

1 Studienüberblick¹

Christina Wallner-Paschon & Ursula Itzlinger-Bruneforth

Österreich beteiligte sich 2016 zum dritten Mal an der internationalen Lesestudie PIRLS zur Erfassung des Leseverständnisses auf der vierten Schulstufe. Dieses Kapitel bietet einen Überblick über wesentliche Aspekte bei der Durchführung von PIRLS. Neben einer Zusammenfassung der Ziele und der für die Studien verantwortlichen internationalen und nationalen Institutionen wird auf die Stichprobe, das Testdesign sowie auf wesentliche Aspekte bei der Interpretation der Ergebnisse eingegangen. Abschließend werden die Qualitätssicherungsmaßnahmen bei der Erstellung der Beiträge in diesem Band dargelegt.

Die Studie PIRLS (Progress in International Reading Literacy Study) erfasst seit 2001 im Fünf-Jahres-Rhythmus das Leseverständnis von Schülerinnen und Schülern der vierten Schulstufe. Initiator dieser Lesestudie ist die unabhängige Forschergemeinschaft IEA (International Association for the Evaluation of Educational Achievement), die 1958 gegründet wurde. Seit über 50 Jahren führt die IEA zahlreiche groß angelegte internationale Vergleichsstudien im Bereich der Bildungsforschung durch, um den politischen Entscheidungsträgern internationale Bildungsindikatoren bereitzustellen. Damit können Stärken und Schwächen in nationalen Bildungssystemen im internationalen Vergleich identifiziert werden.

Ziele von PIRLS

PIRLS misst, wie andere internationale Studien auch, die Qualität von Bildungssystemen zu einem bestimmten Zeitpunkt der schulischen Ausbildung in Form des Ertrags (*Outputs*) und vergleicht diesen über alle Teilnehmerländer hinweg. Darüber hinaus werden wichtige Hintergrundinformationen erhoben, die Hinweise auf mögliche Bedingungsfaktoren guter und schwacher Schulleistungen liefern. Daraus können faktenbasierte Maßnahmen für eine Qualitätsverbesserung im Bildungssystem abgeleitet werden.

Folgende Indikatoren können mithilfe von PIRLS gewonnen werden:

- Basisindikatoren: Die Leistungsdaten aus PIRLS liefern Informationen über den Ist-Stand der Kompetenzen in Lesen am Ende der vierten Schulstufe.
- Kontextindikatoren: Mithilfe von Fragebögen für Schüler/innen, Eltern, Lehrkräfte, Schulleitungen werden Kontextindikatoren erhoben, von denen angenommen wird, dass sie mit dem Lesenlernen zusammenhängen. Sie können wichtige Erklärungsbeiträge für unterschiedliche Schülerleistungen liefern.
- Trendindikatoren: Die regelmäßige Durchführung von PIRLS ermöglicht es, die Leistungen der Schüler/innen im Zeitverlauf zu beobachten und zu analysieren sowie das Zusammenspiel von Leistungsdaten und Hintergrundmerkmalen im Trend zu analysieren.

Organisation von PIRLS 2016

Die internationale Organisation

Die IEA trägt als Initiator der PIRLS-Studie die übergeordnete Projektverantwortung. Grundlegende Entscheidungen, die neben PIRLS auch alle anderen IEA-Studien betreffen, werden in der jährlich stattfindenden *General Assembly* (Generalversammlung) getroffen.

Dabei arbeiten folgende Institutionen an der Planung und Durchführung von PIRLS mit:

- International Study Center (ISC) am Boston College, USA: Verantwortlich für die internationale Organisation und Koordination von PIRLS,
- IEA-Sekretariat (Amsterdam, Niederlande): u. a. Qualitätskontrolle der Übersetzungen aller Testinstrumente der teilnehmenden Länder, Ausbildung der internationalen Qualitätsbeobachter/innen;
- IEA-Datenverarbeitungs- und Forschungszentrum (Hamburg, Deutschland): Verarbeitung, Qualitätskontrolle und Analyse der bei PIRLS erhobenen Daten;
- Statistics Canada (Ottawa, Ontario): Stichprobenziehung, Gewichtung;
- Educational Testing Service (Princeton, New Jersey): Unterstützung in messtheoretischen Belangen.

Für die Entwicklung und Begutachtung der Testaufgaben sowie der Hintergrundfragebögen gibt es jeweils eine Expertengruppe. Somit wird eine breite wissenschaftliche Expertise bei der Erstellung der Erhebungsinstrumente sichergestellt.

¹ Ein Teil dieses Textes basiert auf dem Einleitungskapitel zum Erstergebnisbericht (Wallner-Paschon & Widauer, 2017).

Für die nationale Umsetzung der Studien gibt es in jedem Land ein nationales Projektzentrum. Der *National Research Coordinator* (NRC) ist dabei für die Durchführung der Studie unter Einhaltung sämtlicher internationalen Richtlinien im eigenen Land verantwortlich.

Die nationale Organisation

In Österreich gehört die Durchführung internationaler Assessments zu den gesetzlichen Kernaufgaben des BIFIE (Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens). Aufgabe des BIFIE ist die wissenschaftliche Planung und praktische Umsetzung der Studien. In Zusammenarbeit mit dem ISC wurde am BIFIE das Untersuchungsdesign an die nationalen Gegebenheiten angepasst, die Aufgaben und Fragebögen wurden mitentwickelt, übersetzt und an österreichische Besonderheiten adaptiert. Darüber hinaus wurden die Erhebungen an den Schulen durchgeführt, die Daten elektronisch erfasst und für den ersten Ergebnisbericht analysiert.

Population und Stichprobe

An PIRLS haben sich insgesamt 50 Länder und elf Benchmarking Participants² beteiligt. Die Definition der Stichprobe erfolgt bei PIRLS über die Schulstufe: Zur Grundgesamtheit zählen alle Schüler/innen jener Schulstufe, die dem 4. Schuljahr entspricht³. Das Durchschnittsalter in dieser Schulstufe muss zum Testzeitpunkt mindestens bei 9,5 Jahren liegen. In Österreich sowie in den meisten Teilnehmerländern finden die Tests somit am Ende der 4. Schulstufe (4. Klasse Grundschule) statt.

Die Ziehung der Stichprobe erfolgte in zwei Schritten. Zunächst wurden in jedem Land aus allen Schulen mit Schülerinnen und Schülern der 4. Schulstufe die Testschulen zufällig ausgewählt. Um dabei eine standardisierte Vorgehensweise zu gewährleisten, wurde dieser Schritt für alle Länder von Statistics Canada durchgeführt. Im zweiten Schritt wurde am jeweiligen nationalen Zentrum an diesen Schulen mindestens eine Klasse der 4. Schulstufe – ebenfalls zufällig – ausgewählt. Den Ländern stand es frei, jeweils mehr als eine Klasse pro Schule für die Stichprobe zu ziehen. Alle Schüler/innen der somit festgelegten Klassen waren Teil der PIRLS-Stichprobe.

Sowohl auf Schul- als auch auf Schülerebene ist es nach international definierten Kriterien möglich, Viertklässler/innen vom Test auszuschließen – die Ausschlüsse dürfen in Summe allerdings nicht mehr als 5 % der Population ausmachen. In Österreich wurden auf Schulebene Sonderschulen sowie Schulen mit weniger als drei Schülerinnen und Schülern der 4. Schulstufe im Vorfeld (vor dem ersten Schritt der Stichprobenziehung) ausgeschlossen (1,2 %). Auf Schülerebene werden (nach der Ziehung der Klassenstichprobe) folgende Schüler/innen ausgeschlossen:

- Schüler/innen mit einer andauernden körperlichen Beeinträchtigung,
- Schüler/innen mit einer geistigen Beeinträchtigung, die eine Teilnahme am Test unmöglich macht und
- eine andere Muttersprache, wenn gleichzeitig die Testsprache noch nicht ausreichend beherrscht wird.

Die Ausschlüsse auf Schülerebene betragen insgesamt 4,4 %. Gemeinsam mit dem zuvor beschriebenen Ausschluss sehr kleiner Schulen und Sonderschulen (1,2 % der Zielpopulation) ergibt das eine Ausschlussquote von 5,6 %. Damit hat Österreich bei PIRLS 2016 die international festgelegte Grenze zum Ausschluss von Schülerinnen und Schülern (max. 5 %) knapp überschritten. Für Trendanalysen bedeutet dies allerdings keine merkliche Einschränkung, da bei PIRLS 2006 und 2011 (mit jeweils 5,1 %) die Ausschlussquote nur unwesentlich geringer war.

Die österreichische Stichprobe für PIRLS 2016 besteht aus 150 Schulen, 257 Klassen und 4360 Schülerinnen und Schülern. Mit einer Teilnahmequote von 100 % sowohl auf Schul- als auch auf Klassenebene und 97,6 % auf Schülerebene konnte Österreich die international geforderten Teilnahmequoten (85 % auf Schul- und Schülerebene, 95 % auf Klassenebene) deutlich übertreffen.

Ablauf der Tests und Testinstrumente

Der PIRLS-Test

PIRLS misst die Leseleistung der Viertklässler/innen differenziert für unterschiedliche (1) Leseabsichten und (2) Verstehensprozesse, die beim Lösen der Aufgabenstellung erfor-

2 Benchmarking Participants sind teilnehmende Vergleichseinheiten (Länderregionen oder bestimmte Alters- oder Sprachgruppen der Teilnehmerländer). Sie müssen ebenso wie die regulären Teilnehmerländer die strengen Qualitätsstandards der IEA erfüllen, sodass ihre Daten zur Gänze mit den Daten der Teilnehmerländer vergleichbar sind.

3 Das erste Schuljahr wird dabei über den ISCED-Level 1 definiert. ISCED 2011 ist eine internationale Standardklassifikation der Bildung, die von der UNESCO entwickelt wurde (Details dazu unter: http://uis.unesco.org/sites/default/files/documents/isced-2011-operational-manual-guidelines-for-classifying-national-education-programmes-and-related-qualifications-2015-en_1.pdf). ISCED-Level 1 ist jener Zeitpunkt in der Schullaufbahn, zu dem das systematische Lernen von Lesen, Schreiben und Mathematik sowie der Erwerb eines Basiswissens in anderen Bereichen beginnen.

derlich sind. Darüber hinaus werden (3) das Leseverhalten und die Einstellungen zum Lesen erfasst.

Mithilfe von literarischen Texten und Informationstexten werden beim PIRLS-Test zwei unterschiedliche Leseabsichten generiert: *Lesen, um literarische Erfahrung zu machen*, und andererseits *Lesen, um Informationen zu gewinnen*. Die Lesekompetenzen der Schüler/innen können nach diesen beiden Subskalen getrennt analysiert werden.

Die Verstehensprozesse (kognitive Prozesse, die beim Lösen einer Aufgabe notwendig sind) sollen durch unterschiedliche Aufgaben beziehungsweise Fragestellungen zu den Texten aktiviert werden. Es gibt (1) Aufgaben, bei denen die Schüler/innen explizit angegebene Informationen erkennen und wiedergeben müssen, (2) Aufgaben, die ein Ziehen von einfachen Schlussfolgerungen verlangen, (3) Aufgaben, bei denen ein Interpretieren sowie Verknüpfen von Gedanken und Informationen notwendig ist, um zur richtigen Lösung zu kommen und (4) Aufgaben, bei denen Inhalt, Sprache und einzelne Textelemente untersucht und bewertet werden müssen. Für eine Analyse der Lesekompetenz hinsichtlich der Verstehensprozesse werden jeweils zwei der eben genannten Anforderungen zu Subskalen zusammengefasst: *Wiedergeben und Schlussfolgern* (1 + 2) sowie *Interpretieren und Bewerten* (3 + 4).

Der Aspekt des Leseverhaltens und der Einstellungen zum Lesen wird mittels Hintergrundfragen im Schüler- und im Elternfragebogen erhoben.

Die Messung der Lesekompetenz erfolgt mit einem Papier- und-Bleistift-Test. Insgesamt wurden bei PIRLS 2016 zwölf verschiedene Texte eingesetzt. Um die beiden Leseabsichten abzudecken, gab es fünf literarische Texte sowie fünf Informationstexte mit daran anschließenden Aufgaben. Darüber hinaus wurden ein literarischer Text und ein Sachtext aus PIRLS Literacy 2016⁴ eingesetzt, um die Ergebnisse beider Studien vergleichen zu können. Die Texte wurden zu 16 verschiedenen Testheften kombiniert, wobei jedes Testheft aus 2 Texten besteht.

Für die Bearbeitung des Testhefts hatten die Schüler/innen insgesamt 80 Minuten Zeit. Im Anschluss daran beantworteten sie den Schülerfragebogen. Dieser umfasst neben demographischen Informationen auch Fragen zum familiären Hintergrund, zu den Lesegewohnheiten außerhalb der Schule,

zum Lesen in der Schule sowie zu leistungsrelevanten Aktivitäten außerhalb der Schule. Auch die Eltern sowie die Lehrkräfte der beteiligten Schüler/innen wurden gebeten, einen Fragebogen zu beantworten. Auf Schulebene wurden von den Schulleitungen Informationen zu Schulressourcen, Schulklima etc. erhoben. Der Fragebogen zum Lehrplan in den Teilnehmerländern ergänzt die Informationen auf Schul- und Klassenebene und fokussiert auf die national festgelegten Rahmenbedingungen des Leseunterrichts an den Schulen.

Zur Interpretation der Ergebnisse

Da bei PIRLS, wie bei Large-Scale Assessments üblich, nicht alle Schüler/innen der 4. Schulstufe eines Landes getestet wurden, sondern eine Stichprobe, muss bei der Interpretation der Daten Folgendes beachtet werden:

- Die aus den Daten resultierenden statistischen Kennzahlen (z. B. Mittelwerte) sind Punktschätzungen der tatsächlichen Populationswerte. Sie sind aufgrund der Testung einer Stichprobe mit einem gewissen statistischen Messfehler – dem Standardfehler (SE) – behaftet. Dieser wird unter anderem dafür verwendet, einen Wertebereich (Konfidenzintervall) zu berechnen, innerhalb dessen sich der tatsächliche Populationsmittelwert mit einer bestimmten Wahrscheinlichkeit befindet. Für die vorliegende Publikation wurde diese Wahrscheinlichkeit mit 95 % (d. h. $p < .05$) festgelegt⁵. Das heißt, die Werte der Population liegen mit 95-prozentiger Wahrscheinlichkeit innerhalb der berichteten Konfidenzintervalle.
- Der Standardfehler (statistischer Messfehler) spielt auch bei statistischen Tests eine Rolle, etwa wenn man die Ergebnisse zweier Gruppen miteinander vergleicht (z. B. Mädchen und Buben). Alle Signifikanzprüfungen in dieser Publikation erfolgen, falls nicht explizit anders vermerkt, zweiseitig auf dem 5%-Niveau.
- Da es sich bei PIRLS um sehr große, präzise Stichproben handelt, können teilweise auch sehr kleine Unterschiede (etwa zwischen Mädchen und Buben) statistisch nachgewiesen werden. Das heißt, dass die Unterschiede mit großer Wahrscheinlichkeit in der betreffenden Population existieren, aber relativ klein sein können, sodass sie in praktischer Hinsicht nicht relevant sind. Einige Ergebnisdarstellungen beinhalten deshalb Angaben zur

4 In einigen Ländern gibt es viele Kinder auf der vierten Schulstufe, bei denen der Erwerb der basalen Lesefertigkeiten noch nicht abgeschlossen ist. Für diese Länder bestand die Möglichkeit, die Kinder mithilfe einfacherer und kürzerer Texte und Aufgaben auf der vierten Schulstufe zu testen. Die zugehörige Studie wird als *PIRLS Literacy* bezeichnet. Damit beide Studien auf einer gemeinsamen Skala verankert und direkt miteinander verglichen werden können, wurden bei PIRLS 2016 zwei Texte aus PIRLS Literacy und im Gegenzug bei PIRLS Literacy wiederum zwei Texte aus PIRLS 2016 eingesetzt.

5 Das Konfidenzintervall errechnet sich damit aus dem Wert der Punktschätzung (z. B. Mittelwert) $\pm 1,96 \times \text{SE}$.

Effektstärke eines Unterschieds.⁶ Damit wird zusätzlich zur statistischen Signifikanz auch die Größenordnung von gefundenen Unterschieden quantifiziert. Eine grobe Faustregel für die Einschätzung der Größenordnung sind die Richtwerte nach Cohen (1988), nach denen Effektstärken ab 0.2 als kleine, ab 0.5 als mittlere und ab 0.8 als große Unterschiede eingestuft werden. Demnach kann ab einer Effektstärke von 0.2 (d. h. der Unterschied macht 20 % der Standardabweichung aus) zusätzlich zu einer statistischen Signifikanz auch von einer praktischen Relevanz ausgegangen werden.

Darüber hinaus wird darauf hingewiesen, dass es sich bei den angegebenen Kennwerten (Mittelwerte, Standardfehler, Prozentangaben etc.) um gerundete Werte handelt. Die Werte (inkl. jener in Tabellen und Abbildungen) werden zuerst – unter Berücksichtigung entsprechender Nachkommastellen – berechnet und dann kaufmännisch gerundet. Es kann daher vorkommen, dass beispielsweise die Summe von gerundeten Prozentangaben nicht exakt 100 ergibt oder Summen und Differenzen von Werten geringfügig inkonsistent erscheinen.

(Methodische) Änderungen bei PIRLS 2016

Bei PIRLS 2011 kam für die Skalierung der Subskalen erstmalig ein mehrdimensionales Antwortmodell zur Anwendung (Foy et al., 2012). Um die Vergleichbarkeit zu vorangegangenen Erhebungen sicherzustellen, wurden auch die Daten der Subskalen von PIRLS 2001 und 2006 erneut skaliert. Abweichungen in den Berichten zu PIRLS 2001 und 2006 (Suchań, Wallner-Paschon & Schreiner, 2009; Suchań, Wallner-Paschon, Stöttinger & Bergmüller, 2007; Mullis et al., 2003; Mullis et al., 2007) sind durch diese Veränderung im Skalierungsmodell bedingt.

Bei PIRLS 2016 wurde eine effektivere Methode bei der Schätzung des Standardfehlers für die Lesekompetenz eingesetzt (Foy & LaRoche, 2017). Diese Änderung kann im Vergleich zur bisherigen Methode zu minimal anderen Standardfehlern führen. Abweichungen der Standardfehler in den Berichten zu PIRLS 2006 und 2011 (Suchań, Wallner-Paschon & Schreiner, 2009; Suchań, Wallner-Paschon, Stöttinger & Bergmüller, 2007; Suchań, Wallner-Paschon & Schreiner, 2015; Suchań, Wallner-Paschon, Bergmüller & Schreiner, 2012; Mullis et al., 2003; Mullis et al., 2007) sind durch diese Änderung bedingt.

6 Das hier verwendete Effektstärkemaß für Mittelwertunterschiede (Cohens d) wird nach folgender Formel berechnet:

$$d = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2 + SD_2^2}{2}}}$$

Weitere Informationen und Publikationen zu PIRLS

Folgende Berichte zu PIRLS 2016 wurden bereits publiziert:

Martin, M. O., Mullis, I. V. S. & Hooper, M. (Eds.). (2017). *Methods and Procedures in PIRLS 2016*. Boston College: TIMSS & PIRLS International Study Center. Retrieved from <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>

Mullis, I. V. S. & Martin, M. O. (2015). *PIRLS 2016 Assessment Framework* (2nd ed.). Boston College: TIMSS & PIRLS International Study Center. Retrieved from <http://timssandpirls.bc.edu/pirls2016/framework.html>

Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Boston College: TIMSS & PIRLS International Study Center. Retrieved from <http://timssandpirls.bc.edu/pirls2016/international-results/>

Mullis, I. V. S., Martin, M. O., Goh, S. & Prendergast, C. (Eds.). (2017). *PIRLS 2016 Encyclopedia: Education policy and curriculum in reading*. Boston College: TIMSS & PIRLS International Study Center. Retrieved from <http://timssandpirls.bc.edu/pirls2016/encyclopedia/>

Wallner-Paschon, C. (2017). Austria. In I. V. S. Mullis, M. O. Martin, S. Goh & A. R. Moira. *PIRLS 2016 Encyclopedia. Education Policy and Curriculum in Reading*. Boston College: TIMSS & PIRLS International Study Center. Retrieved from <http://timssandpirls.bc.edu/pirls2016/encyclopedia/>

Wallner-Paschon, C., Itzlinger-Bruneforth, U. (Hrsg.). (2017). *PIRLS 2016. Technischer Bericht*. Salzburg: BIFIE. Verfügbar unter <https://www.bifie.at/material/internationale-studien/pirls/pirls-2016/>

Wallner-Paschon, C., Itzlinger-Bruneforth, U. & Schreiner, C. (2017). *PIRLS 2016. Die Lesekompetenz am Ende der Volksschule. Erste Ergebnisse*. Graz: Leykam.

Alle freigegebenen PIRLS-Texte und die dazugehörigen Aufgaben (Items) sowie die bei PIRLS 2016 eingesetzten Fragebögen sind unter folgendem Link abrufbar: <https://www.bifie.at/material/internationale-studien/pirls/pirls-2016/>

Bibliografie

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum

Foy, P., Brossman, B. & Galia, J. (2012). Scaling the TIMSS and PIRLS 2011 achievement data. In M. O. Martin & I. V. S. Mullis (Eds.), *Methods and Procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Achievement.pdf

Foy, P. & LaRoche, S. (2017). Estimating standard errors in the PIRLS 2016 results. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and Procedures in PIRLS 2016* (S. 4.1–4.9). Boston College: TIMSS & PIRLS International Study Center. Retrieved from <https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-4.html>

Suchań, B., Wallner-Paschon, C., Bergmüller, S. & Schreiner, C. (2012). *PIRLS & TIMSS 2011. Schülerleistungen in Lesen, Mathematik und Naturwissenschaft in der Grundschule. Erste Ergebnisse*. Graz: Leykam.

Suchań, B., Wallner-Paschon, C. & Schreiner, C. (2009). *PIRLS 2006. Die Lesekompetenz am Ende der Volksschule. Österreichischer Expertenbericht*. Graz: Leykam.

Suchań, B., Wallner-Paschon, C. & Schreiner, C. (2015). *PIRLS & TIMSS 2011. Die Kompetenzen in Lesen, Mathematik und Naturwissenschaft am Ende der Volksschule. Österreichischer Expertenbericht*. Graz: Leykam.

Suchań, B., Wallner-Paschon, C., Stöttinger, E. & Bergmüller, S. (2007). *PIRLS 2006. Internationaler Vergleich von Schülerleistungen. Erste Ergebnisse. Lesen in der Grundschule*. Graz: Leykam.

Wallner-Paschon, C. & Widauer, K. (2017). PIRLS 2016 – eine Einführung in die Studie. In C. Wallner-Paschon, U. Itzlinger-Bruneforth & C. Schreiner (Hrsg.). *PIRLS 2016. Die Lesekompetenz am Ende der Volksschule. Erste Ergebnisse*. (S. 7–33). Graz: Leykam.