

PISA 2025

Samuel Greiff
Jennifer Diedrich
Tamara Kastorff
Frank Goldhammer
Olaf Köller
(Hrsg.)

Bildung im
Spannungsfeld von
Innovation und
Transformation

GLOSSAR

Glossar

Begriff	Definition
2-parameter-logistic Modell (2-PL-Modell)	Das 2-PL-Modell umfasst die Modellparameter ► <i>Schwierigkeit</i> und ► <i>Diskrimination</i> . Es kann also das Potenzial einer Aufgabe ausdrücken, zwischen kompetenteren und weniger kompetenten Studienteilnehmer*innen sowie zwischen Studienteilnehmer*innen mit geringen Kompetenzunterschieden zu differenzieren.
Aufgabe	► <i>Item</i>
Agency in the Anthropocene	Agency in the Anthropocene ist ein zentrales Konzept in der Rahmenkonzeption für Naturwissenschaften in PISA 2025. Gemeint ist die Kompetenz von Jugendlichen, in einer Welt zu handeln, die stark vom Menschen geprägt ist (Anthropozän = Zeitalter des menschlichen Einflusses auf Klima, Natur und Erde). In PISA 2025 wird die umweltwissenschaftliche Kompetenz als Teilaspekt von Agency in the Anthropocene erfasst.
Antwortformate (offen, geschlossen, teiloffen)	In PISA werden Testaufgaben in drei Antwortformaten gestellt: Fragen mit Mehrfachwahl (Multiple Choice) bei der eine oder mehrere Antworten gewählt werden muss, Fragen mit einer Kurzantwort (Short Answer) oder mit einer längeren, offenen Antwort (Extended Response).
Bildungsmonitoring	Bildungsmonitoring bezeichnet in Deutschland die systematische, wissenschaftlich abgesicherte Beobachtung des Bildungssystems mit dem Ziel, Bildungsqualität transparent zu machen, Stärken und Schwächen des Bildungssystems zu identifizieren und Ansatzpunkte für Qualitätsentwicklung, Standardsicherung und bildungspolitische Steuerung bereitzustellen. Im Verständnis der Kultusministerkonferenz ist Bildungsmonitoring Teil einer an Ergebnissen von Bildungsprozessen orientierten Steuerung des Bildungswesens. In Deutschland umfasst Bildungsmonitoring insbesondere internationale Schulleistungstudien wie PISA, PIRLS/IGLU und TIMSS, die Überprüfung und Umsetzung von Bildungsstandards, Verfahren der Qualitätssicherung auf Schulebene sowie die gemeinsame Bildungsberichterstattung von Bund und Ländern.
Booklet-Design	Im Rahmen groß angelegter Studien des Bildungsmonitorings wie PISA bezeichnet ein Booklet-Design die strukturierte Zuordnung von Testaufgaben zu verschiedenen Testheften, die den Schüler*innen in zufälliger oder kontrolliert verteilter Form vorgelegt werden. Ziel ist es, eine große Bandbreite an Aufgaben abzudecken, ohne dass jede Person alle Aufgaben bearbeiten muss. Alle Aufgaben werden einer Kompetenzdomäne zugeordnet, die in sogenannten Clustern gruppiert werden, die wiederum in unterschiedlichen Booklets zusammengestellt sind.
Data Adjudication	Als Data Adjudication wird in PISA die internationale Prüfung bezeichnet, ob die Daten eines Staates die technischen Anforderungen an Repräsentativität, Teilnahmequoten und Datenqualität erfüllen. Dabei wird bewertet, ob die Ergebnisse belastbare internationale Vergleiche erlauben oder ob ihre Aussagekraft durch Abweichungen von den Standards substantiell eingeschränkt ist.

Begriff	Definition
Diskrimination, psychometrische	Die psychometrische Diskrimination eines Items beschreibt, wie gut eine Testaufgabe zwischen Personen mit unterschiedlichen Ausprägungen einer Kompetenz oder eines Merkmals unterscheiden kann. In der ► <i>Item-Response-Theorie (IRT)</i> wird die Diskrimination als Itemparameter modelliert. Dieser Parameter hängt mit der Steigung der ► <i>Item Characteristic Curve (ICC)</i> zusammen: Je höher der Diskriminationsparameter, desto stärker verändert sich die Lösungswahrscheinlichkeit bei kleinen Unterschieden im Kompetenzniveau.
Domäne	Als Domäne wird in PISA ein inhaltlich definierter Kompetenzbereich bezeichnet, der mit einer eigenen Rahmenkonzeption, spezifischen Aufgaben und Skalierungen erfasst wird. Zu den Kern-domänen zählen Naturwissenschaften, Lesen und Mathematik; hinzu kommen bei PISA innovative Domänen und Zusatzdomänen. In jeder Erhebungsrunde gibt es eine Hauptdomäne. Für sie wird die Rahmenkonzeption intensiv überarbeitet, mehr Testzeit vorgesehen und eine vertiefte Berichterstattung ermöglicht. In PISA 2025 ist naturwissenschaftliche Kompetenz die Hauptdomäne.
Effektstärke	Die Effektstärke zeigt die praktische Relevanz statistischer Ergebnisse aus sozialwissenschaftlichen Studien an. Sie wird bei einer breiten Streuung der Messergebnisse eher kleiner und bei einer geringen Streuung eher größer (bei gleichbleibendem Unterschied der mittleren Messergebnisse).
Erhebungsrunde	Eine Erhebungsrunde umfasst einen vollständigen Ablauf aller Projektschritte von der Entwicklung der Rahmenkonzeption über Feldtest und Haupterhebung bis zur Berichterlegung. In jeder Erhebungsrunde wird eine neue innovative Domäne entwickelt und es wechselt die Hauptdomäne. Sind einmal alle Domänen Hauptdomäne gewesen, so ist ein ► <i>PISA-Zyklus</i> abgeschlossen. Mit PISA 2025 endet der dritte PISA-Zyklus (2018, 2022, 2025) und die neunte Erhebungsrunde.
Feldtest	Der Feldtest ist die der Haupterhebung vorausgehende Erprobungsphase von PISA. Er dient dazu, Aufgaben, Fragebögen, technische Abläufe und organisatorische Verfahren unter möglichst realistischen Bedingungen zu testen und bei Bedarf zu überarbeiten. Die Haupterhebung ist demgegenüber die eigentliche Datenerhebung, deren Ergebnisse in die internationale Berichterstattung eingehen. Im vorliegenden Berichtsband wird, sofern nicht anders ausgewiesen, die Haupterhebung von PISA 2025 beschrieben.
Frage / Item / Antwortskala / Merkmal	s. Seite 244
Funktionale Grundbildung	Funktionale Grundbildung bezeichnet in PISA anwendungsbezogene Kompetenzen, die Jugendliche für weitere Bildungswege, eine selbstständige Lebensführung und gesellschaftliche Teilhabe benötigen. Im Vordergrund steht nicht curriculares Wissen, sondern Wissen in relevanten Anwendungssituationen zu nutzen, Informationen einzuordnen und begründete Urteile zu treffen.
Grundgesamtheit	Die interessierende Grundgesamtheit in PISA sind Fünfzehnjährige die in den teilnehmenden Staaten eine Schule besuchen. Von der in PISA getesteten Stichprobe wird entsprechend auf diese Grundgesamtheit geschlossen.

Begriff	Definition
Indikator, manifeste (auch: manifeste Variable)	Manifeste Indikatoren sind direkt beobachtbare oder messbare Größen (z. B. das Alter einer Person). Diese können in sozialwissenschaftlichen Untersuchungen eingesetzt werden um ► <i>latente Variablen</i> indirekt zu erfassen.
Innovative Domäne	Die innovative Domäne ist ein zusätzlicher Kompetenzbereich in PISA, der von Erhebungsrunde zu Erhebungsrunde wechselt. Er ergänzt die drei Kerndomänen Naturwissenschaften, Lesen und Mathematik um Bereiche, die vor dem Hintergrund gesellschaftlicher und technologischer Entwicklungen als besonders bedeutsam gelten. In PISA 2025 ist die innovative Domäne Lernen in der digitalen Welt.
Interaktionseffekt	Ein Interaktionseffekt liegt vor, wenn die Wirkung einer unabhängigen Variable auf eine abhängige Variable davon abhängt, welche Ausprägung eine zweite unabhängige Variable hat. Anders gesagt: Zwei Faktoren wirken nicht nur einzeln, sondern beeinflussen sich gegenseitig in ihrer Wirkung.
Internationale Studienleitung	Die internationale Studienleitung von PISA liegt beim OECD-Sekretariat. Es verantwortet die übergreifende Steuerung des Programms, bereitet Entscheidungen des PISA Governing Board vor, koordiniert die strategische Weiterentwicklung der Studie und arbeitet dabei mit den internationalen Auftragnehmern sowie den nationalen Studienleitungen zusammen. Damit liegt die Gesamtverantwortung für Anlage, Governance und internationale Abstimmung der Studie beim OECD-Sekretariat.
Internationaler Auftragnehmer	Als internationaler Auftragnehmer wird in PISA die von der OECD beauftragte Organisation bezeichnet, die zentrale operative und technische Aufgaben der Studie umsetzt. Dazu gehören insbesondere die Entwicklung und Bereitstellung von Instrumenten und Plattformen sowie die technische Unterstützung der nationalen Studienleitungen. In PISA 2025 ist das Australian Council of Educational Research (ACER) internationaler Auftragnehmer.
Item	Ein Item (im Rahmen von PISA auch: Teilaufgabe) ist das einzelne (kleinste) Element in einem Test oder Fragebogen (s. Abbildung unter ► Frage). Bei PISA wird zwischen drei Arten in Bezug auf ihr ► <i>Antwortformat</i> (offen, geschlossen, teiloffen) unterschieden und Antworten zu einem Item anhand eines zwei- oder mehrstufigen Kategoriensystems kodiert (z. B. 0 = „falsch“, 1 = „teilweise richtig“ und 2 = „richtig“).
Item Characteristic Curve	Die Item Characteristic Curve (ICC) beschreibt in der ► <i>Item-Response-Theorie (IRT)</i> den Zusammenhang zwischen dem Kompetenzniveau einer Person und der Wahrscheinlichkeit, ein bestimmtes Item korrekt zu lösen. Die Form der Kurve wird durch Itemparameter wie Schwierigkeit und Diskrimination bestimmt. Die Schwierigkeit gibt an, bei welchem Kompetenzniveau eine bestimmte Lösungswahrscheinlichkeit erreicht wird, während die Diskrimination beschreibt, wie steil die Kurve in diesem Bereich verläuft. Eine steile ICC zeigt an, dass das Item besonders gut zwischen Personen mit leicht unterschiedlichen Kompetenzniveaus unterscheiden kann.

Begriff	Definition
Item-Response-Theorie (IRT)	Die Item-Response-Theorie liefert den theoretischen und methodischen Rahmen für die Untersuchung des Zusammenhangs zwischen manifest beobachteten Variablen und den als dafür ursächlich angenommenen, dahinterliegenden latenten Merkmalen. Die manifesten Variablen (► <i>Indikatoren</i>) wie etwa die Antworten von Jugendlichen auf PISA-Testaufgaben werden mit latenten Merkmalen (► <i>latente Variable</i>) wie etwa der naturwissenschaftlichen Kompetenz der Schüler*innen verbunden.
Kodierung	Kodierung bezeichnet den Prozess, durch den frei formulierte Antworten (► <i>Antwortformate</i>) systematisch in vorab definierte Kategorien oder numerische Werte überführt werden. Intensiv geschulte und fachkundigen Kodierer*innen kategorisieren gemäß der Kodiervorschriften die Antworten der Schüler*innen in „richtig“, „teilweise richtig“ oder „falsch“.
Kohorte	Eine Kohorte bezeichnet eine Gruppe von Personen, die ein gemeinsames, ereignisbezogenes Merkmal teilen, z. B. ein bestimmtes Geburts- oder Einschulungsjahr. Bei PISA wird die Kohorte der 15-jährigen Schüler*innen untersucht, diese Kohorte definiert zugleich auch die ► <i>Grundgesamtheit</i> .
Kompetenzbereich	► Domäne
Kompetenzstufen	Eine Kompetenzstufe beschreibt einen numerisch definierten Wertebereich auf einer Kompetenzskala, in diesem Fall der PISA-Kompetenzskala. Die PISA-Skala ist normiert auf einen Mittelwert von 500 und eine Standardabweichung von 100. Diese Skala wird in der Regel in distinkte Kompetenzstufen von I bis VI unterteilt. Jede Stufe ist mit typischen Aufgabenanforderungen verknüpft und zeigt, welche Arten von Anforderungen Lernende erfolgreich lösen können. Konkret bedeutet das: Kompetenzwerte, die einer bestimmten Stufe zugeordnet werden, sind verbunden mit bestimmten Anforderungen, die typischerweise bewältigt werden können. Anforderungen, die Aufgaben einer höheren Kompetenzstufe zugeordnet sind, werden hingegen bei gleichem Kompetenzniveau mit einer deutlich geringeren Wahrscheinlichkeit gelöst. Das Gleiche gilt umgekehrt für niedrigere Kompetenzstufen.
Kontrollvariable	Eine Kontrollvariable ist eine Variable, die zum Beispiel bei Regressionsanalysen zwischen anderen Variablen statistisch konstant gehalten wird. Dadurch soll verhindert werden, dass durch ein Auslassen der Kontrollvariable Scheinkorrelationen entstehen. Ob eine Variable in einem Regressionsmodell eine Kontrollvariable oder eine tatsächlich interessierende Variable ist, ist eine inhaltliche Frage und muss basierend auf bestehender empirischer Forschung oder theoretisch begründbar sein. Die im vorliegenden Berichtsband verwendeten Kontrollvariablen sind Tabelle 0.2web zu entnehmen.
Large-Scale-Assessment / Studien des Bildungsmonitorings	Large-Scale-Assessment bezeichnet groß angelegte Studien des Bildungsmonitorings mit einer Stichprobe von mehreren Tausend Teilnehmenden, deren Kompetenz und weitere Merkmale systematisch miteinander verglichen werden. Large-Scale-Assessments haben eine eigene Methodik mit ► Plausiblen Werten.

Begriff	Definition
Merkmal	Ein Merkmal im psychologischen Sinne ist ein theoretisches, gedankliches Konzept. Es bezeichnet Dinge, die nicht direkt physisch greifbar oder beobachtbar sind, aber dabei helfen, Phänomene zu beschreiben oder messbar zu machen (z. B. Intelligenz, Glück, Motivation oder Zeit, s. ► Frage beziehungsweise Seite 244.).
Mittelwerte und Streuungsmaße (Standardabweichungen)	Mittelwerte geben an, welches mittlere Niveau in einem Kompetenztest durchschnittlich erreicht wurde; auf ihrer Basis können Rangreihen und Differenzen abgebildet werden. Streuungsmaße, insbesondere die Standardabweichung, beschreiben hingegen, wie stark die individuellen Werte um diesen Mittelwert variieren. In Bildungsstudien wie PISA zeigt sich anhand der Standardabweichung, ob Kompetenzen im Durchschnitt eher homogen und nah am Mittelwert oder stark unterschiedlich verteilt sind.
Modellparameter	Im Rahmen der ► <i>Item-Response-Theorie (IRT)</i> werden Modelle für den Antwortprozess bei Tests oder Fragebögen formuliert. In diesen Modellen werden Annahmen darüber getroffen, wie eine Antwort einer Person mit den Merkmalen der Person zusammenhängt. Ein Modellparameter kann hierbei z. B. die ► <i>Schwierigkeit</i> der Aufgabe sein.
Nationale Studienleitung	Die nationale Studienleitung verantwortet die Durchführung von PISA in Deutschland. Sie koordiniert die nationale Umsetzung der internationalen Vorgaben, steuert Vorbereitung, Feldarbeit, Datenaufbereitung und Berichterstattung und fungiert als zentrale Schnittstelle zwischen OECD-Sekretariat, internationalem Auftragnehmer und den national beteiligten Institutionen. In Deutschland ist diese Funktion im Zentrum für internationale Vergleichsstudien (ZIB) institutionell verankert.
Normalverteilung	Die Normalverteilung ist eine symmetrische Wahrscheinlichkeitsverteilung, bei der die meisten Werte um den Durchschnitt liegen und die Wahrscheinlichkeit für extreme Werte abnimmt. Eine Normalverteilung zeigt sich in sozialwissenschaftlichen Befragungen oder Studien häufig dann, wenn die Stichprobe sehr groß ist, wie z. B. bei PISA.
Oversampling	Oversampling bezeichnet in Studien den gezielten, überproportionalen Einbezug bestimmter Teilgruppen der Gesamtpopulation, um aussagekräftige Analysen auch für kleine oder seltene Gruppen zu ermöglichen. Neben der üblichen Stichprobe wird also zusätzlich eine ergänzende Stichprobe von Teilnehmer*innen gezogen, die ein bestimmtes Merkmal wie beispielsweise einen sonderpädagogischen Förderbedarf haben.
PISA-Zyklus	Eine ► <i>Erhebungsrunde</i> umfasst einen vollständigen Ablauf aller Projektschritte von der Entwicklung der Rahmenkonzeption über Feldtest und Haupterhebung bis zur Berichterlegung. In jeder Erhebungsrunde wird eine neue innovative Domäne entwickelt und es wechselt die Hauptdomäne. Sind einmal alle Domänen Hauptdomäne gewesen, so ist ein PISA-Erhebungszyklus abgeschlossen. Mit PISA 2025 endet der dritte Erhebungszyklus (2018, 2022, 2025) und die neunte Erhebungsrunde.

Begriff	Definition
Plausible Values (PVs)	Plausible Values (PVs) in PISA sind stochastisch erzeugte Schätzwerte für die Kompetenz von Schüler*innen in den getesteten Bereichen (Naturwissenschaften, Lesen, Mathematik). Sie berücksichtigen, dass jede Person nur einen Teil der gesamten Testitems bearbeitet und die gemessene Kompetenz daher mit Unsicherheit behaftet ist. Durch die Verwendung von 10 PVs pro Person können in PISA valide Mittelwerte, Streuungen und Zusammenhänge auf Staatenebene berechnet werden, ohne dabei fälschlicherweise die Messunsicherheit einzelner Schüler*innen zu ignorieren.
Population	► <i>Grundgesamtheit</i>
Psychometrie	Ein Fachgebiet der Psychologie, das sich mit Theorie und Methodik des psychologischen Messens beschäftigt. Dazu gehören beispielsweise psychologische Testverfahren (Intelligenztests, Kompetenz- oder Persönlichkeitstests), aber auch die verwendeten Methoden zur Verarbeitung, Skalierung und Auswertung von Daten.
Quartil	Quartile teilen Werte in vier gleich große Gruppen. Dazu werden die Werte nach der Größe geordnet. Das untere Quartil umfasst die 25 Prozent der Werte mit den niedrigsten Ausprägungen, das obere Quartil die 25 Prozent der Werte mit den höchsten Ausprägungen. Quartile werden genutzt, um Unterschiede innerhalb einer Verteilung zu beschreiben und Gruppen miteinander zu vergleichen. In diesem Berichtsband wird dies vor allem bei der sozialen Herkunft (s. Kapitel 5) angewandt.
Querschnitt	Ein Querschnitt bezeichnet eine Untersuchung, bei der Daten zu einem bestimmten Zeitpunkt von unterschiedlichen Personen oder Gruppen erhoben werden. Im Gegensatz dazu beinhalten Längsschnittstudien mehrere Messzeitpunkte mit denselben Personen oder Gruppen. PISA ist eine Querschnittstudie, da jede Untersuchungskohorte jeweils nur einmal teilnimmt, kann aber auch eingeschränkt längsschnittlich betrachtet werden, da die Kompetenzwerte mehrerer PISA-Erhebungsrunden dieselbe Metrik aufweisen – allerdings nur auf Ebene der teilnehmenden Staaten.
Regression, latente	Eine Regression ist ein statistisches Verfahren, mit dem der Zusammenhang zwischen einer abhängigen Variable und einer oder mehreren unabhängigen Variablen untersucht wird. Sie zeigt, wie sich Änderungen in den Einflussgrößen auf die Zielgröße auswirken. Eine latente Regression erweitert dieses Konzept auf nicht direkt messbare Merkmale (► <i>latente Variable</i>), wie z. B. Kompetenzen in PISA, wobei diese nicht direkt messbaren Merkmale über Modelle (z. B. ► <i>Item-Response-Theorie</i>) operationalisiert werden müssen.

Begriff	Definition
Regressionskoeffizient	Der Regressionskoeffizient ist ein zentrales Konzept der Statistik und Datenanalyse, insbesondere innerhalb der Regressionsanalyse. Er beschreibt die Höhe der Beziehung zwischen unabhängigen Variablen und einer abhängigen Variable und zeigt, wie sich Veränderungen einer Variable auf eine andere auswirken können. In der einfachen linearen Regression stellt der Regressionskoeffizient die Steigung der Regressionsgeraden dar. Er gibt an, wie stark sich die abhängige Variable im Mittel verändert, wenn die unabhängige Variable um eine Einheit zunimmt. Damit hilft der Regressionskoeffizient, die Richtung und Stärke eines Zusammenhangs zwischen Variablen besser zu verstehen und zu bewerten.
Reliabilität	Reliabilität beschreibt die Zuverlässigkeit oder Genauigkeit eines Messinstruments und gibt an, inwieweit ein Test bei wiederholter Anwendung ähnliche Ergebnisse liefert. Sie setzt sich im Wesentlichen aus der ► <i>Varianz</i> der wahren (also systematischen) Merkmalsausprägung (vgl. auch ► <i>latente Variable</i>) und den potenziell messfehlerbehafteten beobachteten Werten zusammen. Je höher der Anteil der wahren Merkmalsausprägung am Gesamtergebnis, desto höher die Reliabilität.
Schulart	Zur PISA-Population in Deutschland gehören alle Schularten der Sekundarstufe I, die von 15-Jährigen besucht werden können. Für die Analysen dieses Berichtsbandes werden sie zu Gymnasien und nicht-gymnasialen Schularten zusammengefasst. Berufs- und Förderschulen können aufgrund kleiner Fallzahlen nicht gesondert ausgewertet werden; weitere nicht-gymnasiale Schularten sind zwischen den Bundesländern in Zugangsvoraussetzungen und angebotenen Schulabschlüssen zu heterogen für differenzierte Analysen. Verschiedene Schularten werden dabei als differenzielle Lern- und Entwicklungsmilieus verstanden, nicht als unmittelbare Ursachen von Kompetenzunterschieden.
Schwierigkeit, psychometrische	Die psychometrische Schwierigkeit ist ein zentrales Merkmal von Testaufgaben (Items). In der ► <i>Item-Response-Theorie (IRT)</i> stellt die Schwierigkeit einen Itemparameter dar. Dieser Schwierigkeitsparameter gibt an, welches Fähigkeitsniveau eine Person aufweisen muss, um eine Aufgabe mit einer bestimmten Wahrscheinlichkeit korrekt zu lösen. Im einfachen Rasch-Modell entspricht die Schwierigkeit dem Fähigkeitswert, bei dem die Lösungswahrscheinlichkeit 50 % beträgt.
Signifikanz	Signifikanz bezeichnet in der Statistik vereinfacht gesagt die Wahrscheinlichkeit, dass ein beobachtetes Ergebnis nicht durch Zufall zustande gekommen ist. Ein Ergebnis wird als signifikant bezeichnet, wenn die Wahrscheinlichkeit, dieses Ergebnis oder ein noch extremeres unter Annahme der Nullhypothese (kein Effekt oder Unterschied zwischen Gruppen) zu beobachten, unter einem festgelegten Schwellenwert (in diesem Berichtsband 5 %) liegt.

Begriff	Definition
Skalierung	Bei der Skalierung werden die Werte der mit unterschiedlichen Aufgaben gemessenen Kompetenzen in standardisierte Skalen überführt, sodass Kompetenzen vergleichbar gemacht werden können. In den PISA-Studien ist Skalierung besonders wichtig, da alle Schüler*innen nur einen kleinen Teil der Aufgaben (► <i>Booklet-Design</i>) bearbeiten und die jeweilige Schwierigkeit der unterschiedlichen Aufgaben berücksichtigt werden muss.
Sozialer Gradient	Der soziale Gradient beschreibt den Zusammenhang zwischen dem sozioökonomischen Status der Schüler*innen und ihren Kompetenzen. Ein hoher sozialer Gradient bedeutet, dass Schüler*innen mit höherem sozioökonomischen Status im Durchschnitt deutlich höhere Kompetenzen erreichen als Schüler*innen mit niedrigerem sozioökonomischen Status. Ein niedriger sozialer Gradient indiziert demgegenüber, dass dieser Zusammenhang schwächer ausgeprägt ist. Der soziale Gradient wird in PISA genutzt, um das Ausmaß sozialer Ungleichheiten zu beschreiben.
Staat	Im PISA-Kontext wird von teilnehmenden <i>Staaten und Ökonomien</i> gesprochen, da nicht alle beteiligten Einheiten in jedem Fall souveräne Staaten im völkerrechtlichen Sinn sind (s. Kapitel <i>Die PISA-Studie 2025</i>). Im Berichtsband werden diese Einheiten vereinfachend als <i>Teilnehmerstaaten</i> beziehungsweise <i>Staaten</i> bezeichnet. OECD-Staaten und EU-Staaten bilden dabei jeweils Teilmengen aller Teilnehmerstaaten. Neben den Durchschnitten für OECD und EU werden im Berichtsband ausgewählte <i>Vergleichsstaaten</i> verwendet. Die Befunde der weiteren Teilnehmerstaaten und Durchschnitte der Partnerstaaten finden sich im Online-Anhang.
Standardfehler	Der Standardfehler ist ein statistisches Maß, das angibt, wie stark ein geschätzter Wert aufgrund von Stichprobenziehung oder Messunsicherheiten vom tatsächlichen Wert in der Population abweichen kann. Durch ihn kann die Messungenauigkeit in der Analyse berücksichtigt werden. Der Standardfehler gibt an, in welchem Wertebereich der entsprechende Kennwert für die Grundgesamtheit mit einer bestimmten Irrtumswahrscheinlichkeit liegt. Über die Höhe dieser Irrtumswahrscheinlichkeit gibt es Konventionen: Oft wird ein Wert von 5 Prozent festgelegt. Bei maximal 5 Prozent Irrtumswahrscheinlichkeit entspricht die Spannweite des Wertebereichs für den tatsächlichen Wert ± 1.96 Standardfehler. Für konkret aus einer Stichprobe berechnete Intervalle lässt sich diese Wahrscheinlichkeitsaussage jedoch nicht treffen.
Stichprobe	Eine Stichprobe bezeichnet die Auswahl einzelner Elemente aus der Grundgesamtheit (vgl. ► <i>Grundgesamtheit</i>) zur Untersuchung in einer wissenschaftlichen Studie. Bestimmte Merkmale der Grundgesamtheit sollten entsprechend ihrem Anteil in der Stichprobe abgebildet werden, damit die Ergebnisse zuverlässig auf die gesamte Zielpopulation übertragen werden können. Verschiedene Verfahren der Stichprobenziehung können helfen, diese Repräsentativität zu gewährleisten.

Begriff	Definition
Stratifizierung	Stratifizierung bezeichnet die Aufteilung einer Population in Untergruppen (Strata) vor der Ziehung einer Stichprobe, um die Repräsentativität sicherzustellen. Bei PISA, werden zum Beispiel Schulen nach Bundesland oder Schultyp stratifiziert, bevor zufällig bestimmte Schulen ausgewählt werden. Dadurch wird sichergestellt, dass Schulen aller Bundesländer und aller Schulformen in der Stichprobe vertreten sind.
Testform	Im Kontext computerbasierter Tests werden die unterschiedlichen Zusammenstellungen von Aufgaben für alle Schüler*innen (vgl. ► <i>Booklet-Design</i>) meist Testformen genannt. Diese Zusammenstellung erfolgt in Aufgabengruppen (sogenannten Clustern), die wiederum aus mehreren Aufgabeneinheiten (Units) mit Aufgaben (Items) bestehen. Alle Schüler*innen bearbeiten also eine Testform am Computer, die mehrere Cluster von Aufgaben umfasst.
Trendanalyse	Analysen im Zeitverlauf (Trendanalysen) untersuchen, wie sich Befunde über mehrere Erhebungsrunden hinweg verändert haben. PISA ist in erster Linie auf den internationalen Vergleich von Bildungssystemen zu einem bestimmten Zeitpunkt angelegt, erlaubt durch aufwendige Verknüpfungen der Erhebungen aber auch Aussagen über Entwicklungen über die Zeit. Besonders gut ist dies für Kompetenzen möglich, eingeschränkter auch für andere Merkmale der Schüler*innen und ihrer Lernkontexte. Aussagen beziehen sich dabei stets auf Bildungssysteme, nicht auf individuelle Entwicklungsverläufe.
Variable, latente (auch: latente Fähigkeitsdimension)	Eine latente Variable ist ein Merkmal, das nicht direkt beobachtbar ist, wie etwa mathematische Kompetenz oder Lesefähigkeit. Es wird durch mehrere beobachtbare Indikatoren (z. B. ► <i>Items</i>) erschlossen und statistisch modelliert.
Varianz	Varianz beschreibt, wie stark die Werte einer Variable um ihren Mittelwert streuen, und ist damit ein Maß für die Unterschiede innerhalb einer Gruppe. Eine hohe Varianz bedeutet, dass die Werte im Mittel weit auseinanderliegen, eine niedrige Varianz weist auf ähnliche mittlere Ausprägungen der Werte hin. Die Quadratwurzel aus der Varianz bildet die ► <i>Standardabweichung</i> .
Weighted Likelihood Estimate (WLE)	Biasreduzierter Schätzwert auf Basis der ► <i>Item-Response-Theorie</i> , der die Ausprägung einer Person auf einem latenten Merkmal beschreibt. In PISA werden WLEs insbesondere zur Skalierung von Fragebogenmerkmalen bzw. Hintergrundindizes verwendet. Viele dieser Skalen werden so transformiert, dass sie in der OECD-Referenzpopulation in einer bestimmten Erhebungsrunde einen Mittelwert von 0 und eine Standardabweichung von 1 aufweisen.

Wenn du über deine Schule nachdenkst: Wie sehr stimmst du den folgenden Aussagen zu?
(Bitte eine Antwort pro Zeile auswählen.)

Frage

Items

	Stimme völlig zu	Stimme eher zu	Stimme eher nicht zu	Stimme überhaupt nicht zu
Ich fühle mich als Außenseiter*in (oder von Dingen ausgeschlossen) in der Schule.	☐	☐	☐	☐
In der Schule finde ich leicht neue Freund*innen.	☐	☐	☐	☐
Ich habe das Gefühl, zu dieser Schule zu gehören.	☐	☐	☐	☐
Ich fühle mich unbehaglich und fehl am Platz in dieser Schule.	☐	☐	☐	☐
Andere Schüler*innen scheinen mich zu mögen.	☐	☐	☐	☐
Ich fühle mich in dieser Schule einsam.	☐	☐	☐	☐

Antwortskala

Skala

Diese Skala erfasst das Merkmal *schulisches Zugehörigkeitsgefühl*