

## Kapitel 2

# Die großen Schulleistungsstudien in Deutschland

**Zusammenfassung** Deutschland beteiligt sich seit Mitte der 1990er Jahre an mehreren Schulleistungsstudien, die teils auf nationaler und teils auf internationaler Ebene die Leistungen von Schülerinnen und Schülern miteinander vergleichen. In diesem Kapitel werden sechs derzeit bundesweit durchgeführte Studien vorgestellt: drei internationale (PISA, TIMSS, IGLU) sowie drei nationale (IQB-Ländervergleich, VERA, NEPS). Herausgearbeitet werden insbesondere der jeweils zu Grunde gelegte Kompetenzbegriff sowie Design und Testkonzeption der Studien. Abschließend fasst eine tabellarische Übersicht die aktuell in Deutschland durchgeführten Schulleistungsstudien zusammen.

Im Zusammenhang mit den Ergebnissen zahlreicher Schulleistungsstudien, die mittlerweile fast jährlich auch in Deutschland veröffentlicht werden, fallen meist auch die Namen zweier Organisationen: die OECD sowie die IEA. Beide treten als Auftraggeber mehrerer internationaler Vergleichsstudien auf. So hat die *Organisation for Economic Co-operation and Development* (OECD) mit Hauptsitz in Paris unter anderem die PISA-Studie initiiert, aber auch internationale Untersuchungen mit der Zielgruppe Erwachsener, etwa TALIS (*Teaching and Learning International Survey*) zu den Bedingungen des Lehrens und Lernens in Schulen oder PIAAC (*Programme for the International Assessment of Adult Competencies*) zu grundlegenden Fähigkeiten und Fertigkeiten Erwachsener. Die *International Association for the Evaluation of Educational Achievement* (IEA) zeichnet beispielsweise für die *Trends in Mathematics and Science Study* (TIMSS) verantwortlich, in welcher Schülerleistungen in den Fächern Mathematik und Naturwissenschaften gegen Ende der Jahrgangsstufen vier und acht sowie am Ende der Sekundarstufe II (*TIMSS Advanced*) erfasst werden. Auch PIRLS (*Progress in International Reading Literacy Study*), in Deutschland *Internationale Grundschul-Lese-Untersuchung* (IGLU) genannt, wird von der IEA aufgelegt. PIRLS bzw. IGLU untersucht und vergleicht

das Leseverständnis von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe.

Vor der Teilnahme Deutschlands an TIMSS im Jahr 1995 galt jahrzehntelang als unbestritten, dass das deutsche Bildungswesen erfolgreich und qualitativ hochwertig ist (Baumert et al. 1997). Abgesehen von der Beteiligung Deutschlands an Vorgängerstudien von TIMSS in den 1960er bzw. 1980er Jahren (vgl. Wendt, Tarelli et al. 2012) war diese Annahme nicht oder nicht deutschlandweit empirisch überprüft worden. Als Folge der enttäuschend durchschnittlichen Ergebnisse der Schülerinnen und Schüler in Deutschland bei TIMSS (Baumert et al. 2000) verabschiedete die KMK (Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland) am 24. Oktober 1997 den sogenannten Konstanzer Beschluss (KMK 1997). Dort wurde unter anderem vereinbart, dass Deutschland künftig durch die Teilnahme an international vergleichenden Schulleistungsstudien regelmäßig vertiefende Befunde zu Kompetenzen von Schülerinnen und Schülern erhalten solle. Seither beteiligt sich Deutschland an mehreren regelmäßig stattfindenden national und international vergleichenden Studien, die in diesem Kapitel vorgestellt werden.

Die Teilnahme Deutschlands an nationalen wie internationalen Vergleichsstudien ist laut Beschluss der Kultusministerkonferenz von 2006 (KMK 2006) zentraler Bestandteil der Gesamtstrategie zum Monitoring des deutschen Bildungssystems und wurde seither mehrfach bekräftigt (vgl. etwa KMK 2015). Monitoring bedeutet dabei das regelmäßige Überprüfen und Bewerten bestimmter Indikatoren für Stärken und Schwächen im Bildungssystem, wozu unter anderem die Leistungsfähigkeit der Schülerinnen und Schüler an Schulen in Deutschland gehört. Die Gesamtstrategie der KMK zum Bildungsmonitoring beruht auf vier Säulen:

- Säule 1: Deutschland beteiligt sich an international vergleichenden Schulleistungsuntersuchungen, insbesondere PISA für die Sekundarstufe I sowie IGLU und TIMSS für die Primarstufe.
- Säule 2: Es werden bundeslandübergreifende, deutschlandweit verbindliche Bildungsstandards in Kernfächern für mehrere Schulabschlüsse definiert. Diese werden regelmäßig im Rahmen von Ländervergleichsstudien überprüft, die nach Möglichkeit an internationale Schulleistungsstudien angekoppelt sind.
- Säule 3: Flächendeckend werden Vergleichsarbeiten (VERA) in den Klassenstufen 3 und 8 durchgeführt, die ebenfalls einen klaren Bezug zu den länderübergreifenden Bildungsstandards haben und insbesondere Anregungen für kompetenzorientierte Unterrichtsentwicklung sowie die diagnostische Kompetenz von Lehrkräften liefern sollen.
- Säule 4: Es gibt eine gemeinsame Bildungsberichterstattung von Bund und Ländern.

Die nachfolgenden Abschnitte greifen die Säulen 1 bis 3 der Gesamtstrategie der KMK zum Bildungsmonitoring in Deutschland auf und stellen die aktuell durchgeführten Vergleichsstudien vor.

Deutschland beteiligt sich derzeit an mehreren national sowie international vergleichenden Schulleistungsstudien. In diesem Buch werden diejenigen beschrieben, die deutschlandweit durchgeführt werden, d. h. unter Beteiligung aller Bundesländer entweder zum innerdeutschen oder zum internationalen Vergleich. Daneben existiert eine Vielzahl weiterer Schulleistungsuntersuchungen, die jedoch regional begrenzt bzw. auf einen Teil der Bundesländer beschränkt bleiben; diese sind nicht Gegenstand dieses Buches. Zunächst werden im Folgenden die international vergleichenden Studien beschrieben, dann die nationalen. Auf internationaler Ebene beteiligt sich Deutschland an PISA, TIMSS und IGLU, während auf nationaler Ebene der IQB-Ländervergleich, VERA sowie NEPS deutschlandweit durchgeführt werden.

## **2.1 PISA: Das *Programme for International Student Assessment***

Mit PISA 2012 hat das *Programme for International Student Assessment* bereits die fünfte Erhebungsrunde abgeschlossen (Sälzer und Prenzel 2013). Nach dem Bekanntwerden der Ergebnisse der ersten PISA-Studie, die im Dezember 2001 der Öffentlichkeit vorgestellt wurden (Baumert et al. 2001), wurde mit dem sogenannten PISA-Schock ein Begriff geprägt, der noch immer sehr präsent ist und in bildungsbezogenen Debatten auch international mit dem unerwartet mittelmäßigen Abschneiden der Schülerinnen und Schüler in Deutschland verbunden wird. Neben den überraschend durchschnittlichen Ergebnissen im internationalen Vergleich trug auch die enorme Diskrepanz der Leistungsfähigkeit innerhalb Deutschlands zum PISA-Schock bei; markante Leistungsunterschiede zwischen Bundesländern, aber auch zwischen Jugendlichen mit und ohne Zuwanderungshintergrund oder aus unterschiedlich wohlhabenden Elternhäusern waren erstaunlich negative Befunde (Baumert et al. 2002; Baumert et al. 2001).

PISA ist international als zyklische Querschnittstudie angelegt, so dass in regelmäßigen Abständen Daten einer bestimmten Teilnehmergruppe erhoben und ausgewertet werden. An PISA sind jeweils alle aktuellen OECD-Staaten beteiligt sowie sogenannte OECD-Partnerstaaten. In PISA 2012 waren dies 34 OECD-Staaten sowie 31 Partnerstaaten (OECD 2014). Die Zielgruppe der PISA-Studie sind Jugendliche, die in ihrem jeweiligen Staat schon möglichst lange zur Schule gehen, diese jedoch noch nicht abgeschlossen haben. Da in den zuletzt mehr als 60 Teilnehmerstaaten (Sälzer und Prenzel 2013) die Festlegung der Schulpflicht bezüglich des Alters der Jugendlichen stark variiert, wurde das Alter der untersuchten Zielgruppe so gewählt, dass sich die Schülerinnen und Schüler möglichst am Ende ihrer Pflichtschulzeit befinden (Heine et al. 2013). Damit wurde staatenübergreifend festgelegt, dass an PISA Schülerinnen und Schüler im Alter von 15 Jahren teilnehmen sollen, die in ihrem jeweiligen Staat eine Schule besuchen. Fünfzehnjährige, die keine Schule besuchen, sind damit nicht Teil der in PISA fokussierten Untersuchungsgruppe.

Die Beteiligung Deutschlands an internationalen Vergleichsstudien wie PISA dient insbesondere dazu, die Leistungsfähigkeit der Schülerinnen und Schüler

am Ende der Sekundarstufe I international zu verankern. Seitens der KMK steht dabei besonders die Vergleichsperspektive im Mittelpunkt, die klar interpretierbare Trendinformationen über das deutsche Bildungssystem im internationalen Vergleich liefert. Auf dieser Grundlage kann über mehrere PISA-Runden abgeschätzt werden, inwieweit bildungspolitische Maßnahmen wirksam geworden sind und wo sich möglicherweise problematische Entwicklungen abzeichnen. PISA liefert ein ganzes Sortiment an Indikatoren für Bildungsergebnisse, die sich im internationalen Vergleich abbilden und einordnen lassen. Als umfangreiche Schulleistungsstudie stellt PISA die Frage, inwieweit es Staaten gelingt, junge Menschen in Schulen auf die Anforderungen des Erwachsenenlebens, das weitere Lernen über die gesamte Lebensspanne sowie Aspekte der gesellschaftlichen Teilhabe vorzubereiten (Sälzer und Prenzel 2013). Die Indikatoren, die aus PISA hervorgehen, müssen also entsprechende Informationen darüber bereitstellen, wie das Wissen und Können der getesteten Jugendlichen durchschnittlich ausgeprägt ist, aber auch, wie sich die Fähigkeiten und Fertigkeiten innerhalb von Staaten verteilen. Das bedeutet allerdings nicht, dass die in PISA gemessenen Kompetenzen zwingend in der Schule erworben worden sind, denn das geht aus den Befunden nicht hervor. PISA erfasst den Ist-Stand der durchschnittlichen Kompetenzen Jugendlicher und kann nicht differenzieren, wo diese erworben wurden. Um jedoch fundiert abschätzen zu können, unter welchen Bedingungen diese Kompetenzen erworben wurden, sind neben dem Kompetenzniveau an sich auch Merkmale der Schülerinnen und Schüler sowie ihres Lebens- und Lernumfelds bedeutsam. Mit ihrer Hilfe können beispielsweise systematische Zusammenhänge zwischen dem Kompetenzniveau und Merkmalen wie dem Geschlecht, der Schulbiographie, der sozialen Herkunft oder bestimmten Einstellungen entdeckt werden.

In Deutschland wurde die Teilnahme an PISA von Beginn an dazu genutzt, neben dem internationalen Vergleich auch auf nationaler Ebene vertiefenden Fragestellungen anhand einer Erweiterung der Schülerstichprobe sowie der eingesetzten Test- und Fragebogeninstrumente nachzugehen. Auf diese Weise konnte bereits in der ersten PISA-Erhebungsrunde im Jahr 2000 ein deutschlandweiter Leistungsvergleich zwischen den Bundesländern angestellt werden. Dabei wurde die international vorgegebene Zufallsstichprobe von fünfzehnjährigen Schülerinnen und Schülern erweitert um eine Stichprobe von Neuntklässlern. Da in Deutschland die meisten Schülerinnen und Schüler im Alter von 15 Jahren nach wie vor die Klassenstufe 9 besuchen, fiel die Wahl auf diese Stufe. Diese sogenannte PISA-Erweiterung (kurz: PISA-E) wurde neben der Teilnahme am internationalen PISA-Programm (kurz: PISA-I) von PISA 2000 bis PISA 2006 durchgeführt (Baumert et al. 2002; Prenzel et al. 2005; Prenzel et al. 2008). Für PISA-E und die darauf aufbauenden Vergleiche zwischen den Bundesländern ist eine sehr viel größere Stichprobe erforderlich als für die gesamtdeutsche Stichprobe, die in den internationalen Vergleich eingeht. Dies liegt unter anderem daran, dass die Repräsentativität bezüglich der vorhandenen Sekundarschularten in den verschiedenen Bundesländern eine gewisse Mindestanzahl an Schulen pro Schulform erfordert. Alle Schulen, die an PISA-I teilnahmen, waren auch in der Stichprobe für PISA-E (vgl. etwa Prenzel et al. 2005) und damit eine Teilmenge der Schulen, die am Bundesländervergleich beteiligt wa-

ren. Seit PISA 2009 ist PISA-E abgelöst durch den Ländervergleich des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) in Berlin, das im Auftrag der KMK Bildungsstandards für verschiedene Schulstufen und Fächer entwickelt und diese regelmäßig anhand von umfangreichen, repräsentativen Schülerstichproben in allen Bundesländern überprüft (vgl. Abschnitt 2.4).

Die Testaufgaben, die in PISA zur Erfassung verschiedener Kompetenzen zum Einsatz kommen, durchlaufen einen langen Entwicklungsprozess. Bereits mehrere Jahre vor einer PISA-Testung wird eine theoretische Rahmenkonzeption von internationalen Expertengruppen aufgesetzt, die als inhaltliche und strukturelle Grundlage für die Entwicklung von Testaufgaben dient. Im Gegensatz zu einigen anderen Studien wie etwa TIMSS, folgt PISA einem eigens definierten Kompetenzbegriff und strebt keine curriculare Validität an. Das bedeutet, dass die Aufgaben ohne Berücksichtigung nationaler Lehrpläne (Curricula) und allein unter Bezug auf die theoretische Rahmenkonzeption konstruiert werden. Es ist also durchaus beabsichtigt, dass manche Themen den Schülerinnen und Schülern noch unbekannt sind. Mehrere Entwurfs- und Überprüfungsstadien werden durchlaufen, in denen etwa die Relevanz für alle beteiligten Staaten, aber auch die sprachliche Übertragbarkeit in verschiedene Staaten und Kulturen überprüft wird (Sälzer und Prenzel 2013). Immer ein Jahr vor der anstehenden PISA-Erhebungsrunde erfolgt ein Feldtest, in dessen Rahmen sowohl bestimmte Prozeduren der Stichprobenziehung und der Testdurchführung als auch die psychometrischen Eigenschaften der Testaufgaben erprobt werden. Letztlich werden nur Aufgaben in der PISA-Haupterhebung verwendet, die eine ausreichende internationale Vergleichbarkeit und zufriedenstellende Messeigenschaften aufweisen. In Ergänzung zu diesen Leistungstests werden in PISA zahlreiche Merkmale der teilnehmenden Jugendlichen sowie ihres häuslichen und schulischen Umfeldes erfasst. Neben einem Schul- und Schülerfragebogen besteht auch die Möglichkeit, einen Eltern- sowie einen Lehrerfragebogen einzusetzen, um möglichst viele Erkenntnisse zu den Entstehens- und Entwicklungsbedingungen von Kompetenzen gewinnen zu können (vgl. etwa Schiepe-Tiska und Schmidtner 2013).

### ***2.1.1 Kompetenzbegriff in PISA: Literacy***

Die drei in PISA untersuchten Kompetenzbereiche, sogenannte Domänen, sind in jeder Erhebungsrunde Lesen, Mathematik und Naturwissenschaften (vgl. z. B. OECD 1999). Im Wechsel ist dabei jede Domäne einmal in drei Erhebungsrounden die Schwerpunktdomäne, das heißt, etwa die Hälfte der eingesetzten Testaufgaben stammt aus dem Bereich dieser Schwerpunktdomäne. Jeweils ungefähr ein Viertel der Testaufgaben wird den beiden jeweiligen Nebendomänen zugeordnet. Ergänzt werden diese drei Domänen jeweils von einer sogenannten übergreifenden Kompetenz, wie etwa dem selbstregulierten Lernen (Baumert et al. 2001), dem Problemlösen (vgl. etwa Leutner et al. 2004; Leutner et al. 2004) oder der Vertrautheit

mit dem Computer. Ein Teil der an PISA teilnehmenden Schülerinnen und Schüler bearbeitet zusätzlich Aufgaben aus dieser fächerübergreifenden Domäne.

Bei der Definition von Kompetenz im Rahmen der PISA-Studie geht die OECD vom sogenannten *Literacy*-Begriff als Konzept einer funktionalen Grundbildung im Sinne einer Allgemeinbildung aus (vgl. etwa OECD 2013a). Bereits seit der ersten Erhebungsrunde im Jahr 2000 wird das *Literacy*-Konzept in der theoretischen Rahmenkonzeption (OECD 1999) definiert als „*knowledge and skills for adult life*“ (S. 7) und damit als Wissen und Fertigkeiten, die für das vor den Schülerinnen und Schülern liegende Erwachsenenleben benötigt werden. Dabei ist die Ausprägung der *Literacy*, im Deutschen auch mit Literalität bezeichnet, als ein Kontinuum gedacht. Damit ist gemeint, dass *Literacy* bei verschiedenen Personen in verschiedenem Maße vorhanden ist, wobei es keinen festgelegten Mindestpunkt gibt, der eine Person mit *Literacy* von einer Person ohne trennt. Vielmehr erlaubt das *Literacy*-Kontinuum einen Vergleich der momentan vorhandenen Grundbildung zwischen mehreren Personen – beziehungsweise wie bei PISA, der durchschnittlich vorhandenen *Literacy* von Fünfzehnjährigen in mehreren Staaten.

Die Teilnehmerinnen und Teilnehmer der PISA-Studie haben im Alter von fünfzehn Jahren selbstverständlich noch nicht alles an Wissen und Fertigkeiten erworben, was sie als Erwachsene brauchen werden (OECD 1999). Das *Literacy*-Konzept, wie es in PISA angewendet wird, geht daher von einem Niveau an Grundkompetenzen aus, das eine gute Prognose für künftiges, lebenslanges Lernen zulässt. Insofern ist *Literacy* als Grundbildung in PISA sowohl funktional als auch anschlussfähig. Funktional heißt hier anwendbar in Bezug auf die momentane und zukünftige gesellschaftliche Teilhabe und anschlussfähig ist im Sinne einer vielversprechenden Grundlage für das kontinuierliche Weiterlernen über die gesamte Lebensspanne gemeint (Sälzer und Prenzel 2013). Diese Fokussierung auf die Anwendbarkeit und die Anschlussfähigkeit ist für die Messung von Bildungsergebnissen in einem bestimmten Alter sehr relevant, denn zum einen bedeutet der Erwerb anwendbaren Wissens, dass Jugendliche bis zum PISA-Testtag schulische wie außerschulische Lerngelegenheiten genutzt haben, um teilhabe- und handlungsfähig zu werden (im Englischen: *literate*). Zum anderen ist der Kenntnisstand in den drei untersuchten Domänen der Mittelpunkt des Interesses, da das Wissen und Können in diesen Bereichen so definiert und ausgewählt wurde, dass man bei gutem Abschneiden im PISA-Test eine solide Grundlage dafür hat, sich lernend weiterentwickeln zu können (Sälzer und Prenzel 2013).

PISA stellt explizit keinen Bezug zu nationalen Curricula her, sondern strebt in Form einer theoretischen Rahmenkonzeption einen internationalen Konsens darüber an, welche Domänen Teil der funktionalen Grundbildung sind und wie diese Domänen jeweils inhaltlich strukturiert sein sollen (OECD 2013a). Damit ist dem Anspruch Rechnung getragen, dass sowohl die Lesekompetenz als auch die Mathematik und die Naturwissenschaften als Kompetenzbereich Anwendungsfelder haben, die deutlich über einzelne Schulfächer hinausreichen. Damit kommt der Begriff *Literacy* dem Konzept einer grundlegenden Allgemeinbildung nahe (vgl. Tenorth, 2004; 2005), auch wenn es sich dabei stets nur um einen Ausschnitt aus dem Spektrum allgemeiner Bildung handeln kann (Sälzer und Prenzel 2013). Lesekompetenz,

Mathematik und Naturwissenschaften bilden als die drei Untersuchungsdomänen und PISA gleichsam eine Stichprobe aus den Inhalts- und Kompetenzbereichen, die weltweit in Schulen vermittelt werden und die als relevant für die weitere Bildungsbiographie der Fünfzehnjährigen sowie ihr Berufsleben und ihre Fähigkeit zur Teilhabe an Kultur und Gesellschaft gelten (ebd.).

Kompetenz in den drei untersuchten Domänen ist als hierarchisches Modell gedacht, das in Form von Kompetenzstufen unterteilt werden kann. Diese Kompetenzstufen werden zum einen durch erzielte Leistungskennwerte (d. h. Punkte auf einer PISA-Kompetenzskala) und zum anderen auf der Basis inhaltlicher Kriterien definiert. So lässt sich eine Kompetenzstufe beispielsweise anhand von Aufgabenanforderungen beschreiben, die typischerweise von Schülerinnen und Schülern auf dieser Kompetenzstufe bewältigt werden. Damit zeigen Kompetenzstufen neben einem Wertebereich auf der Kompetenzskala zugleich auf, was diese Punktwerte auf der PISA-Skala inhaltlich bedeuten. Gleichzeitig ist den Kompetenzstufen zu entnehmen, woran Schülerinnen und Schüler auf einer Stufe mit einer gewissen Wahrscheinlichkeit scheitern würden. Auf dieser Basis kann von den teilnehmenden Bildungssystemen jeweils unter Bezug auf die nationalen Bildungsziele abgeschätzt werden, inwieweit diese erreicht wurden und wo gegebenenfalls Handlungsbedarf besteht (Sälzer und Prenzel 2013). In Deutschland traf dies beispielsweise in PISA 2000 für die Lesekompetenz zu: Fast ein Viertel der Fünfzehnjährigen in Deutschland konnte lediglich auf einem elementaren Niveau lesen und der Abstand zwischen den Ergebnissen der leistungsstärksten Schülerinnen und Schüler und denen der leistungsschwächsten war breiter als in allen weiteren Teilnehmerstaaten (Baumert et al. 2001). Dieses und weitere Ergebnisse nahm die Kultusministerkonferenz zum Anlass, eine Gesamtstrategie zum Bildungsmonitoring (KMK 2006) sowie sieben zentrale Handlungsfelder (KMK 2002) zu entwickeln, um die Bildungsergebnisse und damit letztlich die Zukunftsprognose der Jugendlichen gegen Ende ihrer Pflichtschulzeit zu verbessern.

Beispielhaft sei hier die Definition der mathematischen Grundbildung (*Mathematical Literacy*) genannt, wie sie die OECD für PISA 2012 in der theoretischen Rahmenkonzeption umschreibt (OECD 2013a). *Mathematical Literacy* wird dort definiert als „die Fähigkeit einer Person, Mathematik in zahlreichen Kontexten anzuwenden, zu interpretieren und Formeln zu verwenden. Dazu gehört mathematisches Schlussfolgern ebenso wie die Anwendung mathematischer Konzepte, Vorgehensweisen, Fakten und Werkzeuge, um Phänomene zu beschreiben, zu erklären und vorherzusagen. Mathematische Grundbildung hilft Personen, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und Mathematik in einer Weise zu verwenden, die den Anforderungen des Lebens dieser Person als konstruktivem, engagiertem und reflektiertem Bürger entspricht“ (OECD 2013a, S. 25). Diese Vorstellung mathematischer Grundbildung wird über das Design und die Testkonzeption, die PISA zugrunde liegt, in Testaufgaben überführt und damit messbar gemacht (vgl. Abschnitt 2.1.2).



### ***2.1.2 Design und Testkonzeption von PISA***

PISA ist international mit einem querschnittlichen Design angelegt: Das heißt, dass jede Teilnehmerin und jeder Teilnehmer genau einmal den PISA-Test bearbeitet. Somit gibt es nur einen Messzeitpunkt, weshalb gefundene Zusammenhänge unter anderem nicht kausal zu interpretieren sind (Sälzer und Prenzel 2013). Das internationale Design der PISA-Studie erfordert neben der Testkomponente auch die Administration zweier Fragebögen: Einen für die beteiligten Schülerinnen und Schüler sowie einen für die Schulleitungen. Darüber hinaus werden in Deutschland regelmäßig Fragebögen für Eltern sowie für Lehrkräfte eingesetzt.

Ein wesentlicher Aspekt des internationalen Erhebungsdesigns von PISA ist die sogenannte Aggregation von Daten, die auf der Ebene von Individuen (hauptsächlich Schülerinnen und Schülern) gewonnen werden. Das bedeutet, dass anhand von PISA Aussagen nicht auf individueller Ebene gemacht werden, sondern auf den Aggregationsebenen „Schule“ und „Bildungssystem“ (Sälzer und Prenzel 2013). Die Leistungen der Schülerinnen und Schüler werden also auf der Ebene der Schule bzw. des Bildungssystems zusammengefasst, so dass pro Teilnehmerstaat durchschnittliche Werte für die Ergebnisse resultieren. Dementsprechend ist das internationale Stichprobendesign so angelegt, dass in jedem teilnehmenden Bildungssystem zunächst nach dem Zufallsprinzip Schulen gezogen werden und innerhalb der Schulen wiederum zufallsbasiert fünfzehnjährige Schülerinnen und Schüler. Eine detaillierte Beschreibung des Stichprobendesigns findet sich beispielsweise im nationalen Berichtsband zu PISA 2012 (Heine et al. 2013). Auf diese Weise wird berücksichtigt, dass jede Schülerin und jeder Schüler nicht alle verfügbaren Testaufgaben bearbeiten kann, da Zeit und Konzentrationsfähigkeit begrenzt sind.

Durch die Aggregation und damit die Zusammenfassung der Schülerleistungen auf einer höheren Ebene ist es nicht notwendig, dass jeder an PISA beteiligte Schüler jede vorhandene Testaufgabe bearbeitet hat. Vielmehr ermöglicht der Einsatz eines sogenannten Multi-Matrix-Designs (vgl. auch Abschnitt 4.2) die Zuweisung einer Auswahl von Aufgaben zu den einzelnen Schülerinnen und Schülern. Ein solches Design erlaubt eine möglichst ökonomische Erfassung von Daten, die in mehreren umfangreichen Inhaltsbereichen belastbare Aussagen über die in verschiedenen Bildungssystemen durchschnittlich erreichten Kompetenzen zulassen (Prenzel et al. 2004; Carstensen et al. 2004). In einem Multi-Matrix-Design werden die verschiedenen Testaufgaben zu Aufgabenblöcken (auch Cluster genannt) gruppiert. In PISA sind diese Aufgabenblöcke stets „sortenrein“, d. h. ein Aufgabenblock enthält ausschließlich Aufgaben aus einer der drei untersuchten Domänen. Die Aufgabenblöcke werden systematisch variiert auf mehrere Testhefte verteilt, wobei die Schülerinnen und Schüler jeweils zufällig ein bestimmtes Testheft zugewiesen bekommen (als Rotation bezeichnet). Bis einschließlich PISA 2012 wurden diese Testhefte als Papier-und-Bleistift-Test administriert, mit PISA 2015 findet eine Umstellung auf computerbasiertes Testen statt. Auch bei diesem Modus erfolgt eine Rotation.

In PISA 2012 wurden beispielsweise 13 Testhefte eingesetzt, die aus jeweils 4 Aufgabenblöcken im Umfang von jeweils 30 Minuten Bearbeitungsdauer bestan-



den. Jeder Aufgabenblock kam in mehreren Testheften und dort an jeweils unterschiedlicher Position vor (Heine et al. 2013). Ein Teil der Testhefte enthielt dabei Aufgabenblöcke aus allen drei Domänen, andere hingegen lediglich aus zwei Domänen. Die Aufgaben der Hauptdomäne bestehen zu etwa einem Drittel aus sogenannten Link-Items, d. h. Aufgaben, die bereits in früheren PISA-Erhebungsrunden eingesetzt worden waren und eine Verankerung (Verlinkung) der jeweils aktuellen Erhebungsrunde an früheren Runden ermöglichen, sowie etwa zwei Drittel neu entwickelten Aufgaben. Die beiden Nebendomänen werden ausschließlich anhand von Link-Items getestet (Heine et al. 2013). Auf der Basis der Link-Items können Veränderungen in den Leistungsergebnissen über mehrere Erhebungsrunden hinweg sichtbar gemacht werden. Nach den insgesamt 120 Minuten reiner Testzeit, die von einer Pause nach der ersten Hälfte unterbrochen werden, bearbeiten die Schülerinnen und Schüler den Schülerfragebogen während ungefähr 45 Minuten.

Ein Multi-Matrix-Design sieht vor, dass jeder Untersuchungsteilnehmer nur einen Teil aller verfügbaren Testaufgaben bearbeitet. Allerdings kann es dann vorkommen, dass die Auswahl der Aufgaben, die verschiedenen Schülerinnen und Schülern zugewiesen wird, unterschiedlich schwierig ist. Damit wären auch die erzielten Testleistungen nicht ohne weiteres miteinander vergleichbar. Um die Antworten auf unterschiedlich schwierige Aufgaben dennoch miteinander vergleichbar zu machen, werden die Daten skaliert (z. B. OECD 2012). Das bedeutet, dass für einen fairen und objektiven Vergleich der Kompetenz zwischen verschiedenen Schülern auch die jeweilige Schwierigkeit der unterschiedlichen Aufgaben berücksichtigt wird, wenn ein individueller Kompetenzwert geschätzt wird (Heine et al. 2013). Dafür sind eine ausreichend große Stichprobe (in der Regel mindestens vierstellig) sowie Überlappungen der Testheftinhalte notwendig (vgl. etwa Kolen und Brennan 2004).

Die Testkonzeption in PISA wird seit der ersten Erhebungsrunde in einer theoretischen Rahmenkonzeption festgehalten (z. B. OECD 1999; OECD 2013a). Dort wird beschrieben, auf welchen Konsens sich das international besetzte Expertenkonsortium bezüglich der zu erfassenden Grundbildung geeinigt hat. Im oben beschriebenen *Literacy*-Begriff wird dies deutlich. Es handelt sich bei der theoretischen Rahmenkonzeption um eine inhaltliche Festlegung, was unter der jeweiligen *Literacy* zu verstehen ist. Am Beispiel der mathematischen Grundbildung, die in PISA 2012 die Schwerpunktdomäne war, ist die Testkonzeption etwa folgendermaßen aufgebaut: Zunächst wird definiert, was unter *mathematical literacy* oder mathematischer Grundbildung zu verstehen ist (vgl. Abschnitt 2.1.1). Strukturiert wird die mathematische Grundbildung mit ihren zahlreichen Aspekten anschließend anhand der Elemente *Inhalte*, *Prozesse* sowie *Kontexte* (Sälzer, Reiss et al. 2013). Zunächst werden vier Inhaltsbereiche unterschieden, Veränderung und Beziehungen, Raum und Form, Quantität sowie Unsicherheit und Daten. Der Inhaltsbereich Veränderung und Beziehungen kommt dabei dem schulischen Themengebiet der Algebra nahe und umfasst vorwiegend funktionale und relationale Beziehungen zwischen mathematischen Objekten. Raum und Form entspricht weitestgehend dem Themengebiet der Geometrie, während mit Quantität die Verwendung von Zah-

len zur Strukturierung und Beschreibung von Situationen gemeint ist (Sälzer, Reiss et al. 2013). Unsicherheit und Daten sind der vierte Inhaltsbereich; sie haben als Schwerpunkte den Umgang mit statistischen Daten und Zufallsphänomenen. Verknüpft werden diese vier Inhaltsbereiche in der theoretischen Rahmenkonzeption mit drei mathematischen Prozessen. Diese Prozesse fokussieren grundlegende mathematische Tätigkeiten, die als zentral für eine mathematische Grundbildung ausgewählt wurden (OECD 2013a): (1) Situationen mathematisch formulieren, (2) mathematische Konzepte, Fakten, Prozeduren und Schlussfolgerungen anwenden und (3) mathematische Ergebnisse interpretieren, anwenden und bewerten. Als drittes Strukturierungselement mathematischer Grundbildung werden bestimmte Kontexte definiert, in denen mathematische Probleme verankert werden. In PISA 2003 wurden diese Kontexte mit Situationen bezeichnet (OECD 2004). Diese Kontexte bestimmen zu großen Teilen, welche mathematischen Strategien die Schülerinnen und Schüler jeweils wählen können (oder müssen), um zu einer adäquaten Lösung zu kommen. Dazu gehören beispielsweise Elemente der Lebenswelt der Jugendlichen, die eine direkte Verbindung zum Alltag schaffen und individuelle Interessen der Fünfzehnjährigen ansprechen sollen (OECD 2013a). Damit durch die Kontexte, in denen die Testaufgaben gestellt werden, eine möglichst große Bandbreite an Interessen abgedeckt ist, werden vier Kategorien unterschieden: persönliche, berufliche, gesellschaftliche und wissenschaftliche Kontexte. Dabei zielen persönliche Kontexte auf das direkte Umfeld der Fünfzehnjährigen ab, wie z. B. ihre Familie, ihre Freunde oder ihre Schulklasse. Typische Themen in diesem Kontext sind etwa Einkaufen, Gesundheit, Sport und Freizeit oder Finanzplanung. Berufliche Kontexte betreffen die Arbeitswelt und behandeln Themen wie etwa Dienstpläne, Buchhaltung oder Entscheidungsfindungen. Inhalte des gesellschaftlichen Kontexts sind in sozialen Gefügen im Umfeld von Individuen eingebettet und können lokal, national oder auch global angelegt sein. Mögliche Themen sind demographische Entwicklungen, Wahlsysteme oder Werbung. Wissenschaftliche Kontexte schließlich erfordern die Anwendung von Mathematik auf naturwissenschaftliche oder technologische Themen, wie etwa Wetter und Klima, Medizin, Raumfahrt oder Genetik (Sälzer, Reiss et al. 2013). Anhand der Kontexte wird in PISA angestrebt, mathematische Kompetenz auf möglichst lebensnahe Gegebenheiten zu beziehen, an die die Fünfzehnjährigen in den teilnehmenden Staaten anknüpfen können und die für sie relevant sind. Eine hundertprozentige Übereinstimmung der Lebensnähe in verschiedenen Staaten ist dabei weder notwendig noch sinnvoll. Vielmehr ist davon auszugehen, dass jeder Fünfzehnjährige, der an PISA teilnimmt, einen Teil der Themen in den Testaufgaben noch nicht kennt und auf Kontexte mit höherer und niedrigerer persönlicher Relevanz trifft. Ausschlaggebend ist schlussendlich die Bandbreite an Inhalten, Prozessen und Kontexten, die insgesamt durch alle Aufgaben (und über alle Schülerinnen und Schüler hinweg) abgedeckt wird.

Neben der Mathematik, die hier beispielhaft in ihrer konzeptionellen Struktur dargestellt wurde, werden in PISA auch die weiteren Domänen (Lesekompetenz sowie Naturwissenschaften) aufbereitet und definiert. Dazu gehören auch die jeweils übergreifenden Kompetenzen wie Selbstregulation oder Problemlösen (vgl. z. B. OECD 1999; OECD 2003; OECD 2013a).

## 2.2 TIMSS: Die Trends in International Mathematics and Science Study

Die von der IEA initiierte *Trends in Mathematics and Science Study* (TIMSS) erfasst die Fachleistungen von Schülerinnen und Schülern in den Bereichen Mathematik und Naturwissenschaften. Eine Besonderheit von TIMSS ist, dass die Studie mehrere Jahrgangsstufen von Schülerinnen und Schülern untersucht. TIMSS an Grundschulen ging aus der *Third International Mathematics and Science Study* (TIMS) hervor, die als erste TIMS-Studie in Deutschland im Jahr 1995 mit Schülerinnen und Schülern der Sekundarstufe I und II durchgeführt wurde (Baumert et al. 2000). Auch damals hatte es auf internationaler Ebene einen Grundschulteil gegeben, an dem Deutschland allerdings nicht teilnahm. Erstmals auf der Grundschulstufe beteiligte sich Deutschland im Jahr 2007 an TIMSS (vgl. Bos, Bonsen et al. 2008) und mit TIMSS 2011 erfolgte die Grundschuluntersuchung zum zweiten Mal (Bos et al. 2012). TIMSS wird seit 1995 alle vier Jahre durchgeführt, TIMSS *Advanced* mit Schülerinnen und Schülern am Ende der Sekundarstufe fand bislang dreimal statt (1995, 2008 und 2015). Die derzeit aktuellsten verfügbaren Ergebnisse stammen aus TIMSS 2011. Mit Bezug auf nationale Curricula und weitere Lehr- und Lernbedingungen untersucht TIMSS 2011 die Fertigkeiten von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe, die sie im Rahmen von Aufgaben zur Mathematik und den Naturwissenschaften zeigen. Zentral ist dabei die Idee, grundlegende Kompetenzen zu messen, die nach vier Jahren Schulbesuch erworben wurden; dies ist verbunden mit der Frage, ob diese Kompetenzen sich beim Vergleich der Schülerinnen und Schüler systematisch unterscheiden. Beispiele hierfür sind etwa Unterschiede zwischen Mädchen und Jungen oder bezogen auf die soziale Herkunft oder einen Zuwanderungshintergrund.

An TIMSS 2011 waren weltweit insgesamt 59 Bildungssysteme beteiligt. 50 Staaten oder Regionen führten TIMSS als reguläre Teilnehmer durch, die übrigen unter gesonderten Teilnahmebedingungen als sogenannte Benchmark-Teilnehmer oder mit einer eingeschränkten Stichprobe (Wendt, Bos et al. 2012). Die Rahmenkonzeption, auf deren Grundlage die Aufgaben entwickelt werden, wurde von einer international besetzten Expertengruppe erarbeitet. Neben den Leistungen der Schülerinnen und Schüler werden auch Merkmale der Schülerinnen und Schüler selbst, ihrer Lehrkräfte und der Lernumgebungen erfasst, so dass der Erwerb und die Bedingungen für die Entwicklung mathematischer und naturwissenschaftlicher Kompetenz adäquat nachgezeichnet und analysiert werden können.

Auf konzeptueller Ebene legt TIMSS ein sogenanntes Curriculum-Modell zu Grunde. Dieses Modell unterscheidet drei Ebenen: das Bildungssystem, Schule und Klassenzimmer sowie die Schülerinnen und Schüler (Mullis et al. 2009). Das Modell wurde bereits in früheren TIMSS-Erhebungsrunden verwendet (u. a. Baumert et al. 2000). Gemäß dem Curriculum-Modell hat jedes Bildungssystem ein eigenes, intendiertes Curriculum. Dies steht dem Anspruch einer internationalen Vergleichbarkeit zunächst einmal entgegen, da sich die Curricula verschiedener Staaten vielfach deutlich voneinander unterscheiden. Solche Unterschiede betreffen neben den

Inhalten auch den Zeitpunkt in der Schullaufbahn, zu dem diese gelehrt werden. Zur Erfassung dieses intendierten Curriculums pro Staat werden Experten gebeten, in einem entsprechenden Fragebogen die vorgesehenen mathematischen und naturwissenschaftlichen Inhalte und Prozesse einzutragen, die gemäß curricularer Vorgaben gelernt werden sollen (Wendt, Tarelli et al. 2012). Die an TIMSS teilnehmenden Schülerinnen und Schüler erhalten während des Tests Aufgaben aus einem gemeinsamen Aufgabenpool, der die Vergleichbarkeit von Leistungen zwischen verschiedenen Staaten wiederum ermöglicht. Indem die von den Schülerinnen und Schülern erbrachten Leistungen im jeweiligen curricularen Kontext interpretiert werden, gelingt die Verbindung eines staatenübergreifenden Vergleichs auf der Grundlage gemeinsamer Testaufgaben mit einer curricular verankerten Interpretation auf der Ebene der Teilnehmerstaaten.

Die Ebene der Schule bzw. des Klassenzimmers ist der Ort des sogenannten implementierten Curriculums: Die Umgebung sowie die Organisation einer Schule beeinflussen, inwieweit curriculare Ziele dort umgesetzt und erreicht werden können (Mullis et al. 2009). TIMSS konzentriert sich deshalb auf eine Reihe von Indikatoren, von denen als gesichert gilt, dass sie der Erreichung curricularer Ziele zuträglich sind. Beispielsweise werden Angaben dazu erhoben, wie groß eine Schule ist, wo sie liegt und wie sie ausgestattet ist; auch Aspekte des Schulklimas, des Schulmanagements, der Elternarbeit sowie des Lehrerkollegiums werden berücksichtigt. Besonders relevant sind im Rahmen der TIMSS-Studie Fragen dazu, wie curriculare Inhalte und Vorgaben auf dem Niveau der einzelnen Klasse umgesetzt werden – welcher Lernstoff also wirklich im Unterricht behandelt wird. Durch den expliziten Bezug zum jeweiligen Curriculum der beteiligten Bildungssysteme wird eine Schnittmenge zwischen den gemessenen Kompetenzen in Mathematik und Naturwissenschaften und den Curricula erreicht, die eine gezielte Einordnung der erbrachten Schülerleistungen in die nationalen Anforderungen der Curricula für die jeweilige Altersstufe erlaubt.

Die dritte Ebene, Schülerinnen und Schüler, steht für das sogenannte erreichte Curriculum. Durch die Bearbeitung der TIMSS-Testaufgaben sowie eines Schülerfragebogens zeigen die Schülerinnen und Schüler, was über den Unterricht (bzw. das implementierte Curriculum) von den vorgesehenen Inhalten und Prozessen (also dem intendierten Curriculum) bei ihnen angekommen ist. Neben dem Gelernten werden auch bestimmte Haltungen und Einstellungen der Schülerinnen und Schüler zum jeweiligen Fach erhoben (Wendt, Tarelli et al. 2012).

Die Auswahl der untersuchten Population in TIMSS orientiert sich daran, dass die getesteten Schülerinnen und Schüler in allen Staaten zum Zeitpunkt der Datenerhebung möglichst bereits gleich lange zur Schule gehen. Eines der zentralen Ziele ist die Entwicklung von empirisch fundierten Empfehlungen für die Bildungspolitik, die sich konkret auf schulische und unterrichtliche Praktiken beziehen. Diese hängen oft eng mit der jeweiligen Klassenstufe oder auch mit der besuchten Schulart zusammen. Um möglichst vollständige Schülerjahrgänge miteinander vergleichen zu können, wird als erstes Kriterium die sogenannte ISCED-Klassifizierung bestimmt. ISCED steht für *International Standard Classification of Education* und ist in verschiedene Levels unterteilt (UNESCO Institute of Statistics 2006). Als Zielpo-

pulation für TIMSS wurde festgelegt, dass sich die Teilnehmer im vierten Schuljahr befinden sollen (Wendt, Tarelli et al. 2012), was in den meisten Teilnehmerstaaten der Jahrgangsstufe 4 entspricht. Als zweites Kriterium wurde ergänzt, dass die teilnehmenden Schülerinnen und Schüler im Mittel 9 Jahre und 6 Monate alt sein sollen, damit die eingesetzten Leistungstests sowie die Durchführung der Testsitzung möglichst entwicklungsgerecht sind (Mullis et al. 2009).

Ähnlich wie bei PISA ist bei TIMSS darauf zu achten, dass die eingesetzten Testaufgaben einen Vergleich zwischen unterschiedlichen Bildungssystemen zulassen. Dazu gehört eine gründliche Überprüfung sowohl sprachlicher als auch psychometrischer Eigenschaften der Aufgaben. Psychometrische Eigenschaften umfassen Aspekte psychologischen Messens, also etwa die Genauigkeit oder die Fairness von Aufgaben. Bei TIMSS kommt im Gegensatz zu PISA der Anspruch curricularer Validität hinzu, so dass der Grad der Übereinstimmung zwischen den in TIMSS eingesetzten Tests und den national spezifischen Lehrplänen von Staat zu Staat unterschiedlich ist (Wendt, Tarelli et al. 2012). Um mit dieser Unterschiedlichkeit umgehen zu können, erfolgt zu Beginn eines TIMSS-Zyklus eine sogenannte *Test-Curriculum Matching Analysis* (TCMA; Mullis et al. 2009), in deren Rahmen alle beteiligten Bildungssysteme durch Experten jede Testaufgabe bezüglich ihrer Validität oder Passung für das jeweilige nationale Curriculum beurteilen. Diese Einschätzung wird dann von den nationalen Studienleitungen dazu verwendet, die eigenen Ergebnisse der TIMSS-Erhebung in Bezug auf das nationale Curriculum zu interpretieren. Hierbei kann vergleichend folgender Frage nachgegangen werden: Wie fallen die Ergebnisse eines Staates in Form von der durchschnittlichen Lösungswahrscheinlichkeit der einzelnen Aufgaben im Vergleich zu allen anderen Teilnehmern aus, wenn ausschließlich die Aufgaben ausgewertet werden, die einen direkten Bezug zum nationalen Curriculum haben (Mullis et al. 2009)? Wie auch in PISA wird in TIMSS eine Reihe von Fragebögen eingesetzt, die der Erfassung persönlicher Merkmale wie z. B. Einstellungen, aber auch der Erhebung von Kontextmerkmalen dient. In TIMSS 2011 wurden neben einem Schülerfragebogen auch Fragebögen für Eltern, für die Mathematik- und Sachunterrichtslehrkräfte, die Schulleitung sowie ein Curriculumfragebogen vorgelegt (Wendt, Tarelli et al. 2012).

### 2.2.1 Kompetenzbegriff in TIMSS

In TIMSS stehen mit der Mathematik und den Naturwissenschaften zwei Kompetenzdomänen im Mittelpunkt. Wie auch bei PISA werden beide Domänen in Form einer theoretischen Rahmenkonzeption beschrieben und spezifiziert, so dass auf dieser Basis Aufgaben zur Erfassung der mathematischen und naturwissenschaftlichen Kompetenz entwickelt werden können. Beide Domänen werden in zwei Dimensionen untergliedert: eine inhaltliche und eine kognitive (Mullis et al. 2009). Die inhaltliche Dimension umfasst grundlegende Stoffgebiete (Inhaltsbereiche), während sich die kognitive auf fachspezifische Denkprozesse (sogenannte kognitive Anforderungen) bezieht, die in den Testaufgaben umgesetzt werden (Mullis et al. 2009).

Die kognitiven Anforderungen werden in unterschiedliche, hierarchisch angeordnete Niveaus unterteilt. Auf diese Weise können die Schülerkompetenzen in beiden Domänen differenziert erfasst und beschrieben werden. Zudem können in Verbindung mit dieser relativ allgemein gehaltenen Unterteilung gezielt hinreichende Schnittmengen mit den Curricula der teilnehmenden Bildungssysteme geschaffen werden.

Interessanterweise taucht der Begriff Kompetenz (engl. *competence*) in der theoretischen Rahmenkonzeption für Mathematik und Naturwissenschaften zu TIMSS 2011 nicht explizit auf (Mullis et al. 2009). Dennoch geht aus dieser Rahmenkonzeption klar hervor, dass auch TIMSS als Schulleistungsstudie ein Bild von grundlegenden Kompetenzen zeichnet, welches im Rahmen der Studie erfasst werden soll. Zentral ist laut der Rahmenkonzeption für Naturwissenschaften ein grundlegendes Verständnis von Naturwissenschaften, das es Schülerinnen und Schülern ermöglicht, überlegte Entscheidungen zu treffen (S. 49). Etwas konkreter sollen im schulischen Naturwissenschaftsunterricht die Grundlagen dafür gelegt werden, dass die Schülerinnen und Schüler als Erwachsene das Wissen und Verständnis dafür haben, fundierte Entscheidungen in Lebensbereichen wie Gesundheit, Klima oder Technologie treffen zu können. Konkretisiert wird diese noch sehr allgemein gehaltene Definition in mehreren Schritten. Zunächst werden für die jeweiligen Alterskohorten (Klassenstufe 4 und 8) die Inhaltsbereiche bestimmt und dann die kognitiven Anforderungen. Für die Klassenstufe 4, die in Deutschland an TIMSS 2011 teilnahm, sind drei Inhaltsbereiche vorgesehen: Biologie, Physik/Chemie sowie Geographie. Diese wiederum umfassen eine Vielfalt an Teilgebieten, wie beispielsweise Kennzeichen des Lebendigen, biologische Zusammenhänge in Lebensräumen, Aggregatzustände, Licht, Kräfte oder Wasser auf der Erde und Wetterbedingungen (Kleickmann et al. 2012). Für die Klassenstufe 8 werden die Naturwissenschaften in Biologie, Chemie, Physik sowie Geographie unterteilt. In beiden Kohorten werden drei kognitive Anforderungsniveaus differenziert, die abgestuft nach Komplexität zunächst Reproduzieren, dann Anwenden und schließlich Problemlösen heißen (Kleickmann et al. 2012). Diese kognitiven Prozesse gelten ebenso für die Domäne Mathematik. Auch dort werden für die beiden Alterskohorten jeweils die Inhaltsbereiche differenziert, wobei dies für die Klassenstufe 4 neben Arithmetik die beiden Bereiche Geometrie/Messen sowie Umgang mit Daten sind und für die Klassenstufe 8 Arithmetik, Algebra, Geometrie sowie Daten und Zufall. In Deutschland wurde TIMSS 2011 in der Klassenstufe 4 durchgeführt, wobei die drei für diese Alterskohorte gültigen Inhaltsbereiche in mehrere Teilgebiete gegliedert wurden. Zu Arithmetik gehören Themen wie natürliche Zahlen, Brüche und Dezimalzahlen, Zahlensätze oder Muster und Beziehungen (Selter et al. 2012).

Anders als das *Literacy*-Konzept, welches in PISA und IGLU zugrunde gelegt wird, macht die theoretische Rahmenkonzeption von TIMSS direkt den Schritt von der Bezeichnung einer Domäne zu den Inhaltsbereichen und den jeweiligen kognitiven Anforderungen. Dadurch wird zwar kein konkreter Kompetenzbegriff definiert, aber anhand der beiden Dimensionen Inhalt und kognitive Anforderung heruntergebrochen, was gemäß dem Konsens internationaler Experten an Wissen und Fertigkeiten für die untersuchte Altersgruppe angemessen und erwartbar ist. Zudem wer-



den diese beiden Dimensionen durch die Verankerung in den nationalen Curricula der Teilnehmerstaaten ergänzt, so dass etwa in Deutschland ein unmittelbarer Bezug zu den nationalen Bildungsstandards in Mathematik sowie zu den Lehrplänen für Naturwissenschaften für die Primarstufe hergestellt werden kann, für die es derzeit keine länderübergreifenden Bildungsstandards gibt (Selter et al. 2012; Kleickmann et al. 2012). Die länderübergreifenden Bildungsstandards für Mathematik in Deutschland (KMK 2004b) beschreiben mathematische Kompetenzen, die Kinder am Ende der vierten Jahrgangsstufe typischerweise erreicht haben sollen. Die Bildungsstandards in Mathematik werden in zwei Dimensionen unterteilt, nämlich inhaltsbezogene und allgemeine mathematische Kompetenzen. Die *inhaltsbezogenen* mathematischen Kompetenzen werden konkretisiert, indem sie mit fünf Leitideen verknüpft werden: Zahlen und Operationen, Raum und Form, Muster und Strukturen, Größen und Messen sowie Daten, Häufigkeit und Wahrscheinlichkeit. Innerhalb der Leitideen werden wiederum inhaltsbezogene Kompetenzen formuliert, die im nächsten Schritt als Standards weiter konkretisiert sind. Beispielsweise werden der Leitidee *Zahlen und Operationen* drei Kompetenzen zugeordnet: Darstellungen und Beziehungen von Zahlen verstehen, Rechenoperationen verstehen und beherrschen sowie in Kontexten rechnen. Zur ersten Kompetenz, Darstellungen und Beziehungen von Zahlen verstehen, sind drei Standards formuliert: den Aufbau des dezimalen Stellenwertsystems verstehen, Zahlen bis 1.000.000 auf verschiedene Weise darstellen und zueinander in Beziehung setzen sowie sich im Zahlenraum bis 1.000.000 orientieren (z. B. Zahlen runden oder der Größe nach sortieren). Die fünf *allgemeinen* oder prozessbezogenen mathematischen Kompetenzen umschreiben den Prozess der Entwicklung und des Erwerbs der inhaltsbezogenen Kompetenzen (KMK 2004b). Hierfür ist der schulische Unterricht wesentlich, denn in dessen Rahmen können Anlässe geschaffen werden,

- selbst oder gemeinsam Probleme mathematisch zu lösen
- über das Verstehen und Lösen von Aufgaben zu kommunizieren
- über Vermutungen und mathematische Zusammenhänge zu argumentieren
- Sachsituationen in die Sprache der Mathematik zu übertragen und zu modellieren
- für die Bearbeitung von Problemen geeignete Darstellungen zu entwickeln oder diese auszuwählen (vgl. Walther et al. 2007).

Ebenso wie die inhaltsbezogenen Kompetenzen wurden auch die allgemeinen mathematischen Kompetenzen in Form von Standards konkretisiert. Für die Argumentationskompetenz lauten diese beispielsweise: mathematische Aussagen hinterfragen und auf Korrektheit prüfen, mathematische Zusammenhänge erkennen und Vermutungen entwickeln sowie Begründungen suchen und nachvollziehen (KMK 2004b).

TIMSS verbindet also eine zweidimensionale Vorstellung von Kompetenz (inhaltlich sowie kognitiv) mit den jeweiligen nationalen Curricula für die vierte Jahrgangsstufe. Zur Veranschaulichung der Kompetenz werden im Rahmen von TIMSS für die Gesamtskala Mathematik fünf Kompetenzstufen unterschieden, die von rudimentärem schulischem Anfangswissen über durchschnittliches, grundlegendes Wissen bis hin zu fortgeschrittenen Fertigkeiten reichen (vgl. Selter et al. 2012).



Diese Kompetenzstufen erlauben neben der Darstellung der Leistungsstände von Schülerinnen und Schülern anhand statistischer Kennwerte auch eine inhaltliche Beschreibung dessen, was die Viertklässlerinnen und Viertklässler im Bereich Mathematik beherrschen und was (noch) nicht. Ähnlich sind die fünf Kompetenzstufen für die Domäne Naturwissenschaften aufgebaut. Auch hier wird von rudimentärem schulischem Anfangswissen gesprochen, das über grundlegendes Alltagswissen bis hin zum beginnenden naturwissenschaftlichen Denken reicht (vgl. Kleickmann et al. 2012).

Im nächsten Abschnitt wird die Schulleistungstudie IGLU vorgestellt, die konzeptionell eng mit TIMSS verwandt ist. Daher werden für beide Studien, TIMSS und IGLU, deren Design und Testkonzeption gemeinsam in Abschnitt 2.3.2 beschrieben.

### **2.3 IGLU: Internationale Grundschul-Lese-Untersuchung (engl. PIRLS)**

Die Internationale Grundschul-Lese-Untersuchung (IGLU; engl. *Progress in International Reading Literacy Study*) greift die fundamentale Bedeutung der Lesekompetenz für die individuelle und gesellschaftliche Entwicklung und wirtschaftlichen Wohlstand auf (Campbell et al. 2001). IGLU untersucht, wie gut Viertklässlerinnen und Viertklässler in den teilnehmenden Bildungssystemen verstehend lesen können und inwieweit Verbesserungsbedarf bei der Lesefähigkeit besteht. Ähnlich wie im Rahmen von PISA steht bei IGLU eine bestimmte Gruppe von Schülerinnen und Schülern im Mittelpunkt. Die Definition dieser Gruppe ist jedoch anders als bei PISA nicht altersbasiert, sondern klassenstufenbasiert. Sie entspricht der Grundschulstichprobe in TIMSS und legt fest, dass Schülerinnen und Schüler der vierten Jahrgangsstufe, die zum Testzeitpunkt mindestens 9 Jahre und 6 Monate alt sind, für die Stichprobe in IGLU in Frage kommen. Damit liegt der Testzeitpunkt am Ende der Grundschulphase, was in den meisten Teilnehmerstaaten der vierten Jahrgangsstufe entspricht. Um analog zu TIMSS eine entwicklungsgerechte Passung der Leistungstests zu erreichen, wird auch bei IGLU ein Alterskriterium festgelegt. Das durchschnittliche Alter der Untersuchungspopulation muss zum Zeitpunkt der Datenerhebung mindestens 9 Jahre und 5 Monate betragen (Tarelli et al. 2012).

IGLU folgte der seit 1991 existierenden *Reading Literacy Study*, in deren Verlauf die theoretische Rahmenkonzeption sowie die Testinstrumente für IGLU entwickelt wurden (Campbell et al. 2001). Seit 2001 wird IGLU alle fünf Jahre durchgeführt (Tarelli et al. 2012) und ist ebenso wie TIMSS eine der zentralen Studien der IEA. Das wichtigste Ziel von IGLU ist eine langfristige Dokumentation von Entwicklungen und Fortschritten in den beteiligten Bildungssystemen (Campbell et al. 2001).

Mit der Erhebung im Jahr 2011 fielen die beiden IEA-Studien TIMSS und IGLU erstmals zusammen, so dass Viertklässler in Deutschland sowohl im Bereich der Lesekompetenz als auch in Mathematik und den Naturwissenschaften getestet wurden. Die teilnehmenden Staaten konnten entscheiden, ob beide Studien gemeinsam mit

denselben Schülerinnen und Schülern an zwei unterschiedlichen Schultagen oder parallel mit verschiedenen Klassen durchgeführt werden sollten. In Deutschland wurden IGLU und TIMSS 2011 ebenso wie in 37 weiteren Staaten mit einer gemeinsamen Schülerstichprobe umgesetzt (Tarelli et al. 2012). Damit ist eine gemeinsame Datengrundlage entstanden, anhand derer man die Lesekompetenz der Kinder auch unter Berücksichtigung ihrer Leistungen in den Kompetenzbereichen Naturwissenschaften und Mathematik einordnen und interpretieren und damit in Form von Leistungsprofilen abbilden kann.

Die leitende Fragestellung von IGLU zielt auf die grundlegenden Lesekompetenzen ab, welche Schülerinnen und Schüler in den teilnehmenden Bildungssystemen am Ende der vierten Jahrgangsstufe erworben haben (Mullis et al. 2009; Tarelli et al. 2012). Der Erwerb der Lesekompetenz zu diesem Zeitpunkt dient als Indikator für die voraussichtliche weitere Entwicklung der Schülerinnen und Schüler am Ende der Primarstufe und erlaubt unter Bezug auf theoretische Annahmen und bisherige Forschungsbefunde eine Prognose über die Fähigkeit, sich lesend die Welt zu erschließen. Im Rahmen der Berichterstattung zu den Ergebnissen der ersten IGLU-Studie 2001 in Deutschland wurde ein theoretisches Rahmenmodell vorgestellt, das für die Interpretation der Befunde eine Verortung von Lehren und Lernen in bestimmten gesellschaftlichen Ausgangsbedingungen annimmt (vgl. etwa Baumert und Weiß 2002). Schulinterne Bedingungen des Kompetenzerwerbs werden dabei ebenso berücksichtigt wie bildungspolitische Rahmenbedingungen, aber auch Merkmale der Schülerinnen und Schüler sowie ihrer Elternhäuser.

An IGLU 2011 nahmen weltweit 56 Bildungssysteme teil, die meisten davon als souveräne Staaten (Tarelli et al. 2012). Eine kleine Teilgruppe führte IGLU ausschließlich oder ergänzend an Schülerinnen und Schülern der sechsten Jahrgangsstufe durch, was den Teilnehmerstaaten als internationale Option ermöglicht wird. Ein möglicher Grund für die Wahl dieser Option ist beispielsweise, dass ein Staat begründet davon ausgeht, die Leistungen der Schülerinnen und Schüler der vierten Jahrgangsstufe seien mit dem internationalen Durchschnitt dieser Kohorte nicht vergleichbar. Eine Neuerung in IGLU 2011 war zudem die Einführung von prePIRLS, einem insgesamt einfacheren Test der Lesekompetenz, der für Staaten mit vermutlich unterdurchschnittlicher Lesekompetenz sowohl als Einstieg in IGLU oder zur detaillierteren Beschreibung relativ schwacher Lesefähigkeit genutzt werden kann. Die Aufgaben in IGLU und prePIRLS wurden auf der Grundlage derselben theoretischen Rahmenkonzeption entwickelt, um gezielt den unteren Bereich des Kompetenzspektrums besser beschreiben zu können. Dazu wurden gezielt relativ einfach zu lösende Aufgaben entwickelt. In jedem partizipierenden Bildungssystem nehmen pro Erhebungsrunde etwa 4000 Schülerinnen und Schüler an mindestens 150 Schulen teil.

Ähnlich wie bei PISA wurde auch bei IGLU in Deutschland eine nationale Ergänzungsstudie durchgeführt, die als IGLU-E (Bos, Hornberg et al. 2008) neben der internationalen Verortung der Lesekompetenz bei den Schülerinnen und Schülern in Klasse 4 auch einen bundesweiten Vergleich des Kompetenzniveaus zwischen den Ländern in Deutschland erlaubte. Beide nationalen Ergänzungsstudien wurden mit der Gründung des IQB durch den Ländervergleich abgelöst.

Ergänzend zu den Leistungstests kommen auch im Rahmen der IGLU-Studie Kontextfragebögen zum Einsatz, die auf fünf Ebenen lokalisiert sind: Schülerinnen und Schüler, Eltern, Lehrkräfte, Schule sowie System (Tarelli et al. 2012).

### 2.3.1 *Kompetenzbegriff in IGLU*

Die Anschlussfähigkeit einer Kompetenz ist ein zentrales Element für die Definition von Grundbildung im Sinne des *Literacy*-Konzepts. Eine ausführliche Definition des *Literacy*-Begriffs findet sich in Abschnitt 2.1.1. Das *Literacy*-Konzept wird in IGLU, ähnlich wie in PISA, herangezogen, um zu beschreiben, was die wünschenswerte Lesekompetenz gegen Ende des vierten Schuljahres ausmacht. Lesekompetenz in IGLU ist demnach stark an einer Vorstellung von Grundbildung ausgerichtet, die deutlich über das Entziffern von Buchstabenfolgen hinausgeht. Während Schulkinder zunächst an den grundlegenden Schriftspracherwerb herangeführt werden und lernen, flüssig und sinnverstehend zu lesen (Bremerich-Vos et al. 2012), ist die Idee der Lesekompetenz nach der *Literacy*-Tradition deutlich differenzierter. Lesen können bedeutet, verschiedene Arten von Texten zu verstehen und sie für weiteres Lernen und die Verfolgung eigener Interessen nutzen zu können. Wer kompetent liest, kann lesend lernen und sich die Welt erschließen (Mullis et al. 2009); Lesen als Kulturtechnik ist also ein Schlüssel zur Gesellschaft und wer gut lesen kann, hat auch mehr Freude daran und wird eher um des Lesens willen lesen als jemand, der sich sehr anstrengen muss, um einen Text zu entziffern.

Die Modellvorstellung, die für die Beschreibung der Lesekompetenz in IGLU verwendet wird, unterscheidet drei Bereiche (vgl. Bremerich-Vos et al. 2012):

- Leseintention: Mit welcher Absicht wird ein Text gelesen? Unterschieden wird dabei zwischen literarischen Texten und Informations- bzw. Sachtexten.
- Verständnis von Informationen in einem Text. Unterschieden werden hier textimmanente und wissensbasierte Verstehensleistungen, also das Verstehen von im Text gegebenen Informationen und das Verstehen auf der Basis von Wissen außerhalb des Textes.
- Leseverhalten und Einstellungen gegenüber dem Lesen.

Um die Ausprägung der Lesekompetenz in IGLU vergleichend beschreiben zu können, werden auch hier mehrere Verstehensprozesse unterschieden. Sie ermöglichen es, gute von weniger guten Leserinnen und Lesern zu unterscheiden und dabei auch angeben zu können, womit die getesteten Schülerinnen und Schüler noch Schwierigkeiten beim Lesen haben. Der Begriff der Lesekompetenz in IGLU differenziert vier Verstehensprozesse (vgl. Bremerich-Vos et al. 2012):

- explizit im Text angegebene Informationen finden
- einfache Schlussfolgerungen ziehen
- komplexe Schlussfolgerungen ziehen, Texte interpretieren und Informationen kombinieren

- Textinhalt und Sprachgebrauch prüfen und bewerten.

Diese vier Verstehensprozesse können die an IGLU beteiligten Kinder aufweisen, indem sie entweder textimmanente Information nutzen oder ihr Vorwissen – wie oben in den drei Bereichen der Lesekompetenz beschrieben.

Im Vergleich etwa zu PISA bleibt das Modell und die Definition der Lesekompetenz alles in allem damit zwar allgemeiner, ist dafür jedoch mit den Bildungsstandards der KMK im Fach Deutsch für den Primarbereich vereinbar (vgl. Bremerich-Vos et al. 2012). Die in IGLU eingesetzten Aufgaben lassen sich den in den Bildungsstandards formulierten Leitideen zuordnen und können in diesem Sinne als curricular valide gelten.

Wie auch in PISA und TIMSS üblich, wird die (Lese-)Kompetenz in IGLU in Form eines Stufenmodells hierarchisch abgebildet. Unterschieden werden dabei fünf Kompetenzstufen, die von einem rudimentären Leseverständnis (Stufe I) über die Fähigkeit, „verstreute“ Informationen miteinander zu verknüpfen (Stufe III), bis hin zur selbstständigen Interpretation ganzer Texte und Kombination von Passagen und Aussagen (Stufe V) reichen.

### ***2.3.2 Design und Testkonzeption von TIMSS und IGLU***

Die jeweils letzte Erhebungsrunde im Jahr 2011 fiel für TIMSS und IGLU auf denselben Zeitraum. Deshalb entschieden sich 38 teilnehmende Bildungssysteme, darunter auch Deutschland, für die Durchführung beider Studien an einer gemeinsamen Stichprobe. Demnach hatte jede teilnehmende Schule insgesamt zwei Testtage, die in der Regel unmittelbar aufeinanderfolgten. Die Testkomponenten wurden dabei systematisch so rotiert, dass an etwa der Hälfte der Schulen die IGLU-Testkomponente am ersten Testtag erfolgte und TIMSS am zweiten Tag. An den übrigen Schulen wurden die beiden Erhebungen in umgekehrter Reihenfolge durchgeführt (Tarelli et al. 2012). Die Reihenfolge der Testkomponenten hatte somit keinen Einfluss auf die Leistungsergebnisse der Schülerinnen und Schüler (Foy 2012).

IGLU und TIMSS setzen ebenso wie PISA ein Multi-Matrix-Design und damit eine Rotation mehrerer Testhefte ein (Tarelli et al. 2012; Wendt, Tarelli et al. 2012). In TIMSS enthalten 14 Testhefte jeweils vier Aufgabenblöcke, die aus 10 bis 15 Einzelaufgaben zusammengesetzt wurden. Jedes TIMSS-Testheft umfasst jeweils zwei Blöcke aus Mathematik- und Naturwissenschaftsaufgaben. Ungefähr 60 Prozent der Aufgaben in TIMSS werden aus früheren Erhebungsrunden übernommen, um Veränderungen in den Leistungsergebnissen über die Zeit als Trend abbilden zu können (Mullis et al. 2009). In IGLU werden sogenannte Leistungstests auf 13 verschiedene Testhefte verteilt, so dass ein IGLU-Testheft zwei der insgesamt zehn Leistungstests enthält, die in jeder IGLU-Runde eingesetzt werden. Ein Leistungstest besteht aus jeweils einem kindgerecht gestalteten Text und bis zu 15 ihm zugeordneten Aufgaben. Dabei kommen in jeder Erhebungsrunde sowohl neue als auch bereits in vorherigen Runden verwendete Texte zum Einsatz. Letztere ermöglichen

die Verankerung der Kompetenzwerte der jeweils aktuellen Erhebungsrunde an denjenigen der vorhergehenden Studien, d. h. die Identifikation von Veränderungen in den Leistungsergebnissen über mehrere Erhebungsrunden hinweg (Tarelli et al. 2012).

Die Testkonzeptionen für TIMSS und IGLU werden in sogenannten theoretischen Rahmenkonzeptionen verdichtet und festgehalten. Sowohl TIMSS als auch IGLU legen dabei ein Modell zugrunde, an dem sich die Testkonstruktion ausrichtet. Bei TIMSS ist dies das oben beschriebene Curriculum-Modell (vgl. Abschnitt 2.2), das die drei Ebenen des Bildungssystems, der Schule sowie der Schülerinnen und Schüler differenziert. Dabei steht das intendierte Curriculum auf der Ebene des Bildungssystems und gibt an, welche Inhalte und Fähigkeiten in einer bestimmten Jahrgangsstufe zu vermitteln sind. Diese Ebene des nationalen, sozialen und bildungspolitischen Kontexts ist eng verknüpft mit den nationalen Curricula oder, wie in Deutschland, nationalen Bildungsstandards, die einen Konsens darüber darstellen, welche Kompetenzen eine Gesellschaft von Schülerinnen und Schülern am Ende einer Jahrgangsstufe mindestens erwartet. Das intendierte Curriculum kann detailliert in der sogenannten TIMSS 2011-Enzyklopädie nachgelesen werden, die für jedes teilnehmende Bildungssystem ein entsprechendes Kapitel enthält (Mullis, Martin, Minnich, Stanco et al. 2012). Das implementierte Curriculum transportiert diese Kompetenzen von der Ebene des Bildungssystems hinein in die Schule, zu den Lehrkräften bzw. in das Klassenzimmer. Dort zeigt sich auf der Ebene der Schülerinnen und Schüler das sogenannte erreichte Curriculum anhand der schriftlichen Leistungstests und Schülerfragebogen (Mullis et al. 2009). Das Curriculum ist also in der TIMSS-Testkonzeption der Kern, der diese organisiert und aus dem die Frage abgeleitet wird, inwieweit Schülerinnen und Schüler in ihrem jeweiligen Bildungskontext bestimmte Lerngelegenheiten bekommen und wie sie diese nutzen.

Das in IGLU zugrunde gelegte Rahmenmodell der Bedingungen schulischer Leistungen ist zunächst komplexer als das Curriculum-Modell in TIMSS. Das Modell ist stark von den Schülerinnen und Schülern her gedacht, die mit ihren individuellen Lernvoraussetzungen und Lernprozessen im Zentrum stehen. Diese Voraussetzungen und Prozesse stehen in unmittelbarem Zusammenhang mit ihren Leistungsergebnissen. Ferner werden schulinterne Rahmenbedingungen für den Erwerb der Lesekompetenz modelliert, dazu gehören Lehrkräfte, Klassen und Unterricht. Aus dem Modell geht hervor, dass diese schulinternen Rahmenbedingungen Faktoren sind, die konzeptuell näher an den Leistungsergebnissen der Schülerinnen und Schüler liegen als etwa außerschulische und familiäre Merkmale, wie der sozioökonomische Hintergrund der Kinder (Tarelli et al. 2012). Auch institutionelle Merkmale, z. B. bildungspolitische oder äußere schulische Rahmenbedingungen werden einbezogen; diese haben ihrerseits Einfluss auf die schulinternen Bedingungen des Kompetenzerwerbs. Zur Einbettung der Leistungsergebnisse existiert wie in TIMSS auch für IGLU eine umfassende Enzyklopädie, welche die internationale Berichterstattung der Leistungsergebnisse ergänzt (Mullis, Martin, Minnich, Drucker et al. 2012).

Die theoretische Rahmenkonzeption für TIMSS unterscheidet zur Erfassung mathematischer Kompetenz einerseits Inhaltsbereiche und andererseits kognitive An-

forderungen (vgl. Abschnitt 2.2.1). Die Konzeption greift für die vierte Jahrgangsstufe drei Inhaltsbereiche auf, die anhand der Testaufgaben erfasst werden sollen: Arithmetik, Geometrie/Messen sowie Umgang mit Daten (Selter et al. 2012). Da diese drei Inhaltsbereiche jedoch an sich sehr umfangreich sind, werden sie jeweils in mehrere Teilgebiete gegliedert. Zu Arithmetik gehören etwa die Teilgebiete *Natürliche Zahlen*, *Brüche und Dezimalzahlen* oder *Muster und Beziehungen*. Im Bereich Geometrie/Messen sollen die Schülerinnen und Schüler mit *Punkten, Geraden und Winkeln* sowie mit *zwei- und dreidimensionalen Figuren* umgehen. Zum Umgang mit Daten sind die Teilgebiete *Daten lesen und interpretieren* sowie *Ordnen und Darstellen* zu rechnen. Die kognitiven Anforderungen, die als zweite Dimension neben den Inhaltsbereichen Teil der mathematischen Kompetenz in TIMSS sind, werden als Reproduzieren, Anwenden und Problemlösen bezeichnet. Dabei ist das Reproduzieren der kognitiv am wenigsten anspruchsvolle Bereich, während Anwenden und Problemlösen jeweils höhere kognitive Anforderungen stellen (Mullis et al. 2009). Kognitive Aktivitäten des Anforderungsbereichs *Reproduzieren* sind beispielsweise das Abrufen von Standardwissen, rechnen, Informationen ablesen oder messen. Im Bereich *Anwenden* sollen die Schülerinnen und Schüler u. a. Informationen darstellen, mathematische Operationen ausführen oder auch mathematische Modelle aufstellen. Die anspruchsvollsten kognitiven Aktivitäten werden im Bereich Problemlösen verlangt, wie etwa das Erkennen mathematischer Beziehungen, das Formulieren von Begründungen oder Verallgemeinerungen und Spezialisierungen (Selter et al. 2012). Dieses Verständnis von mathematischer Kompetenz ist zwar nicht vollständig (Selter et al. 2012), jedoch in weiten Teilen mit den Bildungsstandards der KMK im Fach Mathematik für den Primarbereich vereinbar (KMK 2004b).

Die Leistungstests, die in IGLU zum Einsatz kommen, sollen das Leseverständnis der Schülerinnen und Schüler ermitteln. Ein Vergleich des Leistungsstandes in verschiedenen Teilnehmerstaaten ermöglicht ein Benchmarking, d. h. eine systematische Betrachtung von Gemeinsamkeiten und Unterschieden, sowie die Identifikation relativer Stärken und Schwächen. Alle drei bisherigen IGLU-Erhebungsrounden basieren auf einer identischen Rahmenkonzeption (Bremerich-Vos et al. 2012). Die IGLU-Rahmenkonzeption sieht Lesekompetenz ähnlich wie PISA dem angelsächsischen *Literacy*-Konzept verpflichtet. Ein ausführlicher Vergleich der Konzepte der Lesekompetenz findet sich bei Artelt, Drechsel, Bos und Stubbe (2008). Die Lesekompetenz, wie sie in IGLU definiert wird, ist neben dem Bezug zur *Literacy*-Tradition auch mit den Bildungsstandards der KMK im Fach Deutsch für den Primarbereich vereinbar (KMK 2005). Die Bildungsstandards sehen vier Bereiche der grundlegenden sprachlich-literarischen Bildung vor: Sprechen und Zuhören, Schreiben, Lesen sowie Sprache und Sprachgebrauch. Das IGLU-Modell der Lesekompetenz unterscheidet die oben genannten drei Bereiche, von denen zwei in Form von Leistungstests und einer durch einen Schülerfragebogen erfasst werden. Erstens wird ein Text mit einer bestimmten *Leseintention* gelesen. Dabei wird unterschieden, ob ein literarischer Text oder ein informativer Sachtext gelesen wird. Zweitens geht es um das *Verstehen der Informationen* eines Textes, wobei damit sowohl explizit im Text enthaltene Informationen als auch vorwissensbasierte Verstehensleistung-

gen gemeint sind. Drittens gehört zur Lesekompetenz auch eine Reihe von Aspekten des Leseverhaltens sowie Einstellungen zum Lesen (Bremerich-Vos et al. 2012). Lesekompetenz wird in IGLU in vier sogenannte Verstehensprozesse überführt, die unterschiedlich schwierig sind. Dazu wird die Lesekompetenz zunächst aufgespalten in die Nutzung von im Text enthaltener Information sowie die Nutzung externen Wissens. Information innerhalb der Texte kann wiederum in Form von zwei Aspekten verarbeitet werden: entweder indem unabhängige Einzelinformationen gefunden und genutzt werden oder indem Beziehungen zwischen verschiedenen Textteilen hergestellt werden. Auch das externe Wissen, das von den Schülerinnen und Schülern beim Lesen herangezogen wird, kann auf zwei Wegen genutzt werden: entweder, indem über Inhalte oder, indem über Strukturen reflektiert wird. Die Nutzung unabhängiger Einzelinformationen entspricht dabei dem Lokalisieren explizit angegebener Informationen und ist der am wenigsten anspruchsvolle Verstehensprozess. Wenn Schülerinnen und Schüler Beziehungen zwischen Textteilen herstellen, ziehen sie einfache Schlussfolgerungen (zweiter, etwas schwierigerer Verstehensprozess). Der dritte Verstehensprozess, komplexe Schlussfolgerungen ziehen, erfordert das Kombinieren und Interpretieren von Information und gehört zur Reflektion über Textinhalte. Der vierte und anspruchsvollste Verstehensprozess bedeutet, über Strukturen zu reflektieren und damit sowohl den Inhalt als auch den Sprachgebrauch in einem Text zu prüfen und zu bewerten (Bremerich-Vos et al. 2012). Neben diesen international vergleichenden Schulleistungstudien werden derzeit deutschlandweit drei landesweite Untersuchungen mit einem Schwerpunkt auf Kompetenzentwicklung durchgeführt. Der IQB-Ländervergleich zur Überprüfung der Bildungsstandards, VERA sowie NEPS. Diese drei Studien werden in den folgenden Abschnitten beschrieben.

## **2.4 IQB-Ländervergleich zur Überprüfung der Bildungsstandards**

Das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) wurde im Jahr 2004 als wissenschaftliches Institut an der Humboldt-Universität zu Berlin gegründet. Die Gründung geht auf einen Beschluss der Kultusministerkonferenz (KMK) zurück, der nach dem ernüchternden Abschneiden der Schülerinnen und Schüler in Deutschland sowohl bei TIMSS 1995 (Baumert et al. 2000; Baumert et al. 1997) als auch bei PISA 2000 (Baumert et al. 2001) im Hinblick auf eine Gesamtstrategie zum Bildungsmonitoring gefasst wurde (KMK 2006). Zu dieser Gesamtstrategie gehören mehrere Maßnahmen zu Qualitätssicherung und -entwicklung im Bildungswesen, von denen eine besonders wichtige die Einführung und regelmäßige Überprüfung nationaler Bildungsstandards ist. Bereits im Jahr 1997 wurde als Reaktion auf die Befunde aus TIMSS mit dem Konstanzer Beschluss die Sicherung der Qualität schulischer Bildung als langfristiges Ziel ausgegeben (KMK 1997).

Eines der zentralen Erkenntnisse aus TIMSS und PISA waren neben dem im internationalen Vergleich enttäuschend niedrigen mittleren Kompetenzniveau deut-



scher Schülerinnen und Schüler auch die enormen Ungleichheiten, die sich teilweise innerhalb Deutschlands im Vergleich von Bundesländern, aber auch zwischen Gruppen von Schülerinnen und Schülern zeigten (Baumert et al. 2002; Baumert et al. 2003). Große Abstände im mittleren Leistungsniveau zwischen den Bundesländern gaben ebenso Anlass zur Sorge wie das deutlich schwächere Abschneiden von Schülerinnen und Schülern mit Zuwanderungshintergrund im Vergleich zu ihren Klassenkameraden oder der starke Zusammenhang zwischen sozialer Herkunft und der Kompetenzstufe im PISA-Test. Die Entscheidung, nationale Bildungsstandards zu entwickeln und diese regelmäßig zu überprüfen, zielte auf die Schaffung einer gemeinsamen, länderübergreifenden Grundlage der Qualitätsentwicklung und des Bildungsmonitorings ab (Pant, Stanat, Pöhlmann et al. 2013).

Die ersten nationalen Bildungsstandards in Deutschland wurden in den Jahren 2003 und 2004 eingeführt (vgl. etwa KMK 2003; KMK 2004a). Im Einzelnen abgedeckt werden sollen bestimmte Kernfächer sowie Schulabschlüsse, so dass die Bildungsstandards letztlich die Frage beantworten, was Schülerinnen und Schüler zum Zeitpunkt des Erwerbs eines bestimmten Schulabschlusses in Bezug auf diese Kernfächer wissen und können. Das IQB führt seit 2009 in Ergänzung zu den international vergleichenden Bildungsstudien die Ländervergleiche zur Überprüfung der Bildungsstandards in Deutschland durch.

Beim IQB-Ländervergleich besteht ein Zyklus der Studie aus zwei Erhebungs- runden je untersuchter Schulstufe. Auf der Primarstufe erfolgt alle fünf Jahre ein Ländervergleich (Stanat et al. 2012b), auf der Sekundarstufe wie bei PISA alle drei Jahre (Pant, Stanat, Schroeders et al. 2013). Im Wechsel werden dabei die Fächergruppen Sprache (Deutsch, Englisch, Französisch) sowie Mathematik und Naturwissenschaften überprüft. Die erste Erhebungsrunde des Ländervergleichs mit dem Schwerpunkt Sprachen fand 2008/2009 auf der Sekundarstufe I (Jahrgangsstufe 9) statt. Rund 40.000 Schülerinnen und Schüler an 1500 zufällig ausgewählten allgemeinbildenden Schulen nahmen mit ihrer Klasse an diesem ersten Ländervergleich zur Überprüfung der Bildungsstandards teil (Köller et al. 2010). Die zweite Erhebungsrunde des Ländervergleichs im Jahr 2011 fand an Grund- und Förderschulen statt. 27.000 Viertklässlerinnen und Viertklässler an mehr als 1300 Schulen bearbeiteten mit ihrer Klasse die Testaufgaben zu den Fächern Deutsch und Mathematik (Stanat et al. 2012a). Mit dem Ländervergleich 2012, der im selben Zeitraum wie die PISA-Studie 2012 stattfand, erfolgte die zweite Erhebungsrunde auf der Sekundarstufe I. Damit ist für diese Kohorte der erste Ländervergleichs- Zyklus abgeschlossen und mit der nächsten Erhebungsrunde im Jahr 2015 zum Kompetenzbereich Sprachen werden erste Trendanalysen möglich sein (Pant, Stanat, Pöhlmann et al. 2013). Alle an PISA 2012 beteiligten Schulen führten an einem separaten Testtag auch den Ländervergleich 2012 durch, wobei die PISA-Schulstichprobe von ca. 230 Schulen eine Teilmenge der Ländervergleichsstich- probe mit rund 1300 Schulen bildete. Pro Schule bearbeiteten eine bis zwei vollständige Klassen die Aufgaben des Ländervergleichs (Pant, Stanat, Schroeders et al. 2013).

Die Aufgaben, die zur Erfassung der in den Bildungsstandards definierten Kom- petenzen eingesetzt werden, werden vorab von einer Gruppe erfahrener Lehrkräfte

aus dem gesamten Bundesgebiet entworfen. In Zusammenarbeit mit dem IQB und den dort tätigen Wissenschaftlerinnen und Wissenschaftlern aus den entsprechenden Fachdidaktiken der getesteten Kompetenzbereiche sowie Bildungsforschung und Psychologie werden die Aufgaben durch mehrere Entwurfsstadien hindurch stetig weiterentwickelt. Im Verlauf dieses Prozesses erfolgen sprachliche Überprüfungen sowie fachdidaktische und psychometrische Beurteilungen der Aufgabenqualität, ehe die Testaufgaben in mehreren Vorstudien mit Schülerinnen und Schülern erprobt werden. Nur Aufgaben, die sich in diesen Vorstudien bewährt hatten, kommen schlussendlich im Ländervergleich zum Einsatz.

Die im Rahmen des Ländervergleichs erfassten Contextmerkmale der Schülerinnen und Schüler sowie ihres schulischen Umfeldes werden über drei Fragebögen erhoben: Einen Schülerfragebogen sowie einen für Lehrkräfte und einen für Schulleitungen (Siegle et al. 2013).

### ***2.4.1 Kompetenzbegriff im IQB-Ländervergleich***

Der IQB-Ländervergleich begreift die länderübergreifenden Bildungsstandards in Deutschland in Anlehnung an die sogenannte Klieme-Expertise (Klieme et al. 2007) als Kompetenzerwartungen. Damit ist gemeint, dass die in den Bildungsstandards formulierten Kompetenzen als Bildungsergebnisse zu einem bestimmten Zeitpunkt (etwa zum Mittleren Schulabschluss, MSA) von möglichst jedem Schüler und jeder Schülerin beherrscht werden. Dies entspricht dem Bildungsauftrag allgemeinbildender Schulen (Klieme et al. 2007). Konkret werden die hinter den Bildungsstandards angenommenen Kompetenzen der Schülerinnen und Schüler als kontextspezifische, aber auch erlernbare Leistungsdispositionen verstanden, mit denen Fragestellungen eines jeweils spezifischen Inhaltsbereiches gelöst werden können (Koeppen et al. 2008; Pant, Stanat, Pöhlmann et al. 2013). Die Messung von Kompetenzen geschieht dadurch, dass überprüft wird, ob die Bildungsstandards erreicht werden. Das entspricht demnach der Konkretisierung in Form von Aufgaben zu bestimmten Fachgebieten, zu deren Lösung vorab bestimmte Kompetenzen benötigt werden. Schülerinnen und Schüler, die über die jeweilige Kompetenz verfügen, können die entsprechenden Aufgaben mit einer bestimmten Wahrscheinlichkeit lösen. In diesem Sinne drücken die Bildungsstandards, die für die Kernfächer vorliegen, die angestrebten Lernergebnisse als Könnensbeschreibungen (Pant, Stanat, Pöhlmann et al. 2013) aus. Da die Bildungsstandards der Kultusministerkonferenz für Deutschland eine länderübergreifende und verbindliche Orientierung bieten sollen, wurden sie gezielt mit Blick auf Evaluationsprozesse entwickelt und formuliert. Exemplarisch sei hier für das Fach Mathematik die Struktur der Kompetenz auf dem Niveau des Hauptschulabschlusses (HSA) sowie des Mittleren Schulabschlusses (MSA) beschrieben.

Mathematische Kompetenz wurde zuletzt im IQB-Ländervergleich 2012 umfassend überprüft. Ähnlich wie in PISA und TIMSS wird mathematische Kompetenz als mehrdimensionales Konstrukt aufgefasst. Dabei werden drei Dimensionen un-

terschieden: Inhalt, Prozess und Anspruch. Diese drei Dimensionen werden weiter differenziert. So sind der Inhaltsdimension fünf sogenannte Leitideen zugewiesen, die als inhaltsbezogene Kompetenzen gelten können. Der Prozessdimension werden sechs allgemeine mathematische Kompetenzen untergeordnet, während die Anforderungsdimension drei Anforderungsbereiche unterscheidet. Diese drei Dimensionen erlauben die Verortung der Mathematikaufgaben in einem dreidimensionalen Raum, der aus Inhalten, Prozessen und einem Anforderungsniveau besteht.

Die *Inhaltsdimension* mathematischer Kompetenz im Rahmen der Bildungsstandards wird durch fünf Leitideen strukturiert (Roppelt et al. 2013). Diese Leitideen werden bezeichnet als Zahl, Messen, Raum und Form, funktionaler Zusammenhang sowie Daten und Zufall (KMK 2003). Unter der Leitidee *Zahl* werden alle Aspekte zusammengefasst, die mit Quantität zu tun haben. Das bedeutet, dass die Schülerinnen und Schüler Zahlen dazu verwenden, Situationen zu strukturieren und zu beschreiben. Die Leitidee *Messen* beschreibt den Umgang mit Größen, etwa Winkeln, Längen, Flächeninhalten oder Volumina. Die dritte Leitidee, *Raum und Form*, umfasst das Fachgebiet der Geometrie und damit alle Arten ebener und räumlicher Konfigurationen, Gestalten oder Muster (Roppelt et al. 2013). *Funktionaler Zusammenhang* als vierte Leitidee entspricht dem mathematischen Stoffgebiet der Algebra und beinhaltet funktionale bzw. relationale Beziehungen zwischen mathematischen Objekten. Beispiele dafür sind Terme, Gleichungen und Funktionen (Roppelt et al. 2013). Die fünfte Leitidee, *Daten und Zufall*, wird weitgehend durch das mathematische Stoffgebiet der Stochastik abgedeckt. Statistik und Wahrscheinlichkeitsrechnung sind hier zentral.

Die *Prozessdimension* wird dargestellt durch sechs allgemeine mathematische Kompetenzen, die verschiedene kognitive Prozesse in allen definierten Inhaltsbereichen (Leitideen) vorsehen. Die Übergänge zwischen den einzelnen Prozessen sind dabei fließend, zumal auch die meisten mathematischen Tätigkeiten mehrere allgemeine Kompetenzen zugleich ansprechen (Roppelt et al. 2013). Die erste Kompetenz, die im Rahmen der Bildungsstandards Mathematik für die Sekundarstufe I definiert wird, ist das mathematische Argumentieren. Weiterhin gehören sowohl das Verstehen gegebener Argumentationen als auch das Entwickeln eigener situationsangemessener mathematischer Argumentationen dazu.

Die dritte Dimension, *Anspruch*, unterscheidet schließlich drei aufeinander aufbauende Anforderungsniveaus bei der Beschreibung mathematischer Kompetenz. Der am wenigsten anspruchsvolle Anforderungsbereich, Reproduzieren, verlangt von den Schülerinnen und Schülern die Wiedergabe sowie die direkte Anwendung grundlegender Begriffe, Verfahren oder Annahmen in einem abgegrenzten Gebiet und in bekannten Zusammenhängen (KMK 2003). Es geht dabei also um Routinen, die häufig auch mit alltäglichem Wissen und Bezug gezeigt werden sollen. Das zweite Anforderungsniveau, Zusammenhänge herstellen, geht darüber hinaus und stellt die Verknüpfung mathematischer Kenntnisse, Fertigkeiten und Fähigkeiten in den Mittelpunkt. Auch hier handelt es sich im Wesentlichen um den Umgang mit bekannten Sachverhalten. Das höchste Anforderungsniveau schließlich, Verallgemeinern und Reflektieren, verlangt die Bearbeitung komplexer Sachverhalte. Die Schülerinnen und Schüler sollen dabei auch eigene Problemstellungen formulieren

können sowie Lösungen, Begründungen oder Schlussfolgerungen ziehen können. Auch das Abwägen verschiedener Argumente gehört hierzu. Die sechs allgemeinen mathematischen Kompetenzen werden anhand dieser drei Anforderungsbereiche differenziert, so dass jede der Kompetenzen in verschiedenen Ausprägungen beschreibbar wird. Ein Schüler, der lediglich das erste Anforderungsniveau bewältigt, ist dabei weniger kompetent als ein Schüler, der Aufgaben auf Niveau 2 oder 3 lösen kann.

Im Gegensatz zum funktionalen Verständnis mathematischer Grundbildung, wie es etwa im *Literacy*-Begriff in PISA angenommen wird (OECD 2013a; Sälzer, Reiss et al. 2013), zielt das Konstrukt der Mathematikkompetenz im Sinne der Bildungsstandards explizit auch auf die Bearbeitung innermathematischer Problemstellungen ab (Roppelt et al. 2013). Der Unterschied liegt damit beispielsweise darin, dass in PISA ein wesentliches Element mathematischer Kompetenz die Übersetzung eines alltäglichen Problems in die Sprache der Mathematik und zurück ist (Sälzer, Reiss et al. 2013), während bei den Bildungsstandards auch Aufgaben gestellt werden, die rein mathematisch und nicht in einer alltäglichen Situation angesiedelt sind.

#### ***2.4.2 Design und Testkonzeption des IQB-Ländervergleichs***

Der Ländervergleich zur Überprüfung der Bildungsstandards ist als Querschnittsstudie angelegt, bei der die teilnehmende Kohorte genau einmal getestet wird. Damit bundesweit und folglich länderübergreifend gültige Skalen bestimmt werden konnten, auf denen sich die Schülerinnen und Schüler entsprechend ihrer Fähigkeiten sowie die Testaufgaben mit ihrer jeweiligen Schwierigkeit verorten lassen, wurden auf der Basis national repräsentativer Stichproben der 9. und 10. Jahrgangsstufen Kalibrierungs- bzw. Normierungsstudien durchgeführt (Pant, Böhme et al. 2013). Darunter versteht man, dass vor der Durchführung des Ländervergleichs die dafür geplanten Aufgaben von Schülerinnen und Schülern bearbeitet und deren Antworten für eine Schätzung der Itemparameter (Schwierigkeit der Aufgaben) verwendet wurden. Auf diese Weise konnten beispielsweise Aufgaben identifiziert werden, die den statistischen oder psychometrischen Gütekriterien nicht entsprachen. Solche Aufgaben werden in der Regel entweder aus dem Test genommen oder, wo ein Grund für die unzureichende Qualität zu erkennen ist, entsprechend verändert.

Am aktuell letzten Ländervergleich nahmen knapp 45.000 Schülerinnen und Schüler teil, die in die auf Bundeslandebene repräsentativen Stichproben gezogen worden waren (Siegle et al. 2013). Die Schüler besuchten zum Erhebungszeitpunkt im Frühjahr 2012 die 9. Jahrgangsstufe. Damit war in Deutschland die Verknüpfung des Ländervergleichs mit der etwa zeitgleich stattfindenden PISA-Studie möglich, in deren Rahmen auch 2012 neben der international vorgegebenen Stichprobe fünfzehnjähriger Schülerinnen und Schüler auch pro gezogener Schule zwei vollständige neunte Klassen teilnahmen. Bis einschließlich PISA 2006 wurde PISA sowohl international (PISA-I) als auch national für den Bundesländervergleich (PISA-E) durchgeführt, ehe mit PISA 2009 eine Trennung beider Studien er-

folgte und PISA sich ausschließlich auf den internationalen Vergleich des deutschen Bildungssystems (ohne Bundesländervergleich) konzentriert, während der IQB-Ländervergleich sich der innerdeutschen Gegenüberstellung durchschnittlicher Schülerkompetenzen widmet.

Ähnlich wie PISA, TIMSS oder IGLU funktioniert die Messung und Abbildung von Kompetenzen im IQB-Ländervergleich nicht über eine individuelle Leistungsdiagnostik, sondern über die länderübergreifende und ländervergleichende Aggregation der Daten. Dazu soll die untersuchte Stichprobe die Population der Schülerinnen und Schüler der 9. Jahrgangsstufe möglichst genau repräsentieren. Jugendliche, die eine 9. Jahrgangsstufe besuchen, können sich in Deutschland in mehreren Schulformen der Sekundarstufe I befinden. Je nachdem, in welchem Bundesland sie zur Schule gehen, stehen unterschiedliche und auch unterschiedlich viele Schulformen zur Wahl. Das Gymnasium ist die einzige Schulform, die in allen 16 Bundesländern existiert und eine gemeinsame Geschichte hat. Allerdings variiert die Bildungsbeteiligung über die Länder (von 31 Prozent in Bayern bis 43 Prozent in Hamburg) deutlich (Siegle et al. 2013). Sowohl in PISA als auch im Ländervergleich werden die zahlreichen vorliegenden Schulformen der Sekundarstufe I üblicherweise auf sechs (bzw. sieben in PISA, wo die beruflichen Schulen mit zur Schulpopulation gehören) Schulformen verdichtet: Hauptschulen, Schulen mit mehreren Bildungsgängen, Realschulen, Integrierte Gesamtschulen, Gymnasien sowie Förderschulen (vgl. etwa Sälzer und Prenzel 2013; Siegle et al. 2013). Zum Zeitpunkt von PISA 2012 bzw. dem Ländervergleich 2012 existierte in der sogenannten Fachserie 11 des Statistischen Bundesamtes (Statistisches Bundesamt 2012), die als Grundlage für die Bestimmung der Zielpopulation herangezogen wurde, die Schulform Hauptschule, wobei diese Schulart in einigen Bundesländern einer Reform unterzogen wurde und teilweise bereits anders hieß (etwa die Mittelschulen in Bayern). Die Ziehung der stratifizierten Stichprobe erfolgte in drei Schichten, nämlich Förderschulen, Gymnasien und einer dritten Schicht für alle übrigen Schularten (Siegle et al. 2013). Diese drei expliziten Schichten waren auf alle Bundesländer anwendbar und konnten auch bei einer Reform der Schulstruktur erhalten bleiben. Innerhalb der drei expliziten Schichten wurde eine vorab festgelegte Zahl von Schulen gezogen. Innerhalb der dritten expliziten Schicht wurden die Schulen in jedem Bundesland nach Schulform sortiert (implizite Schichtung), sodass sichergestellt werden konnte, dass alle in einem Bundesland vorliegenden Schulformen in die Stichprobe gelangen. In absoluten Zahlen umfasste die Zielpopulation des IQB-Ländervergleichs 2012 insgesamt 859.923 Neuntklässlerinnen und Neuntklässler (Siegle et al. 2013), während die Zielpopulation der Fünfzehnjährigen in PISA 2012 aus 798.136 Jugendlichen bestand (Sälzer und Prenzel 2013).

Wie in zahlreichen Large-Scale-Assessments üblich, setzt auch der IQB-Ländervergleich zur Überprüfung der Bildungsstandards ein sogenanntes Multi-Matrix-Design ein. Da im IQB-Ländervergleich stets mehrere Kompetenzbereiche erfasst werden (2012 zuletzt Mathematik und Naturwissenschaften mit jeweils mehreren Unterbereichen), kann ein einzelner Schüler oder eine einzelne Schülerin in einer einzigen Testsitzung nicht ausreichend viele Aufgaben aus allen erfassten Kompetenzbereichen bearbeiten. Durch *Multi-Matrix-Sampling* (Gonzalez und Rutkow-

ski 2010) werden die einzelnen Aufgaben innerhalb des jeweiligen Kompetenzbereichs gruppiert und damit zu Clustern verbunden, die als Bausteine zu Testheften zusammengefügt werden. So erhält jeder Studienteilnehmer eine Teilmenge aller eingesetzten Aufgaben, die vorab in verschiedenen Testheften zusammengestellt worden sind. Durch eine optimale Zusammenstellung dieser unterschiedlichen Testhefte wird eine zuverlässige Schätzung der Schülerkompetenzen in allen erhobenen Kompetenzbereichen ermöglicht (Siegle et al. 2013). Die Aufgabenblöcke werden dabei so gebildet, dass drei aufeinander folgende Cluster von mindestens 90 Prozent der Schülerinnen und Schüler in 60 Minuten vollständig bearbeitet werden können. Im Ländervergleich 2012 kam ein sogenanntes *Youden-Square-Design* zum Einsatz, nach dem jedes Cluster mit jedem anderen Cluster in genau einem Testheft vorkam und jeder Block an jeder Position im Testheft genau einmal auftrat (vgl. Frey et al. 2009; Siegle et al. 2013). Dieses Design ist benannt nach dem Statistiker William Youden, der im Zusammenhang mit Untersuchungen zur Verbreitung eines Virus bei Tabakpflanzen ein unvollständiges Block-Design einsetzte (Youden 1937). In diesem Kontext wurde die Anbaufläche eines Tabakfeldes in Quadrate unterteilt, innerhalb derer unterschiedliche Treatments verabreicht wurden. Da im Ländervergleich 2012 zwei umfassende Kompetenzbereiche erhoben wurden, Mathematik und Naturwissenschaften, wurden zunächst zwei separate Youden-Squares mit jeweils 31 Aufgabenblöcken in 31 Testheften erstellt (Siegle et al. 2013). Damit die beiden Kompetenzbereiche nicht nur separat erfasst, sondern auch miteinander in Zusammenhang gebracht werden können, wurden jeweils 12 Aufgabenblöcke in acht Testheften kombiniert und so eine teilweise Überschneidung erreicht. Insgesamt kamen im IQB-Ländervergleich 2012 70 Testhefte für Regelschülerinnen und Regelschüler sowie 11 Testhefte für Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf zum Einsatz.

Neben den beschriebenen Kompetenztests werden im IQB-Ländervergleich auch mehrere Kontextfragebögen administriert. Auf der Grundlage der Angaben von Schülerinnen und Schülern, Lehrkräften und Schulleitern sind detaillierte Analysen zu verschiedenen Themen möglich. Beispielsweise können Unterschiede zwischen Mädchen und Jungen, soziale oder zugewanderungsbedingte Disparitäten oder auch Zusammenhänge mit bestimmten Kontextbedingungen oder Einstellungen untersucht werden. Der Fragebogen für Schülerinnen und Schüler im IQB-Ländervergleich 2012 wurde ebenso wie die Testhefte in einem rotierten Design ausgegeben, konkret in acht unterschiedlichen Versionen (Siegle et al. 2013). Dabei enthielten alle acht Versionen zunächst ein gemeinsames Set an zentralen Fragen, das allen Schülerinnen und Schülern vorgelegt wurde. Dazu zählen unter anderem Angaben zum sozio-ökonomischen Hintergrund der Familie oder zum Zuwanderungshintergrund. Anschließend an dieses gemeinsame Fragenset erhielten die Schülerinnen und Schüler eine von acht Versionen, wobei innerhalb einer gezogenen Schulklasse jeweils nur eine Version verwendet wurde. So bearbeitete jeweils die ganze Klasse einen identischen Fragebogen, wobei die Zuweisung der Version zu einer Klasse durch Zufall erfolgte. Unterschiedliche Klassen bearbeiteten also unterschiedliche Themen, so dass die entsprechenden Analysen im Bericht zum



IQB-Ländervergleich jeweils auf einer Teilstichprobe beruhen (etwa zum Thema Testmotivation oder fachspezifisches Selbstkonzept, vgl. Jansen et al. 2013).

Alle Lehrkräfte, die in den an einem Ländervergleich teilnehmenden Klassen die jeweils untersuchten Schulfächer unterrichten, haben die Möglichkeit einen Lehrerfragebogen auszufüllen. Im zuletzt durchgeführten Ländervergleich 2012 waren dies die Fächer Mathematik, Chemie, Biologie oder Physik. Damit variierte die Zahl der infrage kommenden Lehrkräfte je nach Schule abhängig von den unterrichteten Klassen und Fächerkombinationen (Siegle et al. 2013). Für die Lehrkräfte wird also keine Population definiert und demnach auch keine Stichprobe gezogen. Das Hauptanliegen des Lehrerfragebogens war, Angaben über relevante Rahmenbedingungen schulischer Bildungsprozesse zu erhalten. Anhand dieser Einschätzungen durch die Lehrkräfte können etwa Aspekte des Unterrichts, aber auch berufliche Qualifikationen, Berufserfahrung oder kollegiale Zusammenarbeit beschrieben werden. An jeder teilnehmenden Schule wurde zusätzlich die Schulleitung gebeten, einen Fragebogen zu den schulischen Rahmenbedingungen auszufüllen. Inhalte des Schulleiterfragebogens waren unter anderem Schülerzahl, Trägerschaft, Schulprofil oder auch Lehr- und Betreuungsangebot der Schule.

## 2.5 VERA: Vergleichsarbeiten in der Schule

Die mit VERA bezeichneten flächendeckenden Vergleichsarbeiten tragen der oben beschriebenen Säule 3 der Gesamtstrategie der KMK zum Bildungsmonitoring Rechnung. Im Gegensatz zu den bisher beschriebenen Studien kann VERA selbst allerdings nicht als Bildungsmonitoring bezeichnet werden (Prenzel und Seidel 2010). Ein Bildungsmonitoring setzt voraus, dass ein Kompetenzbegriff definiert und operationalisiert wird, der anhand von umfassend entwickelten Testaufgaben als zuverlässigen Leistungsindikatoren gemessen werden kann. Die standardisierte Erhebung solcher Daten durch geschulte Testleiterinnen und Testleiter sowie die anschließende Auswertung der Daten durch Expertengruppen ist ein weiteres Merkmal eines Bildungsmonitorings. Bei VERA hingegen geht es um Testaufgaben, anhand derer den Lehrkräften eine Rückmeldung über den Leistungsstand ihrer Schülerinnen und Schüler in den Fächern Deutsch und Mathematik gegeben wird (Isaac et al. 2006). Eine geeignete Bezeichnung für VERA ist demnach beispielsweise Lernstandserhebung oder Diagnoseinstrument. In einigen Bundesländern werden die VERA-Tests nicht als Vergleichsarbeiten, sondern mit anderen Begriffen bezeichnet, z. B. mit „KERMIT – Kompetenzen ermitteln“ (Hamburg) oder Kompetenztests (Sachsen und Thüringen). Ergänzend hat die KMK in einer Vereinbarung zur Weiterentwicklung von VERA (KMK 2012b) erneut betont, dass die Hauptfunktion der Vergleichsarbeiten im Bereich der Schul- und Unterrichtsentwicklung liegt und dass sie eine Grundlage für Notengebung oder eine Prognose des schulischen Erfolgs darstellen können.

Die Durchführung von VERA erfolgt unter der Zuständigkeit der Länder. Einzelne Aspekte obliegen jedoch dem IQB in Berlin, welches auch den Ländervergleich



zur Überprüfung der Bildungsstandards verantwortlich. Dazu gehören beispielsweise die Aufgabenentwicklung und deren Pilotierung sowie die Bestimmung der Aufgabenschwierigkeiten (Skalierung), die Zusammenstellung der Testhefte sowie die Erarbeitung didaktischer Materialien. Auf Seiten der Länder liegt die Zuständigkeit bei den Landesinstituten, Qualitätsagenturen oder den zuständigen Fachabteilungen der Ministerien. Diese kümmern sich unter anderem um den Druck und die Verteilung der Testhefte, die Testdurchführung sowie die Auswertung und die Rückmeldungen an die Schulen.

VERA wird in zwei Klassenstufen durchgeführt, 3 und 8, und entsprechend VERA-3 und VERA-8 genannt. Wichtigstes Ziel von VERA ist die Initiierung von Unterrichtsentwicklung (Isaac et al. 2006). Im Rahmen von VERA werden unter Bezug auf die Kompetenzbereiche der beiden untersuchten Schulfächer Deutsch und Mathematik verschiedene Teilleistungsbereiche untersucht. Gemessen werden die Leistungen der Schülerinnen und Schüler ähnlich wie beim IQB-Ländervergleich an den Bildungsstandards der KMK. Die Ergebnisse, die aus VERA entstehen, sind zweierlei: Erstens lassen sich Lösungshäufigkeiten auf Aufgabenebene bestimmen und zweitens können für die teilnehmenden Lerngruppen (in der Regel Schulklassen) Kompetenzniveaus definiert werden (Isaac et al. 2006). Solche fachspezifischen Rückmeldungen an die unterrichtenden Lehrkräfte sollen Impulse dafür geben, den Unterricht kompetenzorientiert weiterzuentwickeln. Eine große Herausforderung dabei ist, diese Rückmeldungen auf den eigenen Unterricht zu übertragen. Aus diesem Grund sind etwa die Beschreibungen der einzelnen Kompetenzniveaus möglichst nah am Unterricht gehalten (Isaac und Hosenfeld 2008). Außerdem stehen den Lehrkräften Unterstützungs- und Begleitmaterialien zur Verfügung, wie etwa eine Informationsbroschüre für Eltern, ein Leitfaden zur Unterrichtsentwicklung oder eine Handreichung zur Interpretation und Nutzung der Ergebnisse (Isaac 2013).

Im Gegensatz zu den weiteren hier beschriebenen Studien erfolgt die Lernstandserhebung bei VERA nicht anhand einer Stichprobe von Schülerinnen und Schülern, sondern in Form einer Vollerhebung aller Dritt- und Achtklässler an allgemeinbildenden Schulen in Deutschland. Weitere Unterschiede zu den bisher genannten Systemmonitoring-Studien bestehen beispielsweise in der Testdurchführung, die bei Bildungsmonitorings durch externe Testleiterinnen und Testleiter erfolgt, bei VERA jedoch durch die unterrichtenden Lehrkräfte. Auch die Auswertung findet dezentral durch die Lehrkräfte bzw. die jeweiligen Landesinstitute statt, während sie bei den Studien zum Systemmonitoring zentral durch die entsprechenden Wissenschaftlergruppen erfolgt. Für VERA ergibt sich daraus auch die Möglichkeit, neben sofortigen Rückmeldungen nach der Dateneingabe durch die Lehrkräfte eine differenziertere und innerhalb des Bundeslandes vergleichende Rückmeldung seitens der Landesinstitute nach einigen Wochen zu erhalten. Im Gegensatz zu VERA liegt bei den Studien zum Systemmonitoring zwischen der Datenerhebung und der Veröffentlichung beziehungsweise Rückmeldung der Ergebnisse ein Zeitraum von teilweise mehreren Jahren.

Anders als bei stichprobenbasierten Schulleistungsvergleichsstudien, deren Ergebnisse sich auf Aggregatebenen beziehen (etwa Staaten), liegt der Schwerpunkt

der Leistungserfassung bei VERA auf der Ebene der Einzelschule sowie der Schulklasse (Diemer und Kuper 2011). In diesem Sinne kann die Lernstandserhebung, wie sie in VERA durchgeführt wird, als Teil der Strategie einer sogenannten „neuen Steuerung“ bezeichnet werden (Bellmann 2006; Böttcher 2002). Unter dem Begriff der neuen Steuerung werden zumeist drei Strukturmerkmale gefasst: Standard-, Evidenz- sowie Outputorientierung. Hinter diesen Strukturmerkmalen stehen mehrere Prinzipien, etwa eine Dezentralisierung von Entscheidungen, die auch eine Erweiterung der Autonomie der Einzelschulen zum Ziel hat oder die Forderung, dass Entscheidungen grundsätzlich unter Bezug auf wissenschaftlich gewonnene empirische Befunde getroffen werden sollen (Diemer und Kuper 2011). Das Konzept der Lernstandserhebungen folgt diesen Prinzipien, indem diese Erhebungen das neue Steuerungsparadigma direkt an die Schulen sowie an Lehrkräfte herantragen. Entsprechend sollen an den Schulen die genannten Strukturmerkmale der Standard-, Evidenz- und Outputorientierung umgesetzt werden. In diesem Sinne sind die Lernstandserhebungen im Rahmen von VERA in erster Linie ein Testinstrument zur Erfassung von Schülerleistungen, deren Ergebnisse unmittelbar an die Schulen zurückgemeldet werden können (Kuper und Schneewind 2006). Sie werden erst dann zum Steuerungsinstrument, wenn sie im Schulsystem zu Steuerungszwecken genutzt werden. Dass gerade diese Nutzung jedoch stark von einzelnen Lehrkräften abhängt und daher nicht generell vorausgesetzt werden kann, zeigten beispielsweise Diemer und Kuper (2011) in einer Interviewstudie. Die Autoren fanden, dass die Annahme einer Outputorientierung bei der Einführung von Verfahren zentraler Leistungsmessung und Ergebnisrückmeldung keinesfalls allgemeingültig ist und demnach nicht davon ausgegangen werden kann, dass Lernstandserhebungen schlussendlich flächendeckend für outputorientierte Steuerungsprozesse in Schule und Unterricht genutzt werden. Auch hier zeigt sich ein grundlegender Unterschied zwischen Bildungsmonitoring und Lernstandserhebungen.

## **2.6 NEPS: Das Nationale Bildungspanel – *National Educational Panel Study***

Das Nationale Bildungspanel NEPS (auf Englisch: *National Educational Panel Study*) nimmt im Kontext der Schulleistungsstudien ebenso wie VERA eine Sonderrolle ein. Zum einen handelt es sich bei NEPS um die einzige längsschnittlich angelegte Studie der hier beschriebenen, d. h. die untersuchten Personen werden zu mehreren Zeitpunkten befragt. Darüber hinaus werden im Rahmen von NEPS mehrere Kohorten über eine längere Zeit untersucht, so dass man von einer Mehrkohorten-Längsschnittstudie sprechen kann. Zum anderen erfasst NEPS neben Kompetenzen von Schülerinnen und Schülern Bildungsverläufe über die gesamte Lebensspanne und hat dadurch einen wesentlich breiteren Fokus als die weiteren hier genannten Untersuchungen. Da jedoch die Bildung und damit die sowohl im institutionellen Rahmen als auch außerhalb erworbenen Fähigkeiten und Fertigkeiten von

Menschen in Deutschland eine zentrale Rolle einnehmen, sollte NEPS Teil des Überblicks über große Schulleistungstudien in Deutschland sein.

NEPS ist die umfangreichste sozialwissenschaftliche Studie, die in Deutschland bisher durchgeführt wurde. Verteilt auf die verschiedenen Altersgruppen umfasst die Stichprobe etwa 60.000 Personen (Blossfeld et al. 2011). NEPS setzt sich aus sechs Panelstudien zusammen, die als sogenannte Startkohorten bezeichnet werden. Institutionell ist das NEPS mittlerweile am Leibniz-Institut für Bildungsverläufe (LIfBi) an der Otto-Friedrich-Universität Bamberg angesiedelt, wobei auch Wissenschaftler an anderen Institutionen in Deutschland und dem Ausland am Nationalen Bildungspanel mitwirken.

### ***2.6.1 Kompetenzbegriff in NEPS***

NEPS geht aufgrund seiner Eigenschaft als Längsschnittstudie über den gesamten Lebensverlauf von einem mehrdimensionalen, eigentlich jedoch von mehreren Kompetenzbegriffen aus. Allen gemeinsam ist dabei die Vorstellung, dass Bildung und damit auch Kompetenzentwicklung ein lebenslanger Prozess ist. Anders als bei kohortenspezifischen Schulleistungstudien wie PISA, IGLU oder TIMSS muss der Kompetenzbegriff in NEPS daher auf eine breitere Zielgruppe von Personen anwendbar sein und kann auch nicht punktuelle Meilensteine im Sinne von erworbenen Kompetenzen definieren. Vielmehr muss berücksichtigt werden, dass Bildungsprozesse im Lebenslauf und dabei insbesondere Bildungsentscheidungen eng mit dem Curriculum und dem jeweils vorliegenden Schulsystem in Deutschland verknüpft sind (Neumann et al. 2013). Insofern braucht die theoretische Rahmenkonzeption der Kompetenzbegriffe in NEPS für das Schulalter einen expliziten Bezug zu den jeweiligen Bildungsstandards in Fächern wie Mathematik, Deutsch oder Fremdsprachen, aber auch einen klaren Bezug zu Lebenssituationen im Erwachsenenalter. Exemplarisch werden hier Mathematik und Lesekompetenz kurz dargestellt.

Für Mathematik etwa verknüpft NEPS die theoretischen Rahmenkonzeptionen aus PISA und den Bildungsstandards des IQB-Ländervergleichs (Neumann et al. 2013). Das Konzept mathematischer Grundbildung ist ähnlich wie bei PISA und den Bildungsstandards in mehrere Dimensionen unterteilt, so dass Inhaltsbereiche (so genannte Leitideen) neben kognitiven Komponenten (so genannten allgemeinen mathematischen Kompetenzen) stehen (vgl. z. B. KMK 2012a). Neben diesen beiden Dimensionen, die sich sowohl in den theoretischen Rahmenkonzeptionen zu PISA als auch zu den Bildungsstandards wiederfinden, wird in NEPS noch die Zeit bzw. Alterskohorte als eine dritte, implizite Dimension mitgedacht (Neumann et al. 2013). Wie in der theoretischen Rahmenkonzeption mathematischer Grundbildung in PISA (OECD 2013a) definiert auch NEPS die vier Inhaltsbereiche Quantität, Veränderung und Beziehungen, Raum und Form sowie Unsicherheit und Daten.

Während die Lesekompetenz in NEPS zwar ein ähnliches Verständnis von funktionaler Grundbildung wie in PISA voraussetzt (OECD 2009a), kommen allerdings

im Gegensatz zu PISA keine diskontinuierlichen Texte wie Tabellen oder Diagramme zum Einsatz (Gehrer et al. 2013). Lesekompetenz wird in NEPS weitestgehend als Textverstehen definiert, wobei zur Abbildung der Lesekompetenz über die gesamte Lebensspanne drei Dimensionen unterschieden werden, die sich auch in der Testentwicklung niederschlagen: (a) Textfunktionen und Textsorten, (b) kognitive Anforderungen und (c) Aufgabenformate (Gehrer et al. 2013). Unterschieden werden fünf Textfunktionen bzw. Textsorten, so dass es Aufgaben zu Kommentartexten, Informationstexten, literarischen Texten, Anleitungen und Werbung gibt. Die kognitiven Anforderungen werden differenziert in das Auffinden von Information im Text, das Ziehen textbezogener Schlussfolgerungen sowie in das Reflektieren und Überprüfen von Textinhalten und Schlussfolgerungen (Gehrer et al. 2013).

Insgesamt wurden für das nationale Bildungspanel NEPS neben den genannten Domänen Mathematik und Leseverständnis auch Schrift- und Sprachkompetenz ausgewählt, aber auch naturwissenschaftliche Grundbildung sowie fachübergreifende kognitive Fähigkeiten wie beispielsweise Informationsverarbeitung. Darüber hinaus werden im Rahmen von NEPS auch sogenannte Metakompetenzen sowie soziale Kompetenzen erfasst, etwa Selbstregulation, Selbsteinschätzung der eigenen Fähigkeiten, Umgang mit Informations- und Kommunikationstechnologien oder fachspezifisches Interesse bzw. Motivation. Da das NEPS sich mit Kompetenzen über die gesamte Lebensspanne hinweg befasst, werden auch lebensetappenspezifische Kompetenzen berücksichtigt. Dazu gehören etwa lehrplan- oder berufsbezogene Fähigkeiten ebenso wie deren Entwicklungsbedingungen.

### ***2.6.2 Design und Testkonzeption von NEPS***

Das Design des NEPS baut sich in fünf Säulen auf, die wiederum in acht Bildungsetappen entlang der Lebensspanne unterteilt werden. Säulen und Etappen bilden zusammen die Rahmenkonzeption des Nationalen Bildungspanels. Die fünf *Säulen*, die inhaltlich als Dimensionen von Bildung über die Lebensspanne betrachtet werden können, sind (1) Kompetenzentwicklung (2) Bildungsprozesse in verschiedenen Lernumwelten, (3) Soziale Ungleichheit und Bildungsentscheidungen, (4) Bildungserwerb mit Zuwanderungshintergrund sowie (5) Bildungsrenditen (Blossfeld et al. 2011). All diese Säulen sind jeweils unter Berücksichtigung des gesamten Lebenslaufes zu sehen. Dabei wird die Panelstudie zu jeder thematischen Säule in acht *Bildungsetappen* unterteilt: (1) Neugeborene und Eintritt in frühkindliche Betreuungseinrichtungen, (2) Kindergarten und Einschulung, (3) Grundschule und Übertritt in eine Schulart der Sekundarstufe I, (4) Wege durch die Sekundarstufe I und Übergänge in die Sekundarstufe II, (5) Gymnasiale Oberstufe und Übergänge in (Fach-)Hochschule, Ausbildung oder Arbeitsmarkt, (6) Aufnahme einer beruflichen Ausbildung und der spätere Arbeitsmarkteintritt, (7) (Fach-)Hochschulstudium und Übergang in den Arbeitsmarkt sowie (8) allgemeine und berufliche Weiterbildung (Blossfeld et al. 2011).

Durch die längsschnittliche Anlage des NEPS werden die untersuchten Kompetenzen der Teilnehmerinnen und Teilnehmer nicht nur einmal getestet, sondern in der Regel alle zwei bis sechs Jahre (vgl. etwa Artelt et al. 2013). Auf diese Weise können Entwicklungsverläufe nachgezeichnet, aber auch verschiedene Startkohorten miteinander verglichen werden. Eine besondere Herausforderung ist dabei, dass zur Abbildung der Kompetenzentwicklung über die Lebensspanne nicht derselbe Test mehrmals angewendet werden kann, sondern dass die Tests inhaltlich und in ihrer Schwierigkeit in Bezug auf die jeweils untersuchte Alterskohorte angepasst werden müssen (Pohl und Carstensen 2013).

Im Rahmen von NEPS wird im Gegensatz zu anderen Large Scale Assessments wie PISA oder TIMSS üblicherweise nicht mit einem rotierten Booklet-Design gearbeitet, sondern jeder Testteilnehmer einer bestimmten Kohorte bearbeitet dieselben Aufgaben (vgl. etwa Gehrler et al. 2013). Die Altersadäquatheit der unterschiedlichen Kompetenztests wird jeweils durch Experteneinschätzung, Bestimmung des Schwierigkeitsspektrums der Aufgaben und, am Beispiel der Lesekompetenz, deren Lesbarkeit optimiert. Nach mehreren Pilotstudien werden die am besten für die getesteten Altersgruppen passenden Aufgaben ausgewählt. In NEPS werden darüber hinaus zahlreiche bereits existierende und empirisch bewährte Kompetenztests eingesetzt, beispielsweise das Salzburger Lesescreening (Auer et al. 2005). Diese erprobten Testverfahren sind in aller Regel bereits auf eine bestimmte Altersgruppe normiert, was die Qualität der Kompetenzmessung abstützt.

## 2.7 Zusammenfassender Überblick: Schulleistungsstudien in Deutschland

	PISA	TIMSS	PIRLS / IGLU	IQB-LV	VERA	NEPS
<b>Initiierende Organisation</b>	OECD	IEA	IEA	KMK	KMK	BMBF
<b>Hauptziel</b>	System-monitoring	System-monitoring	System-monitoring	System-monitoring	Unterrichts- und Schulentwicklung	Beschreibung von Bildungsprozessen über die Lebensspanne
<b>Bisherige und geplante Erhebungsrunden</b>	2000, 2003, 2006, 2009, 2012, 2015, 2018	1995, 1999, 2003, 2007, 2011, 2015	2001, 2006, 2011, 2016	2008/2009, 2011, 2012, 2015	2004 - 2006 in 7 Ländern, seit 2007/2008 jährlich in allen Ländern	seit 2010, abhängig von der Kohorte
<b>Population/ Zielgruppe</b>	15-jährige Schüler	Schüler der vierten Jahrgangsstufe, die durchschnittlich 9 Jahre und 6 Monate alt sind und/oder Schüler der achten Jahrgangsstufe bzw. am Ende Sekundarstufe I	Schüler der vierten Jahrgangsstufe, die durchschnittlich 9 Jahre und 5 Monate alt sind	Schüler kurz vor einem bestimmten Schulabschluss (Primarstufe, Sekundarstufe I, Sekundarstufe II)	Schüler, die die Klassenstufen 3 oder 8 besuchen (Vollerhebung)	6 Kohorten: Kleinkinder (7 Monate), 4-Jährige, Schüler der Klassenstufen 5 und 9, Studienanfänger sowie Erwachsene der Jahrgänge 1944 bis 1986
<b>Design</b>	Querschnitt, Trend	Querschnitt, Trend	Querschnitt, Trend	Querschnitt, Trend	Querschnitt, Trend	Längsschnitt mit mehreren Kohorten

	PISA	TIMSS	PIRLS / IGLU	IQB-LV	VERA	NEPS
<b>Turnus</b>	Alle 3 Jahre	Alle 4 Jahre	Alle 5 Jahre	Primarstufe: alle 5 Jahre, Sekundarstufe: alle 3 Jahre	Jährlich	Abhängig von der Kohorte
<b>Domänen</b>	Lesekompetenz, Mathematik, Naturwissen- schaften	Mathematik und Natur- wissenschaften	Lesekompetenz	Sprachen (Deutsch, Französisch, Englisch) und Mathematik/ Natur- wissenschaften	Deutsch, Mathematik	Sprache, Mathematik, Natur- wissenschaften, Meta- kompetenzen
<b>Stichprobe</b>	Ca. 5000 Schüler pro Staat	Ca. 4000 Schüler pro Staat	Ca. 4000 Schüler pro Staat	Primarstufe: ca. 27.000 Schüler, Sekundarstufe I: ca. 45.000 Schüler	Vollerhebung (keine Stich- probe; alle 3. und 8. Klassen an all- gemeinbildenden Schulen)	60.000 Menschen in Deutschland aus verschiedenen Altersgruppen
<b>Teilnehmer</b>	Alle OECD- Staaten plus Partnerstaaten (2012: $N = 65$ )	Bildungssysteme weltweit (2011: $N = 59$ )	Bildungssysteme weltweit (2011: $N = 56$ )	Alle deutschen Bundesländer ( $N = 16$ )	Alle deutschen Bundesländer ( $N = 16$ )	Alle deutschen Bundesländer ( $N = 16$ )
<b>Kontext- fragebögen eingesetzt in Deutschland</b>	Schüler, Eltern, Lehrkräfte, Schulen	Schüler, Eltern, Lehrkräfte, Schulen, Curri- culumexperten	Schüler, Eltern, Lehrkräfte, Schulen, Curri- culumexperten	Schüler, Lehrkräfte, Schulen	Keine	Eltern, Lehrkräfte, Schulen



	PISA	TIMSS	PIRLS / IGLU	IQB-LV	VERA	NEPS
Kompetenz- begriff	Literacy	Inhalte und kognitive Anforderungen	Literacy mit Bezug zum Curriculum	Bildungs- standards der KMK	Leistungsstand	Verknüpfung von Literacy und Bildungs- standards
Theo- retischer Bezugs- rahmen der Kompe- tenzen	Theoretische Rahmen- konzeption (Assessment Framework)	Theoretische Rahmen- konzeption (Assessment Framework)	Theoretische Rahmen- konzeption (Assessment Framework)	Bildungs- standards der KMK	Bildungs- standards der KMK	Assessment Frameworks (PISA, TIMSS, IGLU) sowie Bildungs- standards der KMK

Studienbuch Schulleistungsstudien

Das Rasch-Modell in der Praxis

Sälzer, C.

2016, VI, 89 S. 3 Abb. in Farbe., Softcover

ISBN: 978-3-662-45764-1