

Unterschrift des Betreuers



MASTERARBEIT

# Erstellen einer Standard-Dokumentation für PISA 2003

Eingereicht an der Technischen Universität Wien und  
ausgeführt am Institut für Scientific Computing der Universität Wien

unter der Anleitung von Prof. Dr. Wilfried Grossmann

durch

Florian Schneckenleithner

1210 Wien, Schererstr. 34

[f.schneckenleithner@gmail.com](mailto:f.schneckenleithner@gmail.com)

---

Datum

---

Unterschrift (Student)

## Abstract

Standarddokumentationen sind ein anerkanntes Mittel zur möglichst kurzen und aussagekräftigen Beschreibung der Besonderheiten von Sachverhalten. Die von der OECD initiierte Schulleistungsuntersuchung PISA, ist eine in ihrer Komplexität und ihrem Umfang bisher unerreichte Erhebung des Bildungsstandes von Schülern. Sie findet alle 3 Jahre, in mittlerweile 61 Ländern (2006) statt. Unter Verwendung eines adäquaten Templates der Statistik Austria, wurde im dritten Kapitel eine, ohne den Rest der Arbeit verwendbare "Standard-Dokumentation Metadaten" erstellt. Darin werden allgemeine Informationen, statistische Konzepte und Methoden, Erstellung der Statistik, Datenaufarbeitung und qualitätssichernde Maßnahmen, Publikation und Qualität von PISA in kurzer und prägnanter Weise dokumentiert. In Kapitel 4 gibt ein Ländervergleich zwischen Österreich und Finnland, an Hand derselben Standard-Vorlage, Einblick in die unterschiedlichen nationalen Ausprägungen der Erhebung.

## **Danksagung**

Riina, Ronja und meinen lieben Eltern.

Professor Grossmann für die persönliche Betreuung.

## Inhaltsverzeichnis

|   |           |
|---|-----------|
| Abstract.....   | 2         |
| Danksagung.....   | 3         |
| Inhaltsverzeichnis.....   | 4         |
| Vorwort.....  | 5         |
| <b>1 Einführung.....</b>  | <b>6</b>  |
| 1.1 Motivation .....  | 7         |
| 1.2 Struktur der Arbeit.....  | 7         |
| 1.3 Definitionen .....  | 8         |
| 1.3.1 Standard-Dokumentation.....   | 8         |
| 1.3.2 Metainformationen.....  | 9         |
| 1.4 Ziele .....   | 9         |
| <b>2 Grundlagen.....</b>  | <b>10</b> |
| 2.1 Technischer Bericht der OECD zu PISA.....                                     | 11        |
| 2.1.1 Technical Report PISA 2000.....   | 11        |
| 2.1.2 Technical Report PISA 2003.....   | 11        |
| 2.2 Technische Berichte der Nationalen Projektzentren.....                        | 12        |
| 2.2.1 Technischer Bericht des österreichischen Projektzentrums.....               | 12        |
| 2.2.2 Technischer Bericht des finnischen Projektzentrums.....                     | 12        |
| 2.3 Statistik Austria Standard-Dokumentation Metadaten.....                       | 13        |
| 2.4 Eurostat Handbuch "How to make a Quality Report".....                         | 15        |
| <b>3 Umsetzung Dokumentation.....</b>   | <b>16</b> |
| 1. Wichtige Hinweise.....   | 17        |
| 2. Allgemeine Informationen.....  | 17        |
| 3. Statistische Konzepte, Methodik .....  | 19        |
| 4. Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde Maßnahmen..... | 32        |
| 5. Publikation (Zugänglichkeit).....  | 38        |
| 6. Qualität .....   | 39        |
| 6.1 Relevanz.....   | 39        |
| 6.2 Genauigkeit.....  | 40        |
| 6.2.1 Stichprobenbedingte Effekte, Repräsentativität.....                         | 40        |
| 6.2.2 Nicht-stichprobenbedingte Effekte .....                                     | 41        |
| 6.3 Rechtzeitigkeit und Aktualität.....   | 46        |
| 6.4 Vergleichbarkeit .....  | 47        |
| 6.5 Kohärenz .....  | 47        |
| <b>4 Vergleich zweier Länder .....</b>  | <b>48</b> |
| <b>5 Zusammenfassung .....</b>  | <b>62</b> |
| 5.1 Dokumentation.....  | 62        |
| 5.2 Ländervergleich.....  | 62        |
| 5.3 Resumee .....   | 63        |
| 5.4 Ausblick (PISA06) .....   | 63        |
| <b>6 Literaturverzeichnis .....</b>   | <b>65</b> |

## Vorwort

Als Bachelor der Wirtschaftsinformatik und derzeitiger Student der beiden Master-Studiengänge Informatikmanagement und Wirtschaftsinformatik, ist eine Diplomarbeit über PISA sicher nicht ein typisches Thema. Meine Kernfachkombination "Entscheidungsunterstützung im öffentlichen Sektor", vor allem aber die Hinzunahme von Informatikmanagement in mein Studium, machte Bildung zu einem für mich relevanten Thema.

Aus einer Familie von Lehrern stammend, wurde mir der Blick auf das österreichische Bildungssystem mit in die Wiege gelegt. Im Zuge meines einjährigen Studienaufenthalts in Schweden und durch meine finnische Frau, bekam mein Verständnis von Bildungspolitik eine internationale Dimension. Als bekennender Befürworter eines Europas, dessen Hauptaugenmerk der sozialen Gerechtigkeit und Chancengleichheit gelten muss, empfinde ich die Bildungspolitik als den Schlüssel zu einer ausgewogenen Gesellschaft. Gerade Österreich, das sich in den letzten Jahrzehnten auf seinen bildungspolitischen Lorbeeren ausgeruht hat, tat PISA und die damit verbunden öffentliche Diskussion meiner Meinung nach sehr gut. Die Frage, ob Österreichs Fünfzehnjährige tatsächlich so schlecht sind wie die Tests ergaben, ist da zweitrangig. Wichtig ist die Chance zu nützen und das Österreichische Bildungssystem endlich an die Bedürfnisse junger Menschen im neuen Jahrtausend anzupassen. Die Betrachtung der Bildungssysteme im In- wie im Ausland unter Berücksichtigung ihrer Besonderheiten schafft die Basis für eine Verbesserung des Status Quo. Die PISA-Studie mit ihren umfassenden Ergebnissen im Zusammenhang mit sozio-ökonomischen Faktoren kann zu dieser einiges beitragen. Dazu ist es wichtig tiefere Einblicke in PISA zu erlangen.

## 1 Einführung

In unserer heutigen Welt, in der die Globalisierung immer weiter fortschreitet und räumliche Distanzen an Relevanz verlieren, stehen die Staaten verstärkt in Konkurrenz zueinander. Internationale Vergleiche gewinnen enorm an Wichtigkeit, ob im wissenschaftlichen Diskurs, in kurzen Ranglisten der Boulevardpresse oder für Unternehmen, die über den besten Standort für eine Betriebsansiedlung entscheiden.

Diese Entwicklung macht auch vor der Bildung nicht halt. In der Informationsgesellschaft hat der Kampf um die "Human Resources", die klügsten Köpfe, längst begonnen - sie sind die Schlüsselressourcen des 21. Jahrhundert. Nicht nur um mit der technischen Entwicklung Schritt halten zu können, sondern auch um diese Entwicklung weiter vorantreiben zu können. Grund genug für die Organisation für Wirtschaft und Zusammenarbeit (OECD) Bildung nach ihrer eigenen Definition zu testen.

PISA, "**P**rogram for **I**nternational **S**tudent **A**ssessment", ist ein von der OECD initiiertes Projekt zur internationalen Schülerbeurteilung. Oft auch als PISA-Studie bezeichnet, hat PISA zum Ziel, alltagsrelevantes Wissen und Fähigkeiten von 15-jährigen Schülern zu testen. Seit dem Jahr 2000 wird die Studie im 3-Jahresrhythmus in den meisten OECD-Mitgliedsländern und einer steigenden Zahl von Partnerstaaten durchgeführt. Die Pisa Studie untersucht nicht ein einzelnes Schulfach, sondern die Kompetenzbereiche Lesen, Naturwissenschaften und Mathematik, 2003 auch Problemlösen, wobei jedes Jahr ein anderer Bereich als Schwerpunkt gesetzt wird.

Im Gegensatz zu vorhergegangenen Schulleistungsuntersuchungen hat PISA eine enorme mediale Aufmerksamkeit auf sich ziehen können. Nun ist Bildung nicht ein Produkt, das wie Äpfel greifbar und leicht quantifizierbar ist. Die OECD verwendet komplexe Methoden, um die Schulbildung aller 15 jährigen in den Kompetenzbereichen Lesen, Mathematik, Naturwissenschaften und Problemlösen erfassbar zu machen. Die methodische Vorgehensweise ist derart komplex, dass kaum jemand weiß, wie diese Daten eigentlich zustande kommen, und somit was sie wirklich Aussagen können. Trotzdem finden in vielen

Ländern rege Diskussionen darüber statt, wie man das aktuelle Schulsystem verändern soll, um auf ein unbefriedigendes Abschneiden zu reagieren.

Daher soll PISA 2003 in eine standardisierte Dokumentationsform gebracht werden, die es interessierten Laien, Journalisten und Politikern ermöglicht die wesentlichen Schritte und Transformationen nachvollziehbar zu machen. Erst wenn man verstanden hat was diese Studie tatsächlich aussagt, kann man sich an die Veränderung der bestehenden Systeme machen, sofern man das dann noch für erstrebenswert hält.

## 1.1 Motivation

Im Rahmen meines Masterstudiums Informatikmanagement wurde ich mit den zwei Wissensgebieten Didaktik und Pädagogik näher vertraut, die einem gelernten Wirtschaftsinformatiker sonst eher fremd sind. PISA hat viel Aufmerksamkeit auf das Österreichische Bildungswesen gelenkt. Da in den Medien nur unzureichend über den Test, viel mehr aber über das schlechte Abschneiden berichtet wurde, war es mir ein Anliegen zu verstehen was und wie PISA testet. Schon bald wurde mir bewusst, dass PISA ein sehr komplexes Konstrukt aus statistischen Methoden ist, das für den Normalbürger eigentlich nicht zu durchblicken ist. In Zusammenarbeit mit meinem Betreuer Professor Dr. Grossmann entstand dann die Idee, PISA auf die wesentlichen Schritte zu reduzieren und so verständlicher und übersichtlicher zu machen. Da mit dem Template "Standarddokumentation Metadaten" von der Statistik Austria ein geeignetes Rahmenwerk vorliegt, wurde dieses kurzerhand auch für PISA verwendet.

## 1.2 Struktur der Arbeit

Das erste Kapitel, die Einleitung soll den Leser in das Thema einführen, die Motivation des Schreibers darlegen und außerdem der Definition relevanter Begrifflichkeiten dienen. Im zweiten Teil wird die Konzeption der Arbeit erläutert, also das zu Grunde liegende wissenschaftliche Konzept. Der eigentliche Hauptteil findet sich in Kapitel 3, in welchem die "Standard-Dokumentation Metadaten für PISA 2003" umgesetzt ist, welche als eigenständiges Dokument, ohne den Rest der Arbeit verwendbar sein soll. In Kapitel 4 folgt

ein Vergleich anhand zweier Länder. Darin werden die spezifischen Daten Österreichs und Finnlands, die die nationalen Umsetzungen der internationalen Vorgaben zeigen, in das Schema der Standarddokumentation eingesetzt. Dadurch sollen die in Kapitel 3 dokumentierten Punkte durch praktische Bezüge ergänzt werden. Die Erfassung zweier europäischer Länder soll die Daten in Relation stellen. Das fünfte Kapitel dient der Reflektion und der Zusammenfassung, in der auch offene Fragen für die Zukunft ihren Raum finden. Im sechsten Kapitel finden sich die Literaturquellen.

## 1.3 Definitionen

In diesem Abschnitt sollen dem Leser Annahmen und zentrale Begriffe erläutert werden, um ein besseres Verständnis zu erreichen. Die zwei Schlagworte Metadaten und Standarddokumentation, die dem Template der Statistik Austria entstammen, sollen hier zur Einführung erklärt werden.

Die in Arbeit verwendeten Personenbezeichnungen sind als geschlechtsneutral zu verstehen. Um die Lesbarkeit zu vereinfachen werden spezifische technische Begriffe direkt im Text durch Fußnoten erläutert.

### 1.3.1 Standard-Dokumentation

Unter einer Dokumentation wird landläufig die „Nutzbarmachung von Informationen zur weiteren Verwendung“ [[de.wikipedia.org/Dokumentation](https://de.wikipedia.org/wiki/Dokumentation)] verstanden, deren Ziel es ist, die dokumentierten Objekte ausfindig zu machen. Eine Standard-Dokumentation ist daher eine vereinheitlichte Dokumentation.

Im Fall dieser Arbeit wird eine Vorlage der Statistik Austria für die Standard-Dokumentation von Metadaten herangezogen.

### 1.3.2 Metainformationen

Metainformationen oder zumeist Metadaten sind in ihrer einfachsten Definition "Daten die Informationen über andere Daten enthalten." [de.wikipedia.org/Metadaten] Metadaten erleichtern das Verständnis, Benützbarkeit und das Management von Daten. Typischerweise sind die Metadaten eines Buches: Autor, Erscheinungsjahr, Verlag, ISBN, usw. Diese Kategorisierung ist jedoch nicht allgemein gültig. Für einen Leser wäre die zuvor getroffene Einteilung richtig, da der Inhalt des Buches für ihn die Daten sind, während Autor, Erscheinungsjahr, usw. Metadaten darstellen. Für den Herausgeber eines Bücherkatalogs wären diese jedoch die Daten.

Um zwischen Daten und Metadaten unterscheiden zu können, bedarf es einer Betrachtung des Zusammenhangs. Es gilt: Metadaten befinden sich immer auf einem höheren Abstraktionslevel als die zu beschreibenden Daten.

Im Kontext dieser Arbeit sind Metainformationen also die beschreibenden Kategorien, die aus der Statistik Austria Vorlage "Standard-Dokumentation Metadaten" übernommen werden.

## 1.4 Ziele

Ziel dieser Arbeit ist es, eine den gängigen Normen der Statistik Austria entsprechende Standard-Dokumentation für PISA 2003 zu generieren. Diese soll als eigenständiges Dokument verwendbar sein.

Wichtig bei der Erstellung von Dokumentationen, ist es neben dem Erhalt der Vollständigkeit, diese auch möglichst kurz und prägnant zu halten. Dieser Aspekt war im Bezug auf PISA besonders schwierig war, da die Studie sehr exzessiv dokumentiert ist. Besonders auf Qualitätssicherung und Lenkung wurde sehr viel Augenmerk gelegt. Hier wurden nur die wichtigsten Punkte in diese Arbeit aufgenommen.

## 2 Grundlagen

Grundlage dieser Arbeit bildet die PISA-Studie aus dem Jahre 2003, im Speziellen die sehr ausführliche technische Dokumentation "PISA 2003 Technical Reports" [OECD 2005a]. Zusätzlich wurden für einzelne Details auch die "PISA 2003 Data Analysis Manuals" [OECD 2005b,c], das "The PISA 2003 Assessment Framework - Mathematics, Reading, Science and Problem Solving Knowledge and Skills" [OECD 2003] und der Ergebnisbericht " Learning for Tomorrow's World - First Results from PISA 2003" [OECD 2004] verwendet.

Das zweite Standbein dieser Arbeit ist die von der Statistik Austria erstellte Grundstruktur für *Standard-Dokumentationen von Metainformationen*. Als besonders positives Beispiel aus der Reihe der von Statistik Austria erfolgten Veröffentlichungen, sei hier die "Standard-Dokumentation Metainformation zu EU-SILC 2003" [Statistik Austria 2005] angeführt, die Struktur liegt jedoch allen von ihr erstellten Standard-Dokumentationen von Metainformationen zu Grunde.

Als Leitfaden vor allem in der Dokumentation der Qualitätsmerkmale wurde das von Eurostat herausgegebene Handbuch „How to make a Quality Report“ [Eurostat 2003] verwendet.

Für den in Kapitel 4 abgehandelten Ländervergleich wurde der eigens erstellte nationale technischen Bericht [Haider, et al., 2004] des österreichischen PISA Zentrums herangezogen. Im Fall von Finnland wurden offene Fragen direkt mit dem Leiter des nationalen Projektzentrums in Finnland, Jouni Välijärvi und einer seiner Mitarbeiterinnen geklärt. Aber auch hier wurden die meisten Informationen aus dem internationalen technischen Bericht der OECD entnommen.

## 2.1 Technischer Bericht der OECD zu PISA

Im technischen Bericht findet sich die ausführliche Beschreibung der komplexen Methoden und Details zur Implementation des Projekts. Alle Informationen zum Test und zum Design der Stichprobe, Datenanalysemethoden und Qualitätskontrollmechanismen sind in dem Bericht festgeschrieben. Der technische Bericht wird in englischer Sprache verfasst und alle 3 Jahre zur aktuellen Erhebung neu erstellt. Ziel des Papiers ist es Review und Replikation möglich zu machen.

### 2.1.1 Technical Report PISA 2000

Da PISA 2003 eine Weiterführung von PISA 2000 ist, enthält der "PISA 2000 Technical Report" (OECD 2002) einige aufschlussreiche Informationen zur umfassenden Kenntnis aller Details der Erhebung. Einige Informationen zur PISA Studie sind nur noch in diesem zu finden. Im technischen Bericht 2003 wird dann meist explizit auf den Bericht aus dem Jahr 2000 verwiesen. Besonders hilfreich ist in diesem Zusammenhang die übereinstimmende Reihenfolge und Struktur der Kapitel und Anhänge.

### 2.1.2 Technical Report PISA 2003

Der "PISA 2003 Technical Report" (OECD, 2005a) umfasst 18 Kapitel und 17 Anhänge auf 426 Seiten. Die 18 Kapitel dokumentieren ausführlich alle wichtigen Aspekte und Phasen von PISA.

Die Kapitel 1 bis 18 behandeln der Reihe nach: Überblick über PISA (1), das Test Design und die Testentwicklung (2), die Entwicklung der Kontextfragen (3), das Sampledesign (4), die Übersetzung und kulturelle Angemessenheit des Test- und Umfragematerials (5), die Feld Operationen (6), die Qualitätskontrolle von PISA (7), die Gewichtung und Berechnung der Varianz (8), die Skalierung der kognitiven Daten (9), die Coding-Reliability-Study (10), die Prozeduren zur Datenbereinigung (11), die Erhebungs- (12) und Skalierungsergebnisse (13), Ergebnisse der Coder-Reliability-Study (14), die Daten-Entsprechung (15), Konstruktion der Kenntnis-Skalen (16), Skalierungsprozeduren und Konstruktion der Bewertung der Kontextdaten (17) und die internationale Datenbank (18).

In den 17 Anhängen finden sich Erhebungsformulare (1), PISA Konsortium und Consultants (2), Länderdurchschnitt und Rang nach Testheft (3), Item-Übermittlungs-Richtlinien für Mathematik (4), Item-Review-Richtlinien (5), und ISCED Adaptionen für Partnerstaaten (6), ein fiktives Beispiel der Studien-Programm-Tabelle (7), ein fiktives Beispiel der Kontextfragebogen Adaptions- Tabelle (8), eine Zusammenfassung der Ergebnisse der Qualitätssicherung (9), die Kontrast-Kodierung für die Konditionierungs-Variablen (10), Skalen-Verlässlichkeit (11), Details der bei PISA 2003 verwendeten Mathematik-Items (12), Details der bei PISA 2003 verwendeten Lese-Items (13), Details der bei PISA 2003 verwendeten Naturwissenschafts-Items (14), Details der bei PISA 2003 verwendeten Problemlöse-Items (15), das Level der Elternbildung in Schuljahre konvertiert (16) und das Formular für die Studenten-Listen (17).

## **2.2 Technische Berichte der Nationalen Projektzentren**

### **2.2.1 Technischer Bericht des österreichischen Projektzentrums**

Der nationale technische Bericht des PISA Zentrums in Salzburg erlaubt den Einblick in die Arbeit des österreichischen Projektteams und ist in deutscher Sprache verfasst. In vielen Punkten wird auf den ausführlicheren internationalen technischen Bericht verwiesen. Lediglich nationale Spezifika werden detailliert behandelt.

### **2.2.2 Technischer Bericht des finnischen Projektzentrums**

In Finnland wurde kein nationaler technischer Bericht erstellt. Einige vergleichbare Informationen fanden sich im nationalen finnischen Ergebnisbericht. Was anhand dieser und der im internationalen technischen Bericht verfügbaren Daten nicht beantwortbar war, wurde telefonisch mit dem nationalen Projektleiter und einer Mitarbeiterin abgeklärt.

## 2.3 Statistik Austria Standard-Dokumentation Metadaten

Das Template Standard-Dokumentation Metadaten der Statistik Austria bildet die Grundstruktur dieser Arbeit und liefert somit eine wissenschaftlich fundierte und international anerkannte Basis für die Aufbereitung der PISA Daten und ist in 6 Kapitel gegliedert.

1. Wichtige Hinweise

2. Allgemeine Informationen

- Statistiktyp

- Fachgebiet

- Verantwortliche Organisationseinheit, Ansprechpartner

- Ziel und Zweck, Geschichte

- Periodizität

- Auftraggeber

- Nutzer

- Rechtsgrundlage(n)

3. Statistische Konzepte, Methodik

- Gegenstand der Statistik

- Beobachtungs-/Erhebungs-/Darstellungseinheiten

- Datenquellen

- Meldeeinheit/Respondenten

- Erhebungsform

- Charakteristika der Stichprobe

- Erhebungstechnik/Datenübermittlung

- Erhebungsbogen (inkl. Erläuterungen)

- Teilnahme an der Erhebung

- Erhebungs- und Darstellungsmerkmale, Maßzahlen; inkl. Definition

- Verwendete Klassifikationen

- Regionale Gliederung der Ergebnisse

#### 4. Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde Maßnahmen

Datenerfassung

Signierung (Codierung)

Plausibilitätsprüfung, Prüfung der verwendeten Datenquellen

Imputation (bei Antwortausfällen bzw. unvollständigen Datenbeständen)

Hochrechnung

Erstellung des Datenkörpers, verwendete Rechenmodelle, statistische Schätzmethoden

Sonstige qualitätssichernde Maßnahmen

#### 5. Publikation (Zugänglichkeit)

Revisionen

Vorläufige Ergebnisse

Endgültige Ergebnisse

Publiziert in:

Behandlung vertraulicher Daten

#### 6. Qualität

##### 6.1. Relevanz

##### 6.2. Genauigkeit

###### 6.2.1. Stichprobenbedingte Effekte, Repräsentativität

###### 6.2.2. Nicht-stichprobenbedingte Effekte

Qualität der verwendeten Datenquellen

Abdeckung (Fehlklassifikationen, Unter-/Übererfassung)

Antwortausfall (Unit-Non-Response, Item-Non-Response)

Messfehler (Erfassungsfehler)

Aufarbeitungsfehler

Modellbedingte Effekte

##### 6.3. Rechtzeitigkeit und Aktualität

##### 6.4. Vergleichbarkeit

##### 6.5. Kohärenz

## 2.4 Eurostat Handbuch "How to make a Quality Report"

Das Handbuch der Eurostat "How to make a Quality Report" dient als Methodensammlung für das Reporting der Qualität von statistischen Daten und Produkten. Neben den im Dokument enthaltenen Richtlinien für Qualitätsberichte, sind auch alle wesentlichen Definitionen, die Qualität von Statistiken betreffend, enthalten. Besonders relevant für diese Arbeit ist das Kapitel 2 "Quality dimensions and reporting about them".

Gerade im Bezug auf Standarddokumentationen ist es wichtig sich vorhandener Standards und Richtlinien zu bedienen und diese auch zu erfüllen. Da die Qualität jedoch nur eines von 6 Merkmalen des Statistik-Austria Templates ausmacht, wurden nicht immer alle Punkte (vollständig) behandelt, um die Länge der Dokumentation im Bereich von 30 Seiten zu halten.

### **3 Umsetzung Dokumentation**

Die Umsetzung der Dokumentation ist der Hauptteil dieser Arbeit und soll für sich alleine stehen können. Daher wird in diesem Kapitel die Nummerierung der Abschnitte neu begonnen. Stellt man den Text ab "1. Wichtige Hinweise" bis zum Beginn des übergeordneten 4. Kapitel "Vergleich zweier Länder" frei, so erhält man das eigenständige Dokument "Standard-Dokumentation Metadaten für PISA 2003". Durch den nachfolgenden Seitenumbruch soll das noch einmal verdeutlicht werden.

## 1. Wichtige Hinweise

2003 wurde PISA (**P**rogram for **I**nternational **S**tudent **A**ssessment) bereits zum zweiten Mal durchgeführt. Von 276.165 15-jährigen Schülern in 41 Ländern wurden die Kenntnisse in Lesen, Mathematik, Naturwissenschaften und einmalig auch Problemlösen getestet. Zusätzlich waren von den Schülern Kontextfragebögen auszufüllen. Relevante Informationen über die Schule wurden von den Schulleitern abgefragt. Die Feldarbeit war auf den Zeitraum zwischen 1. März und 31. August 2003 festgesetzt, wobei jedes Land in diesem Testfenster eine 42-Tage-Periode für die nationale Durchführung wählen konnte.

PISA ist eine Trendstudie, die im Zyklus von 3 Jahren durchgeführt wird und deren Ziel es ist, eine Datenbasis zum Wissens- und Kompetenzstand der 15-Jährigen aufzubauen.

Da die Erhebung des Bildungsstands alleine wenig zielführend ist, wurden in besagten Kontextfragebögen sozioökonomische Merkmale erhoben, die die Analyse der Ergebnisse ermöglicht.

## 2. Allgemeine Informationen

### **Statistiktyp**

Primärstatistik

### **Fachgebiet**

Bildung

### **Verantwortliche Organisationseinheit, Ansprechpartner/in**

OECD (**O**rganisation for **E**conomic **C**ooperation and **D**evelopment)

Projektleitung PISA:

OECD Sekretariat:

Andreas Schleicher ( übergreifende Managementverantwortung)

Schnittstelle:

Board of Participating Countries (ein Vertreter jedes teilnehmenden Landes)

Nationale Projektteams:

Leitung durch den Nationaler Projekt Manager (NPM)

Design und Implementierung:

internationales Konsortium:

Leitung: Australian Council for Education Research (ACER)

### **Ziel und Zweck, Geschichte**

PISA steht für "Program for International Student Assessment" - Programm für internationale Studenten Bewertung. Dessen Ziel ist es, die Qualität der nationalen Bildungssysteme zu erfassen und Informationen über qualitätsrelevante Merkmale bereitzustellen. Bei PISA geschieht das durch regelmäßige, international standardisierte Messung von Schülerleistungen und durch die Erhebung von Merkmalen, die als leistungsbeeinflussend gelten. Die OECD möchte mit PISA eine systematische Datenbasis aufbauen und Basis-, Kontext- und Trendindikatoren gewinnen.

Langfristig wird so ein Bestand an Informationen aufgebaut, der eine Beobachtung von Entwicklungstrends im Wissens- und Kompetenzbestand von Schülern aus verschiedenen Ländern und demografischen Untergruppen erlaubt. Durch den Vergleich der Ergebnisse über die Zeit sollen Konsequenzen zwischenzeitlicher Änderungen im System sichtbar gemacht werden können. Wichtig ist hierbei die Untersuchung von Veränderungen in den Zusammenhängen zwischen Leistungs- und Hintergrundvariablen.

PISA wurde im Jahr 1997 offiziell gestartet, die erste Erhebung fand im Jahr 2000 statt. Anfänglich nahmen 43 Länder teil, 2003 dann 41 und 2006 waren es 58.

### **Periodizität**

Zyklische Trendstudie seit dem Jahr 2000 alle 3 Jahre. (2000, 2003, 2006, ...)

**Auftraggeber**

OECD bzw. nationale Bildungs-Ministerien.

**Nutzer**

- OECD
- Öffentliche Verwaltung
- Interessensvertretungen
- Forschungseinrichtungen
- Bildungseinrichtungen
- Öffentlichkeit

**Rechtsgrundlage(n)**

Bilaterale Verträge mit der OECD.

### 3. Statistische Konzepte, Methodik

**Gegenstand der Statistik**

Gegenstand der Statistik sind die Fähigkeiten in den vier (nur 2003) von Pisa als Basiskompetenzen titulierten Domänen:

- Mathematik-Kompetenz
  - Raum und Form (space & shape)
  - Größen (quantity)
  - Veränderung und Zusammenhänge (change & relationships)
  - Ungewissheit (uncertainty)
- Lese-Kompetenz
- Naturwissenschafts-Kompetenz
- Problemlöse-Kompetenz (nur 2003)

Im Zyklus wird eine dieser Kompetenzen in zwei Drittel der Test-Zeit (4 Stunden) als sogenannte Hauptdomäne (major domain) mit 4 Subskalen verstärkt erhoben. (2000 Lesen, 2003 Mathematik, 2006 Naturwissenschaften, 2009 Lesen, usw.) So keine neue Basiskompetenz hinzukommt tritt also alle 9 Jahre dieselbe Hauptdomäne auf.

#### Test-Medium:

PISA ist ein Bleistift und Papier Test.

In Ländern in denen Taschenrechner im Unterricht standardmäßig eingesetzt werden, wurden solche zur Verfügung gestellt. Die Items wurden alle möglichst Taschenrechner-neutral entwickelt, um keine Vor- und Nachteile für einzelne Länder hervorzurufen.

#### Test-Aufgaben-Format (Items):

- Multiple Choice
  - Standard (Wahl der besten Antwort aus meist 4 gegebenen)
  - Komplex (Serie von richtig/falsch Möglichkeiten)
- Kurzantwort (Zahl, Wort, kurze Phrase)
- ausgedehnte (selbst formulierte) offene Beantwortung.

#### Items & Testhefte

Mittels einer Feldstudie, ein Jahr vor PISA, wurden aus 300 Kompetenzfragen die besten 167 Items für den Haupttest ausgewählt. Diese wurden wiederum in 13 Cluster gruppiert. 7 für Mathematik (M1-7) als Hauptdomäne und je 2 für die Nebendomänen (minor domains) (R1, R2, S1, S2, P1, P2), wobei jeder Cluster 30 Minuten Testzeit repräsentiert. Diese 13 Cluster wurden auf 13 verschiedenen Testbögen verteilt, wobei jeder aus 4 Clustern besteht, also 120 Minuten ausmacht. Jeder Cluster kam in allen 4 Positionen des Testhefts einmal vor, insgesamt also in 4 Testheften.

**Anzahl der Items je Domäne:**

Lesen: 28 (60min) [2000 Hauptdomäne 270(!)min]

Mathematik: 85 (210min) [2003 Hauptdomäne 210min]

Naturwissenschaft: 35 (60min)

Problemlösen: 19 (60min)

Für Schüler mit speziellen Bedürfnissen gab es ein einstündiges Testheft: 7 Mathematik Items, 6 Lesen, 8 Science, 5 Problemlösen

Zusätzlich gab es einen Fragebogen zum *selbstgesteuerten Lernen*. Außerdem werden seit 2003 an Schulen und Schüler Fragebögen ausgegeben, die zur Erstellung von Indikatoren dienen, die soziale, ökonomische, kulturelle und erzieherische Aspekte in Zusammenhang mit der Schüler-Performance stellen.

Über die verpflichtende Erhebung hinaus, können Länder auch an diversen Zusatzoptionen teilnehmen, beziehungsweise national weitere Kontextvariablen erheben.

**Beobachtungs-/Erhebungs-/Darstellungseinheiten**

Beobachtungseinheit: Schüler

Erhebungseinheiten: Schule

Darstellungseinheiten: Länder

**Datenquellen**

- Gesamtevidenz der 15 jährigen Schüler des jeweiligen nationalen Bildungsministeriums.
- Primärstatistik

## **Meldeeinheit/Respondenten**

Respondenten: Schüler

Interviewer: Testadministratoren

Meldeeinheit: Schule

## **Erhebungsform**

Altersbasierte mehrstufige Stichprobenerhebung.

## **Charakteristika der Stichprobe**

### Grundgesamtheit und Auswahlrahmen :

Grundgesamtheit sind die 15-Jährigen, die eine Bildungseinrichtung im jeweiligen Land besuchen und sich in der 7. Schulstufe oder höher befinden. Sowohl Schüler auf Voll- und Teilzeitbasis, Schüler in Berufsausbildung, wie auch ausländische Schüler und Schüler aus ausländischen Schulen, die sich im jeweiligen Land befinden, sind in der Erhebung inkludiert. Kein Teil der Zielpopulation sind Personen, die zuhause, in der Arbeit oder im Ausland ausgebildet werden.

Die operationale Definition der Alterspopulation hängt direkt von den Testterminen ab, deren 42-Tagesperiode zwischen dem 1. März und 31. August 2003 liegen musste. Zusätzlich gab es die Einschränkung, dass es sich bei der gewählten Testperiode nicht um die ersten 3 Monate des Schuljahres handeln durfte. Das zugelassene Alter der Schüler der internationalen Zielpopulation ist von 15 Jahren und 3 (vollendeten) Monaten, bis 16 Jahren und 2 (vollendeten) Monaten, am Beginn der jeweiligen Testperiode definiert. Darüberhinaus wurden Abweichungen von bis zu einem Monat dieser Altersdefinition toleriert.

## Samplingverfahren

In fast allen Ländern altersbezogenes 2 stufig stratifiziertes Sample-Design.

1. Stufe: Aus allen Schulen, die von Schülern der definierten Altersgruppe besucht werden, wurde eine Auswahl (von Schulen) anhand der, zur Anzahl dieser Schüler proportionalen Wahrscheinlichkeiten gezogen. Es mussten mindesten 150 Schulen ausgewählt werden.
  - a. Stratifizierung: Die Stratifizierung erfolgt landespezifisch durch eine Teilung in Gruppen. Die Gruppeneinteilung wurde so gewählt , dass die Sampling-Einheiten innerhalb der Gruppen bezüglich der zu messenden Variablen möglichst homogen waren. Dadurch kann diese Methode die Effizienz des Samples deutlich steigern.
  - b. Je 2 Replacement-Schulen je gezogener Schule, für den Fall dass diese dann nicht teilnehmen konnte.
  - c. Die Bereitstellung der für den Samplingvorgang notwendigen Daten über die Population, liegt in der Verantwortung der nationalen Projektzentren. Für den Haupttest 2003 übernahm Westat (als Teil des internationalen Konsortiums) die Ziehung der Schulstichprobe.
2. Stufe: Listenerstellung von Schülern des Zieljahrgangs der gesampelten Schulen. Auswahl von maximal 35 Schülern pro Schule durch systematische Stichprobenziehung (pseudo Zufallsstichprobe), welche eine gleichmäßige Streuung über die Schule garantierte.

In einigen Ländern wurde die Anzahl an Schülern einer gesampelten Schule mit mindestens 20 festgelegt.

In 2 Ländern wurde ein 3 Stufen-Design angewandt. Nach der Wahrscheinlichkeit proportional zur Sampling-Größe wurden:

1. Geografische Regionen gezogen (first-stage units),
2. Schulen in den gewählten Regionen gezogen (second-stage units),
3. Schüler aus den gezogenen Schulen (third-stage units).

Mindestanzahl von untersuchten Schulen: 150

Mindestanzahl von auswertbaren Datensätzen: 4500 Schüler

Bei kleineren Ländern: Vollerhebung

Wurde die within-School sample size ("target cluster size") von einem Land unter (äußerstes Minimum 20) den Wert von 35 Schülern gesenkt so musste die Stichprobengröße der Schulen auf über 150 erhöht werden, um zumindest 4500 Schüler zu testen.

#### Auswahlwahrscheinlichkeit:

Die Anzahl der pro Stratum gezogenen Schülern war direkt proportional zu deren Anzahl in der Grundgesamtheit.

#### Ausschlussmöglichkeiten: (auf 2 Ebenen)

1. Schulebene "school level" (Ausschluss einer ganzen Schule)
  - kleine, entlegene geographische Region wegen Größe bzw. Unerreichbarkeit;
  - Sprachgruppen aus politischen, organisatorischen oder operationalen Gründen
2. Schülerebenen "within-school level"
  - Geistig behinderte Schüler  
(schlechte schulische Leistung oder normale disziplinäre Probleme waren kein Ausschlussgrund!)
  - Funktional behinderte Schüler,  
deren permanente Behinderung es nicht zuließ am Test teilzunehmen.
  - Schüler mit beschränkten Fähigkeiten in der Testsprache  
Normalerweise nur, wenn sie erst seit weniger als einem Jahr in der Sprache unterrichtet wurden (länderspezifisch adaptierbar)
  - Sonstige Gründe – müssen vom Konsortium bestätigt werden

Schulen, die nur von auszuschließenden Schülern besucht wurden, wurden den Ausschlüssen auf Schulebene zugeordnet.

Einschränkungen:

- Insgesamt maximal 5% Ausschlüsse (Schul und Schülerebene) → "defined target population" = 95% der "national desired target population"
- Schullevel-Ausschlüsse auf Grund von Unerreichbarkeit, Machbarkeit, oder anderen Gründen dürfen max. 0,5% der Gesamtanzahl der Schüler der internationalen PISA Zielpopulation ausmachen.
- Schullevel-Ausschlüsse für Schüler mit intellektuellem, funktionellem oder sprachlichem Handikap durften nicht mehr als 2% der Schüler ausmachen.
- Schüler-Level Ausschlüsse für Schüler mit intellektuellem, funktionellem oder sprachlichem Handikap durften nicht mehr als 2,5% der Schüler betragen. Wenn diese Zahl doch höher war, wurden die Schüler mit sprachlichem Handikap aus dieser Quote herausgenommen.

**Erhebungstechnik/Datenübermittlung**

In jeder teilnehmenden Schule wurde die Testdurchführung an einen eigenen Schulkoordinator delegiert, der für diesen Aufwand auch finanziell entschädigt wurde. Der Test fand für alle zu testenden Schüler einer Schule in einem gemeinsamen Test-Raum (meist in der Turnhalle) statt. Die Durchführung erfolgte durch einen speziell geschulten, schulexternen Testadministrator (TA). Waren am Test-Tag mindestens 15% der ausgewählten Schüler nicht anwesend, gab es einen Nachtest.

Die Bearbeitung des PISA Kompetenztesthefts erfolgte in zwei je 60 Minuten dauernden Blöcken. Dazwischen gab es 5 Minuten Pause. Anders als bei PISA 2000 erhielten die Teilnehmer alle Fragen auf einmal und konnten diese in beliebiger Reihenfolge beantworten. Nach einer weiteren Pause von 10 Minuten war der Kontextfragebogen in weiteren 60 Minuten zu beantworten.

Es erfolgte außerdem eine ausführliche Dokumentation durch den Testadministrator.

### Datenübermittlung:

Die Datenübermittlung erfolgt zuerst persönlich, die Testadministratoren brachten nach der Testdurchführung die Testmaterialien zur einer Sammelstelle in der jeweiligen Region, wo sie nach Ende aller Tests von einem Mitglied des nationalen Projektteams abgeholt werden.

Die Übermittlung an das internationale Zentrum musste innerhalb von 12 Wochen nach Ende des nationalen Testzeitraums erfolgen. Sämtliche Daten wurden bereits gründlich kontrolliert und gesäubert, mittels der speziell von ACER entwickelten KeyQuest-Software und einem FTP-Zugang elektronisch übermittelt. Wurde zur Dateneingabe eine andere Software verwendet, musste diese zur Übermittlung in KeyQuest importiert werden.

Bei der so genannten "Data Submission" an das internationale Zentrum musste Folgendes übermittelt werden:

- Eine Kopie der KeyQuest-Dateien „KQdata.mdb“ und „KQadmin.mdb“ mit sämtlichen Daten (Daten für Testhefte, Schülerfragebogen, Schulfragebogen, Multiple Marking, PISA-Schülerlisten, Liste der Schulen, Liste der Schüler/innen, Testsitzungsprotokolle). Da diese Dateien sehr groß (ca. 200 MB) waren, wurde vom internationalen Zentrum ein ftp-Zugang für die Übermittlung zur Verfügung gestellt.
- Eine elektronische Kopie (in PDF-Format) der international vorgeschriebenen und in Abschnitt 5 beschriebenen Kontrollen bzw. der dabei erzeugten Berichte.
- Das komplette Set der nationalen Erhebungsinstrumente (Testhefte und Fragebögen) in PDF-Format.
- Zwei Dokumente mit Informationen zu den übermittelten Daten („Data submission checklist“ und „Essential information required for cleaning data“).

### **Erhebungsbogen (inkl. Erläuterungen)**

#### Test-Medium:

PISA ist ein Bleistift und Papier Test.

In Ländern, in denen Taschenrechner im Unterricht standardmäßig eingesetzt werden, wurden solche zur Verfügung gestellt. Die Items wurden alle

möglichst Taschenrechner-neutral entwickelt, um keine Vor- und Nachteile für einzelne Länder hervorzurufen.

### Test-Aufgaben-Format (Items):

- Multiple Choice
  - Standard (Wahl der besten Antwort aus meist 4 gegebenen)
  - Komplex (Serie von richtig/falsch Möglichkeiten)
- Kurzantwort (Zahl, Wort, kurze Phrase)
- ausgedehnte (selbst formulierte) offene Beantwortung.

### Testhefte:

- 2-Stunden-Testheft (normal)

Nach einer Feldstudie mit 300 Items, ein Jahr vor PISA, wurden 167 davon für PISA ausgewählt. Diese wurden wiederum in 13 Cluster gruppiert. 7 für Mathematik (M1-7) und je 2 für die Nebendomänen (minor domains) (R1,R2, S1,S2, P1,P2), wobei jeder Cluster 30 Minuten Testzeit repräsentiert. Die 13 Cluster wurden in 13 verschiedenen Testbögen präsentiert, wobei jeder aus 4 Clustern besteht, also 120 Minuten ausmacht. Jeder Cluster kommt in allen 4 Positionen des Testhefts einmal vor, insgesamt also in 4 Testheften.
- 1-Stunden-Testheft (spezial)

Für Schüler mit speziellen Bedürfnissen gab es ein einstündiges Testheft: 7 Mathematik Items, 6 Lesen, 8 Science, 5 Problemlösen.

### Beispielfragebögen

Die PISA Fragebögen sind copyrighttechnisch geschützt, um über Jahre hinweg möglichst die gleichen Testfragen stellen zu können und sind daher nicht öffentlich verfügbar.

Die freigegebenen Beispiel-Items finden sich in "The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills" [OECD 2003].

## Teilnahme an der Erhebung

Freiwillig, keinerlei Motivation für die Beteiligten gut abzuschneiden. Bei Abwesenheit zum Testtermin, wird zu einem weiteren Termin ein Ersatztest durchgeführt.

## Erhebungs- und Darstellungsmerkmale, Maßzahlen; inkl. Definition

### 1. Erhebungsmerkmale auf Schulebene: (durch Schulleiter auszufüllen)

- Lage der Schule und Schulgröße: Größe der Stadt/Gemeinde in der sich die Schule befindet, Aufnahmekriterien
- Finanzierungsart: öffentliche/private
- Unterrichtskontext: Unterrichtszeiten für alle Schulprogramme einer Schule, Auswahlverfahren von Schülerinnen, Gruppierung der Schülerinnen nach Fähigkeiten
- Schulressourcen: Wahrnehmung der Schulleiterin bezüglich der Qualität der Schulressourcen (Unterrichtsressourcen, Physiklabore, Lehrermangel), Anzahl der Lehrerinnen, Anzahl der Computer
- Tests und Förderung von Lernen: Zweck von Schülertests, "Beobachtung" von Mathematik-Lehrerinnen, Häufigkeit von Tests, Förderung des Engagements in Bezug auf Mathematik, Existenz und Überprüfung von Bildungszielen
- Schulklima: Konsens der Mathematik-Lehrerinnen, Wahrnehmung des Lernklimas durch die Schulleiterin (lehrer- und schülerbezogene Faktoren des Schulklimas), Wertvorstellungen der Lehrerinnen
- Schulautonomie: Verantwortung für verschiedene Aspekte der Schule, Einfluss von verschiedenen Körperschaften bei Entscheidungen

### 2. Erhebungsmerkmale auf Schülerebene: (180min)

#### a. Kompetenzfragebogen (120min)

- Lese-Kompetenz

*„Lese-Kompetenz“ (reading literacy) heißt, geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potential weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (OECD, 1999; deutsche Übersetzung: Deutsches PISA-Konsortium, 2000).*

– Mathematik-Kompetenz

*„Mathematik-Kompetenz (,Mathematical Literacy‘) ist die Fähigkeit einer Person, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens einer Person als konstruktivem, engagiertem und reflektierendem Bürger entspricht.“ (OECD, 1999; deutsche Übersetzung: Deutsches PISA-Konsortium, 2000)*

Da Mathematik als Hauptdomäne 2003 verstärkt erhoben wurde, ist das Framework in einander überschneidende Leitideen organisiert. Haider S.32

- Raum und Form (space & shape)
- Größen (quantity)
- Veränderung und Zusammenhänge (change & relationships)
- Ungewissheit (uncertainty)

– Naturwissenschafts-Kompetenz

*„Die Fähigkeit, naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen, die die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen“ (OECD, 1999; deutsche Übersetzung: Deutsches PISA-Konsortium, 2000).*

– Problemlöse-Kompetenz

*„Problemlösen ist die Kapazität eines Individuums, kognitive Prozesse zu nutzen, um realen, überdisziplinären (cross-curricular) Situationen, in denen der Lösungsweg nicht unmittelbar sichtbar ist, und in denen die Kompetenzbereiche oder Lehrplanbereiche, die zutreffen könnten, nicht innerhalb einer einzelnen Domäne wie Mathematik, Naturwissenschaft oder Lesen liegen, gegenüberzutreten und sie zu lösen.“ (OECD, 2003).*

b. Schülerfragebogen (60min)

- Schülercharakteristika: Schulstufe, Alter, Geschlecht
- Familiärer Hintergrund: Familienstruktur, Beruf der Eltern, Schul- und Berufsausbildung der Eltern, Geburtsland, Sprache die zu zuhause gesprochen wird, Besitztümer zuhause (kulturelle Besitztümer, Computer, zuhause vorhandene Lernunterlagen, Anzahl der Bücher zuhause)

- Lernen und Unterricht in Mathematik: Unterrichtszeit, Klassengröße, Klassenklima (disziplinäres Klima und Unterstützung durch LehrerInnen), Verwendung von Unterrichtsmitteln
- Engagement der SchülerInnen in Mathematik: selbstbezogene Kognitionen (Angst vor Mathematik und Mathematik-Selbstkonzept, Anstrengung und Ausdauer in Mathematik), Selbstwirksamkeit in Mathematik, motivationale Präferenzen (Interesse und Freude, instrumentelle Motivation), Lernzeit (Zeit für Hausübungen, Zeit mit Nachhilfelehrer oder Förderunterricht, andere Mathematik-Aktivitäten);
- Schulklima: Einstellungen gegenüber der Schule, Schüler-Lehrer-Verhältnis;
- Selbstreguliertes Lernen in Mathematik

### 3. Zusätzlich internationale und nationale Erhebungsmöglichkeiten.

- 2003:
1. Bildungskarriere
  2. Informations- und Kommunikationstechnologie

#### Maßzahlen:

Kompetenzlevel der Schüler in den 4 Bereichen auf den standardisierten Mittelwert 500 mit der Standardabweichung 100.

#### Darstellungsmerkmale:

Mittlere Landes-Scores der einzelnen Bereiche.

#### Gliederungsmerkmale:

- Land
- Soziökonomische Merkmale
- Geschlecht

## Verwendete Klassifikationen

Für PISA wurden folgende internationale Klassifikationen herangezogen:

- International Standard Classification of Occupations (ISCO )  
Um Informationen zur Beschäftigung der Eltern und die Joberwartungen der Schüler selbst zu codieren.
- Socio-Economic Index of Occupational Status (ISEI)  
Um die ISCO Codes wiederum in ein international vergleichbares Maß des Sozio-Ökonomischen Status der Beschäftigung umzucodieren.
- International Standard Classification of Education (ISCED)  
Um eine Klassifizierung der Qualifikation der Ausbildung und des Studienprogrammes zu erreichen und international vergleichbar zu machen. (Details zur ISCED-Klassifizierung finden sich unter [http://www.unesco.org/education/information/nfsunesco/doc/isced\\_1997.htm](http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm). Adaptionen für jene Länder die nicht in ISCED-1997 berücksichtigt wurden, finden sich im technischen Bericht – Anhang 6 [OECD 2005a, S. 383])
- International Standard Organisation (ISO) 3166 Country Codes  
Zur Kodierung der Länder wurde die ISO 3166 Länder Codes herangezogen.

## Regionale Gliederung der Ergebnisse

Die regionale Gliederung der Ergebnisse erfolgt international rein auf nationalstaatlicher Ebene. Ausnahme sind hier lediglich 3 Länder, die sogenannte Substaaten betrachteten:

Italien: 01 Veneto-Nordost 02 Trento-Nordost, 03 Toscana-Zentrum, 04 Piemont-Nordwesten, 05 Lombardei-Nordosten, 06 Bolzano, 07 alle restlichen (nicht zuordenbare) Regionen

Spanien: 01 nicht zuordenbare Regionen, 02 Castilien und Leon, 03 Catalonien, 04 Baskenland

Vereinigtes Königreich (UK): 01 England, Nord Irland und Wales, 02 Schottland

National wird auch nach Stratum aufgeschlüsselt, welches bei nicht-differenzierten Schulsystemen meist regionaler Natur ist.

## 4. Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde Maßnahmen

### **Datenerfassung**

Papier und Bleistifttest

### **Signierung (Codierung)**

#### **I. Schülerebene:**

##### Kontextfragebogen:

- Die Codierung der Kontextfragebögen erfolgte mit Hilfe der unter Klassifikation beschriebenen Standards.
- Codierung der zuhause gesprochenen Sprachen.

##### Testhefte:

Den Testheften wurden PISA-intern die Nummern 1-13 (Testheft 9 als einziges mit allen Kompetenzen) zugewiesen, denen wiederum die Item-Cluster randomisiert zugewiesen sind.

- Mathematik (mathematics): M1-M7
- Lesen (reading): R1, R2
- Naturwissenschaften (science): S1, S2
- Problemlösen (problem solving): PS1, PS2

Weiters erfolgte eine Kodierung der Antworten mittels "double-digit coding", einem 2-stelligen Code bestehend aus Bewertung (score) und Antwortcode(Lösungsweg).

1. Stelle : Bewertung: volle Punkte, keine Punkte, Teilpunkte, ungültig, fehlend, nicht erreicht, nicht vorhanden
2. Stelle: Antwortcode: Zur Kategorisierung des gewählten Lösungsweges

## II. Schulebene:

Codierung Schüler durch den Schulkoordinator:

- SPF-Code (Spalte 8) = Sonderpädagogischer Förderbedarf, mangelnde Deutsch-Kenntnisse
- Teilnahme-Code (Spalte 9)

## III. International:

- Länder-Code: gemäß ISO 3166
- Schul-ID: randomisierte IDs zur Identifizierung der Schule
- Schüler-ID: randomisierte IDs zur Identifizierung des Schülers
- Länderuntercodes (Subnatio): nur in UK, Italien und Spanien – siehe Umsetzung Dokumentation - Kapitel 3: Regionale Gliederung der Ergebnisse
- Stratum-ID: Identifizierung der einzelnen Stratifizierungs-Gruppen

## Plausibilitätsprüfung, Prüfung der verwendeten Datenquellen

Bei demografischen- und Daten wo Abhängigkeiten gegeben sind, wurden standardisierte Plausibilitätchecks durchgeführt. Weiters wurden mittels "Out-of-Range-Checks" auffällig hohe oder niedrige Werte kontrolliert, und Variablen bei denen überdurchschnittlich viele Codes für fehlende oder ungültige Schülerantworten vorkamen.

## Imputation (bei Antwortausfällen bzw. unvollständigen Datenbeständen)

Keine Imputation bei Antwortausfällen.

Das Testdesign sieht eine Imputation jener Domänen, die nicht im vom jeweiligen Schüler bearbeiteten Testheft behandelt wurden, vor. Nur in einem (Testheft 9) der 13 Testhefte sind alle 4 Domänen enthalten, nur diese Daten müssen nicht imputiert werden. Ansonsten wird die Hauptdomäne (2003 Mathematik), die in jedem Testheft vorhanden ist gemeinsam mit sozioökonomischen Faktoren und Schichtzugehörigkeit als Ausgangsdaten für die Imputation verwendet.

## **Hochrechnung**

Die Hochrechnung der geschichteten Zufallsstichprobe (siehe: Charakteristika der Stichprobe - Samplingverfahren) erfolgt anhand der in der Gesamtevidenz aufscheinenden relevanten (15 jährige ab der 7. Schulstufe – siehe: Charakteristika der Stichprobe - Grundgesamtheit und Auswahlrahmen) Schüler. Je nach zugehöriger Schicht werden die erzielten Ergebnisse anhand ein Gewichtungsschemas hochgerechnet.

## **Erstellung des Datenkörpers, verwendete Rechenmodelle, statistische**

### **Schätzmethoden**

Das Modell zur Messung der Kompetenzen hinter PISA ist sehr komplex. Die folgende Aufteilung in 7 Modellkomponenten entstammt sinngemäß "PISA 2000 und PISA 2003 – Vertiefende Analysen und Methoden" [Neuwirth et al. 2006].

#### Modellkomponente MK1: Verteilung der Fähigkeiten in einem Land

Basis für PISA ist ein Regressionsmodell, das heißt, dass die Fähigkeiten der einzelnen Schüler in den Kompetenzbereichen von Erklärungsgrößen abhängig sind. Diese sogenannten Regressionskoeffizienten sind zum Beispiel Geschlecht, Niveau der Schule oder sozioökonomische Faktoren wie Beruf der Eltern. Das Schulniveau wird durch die Mathematik Durchschnittsleistung aller Schüler, die tatsächlich Mathematik Aufgaben gelöst haben, bestimmt. Für Schüler, die in all diesen Eigenschaften übereinstimmen, wird eine Normalverteilung dieser Fähigkeiten angenommen. (->Priorverteilung)

#### Modellkomponente MK2: Stichprobenmodell

Untersucht wird eine geschichtete Stichprobe (min. 4500 Schüler pro Land) der Zielpopulation, wobei das Schichtungskriterium von Land zu Land verschieden sein kann. Zuerst wird eine Auswahl an Schulen gezogen und die Anzahl der Schüler, die der Zielpopulation angehören, berücksichtigt. Dann werden in jeder ausgewählten Schule die zu testenden Schüler zufällig ausgewählt. Die Repräsentativität wird durch ein komplexes Gewichtungsschema gesichert.

### Modellkomponente MK3: Erhebungsinstrumente

Erhebung der Daten:

- Testhefte – Fähigkeiten der Schüler / Schulniveau
- Schülerfragebogen – Soziodemographische Erklärungsgrößen
- Schulleiterfragebogen – Hintergrundinformation

Es gibt unterschiedliche 13 Testhefte die in Summe alle Aufgaben 4-mal enthalten. Nur eines (Nr. 9) enthält Fragen zu alle 4 Kompetenzen, alle enthalten mindestens einen Mathematikblock (Hauptdomäne 2003). Ein 14. einstündiges Testheft, existiert für Schüler denen ein zweistündiger Test nicht zumutbar ist.

Die Testhefte werden randomisiert verteilt, um systematische Verzerrungen zu verhindern.

Die einheitliche Kodierung der Antworten erfolgt durch speziell geschulte Personen.

### Modellkomponente MK4: Raschmodell

Die Schüler-Lösungen werden auf eine Fähigkeitsskala, die die Schwierigkeiten der einzelnen Aufgaben berücksichtigt, transformiert. PISA macht das mittels Raschmodell, für das folgende Voraussetzungen gelten:

- Schwierigkeitsparameter unabhängig von Testperson und Land -> Nationale und internationale Kalibrierungen
- Die Wahrscheinlichkeit eine Aufgabe zu Lösen = Differenz der jeweiligen Individuellen Fähigkeiten und der Schwierigkeit der Aufgabe -> Bestimmung der Schwierigkeit Testdurchlauf in einer Teilstichprobe (500 Personen in jedem teilnehmenden OECD-Land)
- Vergleichbare Testhefte -> Booklet-Effekt, berücksichtigt den Ermüdungseffekt

Die Schätzung der Fähigkeiten erfolgt durch die Maximum-Likelihood-Methode.

Sollten einzelne Aufgaben in manchen Ländern eine abweichende Lösungswahrscheinlichkeit aufweisen, können unter bestimmten Bedingungen einzelne Items für einzelne Länder aus der Score-Berechnung ausgeschieden werden.

### Modellkomponente MK5: Bestimmung der Inputvariablen für die Regression und Schätzung der Regressionskoeffizienten

Das Leistungsniveau der Schüler in Mathematik wird als Indikatorvariable verwendet, welches nach dem Raschmodell (hier ohne sozioökonomische Hintergrundvariablen) mittels Maximum-Likelihood-Verfahren bestimmt wird. Daraus wird dann das gewichtete Mittel der Schüler einer Schule erzeugt, mittels dem grob das Niveau von Schulen verglichen werden kann.

Wenn die erklärenden Variablen in MK1 definiert sind, können für jedes Land die Koeffizienten des Regressionsmodells geschätzt werden. Für jedes Land entsteht so eine Spezifikation des Modells, dessen (multivariate) Verteilung der Kompetenzen auch Aussagen über die Zusammenhänge der Fähigkeitsparameter einzelner Testgebiete macht.

### Modellkomponente MK6: Bestimmung der Posteriorverteilungen der Testpersonen

Nach der Schätzung der Fähigkeitsparameter aus MK1 wird für jede Testperson eine Posteriorverteilung berechnet, welche aus den bekannten Testantworten und Hintergrundvariablen die Wahrscheinlichkeiten der Zahlenwerte für den Fähigkeitsparameter angibt. Das bedeutet, die Verteilung der Kompetenzen einer Testperson hängt stark von der ermittelten Priorverteilung (MK1), also den soziodemografischen Merkmalen (Geschlecht, Status der Eltern), Schulniveau, aber auch dem vom Kontext der Testfragen abhängig ist.

### Modellkomponente MK7: Bestimmung der mittleren Kompetenz eines Landes

Basierend auf diesen Posteriorverteilungen werden die nationalen Kompetenzlevels bestimmt. Für alle Schüler, werden basierend auf den Posteriorverteilungen, für jeden erfassten Kompetenzbereich, 5 plausible Werte ermittelt. Diese statistisch ermittelten Zufallszahlen drücken aus, welche Werte der Fähigkeitsparameter für eine Testperson bei bestimmter sozialer Hintergrundstruktur und bestimmten Testantworten wahrscheinlich sind und werden immer für alle Fähigkeitsdimensionen gemeinsam berechnet. Diese werden dann mittels gewichteter Mittelung zur Bestimmung der mittleren Kompetenzniveaus eines Landes herangezogen.

### **Sonstige qualitätssichernde Maßnahmen**

#### Itementwicklung/-auswahl:

Itementwicklung auf internationaler Ebene mit Hilfe eines genau vorgegebenen Entwicklungsprozesses und vom Konsortium vorgegebener Richtlinien.

Durchführung eines Feldtests in allen Teilnehmer Ländern zur Erprobung der Materialien und Prozeduren. Keine Ergebnisanalyse im eigentlichen Sinn. Auswahl der Items für den Haupttest anhand von Itemanalysen aller Feldtestaufgaben. Erstellung nationaler Feldtestberichte.

#### Testdurchführung:

Qualitätssicherung durch international standardisierte Testsitzungen (Durchführungsobjektivität). Garantiert wurden diese durch externe, speziell geschulte Testadministratoren und ein standardisiertes Manual für die Testdurchführung inklusive standardisierter Skripts.

Weiters ist während jeder Testsitzung vom Testadministrator ein nach internationalen Vorgaben standardisiertes Protokoll über Ablauf und eventuell aufgetretene Probleme zu führen.

Zusätzlich werden im Rahmen des Pisa Quality Monitoring eine repräsentative Auswahl von ungefähr 10% der PISA Schulen, von einem Pisa Quality Monitor (PQM) unangekündigt am Test-Tag besucht. Dieser hat die Aufgabe die Testdurchführung zu beobachten und mit Hilfe eines standardisierten Beobachtungsprotokolls Abweichungen von den Richtlinien zu protokollieren und den Testkoordinator zu interviewen.

Pro Land waren mindestens zwei PQMs im Einsatz, die möglichst viele Schultypen und unterschiedliche Testadministratoren abdecken sollten. Ein Pisa Quality Monitor muss unabhängig von nationalen Projektzentrum und dem Nationalen Projekt Manager sein, sowie eine Vielzahl weiterer Qualifikationen aufweisen.

## 5. Publikation (Zugänglichkeit)

### **Revisionen**

-

### **Vorläufige Ergebnisse**

Keine

### **Endgültige Ergebnisse**

Die Veröffentlichung erfolgte am 7. Dezember 2004, 0:01 Uhr in Paris, unter dem Titel "Learning for Tomorrow's World – First Results from PISA 2003". Ab 6. Dezember stand der Bericht unter Verpflichtung zu Vertraulichkeit für Journalisten zur Verfügung.

Zusätzlich gab es eine Vielzahl an detailliertere Analysen in sogenannten "Thematic Reports", die zu späteren Zeitpunkten erschienen.

**Publiziert in:**

Der Bericht zur PISA Studie 2003 ist auf der Webseite der OECD ([www.pisa.oecd.org](http://www.pisa.oecd.org)) unter [http://www.pisa.oecd.org/document/55/0,2340,en\\_32252351\\_32236173\\_33917303\\_1\\_1\\_1\\_1,00.html](http://www.pisa.oecd.org/document/55/0,2340,en_32252351_32236173_33917303_1_1_1_1,00.html) publiziert. Dort finden sich weitere Berichte, Auswertungen und auch die Datenfiles zur genaueren Recherche.

**Behandlung vertraulicher Daten**

Das Interesse von PISA ist lediglich auf aggregierte Daten beschränkt. Schülernamen werden nicht gespeichert, Schulen bekommen eine Nummer zugeordnet, deren Schlüssel nur dem Nationalen Projektmanager zugänglich ist. Es werden keinerlei Einzeldaten von Schülern und Schulen erhoben.

Das Hauptaugenmerk in Punkto Sicherheit liegt daher bei PISA in der Geheimhaltung der Testfragen. Da die PISA-Studie im Längsschnitt durchgeführt wird und idente Testaufgaben auch in zukünftigen Erhebungen eingesetzt werden, dürfen diese nicht an die Öffentlichkeit gelangen. Da der Testadministrator mit dem gesamten Testmaterial an die Schule kommt und dieses nach der Durchführung der Testsitzung wieder mitnimmt, wird sichergestellt, dass zu keinem Zeitpunkt Material an der Schule gelagert wird.

## 6. Qualität

### 6.1 Relevanz

Die Erfüllung der nationalen Anforderungen ist durch die aktive Mitarbeit jedes teilnehmenden Landes gesichert. Einerseits hat jedes Land einen Vertreter im PISA Governing Board, wo alle wichtigen Entscheidungen getroffen werden, andererseits herrscht ein reger systemisch verankerter, beidseitiger Informationstausch zwischen den nationalen Projektzentren und dem Governing Board.

## 6.2 Genauigkeit

### 6.2.1 Stichprobenbedingte Effekte, Repräsentativität

Die Schätzung der Stichprobenvarianz erfolgt mittels der sehr aussagestarken Fay's Methode der "Balanced Repeated Replication" (BRR).

Der Standardfehler auf der Mathematik-Skala bewegte sich 2003 zwischen 0,96 (Luxemburg) und 6,70 (Türkei).

#### Design-Effekte und Effektive Sampling-Größen

PISA erhebt 5 verschiedene Design-Effekte, die den Einfluss des Erhebungs- und Testdesigns auf den Standardfehler beschreiben.

In den nachfolgenden Formeln werden folgende Variablen verwendet:

$t =$  Statistik

$Var =$  Varianz

$MVar =$  Messvarianz

$SRS =$  Einfache Stichprobenerhebung

$BRR =$  Varianzschätzung mittels Fay's BRR Verfahren

*Designeffekt 1:*

$$Def f_1(t) = \frac{Var_{SRS}(t) + MVar(t)}{Var_{SRS}(t)}$$

Für eine Statistik  $t$  zeigt der Design Effekt 1 die Steigerung (inflation) der totalen Varianz, die aufgrund des Messfehlers aufgetreten wäre, wenn die Erhebung eine einfache Stichprobenerhebung wäre.

*Designeffekt 2:*

$$Def f_2(t) = \frac{Var_{BRR}(t) + MVar(t)}{Var_{SRS}(t) + MVar(t)}$$

Zeigt die Steigerung der totalen Varianz durch den Messfehler.

*Designeffekt 3:*

$$Def f_3(t) = \frac{Var_{BRR}(t)}{Var_{SRS}(t)}$$

Zeigt die Steigerung der Sampling-Varianz allein durch die Verwendung des komplexen Designs.

*Designeffekt 4:*

$$Def f_4(t) = \frac{Var_{BRR}(t) + MVar(t)}{Var_{BRR}(t)}$$

Zeigt die Steigerung der totalen Varianz, die auf den Messfehler zurückzuführen ist.

*Designeffekt 5:*

$$Def f_5(t) = \frac{Var_{BRR}(t) + MVar(t)}{Var_{SRS}(t)}$$

Zeigt die Steigerung der totalen Varianz auf Grund des Messfehlers und des komplexen Designs an.

Außer in Indonesien, Mexiko und der Türkei sind die Design-Effekte relativ gering, in Mathematik definitionsgemäß am höchsten, da hier die meisten Schüler befragt wurden.

In Design-Effekt 1 und 4 ist die Messfehlervarianz immer knapp über 1. Die Design-Effekte 2, 3 und 5, liegen mit Ausnahme jener Länder mit Vollerhebung (Island, Lichtenstein, Luxemburg) bei ungefähr 8 in Mexico, Indonesien, Türkei über 15.

## 6.2.2 Nicht-stichprobenbedingte Effekte

### **Qualität der verwendeten Datenquellen**

Da die Schüler-/Schullisten von den nationalen Ministerien zur Verfügung gestellt werden, handelt es sich um die beste verfügbare Datenquelle. Meist handelt es sich um Daten aus dem Vorjahr, weshalb meist einige nicht systematische Fehler (durch Schulwechsel uä.) auftreten.

### **Abdeckung (Fehlklassifikationen, Unter-/Übererfassung)**

- Abdeckung der national gewünschten Zielpopulation (national desired target population – NDP): 1,0 (Brasilien, Indonesien, Macao, Tunesien, Uruguay) - 0,93 (Spanien, USA, Canada).
- Abdeckung der national (mindestens in der 7. Schulstufe) eingeschriebenen Population (national enrolled population – NEP): 1,0 (Macao, Tunesien, Uruguay - 0,93 (Spanien, USA, Canada).
- Abdeckung der nationalen Population der 15-Jährigen: 1,04 (Schweiz) – 0,36 (Türkei)
- Abdeckung der geschätzten Schulbesucher: 1,09 (Schweiz) – 0,76 (Serbien)
- Abdeckung der Schul-Stichproben Population: 1,12 (Australien) - 0,73 (Indonesien)

### **Antwortausfall (Unit-Non-Response, Item-Non-Response)**

#### **Item-Non-Response**

Ein Schüler beantwortet eine einzelne Frage nicht. Das kann aus folgenden Gründen passieren, welche auch bei der Dateneingabe speziell erfasst wird.

1. Nicht wissen der Antwort
2. Nicht erreichen der Frage (Zeit)

Die durchschnittliche Zahl der fehlenden Items liegt zwischen 1,57 (Niederlande) und 11,24 (Serbien).

Die durchschnittliche Zahl der nicht erreichten Items liegt zwischen 0,33 (Österreich) und 5,56 (Uruguay).

Da es in der Natur einer Kompetenzerhebung liegt, dass einzelne Items nicht beantwortet werden (können), wird nicht speziell verhindert dass ein Item-Non-Response auftritt. Es wird jedoch versucht, dass der Test für alle in der vorhandenen Zeit bewältigbar ist.

## **Unit-Non-Response**

Ein Schüler beantwortet seinen Fragebogen nicht, erscheint nicht zum Test- bzw. Nachtermin oder verweigert die Teilnahme.

Der Antwortausfall oder die Rücklaufquote(PISA) ist limitiert:

1. Schulebene: Um Verzerrungen zu verhindern gibt es auf Schulebene eine vorgegebene Rücklaufquote von mindestens 85% der ursprünglich ausgewählten Schulen und die Möglichkeit nicht teilnehmende Schulen durch Replacement-Schulen (2 je Sample-Schule) zu kompensieren.

Ein Nachweis der Unbedeutsamkeit der Messabweichung durch Schulen die nicht auf der Teilnahmeaufforderung gefolgt sind ist in jedem Fall notwendig.

2. Schülerebene: 80% Rücklaufquote auf nationaler Ebene (über alle teilnehmenden Sample- und Replacement-Schulen hinweg berechnet) inkl. eventueller Nachtests.

Lag die Schülerbeteiligung in einer Schule zwischen 25 und 50%, galt diese Schule nicht als teilnehmende Schule in der Berechnung und Dokumentation der Antwortraten, die Daten wurden jedoch mit in die Auswertung einbezogen. Bei Schulen mit Schülerbeteiligung unter 25%, wurden die Daten nicht in die Datenbank einbezogen und galten ebenfalls als Non-Respondents.

### Schul-Antwortraten (vor replacement):

- Gewichtet: 100% (Indonesien, Liechtenstein, Macao, Serbien, Tunesien) - 64,94% (USA)
- Ungewichtet: 100% (Indonesien, Liechtenstein, Macao, Serbien, Tunesien) – 65,18% (USA)

Schul -Antwortraten (nach replacement):

- Gewichtet: 100% (Finnland, Indonesien, Italien, Korea, Liechtenstein, Macao, Russland, Serbien, Spanien, Thailand, Tunesien, Türkei) - 68,12% (USA)  
(Alle über 90% ausser USA, Niederlande, Kanada und Vereinigtes Königreich)
- Ungewichtet: 100% (Finnland, Indonesien, Italien, Korea, Liechtenstein, Macao, Russland, Serbien, Spanien, Thailand, Tunesien, Türkei) – 65,59% (USA)

Schüler -Antwortraten (nach replacement):

Gewichtet: 98,22% (Liechtenstein) - 77,92% (Vereinigtes Königreich)

Ungewichtet: 98,22% (Liechtenstein) – 73,50% (Österreich)

**Messfehler (Erfassungsfehler)**

Die wichtigsten Ursachen für Messfehler bei PISA sind

- Fragebogen (z.B. schlechte Qualität des Ausdrucks (-> Lesbarkeit), missverständliche Formulierung, Testlänge, usw.)
- Kompetenzskalen
- Respondenten (falsche Angaben)

Um Fehler wie diese zu vermeiden, wurde eine Vielzahl von qualitätslenkenden Vorkehrungen getroffen. Wichtigster Punkt ist dahingehend die Feldstudie, die eine Möglichkeit bot Fehlerquellen zu erkennen und im Haupttest zu vermeiden. Besonders wichtig war die Feldstudie für die Auswahl der Items, um sicherzustellen, dass ihre Übersetzung verständlich erfolgt war und die Antwortwahrscheinlichkeiten der einzelnen Fragen international gleichwertig sind.

Sollten einzelne Fragestellungen (Items) in manchen Ländern im Haupttest immer noch eine abweichende Lösungswahrscheinlichkeit aufweisen, können unter bestimmten Bedingungen einzelne Items für einzelne Länder aus der Score-Berechnung ausgeschieden werden. Insgesamt wurden 22 Items in einem oder mehreren Ländern gestrichen, wobei nur in Tunesien, Italien, Spanien, Russland, Korea und Uruguay mehr als ein Item gestrichen wurde. Zwei Items wurden International gestrichen.

Die Auswahl der finalen Testitems wurde im Rahmen einer Feldstudie durchgeführt und erfolgte mittels Homogenitätsindex.

Finale Zuverlässigkeit der 4 Kompetenzskalen:

| Domäne            | Zuverlässigkeit |
|-------------------|-----------------|
| Mathematik        | 0,918           |
| Lesen             | 0,848           |
| Naturwissenschaft | 0,843           |
| Problemlösen      | 0,874           |

Auch die Testlänge ist eine mögliche Fehlerquelle, da ein zu langer Test (mit zu vielen Items), zu einer schlampigen Arbeitsweise führen kann. Genauso kann ein zu kurzer Test für Unruhe sorgen, was für die noch arbeitenden Schüler störend wäre.

Falsche Angaben von Schülern in den Kontextfragebögen wurden mittels automatisierter Checks ausgemustert. Messfehler durch absichtliches falsch beantworten einzelner Fragen sind kaum quantifizierbar und daher auch nicht konsequent vermeidbar, vermutlich jedoch auch nur sehr selten.

## **Aufarbeitungsfehler**

2 Fehlerquellen:

- Codierung
- Dateneingabe

### Codierung:

Für die einheitliche Kodierung der Antworten wurden Spezialisten ausgebildet. Unterstützend wurden Anleitungen zur Kodierung von Antworten zu jeder einzelnen Frage erstellt, wobei sie zu Frage-Antwortkategorien mit jeweils eigenen Punkte-Codes zusammengefasst wurde. Neben der Beschreibung der Antwortart, die jedem Code zugewiesen werden soll, gab es eine Beispielantwort je Kategorie. Die Quote der manuell zu codierenden Fragen wurde auf rund 40% limitiert. Items mit Schwierigkeiten in der

Kodierung wurden vermieden, zusätzlich Homogenitäts-, Varianz- und Reliabilityanalysen zur Verifizierung eingeführt.

#### Dateneingabe:

Dateneingabe erfolgte durch speziell geschulte Mitarbeiter.

Kontrollmechanismen erfolgten manuell durch Mitarbeiter, automatisiert durch KeyQuest-Software

### **Modellbedingte Effekte**

Fay's BRR: Eine Vielzahl von empirischen Studien bestätigte den Vorteil von Fay's Methode gegenüber der in vorangegangenen internationalen Bildungs- Studien wie TIMMS verwendeten Jackknife-Methode. Fay's BRR steht auf einer festen theoretischen Basis für die Verwendung mit Median und Quantilen.

Rasch Modell: Das Rasch-Modell vereint Aufgabenschwierigkeit und Leistungsfähigkeit auf einer Skala je Kompetenzbereich. Auf Grund der willkürlichen Normierung des Mittelwerts dieser Fähigkeitsskalen auf 500, kann es durch die Hinzunahme einzelner Länder (2003: Türkei) zu einer subjektiven Veränderung der Leistung der anderen Länder kommen, ohne dass diese inhaltlich etwas zur Verbesserung beigetragen haben.

## 6.3 Rechtzeitigkeit und Aktualität

Das Projekt PISA 2003 startete im September 2000 und endete im Dezember 2004. Während dieser Zeit wurde ein strikter Plan verfolgt, um diese groß angelegte internationale Studie sinnvoll durchführen zu können.

Die Phasen Testentwicklung (Framework-Entwicklung, Item-Entwicklung, Nationaler Review, Übersetzung, Feldversuch, Itemauswahl, Coder-Training) bis Testdurchführung (zwischen Februar und Oktober 2003) fanden in genau festgelegten Slots zwischen September 2000 und Oktober 2003 statt.

## 6.4 Vergleichbarkeit

Die Vergleichbarkeit von PISA auf *zeitlicher* Ebene ist durch die normierten Skalen und die gleichbleibenden Fragestellungen gewährleistet.

Auf *räumlicher* Ebene sind die exakten und verbindlichen Vorgaben, die im internationalen Technischen Bericht festgelegt sind, der Garant für die internationale Vergleichbarkeit der Studie.

Im Bezug auf *andere Schulleistungsuntersuchungen* ist die Vergleichbarkeit nicht untersucht.

## 6.5 Kohärenz

Eine Kohärenzprüfung ist auf Grund fehlender geeigneter Datenquellen nicht möglich.

Es gibt jedoch Stimmen, die in den Ergebnissen der PISA-Studie Korrelationen (plus-minus 3 IQ-Punkte) zwischen den PISA-Werten und IQ-Tests erkennen können.

## 4 Vergleich zweier Länder

Der Vergleich von Österreich und Finnland im Template der Statistik Austria, soll die in Kapitel 3 erstellte Standarddokumentation mit den jeweils spezifischen Länderdaten füllen. Diese Einträge sind als Ergänzung, als nationale Spezifikation der internationalen Vorgaben aus Abschnitt 3 zu verstehen. Dabei sollen die Verschiedenheiten oder auch Ähnlichkeiten zweier europäischer Länder in international vorgegebenen Rahmenbedingungen gezeigt werden und in Relation zueinander gestellt werden. Jene Abschnitte in denen es keine spezifischen Daten gibt, sind mit "siehe Kapitel 3" gekennzeichnet. In beiden Fällen ist ein Vergleich mit den Informationen aus Kapitel 3 ratsam.

### 1. Wichtige Hinweise

Siehe Kapitel 3

### 2. Allgemeine Informationen

#### **Statistiktyp**

Primärstatistik

#### **Fachgebiet**

Bildung

#### **Verantwortliche Organisationseinheit, Ansprechpartner/in**

#### **Österreich, Finnland**

OECD (Organisation for Economic Cooperation and Development)

Projektleitung PISA:

OECD Sekretariat:

Andreas Schleicher ( übergreifende Managementverantwortung)

Schnittstelle:

Board of Participating Countries:

ein Vertreter jedes teilnehmenden Landes

**Österreich:** Jürgen Horschinegg

**Finnland:** Jari Rajanen

Nationale Projektteams:

Nationaler Projekt Manager (NPM):

**Österreich:** Günter Haider (Salzburg)

**Finnland:** Jouni Välijärvi (Jyväskylä)

Design und Implementierung:

internationales Konsortium:

Leitung: Australian Council for Education Research (ACER)

### **Ziel und Zweck, Geschichte**

Finnland und Österreich nehmen seit 2000 an PISA teil.

### **Periodizität**

Siehe Kapitel 3

### **Auftraggeber**

Siehe Kapitel 3

### **Nutzer**

Siehe Kapitel 3

**Rechtsgrundlage(n)**

Siehe Kapitel 3

**3. Statistische Konzepte, Methodik****Gegenstand der Statistik**

Siehe Kapitel 3

**Beobachtungs-/Erhebungs-/Darstellungseinheiten**

Siehe Kapitel 3

**Datenquellen**

Gesamtevidenz der 15 jährigen Schüler:

**Österreich:**

(Statistik Austria (Ende 2001) und österreichische Schulstatistik (Schuljahr 2001/02))

**Finnland:**

Statistik Finnland (2001).

**Meldeeinheit/Respondenten**

Siehe Kapitel 3

**Erhebungsform**

Siehe Kapitel 3

**Charakteristika der Stichprobe****Grundgesamtheit und Auswahlrahmen:****Österreich:**

Grundgesamtheit sind die 15-jährigen die im Schuljahr der Erhebung an einer Schule (educational institution) gemeldet sind, inklusive Besucher auf Teilzeitbasis (Österreich

BerufsschülerInnen). Für PISA 2003 Schüler des Jahrgang 1987 (Österreich: 1.1.87 – 31.12.87), die sich mindestens in der 7. Schulstufe befinden.

**Finnland:**

Grundgesamtheit sind die 15-jährigen, die die Pflichtschule (peruskoulu) besuchen und zwischen Februar 1987 und Jänner 1988 geboren sind.

Samplingverfahren:

*Stratifizierungskriterium:*

**Österreich:**

*Allgemeinbildende Pflichtschulen (APS)*

- (1. Volksschule)
2. Hauptschule
3. Polytechnische Schule
4. Sonderschule

*Allgemeinbildende Höhere Schulen (AHS)*

5. Gymnasium
6. Realgymnasium und Wirtschaftskundliches RG
7. Oberstufenrealgymnasium
8. Sonstige Allgemeinbildende Schulen/mit Statut

*Berufsschulen (Berufsbildende Pflichtschulen, BS)*

9. Berufsschule (technisch-gewerblich)
10. Berufsschule (kaufmännisch/Handel und Verkehr)
11. Berufsschule (land- und forstwirtschaftlich) Berufsbildende Mittlere Schulen (BMS)
12. Gewerblich-technisch-kunstgewerbliche Fachschulen
13. Kaufmännische Schulen/Handelsschulen
14. Wirtschaftlich-sozialberufliche Fachschulen
15. Land- und forstwirtschaftliche Fachschulen

*Berufsbildende Höhere Schulen (BHS)*

16. Technische und gewerbliche Höhere Schulen

17. Kaufmännische Höhere Schulen

18. Höhere Schulen für wirtschaftliche Berufe/sozialberufliche Höhere Schulen

19. Land- und forstwirtschaftliche Höhere Schulen

*Anstalten der Lehrer/innen- und Erzieher/innenbildung*

20. Anstalten der Lehrer/innen- und Erzieher/innenbildung

**Finnland:***1. Regionen*

- Uusimaa
- Etelä-Suomi
- Itä-Suomi
- Väli-Suomi
- Pohjois-Suomi

*2. Unterrichtssprache*

- Finnisch
- Schwedisch

*3. Gemeinden*

- ländliche
- städtische (inklusive größerer Gemeinden)

Auf Grund ihrer geringen Anzahl, wurden in allen 52 schwedischen Schulen in Finnland Schüler getestet.

**Erhebungstechnik/Datenübermittlung**

Siehe Kapitel 3.

**Erhebungsbogen (inkl. Erläuterungen)**

**Österreich:** Taschenrechner erlaubt

**Finnland:** Taschenrechner nicht erlaubt

**Teilnahme an der Erhebung**

Siehe Kapitel 3

**Erhebungs- und Darstellungsmerkmale, Maßzahlen; inkl. Definition**3. Zusätzliche internationale und nationale Erhebungsmöglichkeiten

*International:*

1. Bildungskarriere
2. Informations- und Kommunikationstechnologie

**Österreich:** Durchführung beider zusätzlicher Optionen

**Finnland:** Durchführung beider zusätzlicher Optionen

*National:*

**Österreich:**

PISA Plus (nicht zu verwechseln mit dem internationalen PISA Plus, das für Länder steht, die PISA 2000 erst im Jahr 2001 durchführten)

- Nutzung von Informationstechnologien
- Lesegewohnheiten
- Belastung in der Schule
- Befindlichkeit und Schulerfolg
- Qualität in Schulen
- Förderung von Schüler/innen nichtdeutscher Muttersprache

**Finnland:**

- Internetnutzung
- Lese Engagement (mit Norwegen gemeinsam)

**Verwendete Klassifikationen**

Siehe Kapitel 3

**Regionale Gliederung der Ergebnisse**

Siehe Kapitel 3

**4. Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde  
Maßnahmen****Datenerfassung**

Siehe Kapitel 3

**Signierung (Codierung)**

Siehe Kapitel 3

**Plausibilitätsprüfung, Prüfung der verwendeten Datenquellen**

Siehe Kapitel 3

**Imputation (bei Antwortausfällen bzw. unvollständigen Datenbeständen)**

Siehe Kapitel 3

**Hochrechnung**

Siehe Kapitel 3

**Erstellung des Datenkörpers, verwendete Rechenmodelle, statistische  
Schätzmethoden**

Siehe Kapitel 3

## **Sonstige qualitätssichernde Maßnahmen**

### **Österreich:**

15 Schulen wurden von zwei PISA Quality Monitors besucht. Mit Ausnahme von Vorarlberg, wo keine PQM-Besuche stattfanden, waren die ausgewählten österreichischen Schulen über alle Bundesländer verteilt.

### **Finnland:**

2 PISA Quality Monitors besuchten 10% der Schulen in denen Tests stattfanden. Genauere Details waren dem finnischen Projektteam nicht bekannt.

## **5. Publikation (Zugänglichkeit)**

### **Revisionen**

-

### **Vorläufige Ergebnisse**

-

### **Endgültige Ergebnisse**

#### **Österreich:**

Zeitgleich mit dem internationalen Initial Report wurde auch der nationale Ergebnisbericht herausgegeben. In Folge erschienen weitere spezifische Analysen bzw. Ergebnisse von Zusatzprojekten. In Österreich wurde auch ein nationaler Technischer Bericht, der die Methoden und Abläufe von PISA 2003 in Österreich offenlegt, erstellt.

#### **Finnland:**

Zeitgleich mit dem internationalen Initial Report wurde auch der nationale Ergebnisbericht herausgegeben.

**Publiziert in:**

Siehe Kapitel 3

**Österreich:**

Die Veröffentlichung der nationalen Berichte erfolgte über die Webseite [www.pisa-austria.at](http://www.pisa-austria.at).

**Finnland:**

Veröffentlichung auf der Internetseite <http://ktl.iyu.fi/pisa/>.

**Behandlung vertraulicher Daten**

Siehe Kapitel 3

## 6. Qualität

### 6.1. Relevanz

Siehe Kapitel 3

### 6.2. Genauigkeit

#### 6.2.1. Stichprobenbedingte Effekte, Repräsentativität

Standardfehler der PISA 2003 Mathematik- Skala:

|            | <b>SRS</b> | <b>Cluster</b> | <b>Fay's BRR</b> |
|------------|------------|----------------|------------------|
| Österreich | 1.37       | 5.39           | 3.23             |
| Finnland   | 1.10       | 1.79           | 1.78             |

SRS beschreibt den Standardfehler einer einfachen Stichprobenerhebung, Cluster, die eines mehrstufigen Samplingverfahrens ohne komplementäre Informationen und Fay's BRR, die erwartungstreue Schätzung unter Verwendung von Fay's Replikation.

Design-Effekte (DE) und jeweilige Effektive Stichprobengröße (ES) bei mittlerer Leistung (Definition siehe Kapitel 3)

|                          |           | DE1  | DE2  | DE3  | DE4  | DE5  | ES1  | ES2  | ES3  | ES4  | ES5  |
|--------------------------|-----------|------|------|------|------|------|------|------|------|------|------|
| <b>Mathematik</b>        | <b>A</b>  | 1.14 | 4.97 | 5.52 | 1.02 | 5.66 | 4040 | 924  | 833  | 4485 | 812  |
|                          | <b>FI</b> | 1.25 | 2.30 | 2.63 | 1.10 | 2.88 | 4626 | 2519 | 2204 | 5288 | 2011 |
| <b>Lesen</b>             | <b>A</b>  | 1.10 | 5.58 | 6.02 | 1.02 | 6.11 | 4195 | 824  | 764  | 4525 | 752  |
|                          | <b>FI</b> | 1.16 | 2.06 | 2.22 | 1.07 | 2.38 | 5009 | 2820 | 2609 | 5413 | 2437 |
| <b>Naturwissenschaft</b> | <b>A</b>  | 1.09 | 5.29 | 5.69 | 1.02 | 5.78 | 4210 | 868  | 808  | 4524 | 795  |
|                          | <b>FI</b> | 1.28 | 2.04 | 2.33 | 1.12 | 2.60 | 4537 | 2844 | 2492 | 5178 | 2226 |

Die effektive Stichprobengröße liefert die Anzahl an Schülern, die man in einer einfachen Stichproben-Erhebung (SRS – simple random sample) testen müsste, um genauso präzise zu sein wie die PISA 2003 Stichprobe, hier unter Berücksichtigung des jeweiligen Designeffekts.

## 6.2.2. Nicht-stichprobenbedingte Effekte

### Qualität der verwendeten Datenquellen

|            | Ungeeignet | geeignet     | Ausschlüsse within school | Ausschlüsse gesamt |
|------------|------------|--------------|---------------------------|--------------------|
| Österreich | 146 (3.3%) | 6306(96.7%)  | 1.51%                     | 1.62%              |
| Finnland   | 32 (0.52%) | 6314(99.48%) | 1.24%                     | 3.38%              |

### Abdeckung (Fehlklassifikationen, Unter-/Übererfassung)

- *Abdeckungsindex 1:* Abdeckung der national gewünschten Zielpopulation (national desired target population – NDP)
- *Abdeckungsindex 2:* Abdeckung der national (mindestens in der 7. Schulstufe) eingeschriebenen Population (national enrolled population – NEP)
- *Abdeckungsindex 3:* Abdeckung der nationalen Population der 15-Jährigen
- *Abdeckungsindex 4:* Abdeckung der geschätzten Schulbesucher
- *Abdeckungsindex 5:* Abdeckung der Schul-Stichproben Population

|                   | Abdeckungsindizes |      |      |      |      |
|-------------------|-------------------|------|------|------|------|
|                   | 1                 | 2    | 3    | 4    | 5    |
| <b>Österreich</b> | 0.98              | 0.98 | 0.91 | 0.99 | 0.99 |
| <b>Finnland</b>   | 0.97              | 0.97 | 0.95 | 0.98 | 1.00 |

### Antwortausfall (Unit-Non-Response, Item-Non-Response)

#### Item-Non-Response

|                   | Durchschnittliche Anzahl |                        |
|-------------------|--------------------------|------------------------|
|                   | fehlender Items          | nicht erreichter Items |
| <b>Österreich</b> | 5.1                      | 0.33                   |
| <b>Finnland</b>   | 2.85                     | 0.76                   |

**Unit-Non-Response**Schulen:*Schul- Rücklaufquote vor replacement*

|  | <b>Österreich</b>     | <b>Finnland</b> |
|--|-----------------------|-----------------|
| Ungewichtete Anzahl der Antwortenden Schulen / von gesamt<br>(laut nationalem Bericht) | 192/194<br>(193*/193) | 193/197         |
| Ungewichtete Schulbeteiligungsrate<br>(laut nationalem Bericht)                        | 98.97%<br>(100%)      | 97.97%          |
| Gewichtete Schulbeteiligungsrate<br>(laut nationalem Bericht)                          | 99.29%<br>(100%)      | 97.39%          |

\* eine Schule erreichte trotz offizieller Teilnahme eine interne Rücklaufquote von unter 50%, wodurch sie international als nicht teilgenommen gilt. Die erhobenen Daten sind jedoch in der Datenbank enthalten.

*Schul- Antwort-Raten nach replacement*

|   | <b>Österreich</b> | <b>Finnland</b> |
|---|-------------------|-----------------|
| Ungewichtete Anzahl der Antwortenden Schulen / von gesamt | 192/194           | 197/197         |
| Ungewichtete Schulbeteiligungsrate                        | 98.97%            | 100%            |
| Gewichtete Schulbeteiligungsrate                          | 99.29%            | 100%            |

Schüler:*Schüler- Antwort-Rate nach replacement*

|  | <b>Österreich</b> | <b>Finnland</b> |
|--|-------------------|-----------------|
| Anzahl der getesteten Schüler (ungewichtet) / von Gesamt | 4566 / 6212       | 5796 / 6235     |
| Ungewichtete Schülerbeteiligungsrate nach replacements   | 73.5%             | 92.96%          |
| Gewichtete Schülerbeteiligungsrate nach replacements     | 83.56%            | 92.84%          |

### Spezialfall: Problem Rücklaufquote Berufsschüler in Österreich:

In österreichischen Berufsschulen gibt es den Spezialfall, dass Klassen nur geblockt unterrichtet werden und daher während des von PISA vorgesehenen 6-wöchigen Testfensters nicht erreichbar sind und folglich auch nicht getestet werden können. In diesen lehrgangsmäßig geführten Berufsschulen ist es dadurch meist unmöglich die erforderliche schulinterne Rücklaufquote von 50% zu erreichen. Berücksichtigt man die nicht erreichbaren Schüler nicht, erhält man 94.4% Beteiligung. Rechnet man sie ein, liegt die Rücklaufquote bei 37.5%, obwohl nur 48 von 861 Schülern tatsächlich fehlten.

### Berufsschul-Validity-Study

Um die Validität der erhobenen Daten für die Gesamtheit der Berufsschüler zu beweisen, wurden Vergleiche der Häufigkeitsverteilungen ihrer demographischen Daten durchgeführt. Es wurde bewiesen, dass die Verteilung der demographischen Merkmale der erreichten Berufsschüler, nur zufällig unterschiedlich zur Verteilung der Merkmale in der Population aller Berufsschüler ist.

### **Messfehler (Erfassungsfehler)**

Sollten einzelne Fragestellungen (Items) in manchen Ländern eine abweichende aufweisen, können unter bestimmten Bedingungen einzelne Items für einzelne Länder aus der Score-Berechnung ausgeschieden werden.

**Österreich:** Item R055Q03 (wurde international in allen deutschsprachigen Tests gestrichen)

**Finnland:** -

### **Aufarbeitungsfehler**

Siehe Kapitel 3

### **Modellbedingte Effekte**

Siehe Kapitel 3

### 6.3. Rechtzeitigkeit und Aktualität

Siehe Kapitel 3

### 6.4. Vergleichbarkeit

Siehe Kapitel 3

### 6.5. Kohärenz

Siehe Kapitel 3

## 5 Zusammenfassung

### 5.1 Dokumentation

Die Dokumentation von PISA ist sehr umfangreich und ausführlich. Die größte Schwierigkeit war es, dabei den Blick für das Relevante zu behalten und die Zusammenhänge trotzdem schlüssig zu erklären. Denn PISA ist nicht nur eine große (viele Länder umfassende) Erhebung, auch ebenso verbirgt sich dahinter eine Vielzahl an statistischen und qualitätssichernden Methoden, bei denen man schnell den Überblick verliert. Das ist auch mit ein Grund, wieso die Debatte um PISA nur relativ oberflächlich, meist auf den Rankings basierend, betrieben wird.

Die in Kapitel 3 erarbeitete Dokumentation ist auch ohne die vorliegende Arbeit als eigenständige "Standarddokumentation Metadaten für PISA 2003" verwendbar und bietet einen Vergleich mit anderen Erhebungen. Nach dem Template der Statistik Austria umfasst sie mit den Punkten: allgemeine Informationen, Details zu den Statistischen Konzepten und zur Methodik, Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde Maßnahmen, Details zur Publikation und Qualität, alle wesentlichen Aspekte einer wissenschaftlichen Erhebung.

### 5.2 Ländervergleich

Der Vergleich von Österreich und Finnland zeigt sich in Anbetracht der strikten internationalen Vorgaben nur wenig aufschlussreich. Die meisten Punkte sind im Kapitel 3 ausführlich beschrieben und wurden auch in beiden Ländern analog umgesetzt.

Lediglich in den Stratifikations-Kriterien und klarerweise den Qualitätsmerkmalen zeigen sich Unterschiede. Die Testdurchführung wurde in Österreich wie in Finnland sehr professionell und kompetent durchgeführt und auch dokumentiert. Lediglich ein nationaler Technischer Bericht, der im Detail auf die Umsetzung in Finnland eingeht, ist leider nicht vorhanden.

Vermutlich beruht die Tatsache, dass in Österreich ein solcher erstellt wurde, darauf, dass es Probleme mit dem Rücklauf von saisonal geführten Berufsschulen gab.

### 5.3 Resumee

Für die Erstellung von Standarddokumentationen ist ein hoher Grad an Fachwissen, Erfahrung und Einblick in die behandelte Materie notwendig. Ohne die beratende Unterstützung durch meine Betreuer wäre die Arbeit in dieser Qualität nicht möglich gewesen. Trotzdem bin ich mir bewusst, dass die von mir erstellte Arbeit in einigen Punkten durchaus noch verbesserungswürdig ist.

PISA, vor allem seine Methodik und Qualitätssicherungsmaßnahmen sind derart komplex und vielfältig, dass es für einen Laien wie mich schwierig ist, zu entscheiden was in eine Standarddokumentation einfließen soll und was nicht. Auch wenn Standarddokumentationen ihrem Namen getreu als standardisiert gelten, so macht es die Tatsache, dass auch unter den von der Statistik Austria veröffentlichten Dokumenten bemerkenswerte Unterschiede in Ausführung und Qualität waren, leichter, mit meiner Arbeit zufrieden zu sein.

### 5.4 Ausblick (PISA06)

PISA ist eine interessante und gut geplante Studie zur Bildungsermittlung der 15 jährigen. Die Reduktion der öffentlichen Diskussion auf Rankings, ist mit Sicherheit in der Vielschichtigkeit der dahinterliegenden Methodik verbunden, die in ihrer Komplexität öffentlich nicht vermittelbar ist. Da PISA 2003 erst zum zweiten mal durchgeführt wurde, gibt es einerseits noch vereinzelt Probleme, andererseits ist der nationale Vergleich mit vorangehenden Schülergenerationen nur begrenzt (mit dem Jahr 2000) möglich. Über die Jahre wird PISA nicht nur einen internationalen Vergleich zwischen verschiedenen Ländern erlauben, sondern auch landesintern eine bedeutende Kennzahl für die Entwicklung der Schülerleistungen von 15 jährigen bilden.

Da PISA sehr sorgfältig geplant hat, waren im Nachhinein nur noch wenige Adaptionen durchzuführen, was in einer guten Vergleichbarkeit der Studienergebnisse über die Jahre hinweg resultiert. Der kurze Versuch 2003 mit Problemlösen eine zusätzliche Domäne zu etablieren, wurde verworfen. 2006 wurde wie auch schon im Jahr 2000, nur die Kenntnis in Mathematik, Lesen und Naturwissenschaften erhoben. Hauptdomäne waren dabei 2006 die Naturwissenschaften.

In Zukunft wäre es wünschenswert PISA und seine Methodik auch öffentlich ausführlicher zu behandeln. Eine tiefgreifende Diskussion und Verständnis sind einer positiven Entwicklung des Bildungssystems sicher dienlich und somit auch im Sinne der PISA Erfinder.

## 6 Literaturverzeichnis

- Eurostat. 2003.** *Handbook "How to make a quality report"*. Luxembourg : Eurostat, 2003.
- Haider (Hrsg.), Günter, Lang, Birgit und Reiter, Claudia. 2004a.** *PISA 2003 Internationaler Vergleich von Schülerleistungen - Technischer Bericht*. Salzburg : bm:bwk, 2004a.
- Haider, Günter und Reiter, Claudia. 2004b.** *Pisa 2003 - Nationaler Bericht*. Graz : Leykam, 2004b.
- Neuwirth, Erich, Ponocny, Ivo und Grossmann, Wilfried. 2006.** *PISA 2000 und PISA 2003: Vertiefende Analysen und Beiträge zur Methodik*. Wien : Bundesministerium für Bildung, Wissenschaft und Kultur, 2006.
- OECD. 2004.** *Learning for Tomorrow's World - First Results from PISA 2003*. Paris : s.n., 2004.
- **2002.** *PISA 2000 Technical Report*. Paris : s.n., 2002.
- **2005c.** *PISA 2003 Data Analysis Manual - SAS User*. Paris : OECD, 2005c.
- **2005b.** *PISA 2003 Data Analysis Manual - SPSS User*. Paris : OECD, 2005b.
- **2005a.** *PISA 2003 Technical Report*. Paris : OECD, 2005a.
- **2003.** *The PISA 2003 Assessment Framework - Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris : OECD, 2003.
- Statistik Austria. 2005.** *Standard-Dokumentation Metainformationen (Definitionen, Erläuterungen, Methoden, Qualität) zu EU-SILC 2003*. Wien : Statistik Austria, 2005.