Forschungsmonitor Schule
Rezensionen für die Praxis

Johannes Rosendahl

Rezension zu

Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (2013). *PISA 2012*. Münster u. a.: Waxmann.

Kommentierter Kurzbefund

Im Rahmen von PISA 2012 wurden zum fünften Mal Leistungen von Fünfzehnjährigen in Lesen, Mathematik und Naturwissenschaften untersucht: Die Leistungen der Fünfzehnjährigen in Deutschland liegen über dem Durchschnitt der OECD-Staaten und fallen im Vergleich zu früheren Jahren besser aus. Die Anteile der Fünfzehnjährigen, die die gestellten Mindestanforderungen nicht erfüllen (Risikogruppe), sind kleiner geworden und liegen in Lesen bei 15 %, in Mathematik bei 18 % und in Naturwissenschaften bei 12 %. Die Anteile der Fünfzehnjährigen auf den obersten Kompetenzstufen (Spitzengruppe) haben sich kaum verändert und liegen in Lesen bei 9 %, in Mathematik bei 18 % und in Naturwissenschaften bei 12 %.

PISA liefert nur eingeschränkt Erkenntnisse darüber, wie Leistungsunterschiede zwischen Staaten zustande kommen und worauf Veränderungen über die Zeit zurückzuführen sind. Schülerleistungen unterliegen vielfältigen Bedingungen, deren spezifische Bedeutsamkeit aufgrund der (querschnittlichen) Untersuchungsanlage mit PISA nicht sicher bestimmt werden kann. Aus anderen Untersuchungen lässt sich jedoch schließen, dass die Leistungsverbesserungen in Deutschland großteils durch Merkmale erklärbar sind, die nicht unmittelbar mit einer veränderten Schul- und Unterrichtsqualität oder pädagogischen Maßnahmen zusammenhängen, z. B. durch eine gestiegene Gymnasialquote, da die Lernentwicklung an Gymnasien im Durchschnitt günstiger verläuft als an anderen Schulformen, durch zunehmend mehr Fünfzehnjährige in höheren Jahrgangsstufen aufgrund früherer Einschulungen und weniger Klassenwiederholungen und durch veränderte Voraussetzungen der Schülerschaft mit Migrationshintergrund, u. a. ist der Anteil der im Ausland geborenen Fünfzehnjährigen gesunken.

Hintergrund

PISA (Programme for International Student Assessment) ist ein internationaler Schulleistungsvergleich, der Mitte der 1990er Jahre von der OECD (Organisation for Economic Co-operation and Development) initiiert wurde und seit 2000 im dreijährigen Turnus durchgeführt wird. Die deutsche Teilnahme erfolgt auf Beschluss der Kultusministerkonferenz der Länder (KMK) und ist Bestandteil der Gesamtstrategie zum Bildungsmonitoring (KMK 2006, 2015). Im Jahr 2012 wurden zum fünften Mal Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern mithilfe standardisierter Tests erhoben, der Schwerpunkt lag wie 2003 auf Mathematik (Hauptdomäne).

Ziele

PISA ist ein Indikatorensystem, das es ermöglicht, Strukturdaten von Bildungssystemen mit Output-Daten zu verbinden, um Zusammenhänge zu analysieren und gegebenenfalls Schlussfolgerungen über die Effektivität von Bildungssystemen zu ziehen. Die Leitfrage von PISA ist, wie gut es Staaten gelingt, junge Menschen auf die Anforderungen einer Teilhabe an der Gesellschaft, auf das Erwachsenenleben

CC BY NC ND

und das weitere Lernen über die Lebensspanne vorzubereiten. Entsprechend werden Daten zu lernrelevanten Bedingungen in Unterricht und Schule gewonnen und Indikatoren zu Disparitäten bereitgestellt, um systematische Zusammenhänge zwischen Leistungsmaßen und beispielsweise Merkmalen der sozialen Herkunft, der Zuwanderung oder dem Geschlecht berichten zu können.

Anhand der Vergleiche zwischen Staaten, der Veränderungen über die Zeit auf nationaler Ebene und durch Bezug auf inhaltlich definierte Leistungsstandards soll mehr oder weniger ‚gute' Praxis identifiziert und ein Benchmarking zwischen Staaten ermöglicht werden. Mit den Erhebungen im dreijährlichen Rhythmus sollen für die Steuerung des Bildungssystems klar interpretierbare Trendinformationen gewonnen werden, um Wirkungen von Maßnahmen abschätzen und problematische Entwicklungen frühzeitig erkennen zu können. Erwartet wird von den PISA 2012-Befunden für Deutschland u. a. Auskunft darüber, „inwieweit die Ziele von Maßnahmen, die mit der Einführung länderübergreifender Bildungsstandards ab dem Jahre 2003 umgesetzt wurden (vgl. KMK, 2010), inzwischen erreicht worden sind: ein insgesamt höheres Niveau bei deutlicher Reduzierung schwacher Leistungen und die Verringerung von Disparitäten" (Sälzer & Prenzel 2013, S. 11).

Konzepte

Als Testdomänen werden Lesen, Mathematik und Naturwissenschaften berücksichtigt, da sie als primär relevant für die weitere Bildungsbiographie, das Berufsleben wie auch die Partizipation an Kultur und Gesellschaft angesehen werden. Damit werden einerseits „relevante Aspekte einer Grundbildung mit Blick auf Allgemeinbildung erfasst" (Sälzer & Prenzel 2013, S. 15), andererseits wird lediglich „ein Ausschnitt aus dem Spektrum allgemeiner Bildung wie auch allgemeinbildender Schulfächer untersucht", was es zu beachten gilt, um „übergeneralisierende Aussagen über Bildung und die Bildungsqualität der Schule zu vermeiden" (ebd.).

Den mathematischen und naturwissenschaftlichen Tests liegt ein Kompetenzkonzept zugrunde, das nicht curricular ausgerichtet ist, sondern der Idee einer funktionalen Grundbildung folgt. ‚Funktional' meint relevant für die jetzige und spätere Teilhabe an einer Kultur und anschlussfähig für nachfolgendes Lernen. Infolgedessen beziehen sich die Testaufgaben auf verschiedene Inhaltsbereiche, Verfahrensweisen und Situationen, die für ein grundlegendes fachliches Verständnis, die Wissensanwendung in vielfältigen Problemsituationen und für das Weiterlernen im Verlauf der weiteren Bildungsbiographie als bedeutsam eingeschätzt werden.

Die Tests sind so angelegt, dass weniger simple Reproduktion, sondern vielmehr das flexible Anwenden des Wissens und das Lösen domänenspezifischer Probleme gefordert sein sollen. Lesekompetenz wird eine bedeutsame Rolle zugesprochen, da in fast allen Domänen Wissen zum größten Teil über Texte bereitgestellt und angeeignet wird. Die Autorinnen und Autoren weisen auf Gemeinsamkeiten und Unterschiede zwischen den Kompetenzkonstrukten von PISA und den nationalen Bildungsstandards hin: Beispielsweise wird mathematische Grundbildung in PISA nach Inhalten, Prozessen und Kontexten strukturiert. Die Bildungsstandards für Mathematik unterscheiden demgegenüber inhaltsbezogene und allgemeine mathematische Kompetenzen, die drei Anforderungsbereichen zugeordnet werden. Die inhaltsbezogenen Kompetenzen (Raum und Form, Daten und Zufall etc.) entsprechen in etwa den Inhalten, die in PISA verwendet werden (Raum und Form, Unsicherheit und Daten etc.). Die allgemeinen mathematischen Kompetenzen (Argumentieren, Problemlösen etc.) sind im Wesentlichen vergleichbar mit den ‚fundamentalen Fähigkeiten', die mehr oder minder typische Arbeitsweisen meinen, die beim Lösen von Aufgaben benötigt werden (argumentieren, Problemlösestrategien entwickeln etc.). Im Gegensatz zu PISA werden in den nationalen Bildungsstandards die Prozesse (formulieren, anwenden, interpretieren) und Kontexte (persönlich, beruflich, gesellschaftlich, wissenschaftlich) nicht explizit

erwähnt (Sälzer et al. 2013, S. 57).

Forschungslage
In der ersten PISA-Erhebung im Jahr 2000 lagen die Leistungen der Fünfzehnjährigen in Deutschland in allen drei Testbereichen unter dem OECD-Durchschnitt. Bei den folgenden Erhebungen ergaben sich deskriptiv kontinuierlich bessere Leistungen, die für Lesen und Mathematik im Jahr 2009 in Bezug auf den Zeitpunkt der erstmaligen Testung als Hauptdomäne (Lesen 2000, Mathematik 2003) signifikant wurden. In PISA 2009 wurden in Mathematik (erstmals) und Naturwissenschaften (wie 2006) überdurchschnittliche Leistungen erzielt, in Lesen lagen die Leistungen wie bereits 2006 im OECD-Durchschnitt.

Design

Stichprobenziehung
Für die Testteilnahme wurde in einem mehrstufigen Auswahlverfahren eine Zufallsstichprobe gezogen (Heine et al. 2013, S. 309 ff.). Im ersten Schritt wurden aus einer Liste mit allen Schulen, an denen sich in Deutschland potentiell Fünfzehnjährige befinden, 247 Schulen gezogen. Zu diesem Zweck werden die Schulen nach Ländern und Schulformen gruppiert. Die Stichprobenziehung der Schulen erfolgt so, dass der erwartete gruppenbezogene Anteil der Schülerinnen und Schüler näherungsweise dem gruppenbezogenen Anteil der Schülerinnen und Schüler in der Grundgesamtheit entspricht. Anschließend wurden an jeder Schule 25 Fünfzehnjährige per Zufallsauswahl für die Teilnahme bestimmt.

Gewichtung
Da die Wahrscheinlichkeit, in die Stichprobe zu gelangen, aufgrund des mehrstufigen Auswahlverfahrens nicht für alle Schulen und Fünfzehnjährigen identisch ist, werden die Fälle gewichtet. Diese Gewichtung wird mithilfe fünf weiterer Faktoren korrigiert, um beispielsweise Effekte von Ausfällen einzelner Schulen, die Nichtberücksichtigung von Fünfzehnjährigen in Staaten, die nur in einer bestimmten Jahrgangsstufe testen, oder Abweichungen der Schulgröße von den Werten der amtlichen Statistik auszugleichen.

Teilnahmequote
An PISA 2012 beteiligten sich 65 Staaten mit ca. 500.000 fünfzehnjährigen Schülerinnen und Schülern. In Deutschland nahm eine auf Bundesebene repräsentative Stichprobe von 230 Schulen mit 5.001 Schülerinnen und Schülern teil. Die gewichtete Teilnahmequote auf Schulebene beträgt 98,3 %, auf Ebene der Schülerinnen und Schüler 93,2 %.

Skalierung
Um die Kompetenzen der Fünfzehnjährigen in den Testdomänen (Lesen, Mathematik, Naturwissenschaften) möglichst umfassend und präzise zu ermitteln, werden viele verschiedene Testaufgaben eingesetzt. Da nicht jede teilnehmende Person alle Aufgaben einer Testdomäne bearbeiten kann, werden verschiedene Hefte mit unterschiedlichen Aufgaben eingesetzt (Multi-Matrix-Design). Die reine Anzahl der richtig gelösten Aufgaben gibt daher nicht unmittelbar Aufschluss über die Kompetenzausprägung einer Person, sondern die Schwierigkeiten der jeweils bearbeiteten Aufgaben muss berücksichtigt werden. Hierzu dient die sogenannte Skalierung. Das Verfahren gründet auf der Annahme, dass eine individuell ausgeprägte Personeneigenschaft ursächlich für das beobachtbare Antwortverhalten ist. Eine einfache Aufgabe wird von sehr leistungsstarken Personen mit hoher Wahrscheinlichkeit richtig bearbeitet, während eine schwierige Aufgabe von weniger leistungsstarken

Personen eher nicht gelöst wird.

Diese Annahme lässt sich auf Grundlage der Item Response Theory (IRT) mathematisch-formal modellieren. Die Kompetenzen der Schülerinnen und Schüler werden mithilfe des ausgewählten IRT-Modells (z. B. Rasch-Modell) ermittelt, wobei die Modellparameter zuvor so bestimmt werden, dass die Wahrscheinlichkeit für das Auftreten der beobachteten Datenmatrix am größten ist. Da die Ergebnisse der drei Testdomänen zusammenhängen, wurde eine Marginal-Maximum-Likelihood-Schätzung (MML) in einer mehrdimensionalen Rasch-Skalierung mit dem in der Software ConQuest (Adams, Wu & Wilson 2012) implementierten Mixed-Coefficents-Multinomial-Logit-Modell (MCML – Adams & Wu 2007) vorgenommen. Auf diese Weise werden die Zusammenhänge der drei Testdomänen geschätzt, wobei die Minderung der korrelativen Beziehungen aufgrund etwaiger Messungenauigkeiten korrigiert wird. Daneben werden diese Zusammenhänge genutzt, um genauere Schätzungen der Kompetenzwerte zu erhalten. Weiterhin werden Hintergrundvariablen bei der Schätzung der Kompetenzausprägungen berücksichtigt, z. B. Angaben der Schülerinnen und Schüler sowie der Eltern zu schulbezogenen Einstellungen.

Equating

Für die Untersuchung von Veränderungen ist es erforderlich, die einzelnen PISA-Erhebungen mithilfe gemeinsamer Aufgaben zu verbinden. Diese Aufgaben werden in mehreren aufeinanderfolgenden PISA-Erhebungen unverändert eingesetzt. Unter der Annahme, dass diese Aufgaben zu allen Zeitpunkten die gleiche Kompetenz erfassen, wird die Skala der aktuellen Erhebung auf die Referenzskala transformiert (Equating). Diese Transformation ist mit einer gewissen Unsicherheit verknüpft, da die eingesetzten gemeinsamen Aufgaben möglicherweise unterschiedliche Schwierigkeiten zwischen den PISA-Erhebungsrunden aufweisen, d. h. die Transformation könnte geringfügig anders ausfallen, wenn andere gemeinsame Aufgaben eingesetzt würden.

Kompetenzstufen

Kompetenzen werden, vergleichbar der Körpergröße eines Menschen, als kontinuierliche Merkmale aufgefasst. Zur besseren Interpretation und zum leichteren Vergleich werden die mithilfe der Skalierung ermittelten Kompetenzwerte auf eine Skala transformiert, deren Mittelwert bei ihrem ersten Auftreten als Hauptdomäne über alle teilnehmenden OECD-Staaten hinweg auf 500 Punkte bei einer Standardabweichung von 100 Punkten festgelegt wurde. Domänenübergreifend erreichen etwa 2/3 aller Jugendlichen Kompetenzwerte im Bereich von 400 bis 600 Punkten. Zur besseren inhaltlichen Beschreibung werden sechs Kompetenzstufen unterschieden. Die Beschreibungen der Kompetenzstufen basieren auf den Anforderungen der Aufgaben, die bei einem bestimmten Kompetenzniveau mit einer festgelegten Wahrscheinlichkeit (50 %-70 %) gelöst werden.
Die Zuordnung zu einer bestimmten Kompetenzstufe gibt damit an, welche Arten von Aufgaben mit hoher Wahrscheinlichkeit gelöst werden können und welche nicht. Fünfzehnjährige auf Kompetenzstufe I werden als ‚Risikogruppe' angesehen. Diese Jugendlichen verfügen allenfalls über elementares Wissen und sind weit vom Erreichen der Ziele der Sekundarstufe I entfernt. Das niedrige Kompetenzniveau lässt Schwierigkeiten bei der weiteren schulischen oder beruflichen Bildung erwarten.

Ergebnisse

Mathematik

Die Fünfzehnjährigen in Deutschland erreichen überdurchschnittliche 514 Punkte (OECD: 494), wobei einige OECD-Länder besser abschneiden: Korea (554), Japan (536), Estland (521), Niederlande (533)

und die Schweiz (531). Bei der nationalen Follow-up-Erhebung zu PISA 2003 belief sich der Leistungszuwachs zwischen der 9. und 10. Jahrgangsstufe auf knapp 30 Punkte. Die Streuung (Standardabweichung) der Kompetenzwerte (96) liegt im Bereich des OECD-Durchschnitts (92). Einige Staaten weisen überdurchschnittliche Leistungen bei geringen Streuungen auf (Estland 521-81, Dänemark 500-82, Finnland 519-85, Irland 501-85). Ein hohes Leistungsniveau ist demnach nicht zwangsläufig mit großen Unterschieden innerhalb der Schülerschaft verknüpft.

In einigen europäischen Staaten gehören deutlich weniger Jugendliche zur Risikogruppe (höchstens Kompetenzstufe I) als in Deutschland (18 %), beispielsweise in den Niederlanden (15 %), Polen (14 %), in der Schweiz und Finnland (12 %) sowie in Estland (11 %). Jugendliche, die nicht über Kompetenzstufe I hinauskommen, verfügen bestenfalls über elementares mathematisches Wissen und sind weit vom Erreichen der Ziele entfernt, die mit dem Mathematikunterricht in der Sekundarstufe I verbunden sind, ihr niedriges Kompetenzniveau lässt Schwierigkeiten bei ihrer weiteren schulischen oder beruflichen Bildung erwarten.

Im Hinblick auf die Spitzengruppe (Kompetenzstufen V, VI) ist der Anteil in Deutschland (18 %) zwar im OECD-Vergleich (13 %) überdurchschnittlich, aber geringer als in der Schweiz (21 %), den Niederlanden und in Belgien (19 %). Im Vergleich zu 2003 haben sich die Leistungen in Deutschland, Italien, Portugal und Polen verbessert, wohingegen die Mehrzahl der europäischen Länder abnehmende oder gleichbleibende Leistungen zu verzeichnen hat.

Naturwissenschaften

Die Fünfzehnjährigen in Deutschland erreichen überdurchschnittliche 524 Punkte (OECD: 501), wobei einige OECD-Staaten besser abschneiden: Korea (538), Japan (547), Estland (541) und Finnland (545). Bei der nationalen Follow-up-Erhebung zu PISA 2003 belief sich der Leistungszuwachs zwischen der 9. und 10. Jahrgangsstufe auf 21 Punkte. Die Standardabweichung (95) liegt im OECD-Durchschnitt (93), in einigen erfolgreicheren Ländern sind geringere Streuungen zu verzeichnen (Estland: 80; Korea: 82, Polen: 86). Diesen Staaten gelingt es, die naturwissenschaftliche Kompetenz in der Breite zu fördern, ohne die Spitzengruppe zu vernachlässigen.

Der Anteil der Risikogruppe liegt in Deutschland (12 %) unter dem OECD-Durchschnitt (18 %). In anderen Ländern ist die Risikogruppe kleiner: Estland (5 %), Korea (7 %), Finnland (8 %), Japan und Polen (9 %). Jugendliche, die zur Risikogruppe gehören, verfügen über ein geringes naturwissenschaftliches Verständnis, das für viele Ausbildungen unzureichend ist und kaum begründete Entscheidungen zulässt (z. B. zu Fragen der Ernährung und Gesundheit oder auch der Umwelt). Ihnen fehlen grundlegende naturwissenschaftliche Kompetenzen, um einfachste Aufgaben zu lösen. Im Alltag wird es ihnen schwerfallen, einfache naturwissenschaftliche Probleme oder Zusammenhänge zu verstehen oder aus Informationen Schlussfolgerungen zu ziehen.

Die Spitzengruppe (Kompetenzstufen V, VI) ist in Deutschland größer (12 %) als im OECD-Durchschnitt (8 %), ähnlich wie in Australien (14 %), Neuseeland und Estland (13 %), aber kleiner als in Japan (18 %) und Finnland (17 %). Japan und Finnland gelingt es, kompetenzschwache und kompetenzstarke Jugendliche gleichermaßen erfolgreich zu fördern.

Die Verbesserung in Deutschland seit der ersten Testung von Naturwissenschaften als Hauptdomäne 2006 (+9 Punkte) lässt sich nicht statistisch gegen den Zufall absichern, sie wird als eine Stabilisierung auf einem hohen, gegenüber dem OECD-Mittelwert überdurchschnittlichen Niveau bezeichnet.

Lesen

Die Fünfzehnjährigen in Deutschland erreichen überdurchschnittliche 508 Punkte (OECD: 496), wobei einige OECD-Staaten besser abschneiden: Korea (536), Japan (538), Irland (533), Finnland (524), Kanada (523), Polen (518) und Estland (516). Die Standardabweichung (91) liegt signifikant unter dem OECD-Durchschnitt (94). Einige Staaten haben höhere Leistungen bei geringeren Streuungen (Korea: 87, Irland: 86, Polen: 87, Estland: 80). Je kleiner die Streuung ausfällt, umso geringer sind die Unterschiede zwischen den Leistungsstärksten und Leistungsschwächsten. Im Vergleich zu PISA 2000, wo es in Deutschland eine im internationalen Vergleich hohe Leistungsstreuung gab (111), zeichnet sich damit in Deutschland ein Trend in Richtung homogenerer Leistungen ab.

Auf Kompetenzstufe I und darunter befinden sich in Deutschland 15 % der Fünfzehnjährigen, was signifikant weniger ist als im OECD-Durchschnitt (18 %). Für diese Jugendlichen kann ein beträchtliches Risiko für die weitere Schul- und Berufsbiographie wie auch für die gesellschaftliche Teilhabe angenommen werden.

In zwei Staaten liegt der Anteil der Risikogruppe unter 10 % (Korea: 8 %, Estland: 9 %). Das Beispiel dieser Staaten zeigt, dass es prinzipiell möglich ist, den Anteil der schwachen Leserinnen und Leser sehr gering zu halten.

Auf den obersten Kompetenzstufen V und VI liegen im OECD-Durchschnitt 8 % der Schülerinnen und Schüler, der Wert für Deutschland liegt im internationalen Durchschnitt (9 %). Damit fehlt hierzulande eine Spitzengruppe, die vergleichbar ist mit der in Staaten wie Japan (19 %), Korea (14 %) oder auch Neuseeland (14 %), Finnland (14 %), Frankreich und Kanada (13 %) sowie Belgien (12 %).

Die Leseleistungen in Deutschland waren bereits 2009 im Vergleich zu 2000 signifikant höher (484 vs. 497 Punkte), doch auch der Leistungszuwachs von 2009 bis 2012 (508 Punkte) ist statistisch signifikant. Daher wird die Zunahme der Lesekompetenz in Deutschland als substanziell erachtet.

Diskussion und Einschätzung

Mit PISA werden, initiiert und koordiniert durch die OECD, Leistungen von Fünfzehnjährigen aus verschiedenen Bildungssystemen mit Konzepten und Methoden des derzeit vorherrschenden Forschungsparadigmas der empirisch-quantitativ ausgerichteten Kompetenzforschung ermittelt. Diese Projektanlage lässt sich unter verschiedenen Gesichtspunkten (u. a. politisch, gesellschaftlich, wissenschaftlich) hinterfragen und kritisieren (z. B. Bank & Heidecke 2009). An dieser Stelle steht im Fokus, inwiefern die Darstellungen der Schülerleistungen in der vorliegenden Publikation, die für die öffentliche Rezeption der Ergebnisse von PISA 2012 in Deutschland von zentraler Bedeutung ist, aussagekräftig und relevant sind im Hinblick auf die angeführten Ziel- und Fragestellungen: „In Abstimmung mit der KMK und dem BMBF dient der vorliegende Bericht dazu, einen systematischen Überblick über die Ergebnisse für Deutschland zu liefern und soll durch Hinweise auf spezielle Bedingungen und Maßnahmen helfen, die Ergebnisse und vor allem Trends zu interpretieren“ (Sälzer & Prenzel, S. 12).

Ein wesentliches Kennzeichen von PISA ist, dass aufgrund des querschnittlichen Designs und der kaum eingrenzbaren Merkmale, die Schülerleistungen beeinflussen können, keine kausalen Aussagen zu Ursachen von Leistungsunterschieden zwischen Staaten oder zu Leistungsveränderungen auf nationaler Ebene möglich sind, worauf von den Autorinnen und Autoren verschiedentlich hingewiesen wird (Sälzer

& Prenzel 2013, S. 18, S. 42). Doch bereits aus dem Vorwort von KMK und BMBF wird ersichtlich, dass diese Hinweise unbeachtet bleiben: „Die enttäuschenden Ergebnisse von PISA 2000 waren für Deutschland ein bildungspolitischer Weckruf. Seitdem ist viel passiert: Maßnahmen zur Verbesserung der Sprachkompetenz, (…) PISA 2012 zeigt, dass sich diese Anstrengungen lohnen“ (Dorgerloh & Wanka 2013, S. 9).

Befördert werden solche Überinterpretationen durch die Einleitungen der Kapitel zu den einzelnen Testdomänen, in denen zwischenzeitlich ergriffene Maßnahmen aufgeführt werden und beispielsweise bilanziert wird: „Die Anstrengungen um eine verbesserte Bildung insbesondere im letzten Jahrzehnt haben sich folglich gelohnt (…) Möglicherweise hat die Einführung von Bildungsstandards … dazu beigetragen, dass sich die Kompetenz in der Mathematik … positiv entwickelt hat“ (Sälzer et al. 2013, S. 93 f.). Zu kritisieren ist, dass alternative Erklärungsansätze nicht ansatzweise thematisiert werden. Entsprechend wurde die öffentliche Rezeption der PISA 2012-Ergebnisse beispielsweise vom Max-Planck-Institut für Bildungsforschung, das für die Durchführung von PISA 2000 federführend zuständig war, unter der Rubrik ‚Unstatistik des Monats‘ folgendermaßen kommentiert: „Dabei wird aber außer Acht gelassen, dass sich über die Jahre hinweg die Zusammensetzung der Stichprobe verändert hat. Berücksichtigt man die Unterschiede der befragten Schüler in den verschiedenen Stichproben beispielsweise hinsichtlich des Alters, des Geschlechts, des sozioökonomischen Hintergrunds und des Migrationshintergrunds, haben sich die Leistungen der Schüler in Deutschland seit 2003 nicht signifikant verändert“ (Max-Planck-Institut 2013).

Die Stellungnahme des Max-Planck-Instituts entspricht Auswertungen aus dem OECD-Bericht zu PISA 2012 (OECD 2013, S. 313 ff., S. 335, S. 414, S. 428). In anderen Untersuchungen wurden die Leistungsveränderungen in Deutschland mit ähnlichen Ergebnissen analysiert (vgl. Ehmke, Klieme, Stanat 2013, Mildner, Hochweber & Frey 2013). Darüber hinaus betrifft diese Problematik nicht nur die Interpretation von Leistungsveränderungen auf nationaler Ebene, sondern auch den Vergleich der Schülerleistungen zwischen Staaten. Berücksichtigt man beispielsweise den Zuwanderungshintergrund, indem man Jugendliche, die selbst oder deren Eltern zugewandert sind, aus den Analysen ausschließt, schneidet Deutschland vergleichbar mit Finnland ab, dessen besseres Abschneiden in PISA 2000 in Deutschland großes Interesse hervorrief und intensive Reiseaktivitäten auslöste (Kobarg & Prenzel 2009).

Einerseits werden demnach Rückschlüsse auf die Wirksamkeit von Bildungsmaßnahmen und -systemen nahe gelegt, andererseits werden mögliche negative Folgen beispielsweise der Bildungsexpansion und von strukturellen Änderungen im Schulsystem bagatellisiert. So wird die tendenzielle Abnahme der Mathematikleistungen an Gymnasien dahingehend relativiert, dass, wenn man entsprechend dem Anwachsen der Gymnasialquote die gymnasiale Stichprobe um die unteren 5 Perzentile der Leistungsverteilung reduziert, eine mit 2003 vergleichbare Leistung resultiere (Sälzer et al 2013, S. 91). Es ist jedoch nicht auszuschließen, dass zur Abnahme der Mathematikleistungen an Gymnasien die gymnasiale Schulzeitverkürzung und/oder ein wachsender Anteil Leistungsschwächerer und ein darauf angepasster Unterricht beigetragen haben.

Fazit

Die Darstellung der Leistungen von Fünfzehnjährigen in Lesen, Mathematik und Naturwissenschaften durch die nationale Projektleitung ist abgesehen vom deskriptiven Bericht irreführend, denn bei der Diskussion der Ergebnisse wird die Bedeutsamkeit von Stichprobenmerkmalen nicht berücksichtigt. Vielmehr wird der Eindruck erweckt, Maßnahmen der Bildungsadministration hätten die Ergebnisse positiv beeinflusst, mögliche negative Effekte werden hingegen relativiert und ausgeklammert.

Weiterführende Hinweise im Kontext

An den hohen Interkorrelationen insbesondere zwischen Mathematik und Naturwissenschaften entzündete sich eine Diskussion, ob mit den Tests unterschiedliche Konstrukte erfasst werden und ob etwas anderes als kognitive Leistungsfähigkeit (Intelligenz) gemessen wird. Diese Fragen werden beispielsweise in den nachstehenden Artikeln erörtert.

- Baumert, J., Brunner, M., Lüdtke, O. & Trautwein, U. (2007). Was messen internationale Schulleistungsstudien? – Resultate kumulativer Wissenserwerbsprozesse. *Psychologische Rundschau*, 58 (2), 118–145.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57 (2), 69-86.
- Rindermann, H. (2007). Intelligenz, kognitive Fähigkeiten, Humankapital und Rationalität auf verschiedenen Ebenen. *Psychologische Rundschau*, 58 (2), 137-145.

Literatur

- Adams, R. & Wu, M. (2007). A generalized form of the Rasch Model. In M. von Davier & C. H. Carstensen (Hrsg.), *Multivariate and mixture distribution Rasch Models: Extensions and applications*, 57–75. New York: Springer.
- Adams, R., Wu, M. L. & Wilson, M. R. (2012). *ACER ConQuest 3.0 [Computer software].* Melbourne: Australian Council for Educational Research (ACER).
- Bank, V. & Heidecke, B. (2009). Gegenwind für PISA. Ein systematisierender Überblick über kritische Schriften zur internationalen Vergleichsmessung. *Vierteljahresschrift für wissenschaftliche Pädagogik*, 85 (3), 361-372.
- Dorgerloh, S. & Wanka, J. (2013). Vorwort des Präsidenten der Kultusministerkonferenz und der Bundesministerin für Bildung und Forschung. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland*, 9. Münster u. a.: Waxmann.
- Ehmke, T., Klieme, E. & Stanat, P. (2013). Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009. In N. Jude & E. Klieme (Hrsg.), *PISA 2009 - Impulse für die Schul- und Unterrichtsforschung*, 132-150. Weinheim u. a.: Beltz.
- Heine, J.-H., Sälzer, C., Borchert, L., Sibberns, H. & Mang, J. (2013). Technische Grundlagen des fünften internationalen Vergleichs. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland*, 309-346. Münster u. a.: Waxmann.
- Hohn, K., Schiepe-Tiska, A., Sälzer, C. & Artelt, C. (2013). Lesekompetenz in PISA 2012: Veränderungen und Perspektiven. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland*, 217-244. Münster u. a.: Waxmann.
- Kobarg, M. & Prenzel, M. (2009). Stichwort: Der Mythos der nordischen Bildungssysteme. *Zeitschrift für Erziehungswissenschaften*, 597-615.
- Kultusministerkonferenz (KMK) (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Bonn: Wolters Kluwer.
- Kultusministerkonferenz (KMK) (2015). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (überarbeitete Fassung)*. München: Wolters Kluwer.
- Mildner, D., Hochweber, J. & Frey, A. (2013). Vergleichende Analysen der Kompetenzen von Fünfzehnjährigen und Neuntklässlern in den deutschen PISA-Erhebungen 2003 bis 2009. *Zeitschrift für Pädagogik*, Beiheft 59, 151-171.

- OECD (2013). *PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können (Band I): Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften.* Bielefeld: Bertelsmann.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? *Psychologische Rundschau*, 57 (2), 69-86.
- Sälzer, C. & Prenzel, M. (2013). PISA 2012 – eine Einführung in die aktuelle Studie. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland*, 11-45. Münster u. a.: Waxmann.

Reflexionsfragen für die Praxis

Nachfolgende Reflexionsfragen sind ein Angebot, die Befunde der rezensierten Studie auf das eigene Handeln als Lehrkraft oder Schulleitungsmitglied zu beziehen und zu überlegen, inwiefern sich Anregungen für die eigene Handlungspraxis ergeben.

Die Befunde der rezensierten Studien sind nicht immer generalisierbar, was z. B. in einer begrenzten Stichprobe begründet ist. Aber auch in diesen Fällen können die Ergebnisse interessante Hinweise liefern, um über die eigene pädagogische und schulentwicklerische Praxis zu reflektieren.

Reflexionsfragen für Lehrkräfte:

- Welche Leistungsstandards gelten für meine Schülerinnen und Schüler (z. B. nationale Bildungsstandards, länder- oder schulspezifische Vorgaben)?
- Welche Verfahren kenne und nutze ich, um Lernstände bezogen auf diese Standards zu diagnostizieren?
- In welchen Teilbereichen sind die Leistungen meiner Klasse(n) schon gut, wo bestehen noch Defizite?
- Inwiefern weichen in meiner Klasse die Anteile der Schülerinnen und Schüler, die keine akzeptablen Leistungen bzw. Spitzenleistungen erbringen, von den Studienergebnissen ab?
- Welche meiner Schülerinnen und Schüler erreichen keine akzeptablen Leistungen in Hinblick auf die Standards, welche können Spitzenleistungen erzielen?
- Wie passe ich den Unterricht an, um Defizite abzubauen und Spitzenleistungen zu fördern?
- Wie stellen meine Kolleginnen und Kollegen das Erreichen von Standards sicher und wie diagnostizieren sie Schülerleistungen?
- Gibt es abgestimmte Verfahrensweisen zur standardbezogenen Diagnostik an meiner Schule?
- In welchen Gremien (z. B. Fachkonferenzen) kann ich Standards und ihre Erfüllung thematisieren und welche Möglichkeiten der Kooperation bestehen (z. B. Parallelarbeiten)?

Reflexionsfragen für die Schulleitung:

- Wie sind die Schülerleistungen an meiner Schule bezogen auf verbindliche Standards (z. B. nationale Bildungsstandards, länder- bzw. schulspezifische Vorgaben)?
- Inwiefern weichen an meiner Schule die Anteile der Schülerinnen und Schüler, die keine akzeptablen Leistungen bzw. Spitzenleistungen erbringen, von den Studienergebnissen ab?
- Welche Vereinbarungen bestehen zur standardbezogenen Diagnostik an meiner Schule?

- Welche Verfahren zur Leistungsdiagnostik werden eingesetzt und wie werden die Ergebnisse genutzt, um den Unterricht darauf hin anzupassen?
- Inwiefern konnten die Lehrkräfte bisher Informationen aus standardbezogener Diagnostik für ihre Arbeit nutzen, d. h. welche positiven Erfahrungen werden berichtet und welche Hemmnisse bestehen?
- Welche schulischen Gremien (z. B. Fachkonferenzen) beschäftigen sich mit standardbezogener Unterrichtsentwicklung?
- Welche Vorgaben der Schuladministration bestehen in Hinblick auf standardbezogene Diagnostik von Schülerleistungen und ihre Nutzung?
- Welche Unterstützungsangebote gibt es (z. B. Rückmeldeportale zu Ergebnissen aus Vergleichsarbeiten, Informationsseiten und Fortbildungen)?
- Wie gut kennen die Lehrkräfte geeignete Verfahren und für wie bedeutsam halten sie standardbezogene Diagnostik?
- Welche weiteren Voraussetzungen müssen geschaffen werden, damit die Lehrkräfte Schülerleistungen diagnostizieren und die gewonnenen Erkenntnisse für ihre Unterrichtsentwicklung einsetzen?

Rezensent/-in

Dr. Johannes Rosendahl, Dipl. Psych., Referent an der Qualitäts- und Unterstützungsagentur – Landesinstitut für Schule (QUA-LiS NRW), Soest. Arbeitsschwerpunkte: Bildungsmonitoring und Evaluation zentraler Prüfungen, Kooperation mit Wissenschaft zur Gewinnung und Erschließung empirischer Evidenzen

Zitiervorschlag

Rosendahl, J. (2017). Rezension zu Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (2013). PISA 2012. Münster u. a.: Waxmann. *Forschungsmonitor Schule, 1*. Abgerufen von https://www.forschungsmonitor-schule.de/print.php?id=6

Urheberrecht

Dieser Text steht unter der CC BY-NC-ND 4.0 Lizenz. Der Name des Urhebers / der Urheberin soll bei einer Weiterverwendung wie folgt genannt werden: Johannes Rosendahl (2017) für den Forschungsmonitor Schule.