



PISA-Studie

Internationale Schulleistungsstudie der OECD

Vorbereitungsmaterial zur Präsentation im Rahmen des
Bachelor-Seminars Ausgewählte Aspekte der Wirtschafts-
und Sozialstatistik

eingereicht von Jenny Steindl

LMU, Institut für Statistik

Seminarleiter: Prof. Dr. Thomas Augustin

Seminarbetreuer: Eva Endres

München, den 15.05.2015

Inhaltsangabe

1	Beschreibung der PISA - Studie	1
2	Kompetenzbereiche	5
3	Methodik	7
4	Literaturverzeichnis	18

Tabellenverzeichnis

1	An PISA 2012 teilnehmende Staaten	2
---	---	---

Abbildungsverzeichnis

1	In die Stichprobe gezogene Schulen nach Bundesland und Schulart . .	10
2	Stichprobendesign PISA 2012	11
3	Anzahl der an PISA 2012 eingesetzten Units und Items	14
4	In Deutschland eingesetzte Testhefte	15
5	Item-Characteristic Curve	16

1 Beschreibung der PISA - Studie

Das "*Programm for International Student Assessment*" ist eine Studie der "*Organisation for Economic Co-Operation and Development*" - kurz: OECD.

Anliegen

Primäre Aufgabe der PISA-Studie ist es, den OECD-Mitgliedsstaaten Daten vorzulegen, die zu politisch-administrativen Entscheidungen zur Verbesserung der nationalen Bildungssysteme beitragen können. Das zyklische Programm der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung wird von allen Mitgliedsstaaten gemeinschaftlich getragen und verantwortet. Ziel dieser Studie ist es, die Leistungen 15-jähriger Schülerinnen und Schüler in den Kompetenzen Mathematik, Naturwissenschaften, Lesen und fächerübergreifenden Leistungsfähigkeiten zu erfassen. National und international soll so die Leistung der Zielpopulation aufgezeigt werden, um eventuelle Defizite aufzudecken und gegebenenfalls über die nationalen Bildungssysteme zu verbessern. Gemäß einer Vereinbarung zwischen dem Bundesministerium für Bildung und Forschung und der Ständigen Konferenz der Kultusminister der Länder ist die Bundesrepublik Deutschland an dieser Studie beteiligt.

Teilnehmer

Im Jahr 2000 nahmen insgesamt 32 Staaten an der Erhebung der Studie teil, darunter die 28 Mitgliedsstaaten der OECD. Im Jahr 2012 haben 65 Staaten an der Studie teilgenommen, was von einem großen internationalen Interesse der Studie zeugt. Es haben alle 34 OECD Mitgliedsstaaten und 31 Partnerstaaten an der Erhebung teilgenommen. Nachfolgend werden die 65 PISA-Teilnehmerstaaten der Erhebung im Jahr 2012 vorgestellt:

An PISA 2012 teilnehmende Staaten

Albanien*	Jordanien*	Republik Serbien*
Argentinien*	Kanada	Rumänien*
Australien	Kasachstan*	Russische Föderation*
Belgien	Katar*	Schweden
Brasilien*	Kolumbien*	Schweiz
Bulgarien*	Korea	Shanghai (China)*
Chile	Kroatien*	Singapur*
Chinesisch Taipeh*	Lettland*	Slowakische Republik
Costa Rica*	Liechtenstein*	Slowenien
Dänemark	Litauen*	Spanien
Deutschland	Luxemburg	Thailand*
Estland	Macau (China)*	Tschechische Republik
Finnland	Malaysia*	Tunesien*
Frankreich	Mexico	Türkei
Griechenland	Neuseeland	Ungarn
Hongkong (China)*	Niederlande	Uruguay*
Indonesien*	Norwegen	Vereinigte Arabische Emirate*
Irland	Österreich	Vereinigte Staaten
Island	Peru*	Vereinigtes Königreich
Israel	Polen	Vietnam*
Italien	Portugal	Zypern*
Japan	Republik Montenegro*	

Tab. 1: *An PISA 2012 teilnehmende Staaten (Prenzel M., et al. (2013), S. 22)*

* bezeichnet die an PISA 2012 teilnehmenden Partnerstaaten der OECD

Um eine repräsentative Stichprobe zu erhalten, wurden aus den teilnehmenden 65 Staaten 510.000 Schüler auf ihre Leistungsfähigkeit getestet. Die ausgewählten Schüler stehen stellvertretend für insgesamt rund 28 Millionen fünfzehnjährige der 65 teilnehmenden Länder und Volkswirtschaften. In der Bundesrepublik Deutschland wurden hierzu aus insgesamt 247 Schulen zufällig Schüler aus der Grundgesamtheit gezogen.

Durchführung

Die Erhebung der Tests fand im Zeitraum 01. März bis 30. Juni 2012 statt. Insgesamt nahmen circa 500.000 Jugendliche an der PISA-Studie teil. Die Schülerinnen und Schüler hatten zur Bearbeitung der Tests jeweils zwei Stunden Zeit, wobei die Tests an zwei Tagen durchgeführt wurden. Eine Ausnahmeregelung wurde hierbei an Sonderschulen getroffen, in jenen die Schüler eine verkürzte Testversion bekamen und diese an einem Tag bearbeiten sollten.

Den Schülern wurden vorab Instruktionen zur Durchführung der Tests aus einem Skript von geschulten Testleiterinnen und Testleitern vorgelesen, um die Vergleichbarkeit der Durchführungsbedingungen zu gewährleisten. Eine nachträgliche Testerhebung wurde an einer Schule gestattet, da am Tag der Testdurchführung vier oder mehr Schüler fehlten, die zuvor durch das Auswahlkriterium festgestellt wurden.

Die Tests wurden Mithilfe der Datenschutzbeauftragten der Länder zusammen entwickelt. Um zu gewährleisten, dass alle Gesichtspunkte hinsichtlich des Zwecks der Studie berücksichtigt werden, wurden zur Konzipierung der Tests international führende Institutionen und Experten zur Testentwicklung hinzugerufen.

Eltern und Schüler wurden vorab über das Vorgehen, sowie die Ziele der Studie aufgeklärt und erklärten sich mit der Teilnahme schriftlich einverstanden.

Um eine ordnungsgemäße Durchführung der Tests zu gewährleisten fanden an insgesamt 35 Schulen unangemeldete Kontrollen statt, die bestätigten, dass die Untersuchungsbedingungen eingehalten wurden.

Erhebungszyklus

Seit dem Jahr 2000 werden im Dreijahreszyklus PISA-Studien durchgeführt. Einerseits soll so die Entwicklung der Leistungsfähigkeit der Schüler beobachtet werden, um sie sowohl national, als auch international zu vergleichen, andererseits wird

durch die kurzen Abstände der Testdurchführung ein schnelles Eingreifen der Länder möglich.

In jeden Zyklus wird das Hauptaugenmerk auf einen Testschwerpunkt gelegt, der zwei Drittel der Testzeit in Anspruch nimmt. Als Schwerpunkt der PISA-Erhebung 2012 wurden die mathematischen Fähigkeiten festgesetzt.

2 Kompetenzbereiche

Die 15-jährigen Mädchen und Jungen werden in mehreren Kompetenzbereichen überprüft. Zum einen werden Hintergrundfragebögen zur eigenen Person erstellt, zum anderen sollen die Schüler ihr Wissen in den Fachbereichen Lesen, Mathematik und Naturwissenschaften unter Beweis stellen. Die Tests beinhalten sowohl offene Fragen, als auch einen Multiple Choice Teil.

Ziel des Tests ist es die erworbenen Fähigkeiten auf realitätsnahe Herausforderungen anzuwenden. Auch fächerübergreifende Kompetenzen wurden untersucht, wie die Einstellung zum Lernen, dem Umgang mit dem Computer, die Motivation der Schüler und die angewandten Lernstrategien.

Um eine Aussage über die Unterschiede in den verschiedenen Schulformen treffen zu können wurden auch die Schulleiter gebeten einen 30-minütigen Fragebogen zu den Merkmalen der Schule auszufüllen.

Der Test beinhaltete ein breites Spektrum von unterschiedlichen Aufgaben, die alle in schriftlicher Form durchgeführt wurden. Die einzelnen Aufgaben wiesen unterschiedliche Formate auf. Fragen zu den Texten sollten beantwortet werden, indem die Überlegungen hierzu in Verbindung mit den Diagrammen bedacht wurden. Die Testaufgaben in allen Bereichen hatten eine Bearbeitungszeit von sieben Stunden bemessen. Insgesamt wurden neun verschiedenen Testhefte erstellt und die Aufgaben auf mehrere Aufgabenhefte aufgeteilt. Dies stellt sicher, dass jede Aufgabe von einer repräsentativen Stichprobe von Schülern bearbeitet wurde. Im nachfolgendem Abschnitt werden die drei großen Grundkompetenzen der Schüler näher erläutert.

Lesekompetenz

Unter Lesekompetenz versteht man die Fähigkeit geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren. So soll in den verwendeten Tests die schriftliche Information derart genutzt werden um mit diesen die Aufgabenstellung zu lösen. Grundsätzlich war folgendes zu unterscheiden:

- *die Art des Textes:*
- *die Art der Leseaufgabe*
- *die Art des Gebrauchs, für den der Text verfasst wurde:*

Mathematische Grundbildung

Für die mathematische Grundbildung steht nicht einzig die durch den Lehrplan vorgegebene Mathematik im Vordergrund. Vielmehr sollen die Schüler ihr Wissen in alltäglichen Situationen anwenden und zeigen, dass sie mit Hilfe der Mathematik Probleme in Beruf und Alltag lösen können. Die Vertretung eines Standpunkts oder die Einschätzung eines mathematisch ausgedrückten Sachverhalts sollen so angewandt werden.

Die mathematischen Grundkenntnisse wurden in drei Konzepte aufgeschlüsselt:

- *Mathematische Inhalte*
- *Mathematische Prozesse*
- *Situationen, in denen Mathematik angewendet wird*

Naturwissenschaftliche Grundbildung

Die naturwissenschaftliche Grundbildung beinhaltet nicht nur die in den Schulen gelehrteten naturwissenschaftlichen Fächer. Vielmehr soll untersucht werden, ob die Schüler naturwissenschaftliches Wissen handhaben können, indem sie die Konzepte verstanden haben und diese auf alltägliche Problemstellungen anwenden können. Die Schlüsselkonzepte der naturwissenschaftlichen Kompetenzen beinhalten folgende Bereiche:

- *Naturwissenschaftliche Konzepte*
- *Naturwissenschaftliche Prozesse*
- *Naturwissenschaftliche Anwendungsbereiche*

3 Methodik

Stichprobenerhebung

Eine Vollerhebung der PISA-Studie ist aus organisatorischen, zeitlichen und finanziellen Gründen nicht möglich. Aus diesem Grund wurde eine systematische Teilerhebung durchgeführt, welche durch die Ziehung einer Zufallsstichprobe erfolgte. Dadurch soll gewährleistet werden, dass ein Rückschluss auf die Grundgesamtheit zugelassen werden kann. Die international vorgegebene Zielpopulation besteht aus allen fünfzehnjährigen Schülern, wobei die genaue Altersdefinition mit dem internationalen PISA-Konsortium abgestimmt wurde.

Da Schulleistungsstudien, insbesondere *Large-Scale-Assessments*, besondere Anforderungen stellen, werden im Allgemeinen zwei- oder mehrstufige Auswahlverfahren angewendet.

Stichprobenplan und Ziehung der Stichprobe

Die Grundgesamtheit der PISA Studie setzt sich aus der Kohorte der 15-jährigen Schülern ab der siebten Klasse zusammen. Das sind in Deutschland genau die Schüler, die zwischen dem 01. Januar und dem 31. Dezember 1996 geboren wurden. Die Auswahl der teilnehmenden Schüler erfolgt anhand eines zweistufigen Verfahrens. In Deutschland wurde ein erweitertes zweistufiges Verfahren angewandt, welches in den nachfolgenden Abschnitten detaillierter betrachtet wird.

Das grundlegende Verfahren besteht darin, dass zunächst eine Zufallsstichprobe von Schulen gezogen wird. An diesen wird zufällig eine Gruppe von 25 Schülern ausgewählt. Zudem wurden zwei neunte Klassen per Zufallsverfahren ausgewählt, die an der PISA-Studie teilnahmen. Zusätzlich nimmt Deutschland an dem *Computer-based Assessment* (CBA) teil. Die Kompetenzbereiche Lesen, Mathematik und Problemlösung standen den an dem *Computer-based Assessment* teilnehmenden Ländern zur Auswahl, wobei sie sich auf einen Kompetenzbereich konzentrieren mussten. Deutschland fokussierte sich auf den Bereich der Problemlösung. Für das CBA wurden aus den pro Schule ausgewählten 25 Schülern noch einmal zufällig 14 Schüler separiert.

Wie diese mehrstufige Auswahl für Deutschland im Detail erfolgte wird in den nächsten Abschnitten genauer erläutert.

1. Schritt: Erstellung eines Sampling-Frame

Im ersten Schritt wurde die für Deutschland zu analysierende Grundgesamtheit bestimmt. Da bei der späteren Stichprobenziehung zunächst die Schulen ausgewählt wurden war es notwendig, die Grundgesamtheit aller Schulen zu bestimmen, welche potentiell von 15-jährigen Schülern besucht wurden. Diese Grundgesamtheit wird als Sampling-Frame bezeichnet. Hierbei ist es wichtig zu erwähnen, dass bei dem Sampling-Frame keine Schulen mehrfach, fehlerhaft oder unvollständig dokumentiert werden durften. Eine Einrichtung wird gemäß der Richtlinie des KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) als Schule gewertet, wenn diese hinsichtlich des Lehrplans und des Qualifikationsniveaus als eigenständige Einheit verstanden werden kann (vgl. KMK, 2013, S.8). Dies bedeutet, dass sich die Definition einer Schule nicht aus der räumlichen Gegebenheit ergibt. Um das Sampling-Frame aller relevanten Schulen zu erstellen, wurden von den 14 statistischen Landesämtern folgende Informationen auf Basis der Schulstatistik 2010/2011 eingeholt:

- die offizielle Schulnummer
- die Schulart
- die Anzahl an Schüler in den Geburtsjahrgängen 1993 bis 1996
- die Anzahl der Schüler in den 8. bis 10. Klassen
- die Anzahl der Klassen 8. bis 10.
- Informationen über Schulveränderungen (Schließungen, Zusammenlegungen und Schulart)
- für Förderschulen die Informationen über die Förderschwerpunkte

2. Schritt: Ziehung der Schulstichprobe

Nachdem die Grundgesamtheit aller relevanten Schulen erstellt wurde, konnte diese anhand mehrerer sogenannter Stratifizierungskriterien genauer unterteilt werden. Diese Unterteilung bietet mehrere Vorteile:

Die Effizienz der Sampling-Frames wurde durch die Unterteilung erhöht, was bedeutet, dass die Populationsparameter verlässlicher geschätzt werden konnten. Dies ermöglichte es, mit kleineren Stichproben verlässliche Ergebnisse zu erhalten. Zudem

ermöglicht die Stratifizierung zu überprüfen, ob die Charakteristika mit jenen der Grundgesamtheit übereinstimmen. Darüber hinaus kann die Analyse unterschiedlich nach bestimmten Kriterien erfolgen. So können beispielsweise mit Hilfe der Stratifizierung die Ergebnisse einer Hauptschule mit den Ergebnissen einer Realschule verglichen werden.

Stratifizierungen können nach expliziter Stratifizierung und impliziter Stratifizierung unterscheiden werden. Bei der expliziten Stratifizierung wird die Grundgesamtheit zunächst anhand bestimmter Charakteristika in unabhängige Gruppen unterteilt. Aus jeder Gruppe wird anschließend eine separate Stichprobe gezogen. Bei der impliziten Stratifizierung hingegen werden die anhand der expliziten Strata gezogenen Stichproben in einzelne Merkmalsklassen unterteilt. In Deutschland wird die Grundgesamtheit aller Schulen mittels zweier expliziter und zweier impliziter Stratifizierungsverfahren aufgeteilt:

Die Schulen werden zunächst in allgemeinbildende Schulen, Förderschulen und Berufsschulen unterteilt, was einem expliziten Stratifizierungsverfahren entspricht. Anschließend folgt eine weitere explizite Stratifizierung, indem die allgemeinbildenden Schulen nach Bundesländern aufgeteilt werden. Für die implizite Stratifizierung werden die Berufs- und Förderschulen nach Bundesländern aufgeteilt. Zudem werden die allgemeinbildenden Schulen in Hauptschulen, Realschulen und Gymnasien unterteilt. Daraus folgt, dass die Schulen in 18 explizite Strata aufgeteilt werden (16 Bundesländer und 2 Förder- und Berufsschulen) aus denen jeweils separat zufällige Stichproben gezogen werden.

Die Stichproben aus den jeweiligen Strata wurden vom internationalen PISA Konsortium gezogen. Es ergab sich ein Umfang von insgesamt 247 in die Stichprobe gezogenen Schulen. Hierbei wurde darauf geachtet, dass der proportionale Anteil der Schüler in den Stichproben approximativ jenem der Grundgesamtheit innerhalb der Staaten gleicht.

3. Schritt: Ziehung der Schülerstichprobe

Bevor die Schülerstichprobe gezogen werden konnte, mussten bestimmte Informationen der einzelnen Schüler in einer Liste eingetragen werden. Diese Informationen bestanden aus dem Vor- und Nachnamen, dem Geschlecht, dem Geburtsjahr und Geburtsmonat, der Klassenbezeichnung und der Information über einen möglichen

	Haupt- schule	Inte- grierte Gesamt- schule	Schule mit mehreren Bildungs- gängen	Real- schule	Gymna- sium	Förder- schule	Berufs- schule	Gesamt
Baden- Württemberg	2	1	5	13	12	2	4	39
Bayern	10	1	0	13	12	1	6	43
Berlin	0	0	4	0	3	0	0	7
Brandenburg	0	0	2	0	2	1	0	5
Bremen	0	0	1	0	1	0	0	2
Hamburg	0	0	2	0	2	0	1	5
Hessen	2	3	0	5	6	1	1	18
Mecklenburg- Vorpommern	0	0	2	0	1	0	0	3
Niedersachsen	4	2	0	9	8	2	1	26
Nordrhein- Westfalen	10	10	0	15	18	3	2	58
Rheinland-Pfalz	0	2	4	1	4	1	0	12
Saarland	0	0	1	0	1	0	1	3
Sachsen	0	0	3	0	3	1	1	8
Sachsen-Anhalt	0	0	2	0	2	0	0	4
Schleswig- Holstein	0	4	1	0	3	1	0	9
Thüringen	0	0	3	0	1	0	1	5
Gesamt	28	23	30	56	79	13	18	247

Fig. 1: In die Stichprobe gezogene Schulen nach Bundesland und Schulart
(Prenzel M., et al. (2013), S. 316)

Förderbedarf der Schüler. Diese Liste wurde mit der Liste des Statistischen Landesamtes verglichen. Um den Datenschutz einhalten zu können, wurden den Schülern Pseudonyme zugeschrieben. Waren keine Diskrepanzen zwischen den Schülerlisten der Schulen und der des Statistischen Landesamtes erkennbar, wurde mit der Ziehung der Stichprobe begonnen.

Aus den vorher 247 ausgewählten Schulen wird nun mittels einem Zufallsverfahren eine Stichprobe von 25 Schülern pro Schule gezogen. Dies gilt allerdings nur für die allgemeinbildenden Schulen. Zudem wurde an den allgemeinbildenden Schulen eine Stichprobe von zwei 9. Klassen gezogen, die vollständig mit in die Stichprobe aufgenommen wurden. Zusätzlich wurden aus den 25 gezogenen Schülern nochmals 14 Schüler gezogen, die an dem Testprogramm CBA teilnahmen. Bei allen ausgewählten Berufs- und Förderschulen konnte eine Vollerhebung aller neunten Klassen durchgeführt werden. Dieses Verfahren soll durch die unten stehende Grafik veranschaulicht werden.

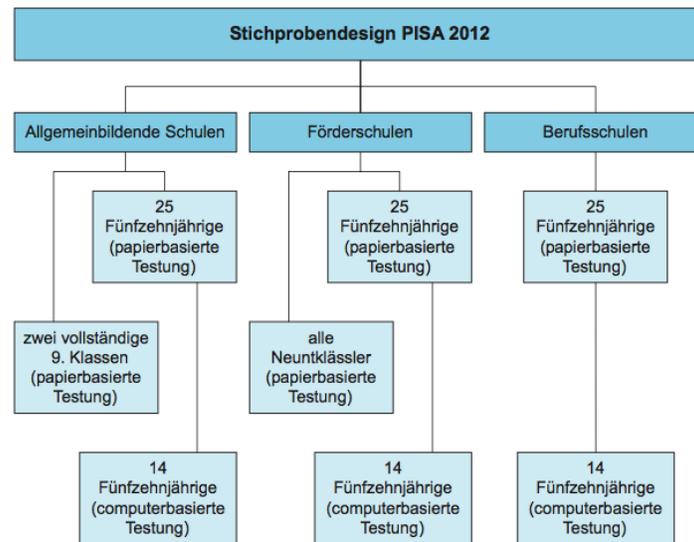


Fig. 2: Stichprobendesign PISA 2012

Realisierte Stichproben

In den vorherigen Abschnitten wurden die Grundgesamtheit, die Schulstichproben und Schülerstichproben genauer erläutert. Nun gilt es, eine weitere Unterteilung zu betrachten. In diesem Abschnitt wird zwischen Brutto- und Nettostichproben unterschieden. Unter Bruttostichproben versteht man alle für die PISA-Studie 2012 ausgewählten Schülerinnen und Schüler, wohingegen man bei der Nettostichprobe nur von der Anzahl von Schülern ausgeht die tatsächlich an der Studie teilnahmen.

In die Stichprobe der PISA-Studie wurden insgesamt 247 Schulen gezogen, wobei sich von diesen fünf Privatschulen gegen die Teilnahme an der Studie aussprachen. Da es den Schulen freigestellt ist an der Studie teilzunehmen, fielen diese fünf Schulen aus der Stichprobe heraus. War dies der Fall, so konnten Ersatzschulen für die entfallenen Schulen bestimmt werden.

Des Weiteren gab es 17 Ausfälle von Schulen, für die keine Ersatzschulen bestimmt werden konnten, da diese entweder nicht vorgesehen waren oder es keine gab. Unter diesen 17 Schulen befanden sich 11 Berufsschulen, die keine 15-jährigen Schüler unterrichteten. Weitere 2 Privatschulen entfielen, die selbst schon als Ersatzschule vorgesehen waren. Zwei der gezogenen Schulen wurden aufgelöst und existierten somit nicht mehr. Für eine ausgefallene Privatschule wurde kein Ersatz gefunden. Eine weitere dänische Schule fiel aus der gezogenen Schulstichprobe, da sie nur vier 15-jährige Schüler an der Schule hatten.

Somit verblieben 230 Schulen, die an der PISA-Studie 2012 teilnahmen. Bei den Schulen wurde eine Teilnehmerquote von 98,3 % erreicht. Bei den einzelnen Schülern betrug die Quote der Teilnehmer 98,2 %. Somit wurde die Mindestvorgabe der OECD Standards erfüllt. Die deutsche PISA-Studie umfasste $n = 5001$ Schüler und $n = 9998$ Neuntklässler.

Gewichtung

Da die Stichproben in mehreren Stufen erfolgte, beeinflusst die Ziehung der Stichprobe die einzelnen Ziehungswahrscheinlichkeiten. So hat nicht jede Schule und nicht jeder Schüler die gleiche Wahrscheinlichkeit in die Stichprobe mitaufgenommen zu werden. Die Gleichheit der Ziehungswahrscheinlichkeiten der einzelnen Schulen wird durch eine Gewichtung wiederhergestellt.

Schulbasisgewichte

Die Sampling-Frames wurden in 18 Strata aufgeteilt und danach nach den Schulformen sortiert. Beispielsweise hat eine Schule in einem Stratum eine niedrigere Wahrscheinlichkeit in die Stichprobe zu gelangen, wenn in diesem Stratum vergleichsweise viele Schulen der selben Schulform vorkommen. Diese Ungleichheit wird durch eine umgekehrt proportionale Gewichtung ausgeglichen, so dass jede Schule *ex post* die selbe Wahrscheinlichkeit hat, in die Stichprobe mitaufgenommen zu werden.

Schülerbasisgewichte und Schülergesamtgewicht

Bei den Ziehungswahrscheinlichkeiten der einzelnen Schüler macht es einen Unterschied, ob es eine große oder eine kleine Schule ist. So hat Beispielsweise ein Schüler einer großen Schule eine kleinere Ziehungswahrscheinlichkeit in die Stichprobe mitaufgenommen zu werden. Bei Schulen mit vielen 15-jährigen Schülern hat jeder einzelne nur eine geringe Wahrscheinlichkeit in die Stichprobe zu gelangen. Umgekehrt gilt das selbe für einen Schüler, der in einer Schule mit wenigen 15-jährigen Schülern unterrichtet wurde. Dieser hat eine höhere Ziehungswahrscheinlichkeit.

Da bei der PISA-Studie gültige Aussagen auf Ebene der Grundgesamtheit getroffen werden sollte, wurde jedem Schüler *post hoc* innerhalb eines Stratum die selbe Ziehungswahrscheinlichkeit zugewiesen. Auch diese wurden umgekehrt proportional zur Ziehungswahrscheinlichkeit angerechnet. Das Schul- bzw. Schülerbasisgewicht lässt sich nach diesen Kriterien berechnen und erlaubt so eine gerechte Ziehung der

Schüler bzw. Klassen in die Stichprobe. Aus diesen einzelnen Gewichten wird nachfolgend das Schülergesamtgewicht berechnet, das als Grundlage der Datenbasis für die PISA-Studie 2012 dient.

Notwendigkeit einer Gewichtung

Neben dem Schulbasisgewicht und dem Schülergesamtgewicht werden noch weitere Gewichtungen vorgenommen. Falls ein in die Stichprobe gezogener Schüler wegen Krankheit am Testtag nicht an der Studie teilnehmen kann, kann es passieren, dass dieser Schüler in der Stichprobe unterrepräsentiert wird. Dies kann zu verzerrten Schätzungen führen. Aus diesem Grund werden weitere Korrekturfaktor eingesetzt.

Korrekturfaktoren für die Gewichtung

Neben den bereits bekannten Gewichtungen der PISA-Studie werden noch fünf weitere Gewichte vorgestellt, welche die Schätzung auf der Ebene der Grundgesamtheit ermöglichen. Zum einen werden *Schulausfälle kompensiert*, um eine Unterrepräsentation zu vermeiden. Als zweiter Faktor werden Schüler in *Staaten* so gewichtet in die Stichprobe mitaufgenommen, wie tatsächlich auf Staatsebene zu erwarten sind. Des weiteren soll eine *Nicht-Teilnahme* von Schülern in den Stichproben ausgeglichen werden. Die beiden letzten Faktoren werden als *Trimming Factors* bezeichnet. Sie sollen zum einen ungewöhnlich hohe Schulbasisgewichte und zum anderen außergewöhnlich hohe Gewichte auf Schülerebene ausgleichen.

Skalierung von Leistungstests zur Modellierung von Kompetenzen

Bei der Entwicklung von Tests kommt es zu einem Konflikt zwischen den psychometrischen Anforderungen aus messtheoretischer Sicht und der Praktischen Durchführbarkeit der Tests. Die Tests von PISA 2012 sollen latente Eigenschaften der zu prüfenden Personen erfassen. Unter latenten Personeneigenschaften der Testtrune versteht man in diesem Fall die Eigenschaften der Schüler in den drei Kompetenzbereichen. Einerseits soll eine möglichst große Anzahl an Fragen und Aufgaben (*Items*) beantwortet werden, andererseits ist die Testzeit begrenzt. Zugleich möchte man mehrere unterschiedliche latente Eigenschaften erfassen. Bei *Large-Scale Assessments* wird zur Gestaltung solcher Fragebögen ein sogenanntes *Balanced Incomplete Block Design* verwendet. Dieses Design wurde entwickelt, um eine randomisierte, balancierte allerdings unvollständige Versuchsplanung durchzuführen. Die Grundüber-

legung hierfür besteht darin, dass nicht jeder Schüler alle Testdurchführung entwickelten Fragen bekommt. Es werden mehrere Testhefte erstellt, die eine unterschiedliche Auswahl von Aufgabengruppen in den einzelnen Testheften enthalten. Die nach dieser Art gestalteten Messinstrumente werden als *Multi-Matrix-Design* bezeichnet. Das *Multi-Matrix-Design* wurde entwickelt, um Daten mit unvollständiger, balancierter Struktur und einer großen Anzahl von Items zur Bestimmung von Populationsschätzwerten auf Basis der Erhebung großer Stichproben zu bestimmen. Bei der Erhebung der PISA-Studie 2012 wurde erstmals ein *Multi-Matrix-Design* verwendet, um einzelne Hintergrundfragen aus dem Schülerfragebogen auf drei Testhefte zu erfassen.

Das PISA-Multi-Matrix-Design und IRT-Skalierung

Beschreibung der PISA-Items als manifestierte Indikatoren kognitiver Kompetenz

Die drei Kompetenzbereiche Lesen, Mathematik und Naturwissenschaften wurden mit Schwerpunkt Mathematik bei PISA 2012 getestet. Hierbei wurden die einzelnen *Items* zu den verschiedenen Bereichen um einen Aufgabenstamm (*Testlet* oder *Unit*) gruppiert. Diese *Testlets* oder *Units* enthalten Textelemente, Grafiken, Tabellen oder Kombinationen hieraus, wobei die Anzahl der *Items* pro *Unit* zwischen einer und sieben Aufgaben variieren konnte. Insgesamt wurden 87 *Units* erstellt, was 207 einzelnen *Items* entspricht. Folgende Grafik soll dies nochmal verdeutlichen.

	Units gesamt	Items gesamt	Link-Units	Link-Items
Mathematik	56	110	25	36
Lesen	13	44	13	44
Naturwissenschaften	18	53	18	53
Gesamt	87	207	56	133

Fig. 3: Anzahl der an PISA 2012 eingesetzten Units und Items
(Prenzel M., (2013), S. 324)

Hierbei beschreiben die *Link-Units* die Aufgaben, die aus früheren PISA-Studien übernommen worden sind. In den Kompetenzbereichen Lesen und Naturwissenschaften wurden ausschließlich *Link-Units* verwendet, wobei für den Kompetenzbereich Mathematik mehr als 50% der Aufgaben neu erstellt worden sind.

Aufbau der Testhefte

Die vorher beschriebenen Units wurden für die drei Kompetenzbereiche in 15 Aufgabengruppen (*Cluster*) unterteilt. Für den Bereich der Mathematik wurden neun

Cluster erstellt (PM1 bis PM5, PM6a, PM6b, PM7a, PM7b). Die *Cluster* PM6a und PM7a wurden hierbei in allen Staaten eingesetzt, die in der PISA-Studie 2009 in dem Bereich Mathematik mindestens 480 Punkte erzielten. Alle anderen Staaten erhielten die *Cluster* PM6b und PM7b, wobei diese Aufgabengruppen leichter sind. In den Kompetenzbereichen Lesen und Naturwissenschaften wurden drei Cluster (PR1 - PR3 bzw. PS1 - PS3) erstellt, wobei hiervon jedes *Cluster* in ca. 30 Minuten von insgesamt 120 Minuten bearbeitbar waren. In der Haupterhebung wurden 15 *Cluster* auf 13 Standardtesthefte verteilt, wobei jedes dieser Hefte vier Aufgabengruppen enthielt mit einer Bearbeitungszeit von insgesamt ca. zwei Stunden. In der nachfolgenden Grafik

Testheft													
Dauer	1	2	3	4	5	6	7	8	9	10	11	12	13
30	M5	N3	L3	M6	M7	M1	M2	N2	L2	M3	M4	N1	L1
30	N3	L3	M6	M7	N1	M2	N2	L2	M3	M4	M5	L1	M1
30	M6	M7	N1	L1	M1	L1	M3	M4	M5	N3	L3	M2	N2
30	N2	L2	M3	M4	M5	M6	M7	N1	L1	M1	M2	N3	L3
45	Fragebogen												

Anmerkung: M = Mathematik, N = Naturwissenschaften, L = Lesen.

Fig. 4: In Deutschland eingesetzte Testhefte (Prenzel M., (2013), S. 325)

Vergleichbarkeit der Schülerleistungen untereinander

Da jeder Schüler ein unterschiedliches Testheft zur Bearbeitung vorgelegt bekommt, die eine unterschiedliche Auswahl an Aufgaben aus einem Aufgabenpool enthält, ist die direkte Vergleichbarkeit zwischen den Individuen nicht gegeben. Obwohl die Aufgabenschwierigkeiten in den einzelnen Testheften weitestgehend gleich verteilt ist, sind die darin enthaltenen Aufgaben dennoch unterschiedlich schwer. Um einen Vergleich zwischen einzelnen Schülern bzw. Gruppen von Schülern anstellen zu können, kann daher nicht die Anzahl an gelösten Aufgaben als Vergleichswert herangezogen werden. Um einen objektiven und fairen Vergleich anstellen zu können, muss die Schwierigkeit der einzelnen Aufgaben mit berücksichtigt werden. Die Schätzung der individuellen Kompetenzwerte wird durch die Verwendung des *Rasch-Modells* gewährleistet.

Latente (kontinuierliche) Personenvariable

Die *Item-Response-Theory* (IRT) wird bei den Auswertungen der PISA-Statistik verwendet. Die latente und individuelle Personeneigenschaften sind anhand des An-

wortverhaltens der Individuen gegeben. Um den Zusammenhang zwischen den Ausprägungen auf der latenten Variable und der beobachtbaren Lösung der Aufgaben herzustellen, wird auf das *Rasch-Modell* (Rasch, 1960) und dessen Erweiterung dem *Partial-Credit-Modell* (Masters, 1962) zurückgegriffen.

Das Rasch-Modell

Das *Rasch-Modell* basiert auf zwei grundlegenden Annahmen. Der erste Parameter wird als *Personenparameter* bezeichnet, wobei dieser die Fähigkeit der getesteten Person wiedergeben soll. Als zweiten Parameter wird der sogenannte *Itemparameter* benötigt, der den Schwierigkeitsgrad der Aufgabe beschreiben soll. Das *Rasch-Modell* beschreibt neben den beiden oben genannten Parametern zusätzlich eine psychologisch plausible, stochastisch-probabilisierte Beziehung der beiden Parameter und die Wahrscheinlichkeit zur richtigen Lösung einer Aufgabe. Die Beziehung der Parameter fließt als logistische Funktion in das Modell mit ein. Zur grafischen Darstellung verwendet man die *Item-Characteristic Curve* (ICC), welche wie folgt interpretiert werden kann:

Mit zunehmender positiver Differenz zwischen dem *Personenparameter* und dem *Itemparameter* nimmt die Wahrscheinlichkeit der Lösung einer Aufgabe oberhalb der 50% Lösungswahrscheinlichkeit zu. Dies bedeutet, dass die Lösungswahrscheinlichkeit des betreffenden *Items* größer als 50% ist, wenn der *Personenparameter* größer ist, als der *Itemparameter*.

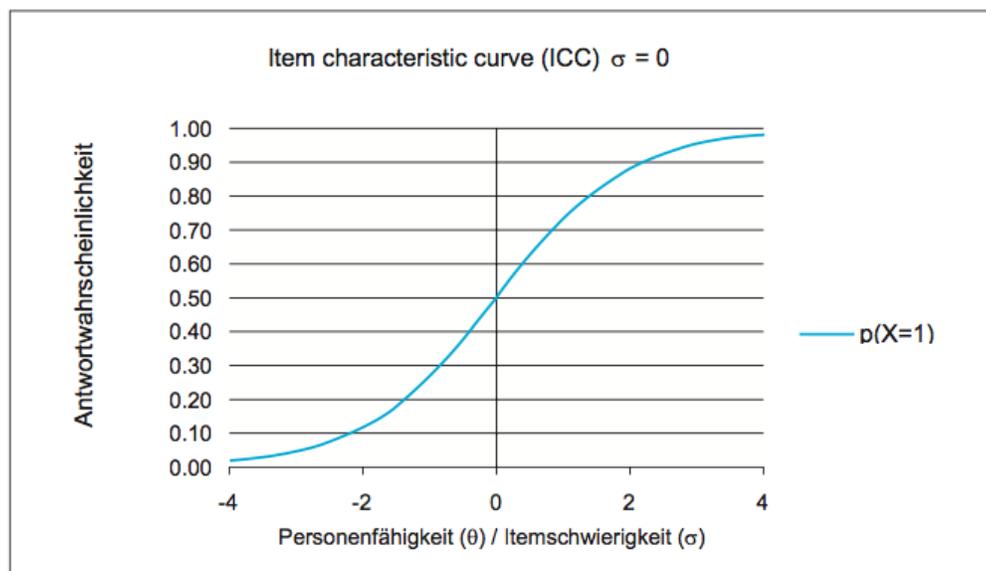


Fig. 5: Item-Charakteristik Curve (ICC) (Prenzel M., (2013), S. 327)

Formal:

σ Itemparameter

θ Personenparameter

$p(X)$ Lösungswahrscheinlichkeit

Person v

Item i

$$p(X_{vi}) = \frac{\exp(X_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)}, X \in 0, 1 \quad (1)$$

Hierbei bedeutet $X_{vi} = 0$, dass eine falsche Antwort gegeben wurde und $X_{vi} = 1$, dass die gegebene Antwort richtig war. Da das *Rasch-Modell* keine teilweise richtigen Antworten zulässt, wird hierfür auf die Erweiterung des *Rasch-Modells* zurückgegriffen, das sogenannte *Partial-Credit-Modell*.

Das Partial-Credit-Modell

Das *Partial-Credit-Modell* lässt es zu mehrstufige Antwortformate mit zu berücksichtigen. So werden auch teilweise richtig gegebene Antworten von Schülern in der Testauswertung berücksichtigt. Um dies zu ermöglichen, wird der *Itemparameter* in einzelne *Schwellenparameter* zerlegt, die ein mehrstufiges, ordinales Antwortformat der Schüler zulassen.

Formal:

σ Itemparameter

θ Personenparameter

s Schwellenwerte

σ_{ix} realisierter Schwellenparameter

σ_{is} möglicher Schwellenparameter

$m + 1$ mögliche Kategorien

$$p(X_{vi} = x) = \frac{\exp((x\theta_v) - \sigma_{ix})}{\sum_{s=0}^m \exp((s\theta_v) - \sigma_{is})}, x \in 0, 1, \dots, m \quad (2)$$

Die Kategorien müssen aufsteigend von 0 bis m kodiert werden. Nun können die Antworten der Schüler in richtig/falsch und teilweise richtig übertragen werden.

4 Literaturverzeichnis

ARTELT C., BAUMERT J., KLIEME E., NEUBRAND M., PRENZEL M., SCHIEFELE U., SCHNEIDER W., SCHÜMER G., STANAT P., TILLMANN K., WEIß M. (HRSG.) (2001): *PISA 2000 Zusammenfassung und zentrale Befunde* Max-Planck-Institut für Bildungsforschung, Berlin

OECD (HRSG.). (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD. (ISBN 92 - 64 - 19671 - 4)

OECD (HRSG.). (2013). *PISA 2012 Ergebnisse, Band I, Was Schülerinnen und Schüler wissen und wie sie dieses Wissen einsetzen können*

PRENZEL M., SÄLZER C., KLIEME E., KÖLLER O. (HRSG.)(2013): *PISA 2012 Fortschritte und Herausforderungen in Deutschland* Waxmann Verlag, Münster/New York, München/ Berlin

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2013). Definitionenkatalog zur Schulstatistik 2012. Zugriff am 14.05.2015. Verfügbar unter <http://www.kmk.org/statistik/schule/statistische-veroeffentlichungen/definitionenkatalog-zur-schulstatistik.html>.

RASCH, G. W. (1960). Probabilistic models for some intelligence and attainment tests (Studies in mathematical psychology). Chicago: The University of Chicago Press.

MASTERS, G. (1982). A Rasch Model for partial credit scoring. *Psychometrika*, 47 (2), 149-174.