen.

Es werden regelmäßig empirisch fundierte *Informationen* über *aktuelle Zustände* (z. B. Wie groß ist das Angebot an U3-Plätzen in Deutschland?) und *Entwicklungen im Zeitverlauf* (z. B. Wie haben sich die Bildungsausgaben in den letzten 10 Jahren verändert?) im gesamten Bildungswesen berichtet, um Problemlagen aufzuzeigen und aktuelle und zukünftige Herausforderungen zu benennen.

Konkrete Empfehlungen für die Bildungspolitik werden hingegen *nicht* abgegeben. Vielmehr werden Informationen über das Bildungssystem *transparent* gemacht, sodass eine wissenschaftlich fundierte Grundlage für Diskurse über Bildungsfragen und bildungspolitische Entscheidungen zur Verfügung steht. Der Bildungsbericht richtet sich folglich an Entscheidungstragende in Bildungspolitik und -administration, aber auch an die interessierte Öffentlichkeit und ist insofern ein Dokument der *Information* und der *Rechenschaftslegung* gleichermaßen. So kann er mitunter auch das (Nicht-)Erreichen bildungspolitischer Ziele belegen – so wurde beispielsweise die im Jahr 2008 vereinbarte bildungspolitische Zielvorgabe, den Anteil der Schulabgängerinnen und -abgänger ohne Schulabschluss bis zum Jahr 2015 zu halbieren, nicht erreicht (weder bis 2015, noch bis heute). Der nationale Bildungsbericht folgt weitgehend dem gleichen Aufbau:

- Den Ausgangspunkt bilden demografische, ökonomische sowie weitere gesellschaftliche *Rahmenbedingungen*, in die das Bildungswesen eingebettet ist, z. B. Bevölkerungsentwicklung (z. B. Geburtenzahl, internationale Zu- und Abwanderung), wirtschaftliche Entwicklung, Familien- und Lebensformen (z. B. Anteil Alleinerziehende, Kinder in Risikolagen).
- Anschließend werden bereichsübergreifende oder -vergleichende *Grundinformationen zu Bildung in Deutschland*, etwa zu den Bildungsinvestitionen oder zu Bildungsbeteiligung und Bildungsstand der Bevölkerung gegeben.

- Anschließend fokussiert der Bericht auf Trends und Problemlagen, die sich in den einzelnen *Bildungsbereichen* abzeichnen – von der frühen Bildung bis zur Weiterbildung.
- Jeder Bildungsbericht widmet sich darüber hinaus einem eigenen *Schwerpunktthema*, das zum jeweiligen Zeitpunkt von besonderer bildungspolitischer und bereichsübergreifender Relevanz ist, wie beispielsweise Digitalisierung (2020), Migration (2006, 2016) oder Inklusion (2014).
- Zuletzt werden *Wirkungen und Erträge von Bildung* in den Blick genommen (vgl. Kühne 2022).

Hinweis: Zugang zu den Bildungsberichten im Internet

Unter www.bildungsbericht.de sind alle seit 2006 veröffentlichten nationalen Bildungsberichte frei zugänglich – ebenso wie zusätzliches Datenmaterial in Form von informationsreichen Tabellen, da nicht alle verfügbaren Daten auch berichtet werden. Seit 2018 wird neben dem mehrere hundert Seiten umfassenden Gesamtbericht auch eine Kurzfassung zur Verfügung gestellt, die die wichtigsten Ergebnisse zusammenfasst und anschaulich in Form möglichst einfach zugänglicher Infografiken aufbereitet (vgl. Kühne 2022).

Eine umfassende Bildungsberichterstattung hat sich in Deutschland auf nationaler, aber auch in vielen Bundesländern, Regionen (z.B. der Bildungsbericht Ruhr) und einzelnen Kommunen etabliert. Bildungsberichte gelten dabei als ein zentrales Element des Bildungsmonitorings mit dem Ziel, wissenschaftlich fundierte Informationen zu aktuellen Entwicklungen und langfristig bedeutsamen Trends im gesamten Bildungswesen bzw. in einem Teilsystem adressatengerecht aufzubereiten und darzustellen.

Allerdings weist die Bildungsberichterstattung auch Grenzen auf: Bildungsberichte generieren *Beschreibungswissen*, Fragen nach Ursachen und Wirkungszusammenhängen können oftmals nicht beantwortet werden, etwa im Hinblick auf die Wirksamkeit bereits umgesetzter bildungspolitischer Maßnahmen. In vielen Bereichen (insbesondere im Schulbereich) gibt es zudem nur eine *unzureichende bundesweit verfügbare Datenlage*, sodass nicht alle steuerungsrelevanten Fragen beantwortet werden können. Auch kann diese deskriptive Gesamtschau des Bildungssystems weder die kleinräumliche noch die pädagogische (vor allem instruktionale) Vielfalt des Bildungsgeschehens im Detail abbilden (vgl. Kühne 2022).

5.2 Schulqualität erfassen, sichern und weiterentwickeln auf der Ebene der Einzelschule

Die im vorherigen Abschn. 5.1 dargestellten Instrumente und Maßnahmen beziehen sich auf das Bildungs- bzw. Schulsystem als Ganzes und entsprechen – mit Ausnahme der Gemeinsamen Abituraufgabenpools der Länder – dem Systemmonitoring bzw. einem Bildungsmonitoring im engeren Sinne. Im folgenden Abschn. 5.2 geht es hingegen um *Verfahren der Qualitätssicherung und -entwicklung auf Ebene der einzelnen Schulen* und damit um Verfahren und Instrumente des Monitorings im weiteren Sinne (Schulevaluation).

Den Erläuterungen zur Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (vgl. Kultusministerkonferenz 2015, S. 13) folgend gibt es in diesem Kontext in fast allen Bundesländern neben den s.g. *Vergleichsarbeiten* (Abschn. 5.2.1) sowohl *Verfahren zur externen Evaluation der einzelnen Schule* (auch bekannt unter dem Begriff *Schulinspektion*; Abschn. 5.2.2.3) als auch zur *schulinternen Evaluation* (Abschn. 5.2.2.2), für die s.g. *Referenzrahmen für Schulqualität* (Abschn. 5.2.2.1) als Grundlage zur Verfügung stehen.

5.2.1 Landesweite Vergleichsarbeiten

Vergleichsarbeiten (geläufige Abkürzung: VERA) sind neben internationalen und nationalen Schulleistungsvergleichen (vgl. Abschn. 5.1.1 und Abschn. 5.1.2.2) ein weiteres Element der KMK-Gesamtstrategie zum Bildungsmonitoring. Die Vergleichsarbeiten sind in einigen Bundesländern unter anderen Namen (Lernstandeserhebungen, Kompetenztests, KERMIT) bekannt – trotz unterschiedlicher Bezeichnungen verbirgt sich dahinter Vergleichbares, bundesweit einheitliche Testverfahren: Vergleichsarbeiten sind

- jährlich verbindlich
- in der Jahrgangsstufe 3 (VERA 3: Deutsch, Mathematik) und
- in der Jahrgangsstufe 8 (VERA 8: Deutsch, Mathematik, Erste Fremdsprache)
- in allen Ländern (mit Ausnahme von Niedersachsen)
- an allen allgemeinbildenden Schulen
- unter Einbezug aller Schülerinnen und Schüler durchzuführen –
- es handelt sich also um eine Vollerhebung.

Die Länder haben sich darauf geeinigt, dass VERA jeweils in mindestens einem Fach obligatorisch durchgeführt wird. Den Ländern steht es darüber hinaus frei,

mehr als ein Fach bzw. einen Kompetenzbereich landesweit für alle Schulen verpflichtend überprüfen zu lassen (vgl. Hunger 2021).

- Vergleichsarbeiten sind schriftliche standardbasierte Tests, d.h. sie basieren auf den Bildungsstandards und den zugehörigen Kompetenzstufenmodellen. Die Ergebnisse werden *auf der Ebene einzelner Schulen und Klassen* berichtet und zeigen, über welche Kompetenzen Schülerinnen und Schüler in der dritten bzw. achten Jahrgangsstufe verfügen.

Sie richten sich primär an die Verantwortlichen jeder einzelnen Schule (Lehrkräfte, Schulleitung) und stellen eine Art ‚Frühwarnsystem‘ dar, weil sie bereits ein Jahr vor dem jeweiligen Abschluss, für den die Bildungsstandards definiert sind, Informationen liefern, inwieweit die Anforderungen der Bildungsstandards von den Schülerinnen und Schülern bereits erreicht werden und wo noch Handlungsbedarfe bestehen. Lehrkräften steht also noch Zeit zur Verfügung, um darauf zu reagieren und auf die Kompetenzziele hinzuwirken, die am Ende der vierten Klassen bzw. am Ende der Sek. I erreicht sein sollen.

Die Ergebnisse werden hingegen *nicht* für einen Vergleich zwischen den Ländern genutzt, auch wird *kein* Schulranking in den einzelnen Ländern vorgenommen – die Ergebnisse auf Schul- und Klassenebene werden *nicht* veröffentlicht (vgl. Hunger 2021).

Die einzelnen Länder haben die Gesamtverantwortung für die Durchführung der Vergleichsarbeiten. Sie organisieren die Vorbereitung, den Ablauf, die Auswertung und die Ergebnismeldung jeweils in eigener Verantwortung anhand länderspezifischer Regelungen, wobei einige konzeptionelle Unterschiede erkennbar sind, etwa hinsichtlich des Verpflichtungsgrades, der Gestaltung der schriftlichen Ergebnismeldungen an die Schulen und der Unterstützungsangebote zur Weiterarbeit mit den Ergebnissen (vgl. Tarkian, Maritzen et al. 2019).

Die Federführung für die Entwicklung der Testaufgaben liegt hingegen beim IQB (vgl. Abschn. 5.1.2.1). Die Aufgaben werden von Lehrkräften aus unterschiedlichen Schulformen aus allen Bundesländern in Zusammenarbeit mit Fachdidaktikerinnen und Fachdidaktikern und anderen Wissenschaftlerinnen und Wissenschaftlern erarbeitet und mehrfach erprobt, bevor sie tatsächlich zum Einsatz kommen.

Die Durchführung der Tests erfolgt in den Schulen durch Lehrkräfte, die dazu Testmaterialien sowie Informations- und Instruktionmaterial erhalten. In den meisten Ländern korrigieren die Lehrkräfte die Testhefte und geben die Ergebnisse der Schülerinnen und Schüler in ein Auswertungssportal ein. Die Auswertung erfolgt landeszentral durch eine vom Land beauftragte Einrichtung

(vgl. im Überblick Tarkian, Maritzen et al. 2019, S. 47); von dort erhalten die Schulen auch die Ergebnismeldung, die i. d. R. die Testergebnisse jeder einzelnen Schülerin bzw. jedes einzelnen Schülers sowie der einzelnen Klassen an einer Schule, aber auch landesspezifische Vergleichswerte umfassen.

Hinweis: Aufgabenentwicklung und Beispielaufgaben

Aufgabenentwicklung und Beispielaufgaben: Auf der Homepage des IQB wird der komplexe Prozess zur Entwicklung von Testaufgaben detailliert beschrieben (<https://www.iqb.hu-berlin.de/vera/ueberblick/testentwicklung/>). Zudem finden sich dort auch Beispielaufgaben aus vergangenen Durchgängen mit den dazugehörigen Lösungen und den didaktischen Kommentaren (<https://www.iqb.hu-berlin.de/vera/aufgaben/>).

Die Ergebnisse der Vergleichsarbeiten zeigen v. a. die prozentuale Verteilung der Schülerinnen und Schüler einer Klasse auf die Kompetenzstufen und geben an, welche Kompetenzstufe einzelne Schülerinnen und Schüler in der jeweiligen Klasse erreichen. Ferner kann ein Vergleich des Lernstandes der eigenen Klasse erfolgen, etwa mit den Ergebnissen der Parallelklassen derselben Schule, mit dem Landeswert oder als sogenannter *fairer Vergleich* mit Schulen bzw. Vergleichsgruppen, die vergleichbare Rahmen- und Kontextbedingungen (z. B. ähnliches schulisches Einzugsgebiet oder vergleichbare Klassenzusammensetzung) aufweisen.

Die Ergebnisse geben den Lehrkräften nicht nur eine Rückmeldung darüber, in welchen Bereichen ihre Schülerinnen und Schüler bzw. Klasse besondere Stärken und Schwächen aufweisen, sondern auch wie deren Leistungen relativ zu denen anderer Klassen und Schulen einzuordnen sind (vgl. z. B. Isaac und Hosenfeld 2008).

- ▶ In der Gesamtschau sollen Vergleichsarbeiten eine Bestandsaufnahme fachlicher Kompetenzen und erreichter Lernstände anhand eines *kriterialen* Vergleichsmaßstabs – den in den Bildungsstandards festgelegten Kompetenzziele – sowie durch multiple Vergleichsmöglichkeiten (bezugsgruppenorientierter, *sozialer Vergleich*) ermöglichen.

Mit Vergleichsarbeiten wird zudem versucht, direkter als es über die auf der Systemebene verorteten IQB-Bildungstrends (vgl. Abschn. 5.1.2.2) möglich und intendiert ist, am Schul- und Unterrichtsgeschehen anzusetzen und darüber die *Kompetenzorientierung* zu stärken. Insofern sollen Vergleichsarbeiten die Implementation der Bildungsstandards unterstützen – die im Kontext der Vergleichs-

arbeiten entwickelten Aufgaben und Materialien sowie weitere Fortbildungen und Unterstützungsangebote des IQB und der Länder sollen die Lehrkräfte an den Schulen unterstützen, kompetenzorientiert und auf die Bildungsstandards bezogen zu unterrichten.

Wichtigstes Ziel der Vergleichsarbeiten ist die Unterstützung der *Unterrichts- und Schulentwicklung* an jeder einzelnen Schule: Die Ergebnisse der Vergleichsarbeiten liefern insbesondere differenzierte Informationen über Stärken und Schwächen einzelner Klassen und bieten damit Lehrkräften wichtige Hinweise für die Planung und Gestaltung des Unterrichts im letzten Jahr der Grundschule bzw. bis zum Erreichen der Schulabschlüsse in der Sekundarstufe I. Die Rückmeldungen können nicht nur für den eigenen Unterricht genutzt werden, sondern sollen auch kollegiale Unterrichtsentwicklung anstoßen, indem etwa pädagogische Interventionen und Fördermaßnahmen auf der Basis der Ergebnisrückmeldungen kooperativ begründet und geplant werden (vgl. Tarkian, Maritzen et al. 2019).

Die Bundesländer stellen z. T. umfangreiche Materialien und Unterstützungsangebote (Beratung, schulinterne Fortbildungen) zur Verfügung, um Schulen und Lehrkräfte im Umgang mit den Ergebnissen zu unterstützen und *Impulse für die Schul- und Unterrichtsentwicklung* zu geben. Um zu unterstreichen, dass Vergleichsarbeiten primär der Schul- und Unterrichtsentwicklung dienen, hat die Kultusministerkonferenz vereinbart, dass die Ergebnisse aus den Vergleichsarbeiten weder benotet noch als Grundlage für Übergangsempfehlungen/-entscheidungen genutzt werden sollen. Ob und inwieweit die Ergebnisrückmeldungen tatsächlich für die Unterrichtsplanung und schulische Entwicklungsprozesse genutzt werden, entscheidet jede Lehrkraft und Schule eigenständig (vgl. Henschel und Stanat 2019).

Aus der Forschung zu den Wirkungen der Vergleichsarbeiten wird insgesamt deutlich, dass die tatsächlichen Wirkungen hinter den Erwartungen zurückbleiben (vgl. van Ackeren et al. 2017). Insbesondere werden nicht alle Ziele als solche wahrgenommen – die Vergleichsarbeiten werden primär als Instrument der Standortbestimmung und Rechenschaftslegung betrachtet und weniger als Instrument zur Unterstützung der Schul- und Unterrichtsentwicklung.

Ogleich die Vergleichsarbeiten schon länger implementiert sind, bestehen aufseiten vieler Lehrkräfte Unklarheiten über deren Funktion. Insgesamt zeigt sich unter den Lehrkräften eine abnehmende Akzeptanz der Vergleichsarbeiten – die Einstellungen der Schulleitungen stellen sich demgegenüber positiver dar, dies gilt auch für die wahrgenommene Nützlichkeit sowie die Nutzung.

Grundsätzlich besteht zwar eine Bereitschaft, die Ergebnisrückmeldungen zur Kenntnis zu nehmen, auszutauschen und diese auch zu reflektieren, etwa im Rahmen von Schul- und Fachkonferenzen, aber sie sind nur selten Ausgangspunkt für

Überlegungen zur Schul- und Unterrichtsentwicklung. Insbesondere werden die Ergebnisse selten für die Weiterentwicklung schulischer Konzepte, die gezielte Fortbildungsplanung und die Personalentwicklung genutzt. Ein Mehrwert wird vorrangig in der individuellen Leistungsdiagnostik und mit Blick auf eine verstärkte Einzelförderung gesehen.

Insgesamt profitieren Schulen mit hoher Expertise in der Nutzung von Datenrückmeldungen (z. B. vorhandene Kompetenzen, um die Ergebnisrückmeldungen zu verstehen) stärker als Schulen, die in diesem Bereich über weniger Kompetenzen verfügen.

Insofern bedarf es eines Unterstützungssystems, das Schulen bedarfsorientiert darin berät und begleitet, sich datengestützt zu entwickeln. Der Forschungsstand zu den unterrichtlichen Wirkungen ist uneinheitlich: Eine Studie konnte zeigen, dass zunehmend den Vergleichsarbeiten ähnliche Aufgaben stärker in den Unterricht integriert werden – eine andere Studie konnte hingegen keine Veränderungen im Unterrichtshandeln, z. B. stärkere Differenzierung oder Kompetenzorientierung im Unterricht, feststellen (ebd.).

5.2.2 Verfahren und Instrumente zur einzelschulischen Qualitätsentwicklung

In der erstveröffentlichten Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring aus dem Jahr 2006 wurden die Vergleichsarbeiten (vgl. Abschn. 5.2.1) explizit im Sinne einer Einzelmaßnahme benannt – in der überarbeiteten Fassung aus dem Jahr 2015 wird deutlich allgemeiner von *Verfahren zur Qualitätssicherung auf Ebene der Schulen* gesprochen und damit die Bedeutung der Einzelschule als zentraler Ansatzpunkt für die Verbesserung der Qualität des Gesamtsystems betont. Insofern sind landesweite Vergleichsarbeiten lediglich *ein* Element eines ganzen Maßnahmenbündels, mit denen die Länder eine evidenzbasierte Qualitätsentwicklung und -sicherung auf Ebene der einzelnen Schule gewährleisten möchten (vgl. Kultusministerkonferenz 2015, S. 13).

Zu diesem Maßnahmenbündel gehören in fast allen Bundesländern (vgl. auch Thiel et al. 2019)

- *Verfahren zur externen Evaluation der einzelnen Schule* (auch: ‚Schulinspektion‘), in deren Rahmen die einzelne Schule regelmäßig von schulexternen Evaluationsteams besucht und systematisch begutachtet wird und Rückmeldungen über ihre Stärken, aber auch zu Handlungsbedarfen erhält (vgl. Abschn. 5.2.2.3).

- Zusätzlich unterstützen die Länder die *interne Evaluation von Schulen*, also die Evaluation durch die Mitglieder der Schule selbst, durch die Bereitstellung entsprechender Instrumente und Beratungsangebote (vgl. Abschn. 5.2.2.2).
- Grundlage für die interne und externe Evaluation bilden die s.g. *Referenzrahmen für Schulqualität* (vgl. Abschn. 5.2.2.1), die Erwartungen und Anforderungen an die Qualität von Schule und Unterricht beschreiben.

Für eine Evaluation werden Daten (z.B. zur Wahrnehmung des Schulklimas aus Sicht von Schülerinnen und Schülern oder Einschätzungen zur Lehrkräftekooperation) organisiert erhoben (z.B. mittels Fragebögen, Interview- oder Beobachtungsleitfäden) und transparent und nachvollziehbar dokumentiert. Evaluation soll einen konkreten Nutzen haben und zu Konsequenzen führen.

- ▶ **Evaluation:** Allgemein formuliert bezeichnet Evaluation die *systematische Bestandsaufnahme, Analyse und Bewertung* schulischer Arbeit im Sinne eines Soll-Ist-Vergleichs.

Unterschieden werden können drei *Evaluationsformen*:

- *Selbstevaluation* bedeutet, das *eigene* professionelle Handeln systematisch zu beobachten, zu analysieren und zu bewerten, um es zu stabilisieren oder zu verbessern. So stehen Lehrkräften zahlreiche etablierte Instrumente zur Evaluation ihres eigenen Unterrichts zur Verfügung (z. B. EMU: Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung).
- Bei einer *internen Evaluation* beobachten, analysieren und bewerten die Mitglieder einer Schule schulinterne (= ihre eigenen) Arbeitsprozesse und Ergebnisse mit dem Ziel der Sicherung oder Weiterentwicklung der Qualität der eigenen Schule.
- Im Rahmen einer *externen Evaluation* (auch: Fremdevaluation) beobachten, analysieren und bewerten Schul-Externe (z. B. die Schulaufsicht, Wissenschaftlerinnen und Wissenschaftler...) die Arbeit und die Arbeitsergebnisse einer Schule.

Zudem gibt es zwei zentrale *Evaluationstypen*:

- Die *formative* (= prozessbezogene) Evaluation zielt auf eine entwicklungsbegleitende Optimierung beginnender oder laufender Angebote oder Maßnahmen. Die Ergebnisse der Evaluation fließen direkt in die Optimierung des jeweiligen Angebots zurück.

- Die *summative* (= produktbezogene) Evaluation beurteilt den Erfolg oder die Wirksamkeit eines Angebots oder einer Maßnahme. Die Ergebnisse können zeigen, ob das Angebot oder die Maßnahme tatsächlich so erfolgreich ist, wie man es erwartet hat, und Entscheidungshilfe sein, diese/-s fortzuführen, einzustellen, einzuschränken oder auszuweiten.

Im Fokus der verschiedenen Evaluationsverfahren steht die Einzelschule. Evaluation soll helfen, Schul- und Unterrichtsqualität sichtbar zu machen – insbesondere die Qualität von Prozessen (z.B. Lehrkräftekooperation, Schulleitungshandeln, Schulklima, Partizipation von Schülerinnen und Schülern) – und die Evaluationsergebnisse sollen für die einzelschulische Qualitätsentwicklung genutzt werden.

5.2.2.1 Referenzsysteme zur Schulqualität

In allen Bundesländern gibt es sogenannte *Referenzrahmen zur Schulqualität* als Unterstützungsinstrument für die systematische Qualitätsentwicklung an Einzelschulen. Die Bezeichnung variiert in den Bundesländern, neben ‚Referenzrahmen‘ werden auch die Begriffe *Qualitätsrahmen*, *Qualitätsstableau*, *Handlungsrahmen* oder *Orientierungsrahmen* verwendet (vgl. Thiel und Tarkian 2019, S. 16). In mehr oder weniger umfangreichen Dokumenten – der Seitenumfang liegt je nach Bundesland zwischen zwei und mehr als hundert Seiten – werden Qualitätskriterien zu bestimmten Qualitätsdimensionen/-bereichen benannt, um allen an Schule Beteiligten in einem Bundesland transparent zu machen, was unter ‚guter‘ Schule (*Schulqualität*) und ‚gutem‘ Unterricht (*Unterrichtsqualität*) zu verstehen ist.

Auch ihr Aufbau folgt zumeist dem etablierten Kontext-Input-Prozess-Output/Outcome-Modell von Schulqualität (vgl. Abschn. 4.3.3), wobei nicht alle Rahmenkonzepte explizit auch Inputs und Kontextbedingungen berücksichtigen. Neben Output-Kriterien (z. B. fachliche und überfachliche Kompetenzen der Schülerinnen und Schüler, erworbene Abschlüsse) stehen primär *Prozesse auf Schulebene* und *Prozesse auf Unterrichtsebene* im Fokus der Referenzrahmen und damit die Merkmale, die von den Mitgliedern einer Schule selbst gestaltet werden können. So sind z. B. Kooperation, Leitungshandeln/Schulmanagement, Qualitätsentwicklung und Schulklima Qualitätsmerkmale auf Schulebene und Kompetenz- bzw. Standardorientierung, Individualisierung/Umgang mit Heterogenität und Unterrichts-/Lernklima Merkmale von Unterrichtsqualität (vgl. Thiel und Tarkian 2019).

Hinweis: Referenzrahmen aller Länder im Internet

Über den Deutschen Bildungsserver (<https://www.deutscher-bildungsserver.de>) können die Referenzrahmen aller Länder eingesehen werden. Viele Bundesländer bieten sehr umfangreiche Online-Unterstützungsportale, Anwendungsmaterialien, Fortbildungen und individuelle Beratungsangebote zum Umgang mit den Referenzrahmen und deren Einbindung in Schulentwicklungsprozesse, z.B. Hessen oder Nordrhein-Westfalen.

Das zugrunde liegende Qualitätsverständnis der Referenzrahmen fußt auf verschiedenen Quellen, Referenzen und Bezugsbereichen, v.a. auf gesetzlichen Vorgaben, wissenschaftlichen Erkenntnissen (z.B. theoretische Modelle zur Schulqualität, Ergebnisse der empirischen Schul- und Unterrichtsforschung) aber auch der Perspektive der Schulpraxis. Die Referenzrahmen sind Ergebnis eines *partizipativen Entwicklungsprozesses*, an dem – je nach Bundesland – zahlreiche Akteure beteiligt sind (u.a. Lehrkräfte, Schulleitungen, Schülerinnen und Schüler sowie deren Eltern, die Schulaufsicht, Verbände, Wissenschaftlerinnen und Wissenschaftler).

Referenzrahmen sind *entwicklungsoffene Konzepte*, die es ermöglichen, schulpolitische Entwicklungen und Entscheidungen sowie aktuelle gesellschaftliche Entwicklungen ebenso zu integrieren wie neuere wissenschaftliche Erkenntnisse. Viele Referenzrahmen liegen mittlerweile in (mehrfach) überarbeiteter Fassung vor (z. B. der rheinland-pfälzische Orientierungsrahmen Schulqualität in der 5. Auflage). Vergleicht man die Referenzrahmen der Bundesländer, so scheint es zwar ein *geteiltes Grundverständnis von Schul- und Unterrichtsqualität* zu geben (Elsing und van Ackeren 2017), aber insgesamt sind Aufbau und Detailliertheitsgrad in den länderspezifischen Referenzrahmen sehr unterschiedlich (Thiel und Tarkian 2019).

Referenzrahmen erfüllen verschiedene Funktionen:

- Sie sind Orientierungsmaßstab für die Entwicklung der Einzelschule (*Entwicklungsfunktion*) – auf Basis der in den Referenzrahmen formulierten Ansprüche und Erwartungen an die Qualität von Schule und Unterricht können mittels *interner Evaluation* (vgl. Abschn. 5.2.2.2) einzelschulische Stärken und Schwächen identifiziert, Handlungsbedarfe formuliert und Qualitätsentwicklungsmaßnahmen abgeleitet werden.

- Referenzrahmen sind zugleich auch Arbeitsgrundlage für die *externe Evaluation* (vgl. Abschn. 5.2.2.3), wenn eine Schule durch schulexterne Personen auf Basis der in den Referenzrahmen definierten Merkmale von Schul- und Unterrichtsqualität evaluiert wird. Insofern dienen Referenzrahmen auch der Rechenschaftslegung, ob die Schule ihren gesellschaftlichen Auftrag und die Anforderungen an ‚gute Schule‘ und ‚guten Unterricht‘ erfüllt (*Rechenschaftsfunktion*).

5.2.2.2 Interne Evaluation

Unter *interner Evaluation* einer Schule werden zahlreiche Verfahren subsumiert, die durch Personen, die in dieser Schule tätig sind (i.d.R. Schulleitungen und Lehrkräfte) geplant, durchgeführt und genutzt werden, also für deren Durchführung die Einzelschule verantwortlich ist und die diese eigenständig gestaltet.

- ▶ Bei einer *internen Evaluation* beobachten, analysieren und bewerten die *Mitglieder einer Schule* schulinterne (= ihre eigenen) Arbeitsprozesse oder Ergebnisse – je nach Erkenntnisinteresse kann es darum gehen, Aspekte des Unterrichts (z.B. Umgang mit Heterogenität), Teilbereiche der Schule (z.B. Arbeitsprozesse in Fachkonferenzen) oder der Schule als Ganzes (z.B. das Ganztagsangebot) näher in den Blick zu nehmen.

Grundsätzlich dienen interne Evaluationen der Bewertung von schulinternen Arbeitsprozessen und Ergebnissen, und *nicht* der Beurteilung von Personen (z. B. einzelner Lehrkräfte). So erhält eine Schule Impulse für Qualitätsentwicklungsmaßnahmen und für die Förderung der Professionalität schulischer Akteure. Schulinterne Evaluation hat bildungspolitisch einen hohen Stellenwert – in vielen Ländern gibt es eine *gesetzliche Verpflichtung* für Schulen, eigenverantwortlich interne Evaluationen durchzuführen. Die Mitwirkung an Verfahren zur schulinternen Evaluation ist folglich *integraler Bestandteil der professionellen Tätigkeit von Lehrkräften*.

Schulinterne Evaluation stellt ein zentrales Instrument für die schulische Selbststeuerung dar und soll helfen, Entwicklungsbedarf im schulischen Alltag zu erkennen, Ziele für die weitere Arbeit festzulegen und so schulische Entwicklungsprozesse anzustoßen (*Entwicklungsfunktion*). Sie können auch der *Rechenschaftslegung* dienen, wenn schulische Arbeit transparent gemacht und zur Diskussion gestellt wird (z. B. gegenüber der gesamten Schulgemeinschaft, der Öffentlichkeit oder auch gegenüber der Schulaufsicht). Eine Handreichung zur schulinternen Evaluation aus Schleswig-Holstein betont beispielsweise, dass interne Evaluation

„der Vergewisserung nach innen, aber auch der Rechenschaftslegung nach außen“ dient (IQSH 2014, S. 7; zit. nach Tarkian, Riecke-Baulecke et al. 2019, S. 199).

Idealtypisch lassen sich zwei zentrale Evaluationsfelder unterscheiden:

- Zum einen die *Evaluation der Schul- und Unterrichtsqualität* (bzw. bestimmter Qualitätsbereiche), die auf der Grundlage der Referenzrahmen zur Schulqualität (vgl. Abschn. 5.2.2.1) stattfinden.
- Darüber hinaus können auch konkrete schulische, unterrichtliche und außerunterrichtliche Programme, Projekte, Produkte oder Maßnahmen einer Schule evaluiert werden.

Schulinterne Evaluationen sind alle Formen der systematischen Informationsbeschaffung und -bewertung in der Verantwortung schulischer Akteure und dienen insbesondere

- der innerschulischen *Bestandsaufnahme* (Ist-Analyse) zu Beginn eines Entwicklungsprozesses, um die Ausgangssituation an der eigenen Schule zu erfassen, Handlungsbedarfe zu identifizieren und Qualitätsentwicklungsprozesse anzustoßen,
- der *abschließenden Bilanz* (summative Evaluation/Ergebnisevaluation) im Sinne einer Erfolgskontrolle, indem der Erfolg eines Angebots oder die Wirksamkeit einer Maßnahme beurteilt werden, um auf Basis der Ergebnisse festzustellen, ob das Angebot oder die Maßnahme tatsächlich so erfolgreich ist, wie man es erwartet hat, und so zu entscheiden, diese/-s fortzuführen, einzustellen, einzuschränken oder auszuweiten, oder
- der *Prozessanalyse* (formative Evaluation/Prozessevaluation), die der entwicklungsbegleitenden Optimierung beginnender oder laufender Angebote oder Maßnahmen dient, d. h. die Ergebnisse der Evaluation fließen direkt in die Optimierung des jeweiligen Angebots zurück.

Schulen können mittlerweile auf einen großen Pool von Praxisleitfäden und standardisierte und etablierte Instrumente zur internen Evaluation (z.B. Fragebögen, Beobachtungsbögen, Gesprächsleitfäden) zurückgreifen, die z.T. auch online zur Verfügung stehen. Nicht jede Schule muss also eigene Instrumente und Techniken entwickeln. Gleichwohl ist nicht jeder im Internet verfügbare Fragebogen oder Gesprächsleitfaden geeignet, wenn es um sehr spezifische Projekte oder Maßnahmen an einer Schule geht. Dann kann auch die *Konstruktion eigener Erhebungsinstrumente* sinnvoll sein.

In nahezu allen Bundesländern werden standardisierte Instrumente für die interne Evaluation bereitgestellt, zumeist auch als Online-Variante, um Befragungen webbasiert durchführen und auswerten zu können. *Landeseigene* Instrumente sind i. d. R. an die jeweiligen Referenzrahmen für Schulqualität angepasst, d. h. die Instrumente (z. B. Fragebögen) sind systematisch entlang der Qualitätsdimensionen und -bereiche der Referenzrahmen gegliedert. Die meisten Länder stellen auch Instrumente zur Verfügung, die über die im Referenzrahmen beschriebenen Qualitätsdimensionen hinausgehen bzw. dort eher integrativ behandelt werden (wie z. B. zu den Themen *Zufriedenheit* oder *Ganztagsschule*) bzw. ermöglichen es, durch die Formulierung eigener Fragen auch maßgeschneiderte Befragungen selbst- bzw. mitzuentwickeln, um den besonderen einzelschulischen Bedürfnissen gerecht zu werden (vgl. Tarkian, Riecke-Baulecke et al. 2019, S. 220).

Zudem gibt es *länderübergreifende* Angebote zur schulinternen Evaluation – sowohl zur *Evaluation der Unterrichtsqualität* (z. B. EMU: Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung, oder SefU: Schülerinnen und Schüler als Experten für Unterricht) als auch zur *Evaluation der Schulqualität* (z. B. SEIS: Selbstevaluation in Schulen). Diese Angebote stehen grundsätzlich allen Schulen in allen Bundesländern zur Verfügung.

Trotz der großen bildungspolitischen Bedeutung liegen nur wenige empirische Befunde zur Evaluationspraxis an deutschen Schulen vor, wengleich der Verbreitungsgrad generell hoch eingeschätzt wird. Allerdings wird so lediglich ausgesagt, dass schulinterne Evaluation praktiziert wird – aber nicht in welcher Art und Weise und wie gut.

Interne Evaluation wird generell – und im Vergleich zu anderen Evaluationsverfahren (z. B. Vergleichsarbeiten, vgl. Abschn. 5.2.1) – von Schulleitungen und Lehrkräften als nützlicher bewertet. Die Ergebnisse aus internen Evaluationen werden zudem umfassend rezipiert und sind häufiger Ausgangspunkt von Entwicklungsvorhaben. Empirische Studien zeigen, dass die Evaluationsergebnisse von der überwiegenden Mehrzahl der Lehrkräfte als Ausgangspunkt für Schul-/Unterrichtsentwicklung verwendet werden. Schulleitungen und Lehrkräfte berichten sowohl von initiiertem Schulentwicklung (z. B. die Intensivierung von Kommunikation im Kollegium) als auch Unterrichtsentwicklung (z. B. Einsatz neuer Unterrichtsmethoden und -materialien); gleichwohl gibt es in bestimmten Bereichen keine Entwicklung. Als förderlich für die Ableitung von Maßnahmen zur Qualitätsentwicklung aus den Ergebnissen schulinterner Evaluation sind die Informiertheit über das Verfahren, die eingeschätzte Nützlichkeit des Verfahrens und der aus den Ergebnissen geschlussfolgerte Veränderungsanlass zu nennen (vgl. Wurster 2022).

5.2.2.3 Externe Evaluation

Externe Evaluation ist zunächst einmal ein Sammelbegriff für Evaluationsverfahren, die von schulexternen Personen durchgeführt werden. So sind z.B. auch Vergleichsarbeiten (vgl. Abschn. 5.2.1) ein externes Evaluationsverfahren, da die Testaufgaben zentral gestellt werden und auch die Auswertung der Ergebnisse und die Datenrückmeldung an die Schulen von Extern erfolgt. In diesem Kapitel geht es aber um ein ganz konkretes Verfahren, das in den Ländern unterschiedliche Bezeichnungen hat – neben *Externe Evaluation* werden je nach Bundesland auch die Begriffe *Schulinspektion*, *Schul-TÜV*, *Schulvisitation*, *Qualitätsanalyse*, *Qualitätsfeststellung an Schulen*, *Fremdevaluation* oder *Fokusevaluation* verwendet (vgl. Tarkian, Lankes et al. 2019, S. 106).

- ▶ *Externe Evaluation* bezeichnet hier die regelmäßige Erfassung und Analyse der Schul- und Unterrichtsqualität einer Einzelschule im Rahmen eines standardisierten Verfahrens, das von Schulexternen – professionell geschulten Evaluationsteams – durchgeführt wird. Sie soll also die Einzelschulentwicklung durch einen ‚neutralen Blick von außen‘ unterstützen.

Das Verfahren gehört in den Ländern meist zum zentralen Aufgabenbereich von *Qualitätsagenturen*, die in den vergangenen Jahren in den Bundesländern aufgebaut wurden. Sie sind nicht Teil der klassischen Schulaufsicht, sondern von dieser institutionell und personell unabhängig.

Gegenstand der Externen Evaluation ist die *Schule als Ganzes* und nicht einzelne Projekte bzw. Maßnahmen oder das Handeln einzelner Lehrkräfte. Das Evaluationsteam besteht aus Personen, die *nicht* der zu evaluierenden Schule angehören, wobei Größe und Zusammensetzung der Evaluationsteams länderspezifisch variieren. Zumeist handelt es sich um eine Gruppe von zwei bis vier Personen, die selbst lange Zeit in der Schule tätig waren (v. a. erfahrene Schulleitungsmitglieder und Lehrkräfte) – Mitglied eines Evaluationsteams können aber auch Personen aus der Schulaufsicht oder den Studienseminaren sowie (je nach Bundesland) Externe, d. h. Vertreter von Eltern- oder Schülerverbänden, der Wirtschaft oder der Wissenschaft, die das Verfahren um eine außerschulische Perspektive bereichern, sein. Um die Rolle des Evaluierenden professionell ausüben zu können, durchlaufen die Mitglieder der Evaluationsteams eine Qualifizierung, die je nach Bundesland sehr unterschiedlich ausfällt (vgl. Tarkian, Lankes et al. 2019). Grundlage der Externen Evaluation sind i. d. R. die in den landeseigenen Referenzrahmen für Schulqualität (vgl. Abschn. 5.2.2.1) enthaltenen Qualitätsdimensionen, wobei insbesondere die *Qualität schulischer und unterrichtlicher Prozesse* im Fokus steht.

Der *Ablauf einer Externen Evaluation* umfasst üblicherweise drei Phasen (Vorbereitung, Schulbesuch, Rückmeldung), wobei die Verfahren länderspezifisch variieren (vgl. Tarkian, Lankes et al. 2019):

Kern der Externen Evaluation ist der *Schulbesuch* durch das Evaluationsteam: Je nach Schulgröße und den Untersuchungsschwerpunkten werden die Schulen für einen Zeitraum zwischen zwei und fünf Tagen vom Evaluationsteam besucht. Ein *Rundgang* durch die Schulgebäude und schulische Flächen ermöglicht dem Evaluationsteam einen Eindruck von den räumlichen Voraussetzungen und der Ausstattung.

Kernbereich der Externen Evaluation sind die *Unterrichtsbesuche*, d. h. die Mitglieder des Evaluationsteams hospitieren im Unterricht um auf der Basis von mehr oder weniger standardisierten Unterrichtsbeobachtungsbögen Aussagen zur Qualität des Unterrichts machen zu können. Sowohl der zeitliche Umfang der Hospitation (von 20 bis 90 Minuten) als auch das Ausmaß der Hospitationen variieren länderspezifisch. So wird z. B. im Saarland im Unterricht des gesamten Kollegiums hospitiert, während in Nordrhein-Westfalen wenigstens die Hälfte des Kollegiums beim Unterrichten gesehen werden soll. Zudem finden *Vor-Ort-Gespräche* mit ausgewählten Vertretern verschiedener schulischer Gruppen (Schulleitung, Lehrkräfte, Schülerinnen und Schüler sowie deren Eltern, weiteres schulisches Personal oder auch Kooperationspartner), die zumeist auf Basis von Leitfäden, die mehr oder weniger stark standardisiert sind, erfolgen.

Zur *Vorbereitung des Schulbesuchs* werden im Vorfeld verschiedene Unterlagen von den Schulen angefordert, v. a. Dokumente (wie etwa das Schulprogramm, schulinterne Lehrpläne, Fortbildungsplanung), statistische Angaben (z. B. Informationen über Klassenfrequenzen, Schulgebäude und Ausstattung der Schule, Betreuungsangebot) sowie Entwicklungsberichte, also Maßnahmen, die seit der letzten Externen Evaluation umgesetzt wurden. Der Umfang dessen, was für eine *Daten- und Dokumentenanalyse* angefordert wird, ist länderspezifisch sehr unterschiedlich und wurde zur Entlastung der Schulen mittlerweile stark reduziert. Darüber hinaus erfolgt im Vorfeld des Schulbesuchs eine *standardisierte Befragung aller schulischen Akteursgruppen* mittels Fragebögen, wobei neben Lehrkräften, Schülerinnen und Schülern sowie deren Eltern zum Teil auch weiteres pädagogisches sowie nicht-pädagogisches Schulpersonal wie Kooperationspartner befragt werden. Die Erhebung erfolgt anonym – entweder papier- oder computerbasiert.

Im Anschluss an den Schulbesuch werden die erfassten Daten im Nachgang aufbereitet und den Schulen zurückgemeldet. Die Rückmeldung erfolgt in allen Ländern grundsätzlich schriftlich durch die Zustellung eines entsprechenden

Ergebnisberichts. Diese unterscheiden sich in den Ländern sowohl in Umfang, Struktur und Detaillierungsgrad als auch mit Blick auf die rückgemeldeten Inhalte. In einigen Bundesländern gibt es zudem ein *Rückmeldegespräch* mit dem Evaluationsteam in zeitlichem Abstand zum Schulbesuch (verpflichtend oder freiwillig) sowie *Zielvereinbarungen* zwischen der Schulaufsicht und jeder evaluierten Schule – entweder in Form schriftlicher Vereinbarungen und/oder Bilanz- bzw. Zielvereinbarungsgespräche auf Basis der Evaluationsergebnisse zu Entwicklungszielen und geplanten Maßnahmen an der Einzelschule.

Bei besonderen Ergebnisauffälligkeiten sind in manchen Ländern Nachinspektionen möglich, sodass hier besonders entwicklungsbedürftige Schulen schon zu einem früheren Zeitpunkt als turnusmäßig geplant, in der Regel nach ein bis zwei Jahren, erneut besucht werden. Bayern verfolgt hingegen eine stärker entwicklungsorientierte Perspektive: Bei besonders positivem Abschneiden einer Schule in ausgewählten Prozessmerkmalen kann die Schule den Status einer MODUS-Schule erwerben (auf Antrag beim zuständigen Ministerium). Diese Schulen sind berechtigt, Weiterentwicklungsmaßnahmen zu erproben, insbesondere in den Arbeitsfeldern Unterrichtsentwicklung, Personalentwicklung und Personalführung sowie inner- und außerschulische Partnerschaften. Die Schulen mit zuerkanntem MODUS-Status dürfen von den Schulordnungen abzuweichen, soweit sichergestellt ist, dass die Lehrplanziele erreicht und den Schülerinnen und Schülern keine Nachteile bei Abschlüssen oder beim Erwerb schulischer Berechtigungen entstehen (vgl. Bayerisches Landesamt für Schule o. J.).

In der Gesamtschau ist Externe Evaluation sehr aufwendig und ressourcenintensiv. Gleichwohl waren mit der Einführung der Externen Evaluation große Erwartungen an die Qualitätssicherung und Qualitätsentwicklung verbunden: Genuines Ziel der Externen Evaluation ist es, Schulen über Rückmeldungen, die an landesspezifisch festgelegten Merkmalen von Schul- und Unterrichtsqualität ausgerichtet sind, bei der Schul- und Unterrichtsentwicklung zu unterstützen (*Entwicklungsfunktion*) – Rürup (2008, S. 476) spricht von „selbstgesteuerte[r] Qualitätsentwicklung durch die Konfrontation der Schule mit möglichst objektiven und validen Daten“.

Von Schulexternen werden gleichwohl lediglich Impulse und Entwicklungsanstöße für Maßnahmen zur Schul- und Unterrichtsentwicklung gegeben, die von den Einzelschulen in eigenständiger Weise umgesetzt und weiterverfolgt werden können. Darüber hinaus kann Externe Evaluation noch weitergehende Funktionen haben, die aber je nach Bundesland sehr unterschiedlich berücksichtigt werden, und zwar (vgl. Tarkian, Lankes et al. 2019):

- *Controlling*: Einige Bundesländer sehen eine aktive Einbindung der Schulaufsicht am Verfahren der Externen Evaluation im Sinne eines Prozess- und Ergebniscontrollings vor (z. B. Einflussnahme auf die Teilnahme, durch Begleitung der Schule im Prozess der Datenerhebung, Einsichtnahme in die schulspezifischen Evaluationsergebnisse, Mitwirkung am Rückmeldegespräch u.ä.). Üblich ist in fast allen Verfahren, dass die Schulaufsicht regulär den Ergebnisbericht aller evaluierten Schulen erhält; auch ein Bilanz- bzw. Zielvereinbarungsgespräche oder schriftliche Zielvereinbarungen zwischen der Schulaufsicht und den Schulen sind üblicherweise vorgesehen.
- *Monitoring*: Zusammengefasste – *nicht* auf eine einzelne Schule rückbeziehbar – Evaluationsergebnisse (z. B. landesweit oder regional aggregiert) können auch für eine datengestützte (regionale) Schulsystemsteuerung im jeweiligen Land vorgesehen werden. In Mecklenburg-Vorpommern werden z. B. zusammengefasste Teilergebnisse im Rahmen der Landesbildungsbericht-erstattung (vgl. Abschn. 5.1.3) veröffentlicht.
- *Information*: Durch Offenlegung des einzelschulischen Ergebnisberichts können verschiedene Akteursgruppen über die Qualität der jeweiligen Schule informiert werden. Durch *inerschulische Offenlegung*, etwa im Rahmen von Konferenzen oder auch in schriftlicher Form, werden die Mitglieder der Schule (z. B. Gremienmitglieder, Schülerinnen und Schüler sowie deren Eltern) und schulische Kooperationspartner über das Evaluationsergebnis informiert. Eine *Veröffentlichung über die schulischen Organisationsgrenzen hinaus* (z. B. auf der Schulhomepage) ist den Schulen zum Teil freigestellt, teilweise verpflichtend vorgesehen – oder auch explizit untersagt.

Insgesamt gibt es nur wenige empirische Studien zu den Wirkungen von Externer Evaluation – im Kern zeigt sich aber, dass eine auf die Ergebnisse der Externen Evaluation bezogene Schul- und Unterrichtsentwicklung häufig *nicht* stattfindet. Stattdessen finden häufig Maßnahmen *vor* und nicht nach der Externen Evaluation statt mit dem Ziel einer überzogenen positiven Selbstdarstellung der Schule (*window dressing*).

Unter Abwägung von Kosten und Nutzen haben mittlerweile mehrere Länder ihre Verfahren der Externen Evaluation nach vergleichsweise kurzer Laufzeit wieder vom Regelbetrieb ausgesetzt (z. B. Rheinland-Pfalz, Baden-Württemberg) oder durch ein alternatives Instrument ersetzt (Externe Evaluation auf Abruf in Hessen und Schleswig-Holstein). Nur in der Hälfte der Bundesländer gibt es noch landesweit verbindliche Verfahren zur Externen Evaluation (Regelbetrieb) (Tarkian, Lankes et al. 2019).

5.3 Evidenzbasierter Steuerungsanspruch

In diesem Kapitel wurden verschiedene Instrumente und Verfahren vorgestellt, die die Qualität im Schulwesen auf unterschiedlichen Ebenen sichern und weiterentwickeln sollen. Mit den Befunden der internationalen und nationalen Schulleistungsstudien, den Vergleichsarbeiten, mit den unterschiedlichen Varianten schulinterner und -externer Evaluation sowie mit bundes- und landesweiten sowie kommunalen Bildungsberichtssystemen ist eine *Vielzahl an Daten bzw. Informationen zum deutschen Bildungssystem*, zum Teil mit internationalen Vergleichsperspektiven, für unterschiedliche Adressaten und Akteure im Bildungswesen verfügbar.

- ▶ Mit dem Konzept der *Evidenzbasierung* ist die Hoffnung verbunden, das Handeln in Bildungspolitik und -administration sowie das Handeln schulischer Akteure durch bereitgestellte Daten effektiver und effizienter zu gestalten und so zur Entwicklung von Schule und Unterricht beizutragen.

Mit dem Begriff ‚Evidenz‘ ist *empirisch gesichertes Wissen* gemeint, also Befunde, die mit empirisch-wissenschaftlichen Methoden gewonnen wurden und mit denen Theorien bekräftigt oder widerlegt werden können (vgl. Bromme et al. 2016, S. 131). Es ist also vom eigenen Handlungs- und Erfahrungswissen schulischer Akteure zu differenzieren.

Im Sinne eines breiten Begriffsverständnisses verstehen Demski et al. (2012, S. 132) Evidenzen als „systematisch generierte, verobjektivierte und explizierte Informationen und Wissensbestände zur Wirksamkeit von Bildungsprozessen und ihren spezifischen Rahmenbedingungen“ und unterscheiden zwischen

- *Evidenzquellen im engeren Sinne* (eher formalisiert, eher ausgeprägte Wissenschaftsorientierung, eher externen Impuls wie z. B. Ergebnisse aus Schulleistungsstudien, Bildungsbericht, Rückmeldungen aus Vergleichsarbeiten und schulexterner Evaluation) und
- *Evidenzquellen im weiteren Sinne* (eher geringe explizite Wissenschaftsorientierung, eher interner Impuls, z. B. Ergebnisse der internen Evaluation, Schülerfeedback, Feedback aus kollegialen Hospitationen, innerschulisch durchgeführte Parallelarbeiten) (Demski et al. 2012, S. 140).

Zur Nutzung evidenzbasierter Wissensbestände durch schulische Akteure ist nur wenig bekannt. Ob und inwieweit gewonnene Daten auch genutzt werden, hängt

von vielfältigen Faktoren ab. Dazu gehören v.a. die Qualität der Daten und der Rückmeldung (z.B. Verständlichkeit, eingeschätzte Relevanz und Nützlichkeit), vorhandene Unterstützungsangebote zum professionellen Umgang mit den Daten sowie individuelle (z.B. Berufserfahrung, Erfahrungen mit Evidenzquellen) und organisationsstrukturelle und organisationskulturelle Merkmale, z.B. Schulleitungshandeln, Schulkultur (vgl. Demski 2017).

Demski (2017) konnte zeigen, dass prozessbezogene Informationsquellen, die einen *engen Bezug zu der eigenen Praxis im Unterricht* haben, von schulischen Akteuren vergleichsweise intensiv genutzt werden. Dabei werden von den Lehrkräften das Schülerfeedback und von den Mitgliedern der Schulleitung kollegiale unterrichtsbezogene Entwicklungsmaßnahmen besonders stark genutzt. Demgegenüber fällt die Nutzung von Instrumenten, mit denen ein *bildungspolitisch gesetzter expliziter Steuerungsanspruch* verbunden ist (Rückmeldungen aus Schulinspektionen/Externen Evaluationen, Vergleichsarbeiten und Schulleistungsvergleichen), deutlich geringer aus. Insofern zeigt sich, dass solche Evidenzquellen, mit denen die Ziele des Monitorings und der Rechenschaftslegung verknüpft sind und die einen verpflichtenden Charakter haben, nicht per se stärker von den schulischen Akteuren wahrgenommen und genutzt werden.

Eine häufig vorgebrachte Kritik ist, dass die zahlreichen Instrumente und Verfahren zur Qualitätssicherung und Qualitätsentwicklung lediglich deskriptives *Beschreibungswissen* im Sinne einer Beschreibung des Ist-Zustandes liefern, aber kaum *Erklärungswissen*, also Aussagen über mögliche Ursachen und Zusammenhänge, bereitstellen, welches wiederum Hinweise auf Ansatzpunkte für Veränderungen und letztlich *Handlungswissen* für praktische Schlüsselfragen der Bildungspolitik, Bildungsadministration und Schulpraxis liefert. Insofern hat die KMK in ihrer überarbeiteten Gesamtstrategie zum Bildungsmonitoring betont, dass es zukünftig stärker darum gehen muss, in noch größerem Maße Erklärungs- und anwendungsbezogenes Wissen für die Schulpraxis und die Bildungsverwaltung bereitzustellen (vgl. Kultusministerkonferenz 2015, S. 3 und S. 15). Es bleibt abzuwarten, ob dieser Anspruch auch erfüllt werden wird.

5.4 Anregungen zur Wiederholung und Reflexion

1. Stellen Sie zentrale Unterschiede zwischen Aufgaben aus Schulleistungstudien und aus Klassenarbeiten heraus.
2. Verschaffen Sie sich einen Überblick über die Maßnahmen, die in Folge der Schulleistungstudien auf der Ebene des Schulsystems in Deutschland implementiert wurden.

3. Welche Instrumente sind auf ein Systemmonitoring ausgerichtet, welche auf die Schul- und Unterrichtsentwicklung an der Einzelschule?
4. Machen Sie sich für ein von Ihnen selbstgewähltes Fach mit den KMK-Bildungsstandards für die Schulstufe, in der Sie später einmal unterrichten möchten, vertraut. Sofern für die Fächer, die Sie später unterrichten werden, keine Bildungsstandards seitens der KMK vorliegen, wählen Sie ein Fach, das Bezüge zu den von Ihnen studierten Unterrichtsfächern hat.
5. Inwiefern ist es wichtig, dass Lehrkräfte über forschungs- und evaluationsmethodische Kenntnisse verfügen?
6. Arbeiten Sie die zentralen Gemeinsamkeiten und Unterschiede zwischen den IQB-Bildungstrends und Vergleichsarbeiten heraus.
7. Recherchieren Sie im Internet exemplarisch einen Bildungsbericht aus Ihrem Bundesland (oder einem Nachbarland) oder einer Region in Ihrem Bundesland und vergleichen Sie diesen mit der Struktur des Nationalen Bildungsberichts.
8. Informieren Sie sich, welche Instrumente zur Beschreibung und Erfassung der Schul- und Unterrichtsqualität es in dem Bundesland, in dem Sie später einmal unterrichten möchten, gibt und verschaffen Sie sich einen ersten Überblick.
9. Erfassen Sie die Qualität Ihres eigenen Unterrichts (z. B. im Rahmen eines Praktikums oder einer Vertretungstätigkeit), etwa auf der Basis von EMU (Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung). Welche Konsequenzen können Sie aus den Ergebnissen ableiten?

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) 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 Kapitel 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.



DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

ub | universitäts
bibliothek

Dieser Text wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

DOI: 10.1007/978-3-658-43348-2_5

URN: urn:nbn:de:hbz:465-20240704-171001-2



Dieses Werk kann unter einer Creative Commons Namensnennung 4.0 Lizenz (CC BY 4.0) genutzt werden.