

3.6 Prozesse zur Stichprobenziehung und Datenerhebung

Nina Johanna Welsandt & Daniel Deimel

Im folgenden Unterkapitel werden die Methoden erläutert, mit denen die repräsentativen Stichproben von ECON 2022 und der parallel stattfindenden *International Civic and Citizenship Education Study*¹ (ICCS 2022) konstruiert wurden. Die Datenerhebung von ECON 2022 erfolgt als nationale Zusatzerhebung im unmittelbaren Zusammenhang mit ICCS 2022. Als Studie unter der Gesamtverantwortung der IEA beachtet ICCS 2022 etablierte technische Standards für repräsentative Schulleistungsstudien (Martin et al., 1999).

Kapitelauflaufbau

Weiterhin werden die Verfahren umrissen, die zur Datenaufbereitung und -skalierung genutzt wurden. Methodische Herausforderungen, die sich aus den genutzten Verfahren ergeben, werden vorgestellt. Es folgt die Vorstellung der Prozesse zur Qualitätssicherung und ein Einblick in die Erhebungssituation.

3.6.1 Stichprobendesign

Repräsentativität in der Sozialforschung bedeutet, dass nicht nur Aussagen über die spezifische Stichprobe der Studie getroffen werden können, sondern reliable (zuverlässige) Schätzungen für die Population im Sinne der Gesamtheit einer Gruppe gemacht werden. Die Zielpopulation der Schüler*innen von ECON 2022 umfasst Schüler*innen der achten Jahrgangsstufe, sofern dies die Stufe ist, die von 13.5 Jahre alten Schüler*innen zum Erhebungszeitpunkt durchschnittlich am häufigsten besucht wird. Ergänzt wird die Schülerperspektive in ICCS 2022 durch eine Stichprobe von Lehrkräften, welche zum Erhebungszeitraum in dem entsprechenden Jahrgang unterrichten sowie durch Schulleitungen der teilnehmenden Schulen. Für ECON konnten für diese Befragungsgruppen nur einzelne Fragen ergänzt werden.

Zielpopulation
von ECON 2022

Ziel von ECON ist es, mithilfe der Schülerstichprobe möglichst genaue, repräsentative Aussagen über die Zielpopulationen in NRW zu treffen. Hierzu werden verschiedene Verfahren des Stichprobendesigns angewendet, die im Folgenden umrissen werden.

Der Ansatz von ICCS 2022 (als internationales Large-Scale-Assessment) und somit ECON 2022 ist es, Schüler*innen als Teil ihrer Klasse bzw. Lerngruppe repräsentativ für ihre Schule zu erfassen. Jedoch ist es aufgrund der Komplexität von Schulrealitäten, welche sich etwa durch ein mehrgliedriges Bildungssystem und eine ungleiche Verteilung von Schüler*innen mit Zuwanderungsgeschichte konkretisiert, nicht hinreichend, eine einfache Zufallsstichprobe zu ziehen. Um diese Komplexität zu berücksichtigen und die Genauigkeit der Parameterschätzungen zu erhöhen, entspricht das Stichprobendesign einer stratifizierten zweistufigen Cluster-Stichprobe (Bortz & Döring, 2016). Die Stichprobe wurde auf Grundlage gemeinsamer Planungen mit dem nationalen Studienzentrum durch die Sampling-Unit der IEA Hamburg gezogen.

1 Unter Beachtung von Aspekten, die für ECON 2022 spezifisch sind, orientiert sich dieses Kapitel an der Darstellung der einschlägigen Methodologie in: Abs, H. J., Hahn-Laudenberg, K., Deimel, D. & Ziemes, J. F. (2024). *ICCS 2022. Schulische Sozialisation und politische Bildung von 14-Jährigen im internationalen Vergleich*. Waxmann. Kap. 19 und 20.

Genutzte Stratifizierungsverfahren

In einem ersten Schritt wurde die Zielpopulation der Schülerschaft in mehrere Untereinheiten eingeteilt (*Stratifizierung*). In der *expliziten Stratifizierung* wurden Schulen mit ähnlichen Merkmalen zu sinnvollen Einheiten (*Strata*) zusammengefasst. Die *Strata* ergeben sich zunächst durch die Schulformen der gegliederten Schulsysteme. Als weiteres Stratum wurde der Anteil der Schüler*innen mit Migrationshintergrund herangezogen; hier wurde der *Cut-off*-Wert von 33 % innerhalb der Schülerschaft gewählt². Ziel der Stratifizierung ist es, die Verteilung der Schüler*innen in der Zielpopulation auf die verschiedenen Schultypen innerhalb der Stichprobe abzubilden. Innerhalb dieser *Strata* erfolgte in einem zweiten Schritt eine Zufallsauswahl von Schulen unter Berücksichtigung der Anzahl der Schüler*innen im jeweiligen Jahrgang (*probability proportional to size, PPS sampling*). Dies soll ausgleichen, dass Schüler*innen an großen Schulen eine geringere Wahrscheinlichkeit haben, Teil der Stichprobe zu sein, als Schüler*innen an kleinen Schulen. Bei den Förderschulen wurde die explizite Stratifizierung nach Migrationshintergrund aufgrund der geringen Größe von manchen sich daraus ergebenden *Strata* durch eine implizite Stratifizierung ersetzt, in der neben Schulgröße auch Migrationsanteil laut Schulstatistik bei der Sortierung der Schulen für die Ziehung nach PPS-Verfahren berücksichtigt wurde.

Oversampling zur Präzisionserhöhung einzelner Teilstichproben

Um Effekte, die in Zusammenhang mit der Zuwanderungsgeschichte von Schüler*innen stehen können, mit höherer Genauigkeit schätzen zu können, wurde zudem ein *Oversampling* von Schulen mit einem hohen Anteil an Schüler*innen mit Zuwanderungsgeschichte vorgenommen. Von diesen Schulen wurden also mehr erhoben, als eine rein proportionale Stichprobe vorgesehen hätte. Die beschriebenen Verfahren werden anschließend im Prozess der Gewichtung so berücksichtigt, dass Aussagen für die Population repräsentativ sind.

In einem letzten Schritt der Stichprobenziehung erfolgte eine zufällige Auswahl einer achten Klasse innerhalb jeder ausgewählten Schule. In dieser Klasse werden alle Schüler*innen erhoben, welche keine Ausschlusskriterien erfüllen. An Förderschulen wurden zudem, aufgrund der geringeren Anzahl von Schüler*innen pro Klasse, alle Klassen eines Jahrgangs zu *Pseudo-Klassen* zusammengefasst und für alle weiteren Verfahren wie die Schüler*innen derselben Klasse behandelt.

Aufgrund der Clusterung der Schüler*innen in Schulen und Schulklassen sowie der genutzten Verfahren der Stichprobenziehung ist eine größere Stichprobe erforderlich, um eine ähnliche statistische Aussagekraft wie eine einfache Zufallsstichprobe geringerer Größe zu erreichen. Die angestrebte Präzision des Samples soll derselben Präzision einer hypothetischen einfachen Zufallsstichprobe (*effective sample size*) von 400 Schüler*innen entsprechen; mit einem 95 %-Konfidenzintervall von ± 0.1 Standardabweichungen (Martin et al., 1999). Als angestrebter Wert für IEA-Studien hat sich hierfür eine intendierte Stichprobengröße von 150 Schulen pro Bildungssystem etabliert, da diese in der Regel in einer realisierten Stichprobe von mindestens 3.000 Schüler*innen mündet.

Intendierte und realisierte Stichprobe

In jeder empirischen Studie ist davon auszugehen, dass die *realisierte Stichprobe* der *intendierten Stichprobe* nicht vollständig entspricht. Verweigert eine Schule ihre Teilnahme, kann hierfür in einer der hinsichtlich der Größe ähnlichen Ersatzschulen desselben Stratum eine zufällig ausgewählte Klasse erhoben werden. Steht eine Schule überhaupt nicht zur Verfügung (etwa aufgrund einer Schulschließung oder weil keine Klasse im entsprechenden Jahrgang vorhanden ist),

2 Maßgeblich ist hier die spezifische Definition des Bundeslands NRW im Rahmen der Schulstatistik. Zur vertieften Auseinandersetzung siehe Abs, H. J., Hahn-Laudenberg, K., Deimel, D. & Ziemes, J. F. (2024). *ICCS 2022. Schulische Sozialisation und politische Bildung von 14-Jährigen im internationalen Vergleich*. Waxmann. Kap. 18.

wird sie nicht ersetzt. Während die öffentlichen Schulen aus Gründen der schulischen Qualitätssicherung zur Teilnahme verpflichtet waren, stand es Schulen in nichtöffentlicher Trägerschaft frei.

Die Teilnahmeentscheidung für individuelle Schüler*innen erfolgte für die verschiedenen Teile der Befragung unterschiedlich. Während die Teilnahme am ECON-Test von Seiten des Landes verbindlich gemacht wurde, konnte eine Teilnahme am Fragebogen zu Einstellungen und Kontexten nur freiwillig und mit Elterngenehmigung erfolgen.

Freiwilligkeit und
Verpflichtung der
Teilnahme für
Schüler*innen

Da die Stichprobenkriterien in ICCS an internationale Vorgaben gebunden waren und für ECON und ICCS identische Stichproben angestrebt wurden, greifen die internationalen Vorgaben auch für ECON. Das heißt, es werden nur Klassen berücksichtigt, bei denen die Teilnahmequote auf Schülerebene über 50 % lag, weil ansonsten systematische Verzerrungseffekte zu befürchten sind. Ein Beispiel dafür ist die Annahme, dass die Nicht-Teilnahme, etwa aufgrund fehlender Elterngenehmigung, möglicherweise nicht zufällig, sondern systematisch mit weiteren Merkmalen der Schüler*innen verbunden ist. Der Ausschluss hängt möglicherweise mit Merkmalen der Schüler*innen zusammen, die ebenfalls Einfluss auf die Bearbeitung des Tests oder auf die im Fragebogen erfragten Inhalte haben.

Darüber hinaus bestehen internationale Vorgaben darüber, unter welchen Bedingungen und in welchem Ausmaß Schulen oder Klassen bereits im Vorfeld der Zufallsauswahl ausgeschlossen werden dürfen. Auf Schulebene dürfen spezielle Schulformen (wie Förderschulen) ausgeschlossen werden, soweit sie sich curricular und unterrichtsmethodisch von Regelschulen bedeutsam unterscheiden. Dies betrifft in Deutschland die Förderschwerpunkte *Geistige Entwicklung*, *Hören & Kommunikation*, *Sehen*, *Körperliche Entwicklung* sowie *Schulen für Kranke*. Auf Klassenebene sind weitere Ausschlussgründe international definiert: Sofern eine Klasse ausschließlich aus Schüler*innen besteht, die aus körperlichen, emotionalen oder kognitiven Gründen nicht befähigt sind, an der Studie teilzunehmen, oder ausschließlich aus Schüler*innen, die die Sprache der Erhebungsinstrumente seit weniger als einem Jahr sprechen, wird diese Klasse vor der Zufallsauswahl ausgeschlossen. Innerhalb der ausgewählten Klassen wurden Schüler*innen ausgeschlossen, denen eine Teilnahme aufgrund der vorgenannten Merkmale nicht möglich war. Auch testökonomische Aspekte spielen hier eine Rolle, da für eine relativ geringe Anzahl von Schüler*innen ein hohes Ausmaß an Anpassungen erforderlich wäre, ohne dass die Vergleichbarkeit der Befragungsinstrumente hinreichend empirisch abgesichert werden kann.

Die internationalen Vorgaben legen fest, dass die Ausschlussquote 5 % der Schüler*innen für eine uneingeschränkte Vergleichbarkeit nicht übersteigen darf. Die Ausschlussquote in NRW liegt bei 3,4 %.

3.6.2 Intendierte und realisierte Stichproben

Im nächsten Abschnitt werden die geplanten und realisierten Stichproben für Schüler*innen, Lehrkräfte und Schulleitungen vorgestellt. Die intendierte und realisierte Stichprobe von Schüler*innen, Lehrkräften und Schulleitungen findet sich in Tabelle 3.6.1.

Tabelle 3.6.1: Stichproben und geschätzte Population

	<i>Intendiert</i>	<i>Realisiert</i>	<i>Geschätzte Population</i>
Schüler*innen	3.998	3.020	166.438
Lehrkräfte in Jahrgang 8 ¹	-	2.916	46.349
Lehrkräfte in Jahrgang 8 (Sozialwissenschaften) ^{1,2,3}	-	1.012	16.230
Schulleitungen ²	152	141	1.945
Schulen (gültige Schülerdaten) ³	152	145	1.875
Schulen (gültige Lehrerdaten) ³	152	129	1.875

1 Listung der Anzahl der Lehrkräfte der Zielpopulation erfolgt erst durch die Schulen und ist nicht Teil der Stichprobenplanung.

2 Die Samplingprozesse streben eine repräsentative Stichprobe für Schüler*innen an. Gewichtete Schätzungen der Population der Schulleitungen sind daher von größerer Ungenauigkeit betroffen.

3 Die realisierte Stichprobe bezieht sich auf Schulen, mit wenigstens 50 % Teilnahme auf Ebene der Schüler*innen bzw. Lehrkräfte.

Die Stichprobe der Schüler*innen in NRW wurde zufriedenstellend ausgeschöpft. Je verbindlicher die Datenerhebung für die Schulen gemacht werden kann, desto valider sind die Ergebnisse. In NRW war die Teilnahme für Schüler*innen und Lehrkräfte verpflichtend. So konnte die Schulstichprobe gut ausgeschöpft werden. Insgesamt liegen für 145 Schulen und für 3.020 Schüler*innen gültige Daten vor. Bei den Lehrkräften ist die Ausschöpfung nicht ganz so stark, gleichwohl wurde immer noch eine sehr seltene repräsentative Stichprobe für diesen Inhaltsbereich realisiert. Die Stichprobe der Lehrkräfte in NRW ist geeignet, um repräsentative Aussagen über Lehrkräfte, die im Jahrgang 8 unterrichten, zu treffen, wenn auch mit kleineren Einschränkungen. Die Stichprobe umfasst 2.916 Lehrkräfte an 129 Schulen.

Für die Befragung der Schulleitungen war jeweils ein Mitglied der Schulleitung dazu aufgefordert, den Fragebogen zu bearbeiten. Die intendierte Stichprobe wurde durch die verpflichtende Teilnahme fast vollständig (93 %) ausgeschöpft. Es ist zu beachten, dass die Stichprobenplanung vor allem eine Repräsentativität für Schüler*innen anstrebt und sich alle Schätzungen für Schulleitungen durch eine größere Unsicherheit und größere Standardfehler äußern.

Tabelle 3.6.2 gibt ergänzend den Anteil der Schüler*innen sowie die geschätzte Population in den verschiedenen Referenzkategorien für Geschlecht und Zuwanderungsgeschichte an.

Tabelle 3.6.2: Schüler*innen der achten Jahrgangsstufe

	<i>N</i> ¹	<i>Est.</i> ²	% ³
Geschlecht			
weiblich	1.278	74.607	45
männlich	1.527	88.917	54
anders	42	2.616	2
fehlend	5	211	0
Zuwanderungsgeschichte			
min. ein Elternteil in DE geboren	1954	100.145	67
Schüler*in in DE geboren, Eltern zugewandert	619	32.870	22
Schüler*in und Eltern zugewandert	319	15.606	11

1 Absolute Fallzahl.

2 Gewichteter Anteil der Schüler*innen an der Population.

3 Anteil an der gewichteten Population in Prozent.

3.6.3 Gewichtung und Schätzung von Stichproben und Messfehlern

Durch die IEA-Sampling-Unit wurden den Datensätzen Gewichte hinzugefügt. Bei einer Gewichtung werden für statistische Auswertungen Antworten von Schüler*innen in dem Maße stärker berücksichtigt, indem sie in der realisierten Stichprobe im Vergleich zur intendierten Stichprobe unterrepräsentiert sind. Hierdurch können anhand der Daten eher Aussagen über die Population als nur über die spezifische Stichprobe gemacht werden. Für jede Klasse wurde ein individueller Gewichtungsfaktor anhand eines Basisgewichts (inverse Selektionswahrscheinlichkeit separat für jeden Schritt der Stichprobenziehung) und eines Adjustierungsfaktors generiert, welcher anhand der tatsächlich teilgenommenen Schüler*innen aus allen Selektionsstrata berechnet wurde. Im Gewichtungsprozess wurde angenommen, dass fehlende Werte sich über die berücksichtigten Strata hinaus nicht systematisch erklären lassen (*missing completely at random*).

Durch Gewichtungen, stratifizierte Stichprobenziehung und Unterschiede zwischen der intendierten und realisierten Stichprobe droht eine Unterschätzung der Varianzen und damit der Standardfehler. Einige statistische Verfahren versuchen, das Maß der Verzerrung anhand der realisierten Stichprobe zu schätzen und realistischere Standardfehler zu präsentieren: Die *Jackknife*-Methode nutzt die Selektionsstrata, die eingangs beschrieben worden sind. Innerhalb der expliziten Strata werden Schulen in Paaren angeordnet, die sich im Hinblick auf vordefinierte Merkmale wie etwa Schulform und Anteil an Schüler*innen mit Migrationshintergrund ähneln. So entstanden für ICCS 2022 75 *Jackknife*-Zonen. Zur Schätzung des Standardfehlers wird jede Berechnung 75-mal durchgeführt. In jeder Durchführung der Berechnung wird eine Schule ausgeschlossen und die gepaarte Schule dafür doppelt gewichtet. Aus der Varianz der Ergebnisse werden die Standardfehler der Populationsschätzung berechnet. ECON 2022 greift aufgrund derselben Stichprobe auf die gleiche Gewichtung zurück.

Umso mehr die realisierte Stichprobe von der intendierten abweicht, desto intensiver muss dies im Gewichtungsprozess berücksichtigt werden (durch höhere Gewichte) und desto größer wird der Standardfehler. Der Standardfehler (*SE*) ist

Gewichtung als Verfahren zur Absicherung von Repräsentativität

Schätzung von Standardfehlern

Zusammenhang zwischen Ausschöpfung von Stichproben, Gewichtung und Standardfehlern

ein Maß für die Genauigkeit einer Schätzung und wird insbesondere bedeutsam, wenn etwa die Signifikanzen von Gruppenunterschieden berechnet werden sollen. Der größere Standardfehler bedeutet eine größere Ungenauigkeit der Schätzung und führt weiterhin dazu, dass Unterschiede zwischen den Ergebnissen von (Teil-)Gruppen größer sein müssen, um statistisch signifikant zu werden. Mit einer 95-prozentigen Wahrscheinlichkeit liegt der wahre Wert einer Messung (also hier der tatsächliche Wert in der Population) innerhalb des Intervalls von 1.96 Standardfehlern über oder unter dem berichteten Wert. Ein Unterschied wird als statistisch signifikant bezeichnet, wenn es sehr wahrscheinlich ist, dass sich dieser von einem vorher festgelegten Wert, bspw. Null oder dem Wert einer Vergleichsgruppe, unterscheidet. In der Regel wird in den Sozialwissenschaften eine maximale Irrtumswahrscheinlichkeit von 5 % als Richtwert genommen. Dies bedeutet, dass es eine Wahrscheinlichkeit von weniger als 5 % gibt, dass das beobachtete Ergebnis in der Stichprobe zufällig auftritt, wenn kein tatsächlicher Effekt vorliegt.

Um Prozesse der Gewichtung zum Ausgleich der heterogenen Abdeckung einzelner Strata zu rechtfertigen, wird davon ausgegangen, dass die vergleichsweise geringere Beteiligung in einzelnen Strata nicht auf einer systematischen und inhaltlich bedeutsamen Beeinflussung beruht. Inhaltlich relevant und systematisch wäre etwa, wenn an bestimmten Schulen innerhalb nur eines Stratums deutlich mehr finanzielle Mittel zur Verfügung stehen, die dazu führen, dass diese Schulen eher teilnehmen und Schüler*innen gleichzeitig besser vorbereitet sind. Es liegen keine Hinweise auf eine solche Beeinflussung der Teilnahmeentscheidung von Schulen vor.

3.6.4 Datenerhebung und Qualitätssicherung

Durchführung eines Feldtests

Die Datenerhebung für ECON 2022 erfolgt als nationale Zusatzerhebung im unmittelbaren Zusammenhang mit ICCS 2022. Der Feldtest beider Studien wurde zwischen dem 20.04.2021 und dem 25.06.2021 durchgeführt. In diesem Zusammenhang wurde für ECON 2022 ein Readiness-Tool entwickelt. Dieses Tool überprüfte, ob Schulcomputer die Anforderungen für das computerbasierte Assessment erfüllen. Weiterhin wurde eine Expertenbefragung (vgl. Kapitel 4.2) durchgeführt, um ein Assessment mittlerer Schwierigkeit zu konstruieren. Durch Skalen- und Faktorenanalysen konnte das Test- und Fragebogenmaterial weiter verfeinert und auf die Zielgruppe angepasst werden (vgl. Kap. 4.3 und 4.4).

Haupterhebung nach internationalen Standards

Die Haupterhebung wurde zwischen dem 14.03.2022 und dem 23.06.2022 durchgeführt und orientierte sich in der Umsetzung an den international standardisierten Abläufen von ICCS 2022. Die Erhebungen von ECON 2022 und ICCS 2022 wurden am gleichen Tag nacheinander durchgeführt. In der Befragung ganzer Klassen anstatt ausgewählter Schüler*innen folgt ECON 2022 der Erhebung in ICCS 2022. Diese Erhebung im Klassenverband erweist sich für die praktische Umsetzung in der Schule als günstiger und ist zudem bedeutsam, wenn spezifische Bedingungen der Lernumgebung auf Klassenebene erfasst werden sollen. Die Befragungen zu ECON 2022 werden nach einer Pause im Anschluss an ICCS 2022 durchgeführt. Tabelle 3.6.3 visualisiert den Ablauf des Testtags.

Tabelle 3.6.3: Ablauf am Testtag

Ablauf am Testtag	Dauer
Begrüßung, Login der Schüler*innen, Einweisung in die Bearbeitung der ICCS-Aufgaben	ca. 20 Minuten
Bearbeitung der ICCS-Testaufgaben am Computer	45 Minuten
Pause	5 Minuten
Bearbeitung des internationalen ICCS-Schülerfragebogens am Computer	40 Minuten
Pause	10 Minuten
Bearbeitung des europäischen ICCS-Schülerfragebogens am Computer	20 Minuten
Pause	20 Minuten
Erneuerung Login der Schüler*innen, Einweisung in die Bearbeitung des ECON-Tests	10 Minuten
Bearbeitung der ECON-Testaufgaben am Computer	60 Minuten
Pause	5 Minuten
Bearbeitung des ECON-Schülerfragebogens	25 Minuten
Beenden der Testsitzung und Einsammeln aller Materialien	ca. 10 Minuten
Bearbeitungszeit	195 Minuten
Gesamtdauer der Testsitzung	275 Minuten

Die Organisation und Durchführung der Erhebungen an den Schulen wurden im Vorfeld von Schulkoordinator*innen unterstützt. Diese unterstützten bei der Auswahl der teilnehmenden Schüler*innen und Lehrer*innen und erhielten von der IEA ein ausführliches Manual mit sämtlichen Abläufen. Die Schulkoordinator*innen unterstützten zudem bei der Einholung der Einverständniserklärungen und standen den Testleiter*innen am Tag der Erhebung zur Seite.

Unterstützung durch Schulkoordinator*innen

Testleiter*innen, meist Lehramtsstudierende oder andere Personen mit einem akademischen pädagogischen oder psychologischen Hintergrund, wurden vorab von der IEA Hamburg ausgewählt und geschult. Auf Basis detaillierter Manuale und einer virtuellen Testleiterschulung übernahmen sie im Vorfeld der Erhebung grundlegende organisatorische Aufgaben in Absprache mit den Schulkoordinator*innen. Am Testtag selbst führten die Testleiter*innen die Studien durch. Sofern das Readiness-Tool im Voraus identifizierte, dass in den ausgewählten Schulen externe Laptops verwendet werden müssen, brachten die Testleiter*innen diese am Testtag mit. Die Testumgebungen konnten über einen USB-Stick gestartet werden. Die Schüler*innen wurden von den Testleiter*innen mit individuellen Zugangscodes angemeldet. Die Ergebnisse wurden automatisch auf den USB-Sticks gespeichert und später von der IEA ausgelesen.

Durchführung der Studien durch Testleiter*innen

Durchführung über USB-Sticks

Für ECON 2022 wurde ein eigenständiger Test entwickelt, der inhaltlich unabhängig von ICCS 2022 ist. Schüler*innen erhielten nach einer 20-minütigen Pause im Anschluss an den ICCS-Teil am Testtag zunächst das *Technology Based Assessment – Economic Literacy* zu Kompetenzen im Bereich der ökonomischen Bildung, der mit dem Lehrplan abgestimmt ist. Nach einer weiteren Pause folgte die Bearbeitung des ECON-Schülerfragebogens. Darin wurden Fragen zum individuellen ökonomischen Kontext der Schüler*innen gestellt (z.B. genutzte Finanzinstrumente, ökonomische Interaktionspartner*innen) sowie Fragen zu ökonomischen Einstellungen (z.B. Vorsorge, Risikobereitschaft) und Merkmalen der schulischen Lernumgebung. ECON 2022 greift hinsichtlich weiterer Hintergrundvariablen der Schüler*innen sowie der Schulmerkmale auf die erhobenen Daten

in ICCS 2022 zurück, wodurch die Befragung kürzer sein kann und stärker auf Inhalte der ökonomischen Bildung fokussiert werden konnte.

Ergänzungen in
Lehrkräfte- und Schul-
leitungsbefragungen

Die Datenerhebung der Lehrer*innen (TcQ) und Schulleiter*innen (ScQ) erfolgte bei ICCS 2022 über ein Online-Tool mit persönlichen Login-Daten und einer Bearbeitungszeit von 20–30 Minuten. Für ECON 2022 wurde keine separate Lehrkräftebefragung durchgeführt. Innerhalb der ICCS-2022-Fragebögen war es möglich, einzelne Fragen zu Qualifikationen und Qualifikationsbedarfen von Lehrer*innen, die das Fach Wirtschaft unterrichten, zu ergänzen. Weiterhin war es auch bei der Schulleiterbefragung möglich, einzelne Fragen zur Verankerung ökonomischer Bildung in der Schule zu ergänzen.

Einblick in die
Qualitätssicherung

Zur Qualitätssicherung³ der Erhebungssituationen von ECON 2022 und ICCS 2022 wurden zwei unabhängige Kontrollinstanzen eingesetzt: Beobachter*innen des nationalen Studienzentrums (NQO) und externe Beobachter*innen (IQO). Diese führten stichprobenartige, unangekündigte Überprüfungen durch. Insgesamt wurden etwa zehn Prozent der Testdurchführungen nach einheitlichen internationalen Richtlinien beobachtet. Dabei wurden Aspekte wie die Vollständigkeit der Dokumente, der Transport der Testmaterialien, das korrekte Ausfüllen der Teilnehmerlisten, die Einhaltung der vorgegebenen Zeitlimits und die standardisierten Instruktionen an die Schüler*innen dokumentiert. Störungen jeglicher Art während der Testsitzungen wurden ebenfalls erfasst. Darüber hinaus wurden Interviews mit den Schulkoordinator*innen und Testleiter*innen durchgeführt, um Einblicke in die Vorbereitung und Durchführung der Erhebung zu erhalten und Verbesserungsvorschläge an die IEA aufzunehmen. Sowohl das NQO-Team als auch das IQO-Team bestätigten, dass die internationalen Erhebungsstandards eingehalten wurden.

3.6.5 Hinweise zur Datenauswertung

Einführung in
verwendete Software

Abschließend werden methodische Entscheidungen vorgestellt, die für ein Nachvollziehen der quantitativ-empirischen Berechnungen in diesem Bericht relevant sind. Für die Berechnungen aller Ergebnisse wurde der IDB-Analyzer Version 5 der IEA (IEA, 2023) in Verbindung mit R Version 4.3.1 (R Core Team, 2023) sowie SPSS Version 29 (IBM, 2022) genutzt. Der IDB Analyzer ist eine frei zugängliche Analysesoftware, die Makros und einen Syntaxgenerator zur Verfügung stellt, mit denen notwendige Verfahren zur Berechnung von Schätzungen und Standardfehlern korrekt angewendet werden.

Verschiedene Kategorien
fehlender Werte

Fehlende Werte (vgl. Kapitel 4.3.2) erhöhen die Unsicherheit beim Schätzen von Testergebnissen oder anderen Skalenwerten für (Teil-)Populationen. Wenn Schüler*innen ein Item nicht beantwortet haben, wurde dies im Wissenstest als eine fehlende Antwort codiert. Fehlende Werte in Hintergrundvariablen wie Geschlecht, Alter oder sozioökonomischer Status der Eltern wurden für den vorliegenden Bericht nicht imputiert, also nicht anhand anderer vorhandener Werte geschätzt und ersetzt. Dies führt dazu, dass bei Berechnungen, die Unterschiede zweier Schülergruppen hinsichtlich eines bestimmten Merkmals betrachten, die

3 Die Qualitätssicherung wurde maßgeblich als Teil von ICCS 2022 durchgeführt. Für ECON 2022 konnten einzelne Fragen ergänzt werden. Zur detaillierten Darstellung siehe Abs, H. J., Hahn-Laudenberg, K., Deimel D. & Ziemes, J. F. (2024). *ICCS 2022. Schulische Sozialisation und politische Bildung von 14-Jährigen im internationalen Vergleich*. Waxmann. Kap. 19.

Schüler*innen nicht berücksichtigt werden, die keine Angaben zu dieser Hintergrundvariable gemacht haben (Ausschluss *listwise*).

Literatur

- Abs, H. J., Hahn-Laudenberg, K., Deimel, D. & Ziemes, J. F. (2024). *ICCS 2022. Schulische Sozialisation und politische Bildung von 14-Jährigen im internationalen Vergleich*. Waxmann
- Bortz, J. & Döring, N. (2016). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (5. vollständig überarb., akt. Aufl.). Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- IBM. (2022). IBM SPSS Statistics for Windows, Version 29.0. IBM Corp.
- IEA. (2023). IEA IDB Analyzer (Version 5.0). IEA.
- Martin, M. O., Rust, K. & Adams, R. J. (1999). *Technical Standards for IEA Studies*. IEA.
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing: Version 4.3.1*. <https://www.R-project.org>

Bildnachweis S. 55:

Edward Burtynsky, Phosphor Tailings Pond #4, Near Lakeland, Florida, USA, 2012

© Edward Burtynsky, courtesy Flowers Gallery, London / Nicholas Metivier Gallery, Toronto